MINISTRY OF HIGHER EDUCATION, SCIENCE AND INNOVATIONS OF THE REPUBLIC OF UZBEKISTAN

NATIONAL UNIVERSITY OF UZBEKISTAN NAMED AFTER MIRZO ULUGBEK

RAQAMLI TEXNOLOGIYALAR VA SUN'IY INTELLEKTNI RIVOJLANTIRISH ILMIY-TADQIQOT INSTITUTI

INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE

PROCEEDINGS
of materials on topic:

«CONTEMPORARY TECHNOLOGIES OF COMPUTATIONAL LINGUISTICS - CTCL.2024»

Crossref we use DOIs doi Google Scholar OPEN ACCESS

TASHKENT - 2024

“KOMPYUTER LINGVISTIKASINING ZAMONAVIY TEXNOLOGIYALARI – CTCL.2024” NOMLI XALQARO ILMIY-AMALIY KONFERENSIYASINING ILMIY MAQOLALARI VA TEZISLAR TO‘PLAMI. T.: O‘zMU, 2024.
2024-yil 22-aprel

TASHKILIY QO‘MITA TARKIBI:

N.Z.Abduraxmonova.	O‘zMU Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi mudiri, f.f.d., professor.
Z. T.Taxirov.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi professori, f.f.n.
A.R.Iskandarova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi dotsenti, PhD.
D.A.Fattaxova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi katta o‘qituvchisi, PhD.
U.R.Xalmuxamedova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi katta o‘qituvchisi.
N.Sh.Inogamova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi katta o‘qituvchisi.
F.R.Topildiyeva.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi katta o‘qituvchisi.
D.K.Xudoyberganova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi katta o‘qituvchisi, PhD.
Z.M.Kasimova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi katta o‘qituvchisi, PhD.
S.M.Po‘latova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi katta o‘qituvchisi, PhD.
G.G‘.Xoliyorova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
O.O.Yulbarsov.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
F.M.Aliyev.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
M.S.Jo‘raqulova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
Z.G.Qodirova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
G.I.Kuvondikova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
M.F.Abdunazarova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
D.Xolbekova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
D.A.Mirvaliyeva.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
M.H.Eshmamatova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
Q.R.Ibodullayeva.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
U.M.Askarova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
S.A.Abduraxmonova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
N.O‘rinboyeva.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi magistranti.
M.Xudoyberganova.	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi magistranti.

Mazkur xalqaro ilmiy-amaliy konferensiya to‘plami kompyuter lingvistikasining ayrim yo‘nalishlari, jumladan, korpus lingvistikasi, til ta’limida kompyuter texnologiyalari, NLPda til va nutq tahlili (lingvistik analizator: morfologik, sintaktik va semantik tahlil, nutq analizi va sintezi), mashina tarjimasining lingvistik va dasturiy ta’minoti, kompyuter leksikografiyasi: ontologiya, tezaurus, shuningdek, zamonaviy tilshunoslik masalalari kabi mavzular tahlilga tortilgan maqola va tezislardan tashkil topgan.

Ushbu ilmiy maqola va tezislar to‘plami kompyuter lingvistikasi, kompyuter texnologiyalari, amaliy filologiya kabi yo‘nalishlarda ilmiy izlanishlar olib borayotgan professor-o‘qituvchilar, doktorantlar, mustaqil izlanuvchilar, magistr va talabalar uchun mo‘ljallangan.

To‘plamga kiritilgan maqolalar va tezislar mazmuni hamda sifatiga mualliflar mas’uldir.

MAS’UL MUHARRIR:
prof., f.f.d. N.Z.Abduraxmonova

TAHRIR HAY’ATI:
f.f.n., prof. Z. T.Taxirov
f.f.n., dots. A.R.Iskandarova
ped.f.n., dots. D.A.Fattaxova
PhD D.K.Xudoyberganova
PhD Z.M.Kasimova

Ushbu xalqaro ilmiy-amaliy anjuman to‘plami Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti Ilmiy-texnik kengashi tomonidan (2024-yil 27-martdagi 3-sonli majlis bayonnomasi) nashrga tavsiya qilingan.

СБОРНИК НАУЧНЫХ СТАТЕЙ И ТЕЗИСОВ
МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ
“СОВРЕМЕННЫЕ ТЕХНОЛОГИИ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ – CTCL.2024”.
Т.: НУУз, 2024.
22 апреля 2024 г.

СОСТАВ ОРГАНИЗАЦИОННОГО КОМИТЕТА:

Н.З. Абдурахмонова.	Заведующий кафедрой компьютерной и прикладной лингвистики НУУз, д.ф.н., профессор.
З. Т. Тахиров.	Профессор кафедры компьютерной и прикладной лингвистики, к.ф.н.
А.Р.Искандарова.	Доцент кафедры компьютерной и прикладной лингвистики, к.ф.н.
Д.А. Фаттахова.	Старшая преподавательница кафедры компьютерной и прикладной лингвистики, к.пед.н.
У.Р. Халмухамедова.	Старшая преподавательница кафедры компьютерной и прикладной лингвистики.
Н. Ш. Иногамова.	Старшая преподавательница кафедры компьютерной и прикладной лингвистики.
Ф.Р. Топилдиева.	Старшая преподавательница кафедры компьютерной и прикладной лингвистики.
Д.К. Худойберганова.	Старшая преподавательница кафедры компьютерной и прикладной лингвистики, к.ф.н.
З.М. Касимова.	Старшая преподавательница кафедры компьютерной и прикладной лингвистики, к.ф.н.
С. Пулатова.	Преподавательница кафедры компьютерной и прикладной лингвистики, к.ф.н.
Г.Г. Холёрова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
О.О. Юлбарсов.	Преподаватель кафедры компьютерной и прикладной лингвистики.
Ф.М. Алиев.	Преподаватель кафедры компьютерной и прикладной лингвистики.
М.С. Джуракулова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
З.Г. Кадирова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
Г. И. Кувондикова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
М.Ф. Абдуназарова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
Д. Холбекова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
Д.А. Мирвалиева.	Преподавательница кафедры компьютерной и прикладной лингвистики.
М.Х. Эшматматова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
К.Р. Ибодуллаева.	Преподавательница кафедры компьютерной и прикладной лингвистики.
У.М. Аскарова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
С.А. Абдурахмонова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
Н.З. Уринбоева.	Магистрант кафедрой компьютерной и прикладной лингвистики.
М. Худойберганова.	Магистрант кафедрой компьютерной и прикладной лингвистики.

Материалы международной научно-практической конференции, включающие статьи и тезисы, посвящены некоторым направлениям компьютерной лингвистики, в том числе корпусной лингвистике, компьютерным технологиям в языковом образовании, анализу языка и речи в NLP (Лингвистический анализатор: морфологический, синтаксический и семантический анализ, анализ речи и синтез), Лингвистические базы данных и программное обеспечение машинного перевода, Компьютерная лексикография: онтология, тезаурус, а также проблемы современной лингвистики.

Сборник научных статей и диссертаций предназначен для профессоров и преподавателей, докторантов, независимых исследователей, магистров и студентов, ведущих научные исследования в таких областях, как компьютерная лингвистика, компьютерные технологии и прикладная лингвистика.

Авторы несут ответственность за содержание и качество статей и тезисов, включенных в книгу.

ОТВЕТСТВЕННЫЙ РЕДАКТОР:
д.ф.н., проф. Н.Абдурахмонова

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:
PhD, проф. З.Тахиров
PhD, доц. А.Искандарова
PhD З.Касимова
PhD Д.Фаттахова
PhD Д.Худойберганова

Сборник международной научно-практической конференции рекомендован к публикации Научно-технической комиссией Национального университета Узбекистана имени Мирзо Улугбека (протокол № 3 от 27 марта 2024 года).

PROCEEDINGS OF ARTICLES AND THESIS
OF THE INTERNATIONAL SCIENTIFIC -PRACTICAL CONFERENCE TITLED
«CONTEMPORARY TECHNOLOGIES OF COMPUTATIONAL LINGUISTICS-CTCL.2024»
T.: NUUz, 2024.
April 22, 2024.

ORGANIZING COMMITTEE:

N. Z. Abdurakhmonova.	Head of Department of Computational and Applied Linguistics of UzMU, DSc, Professor.
Z. T. Takhirov.	Professor of the Department of Computational and Applied Linguistics, PhD.
A.R. Iskandarov.	Associate Professor of the Department of Computational and Applied Linguistics, PhD.
D.A. Fattakhova.	Senior teacher of the Department of Computational and Applied Linguistics, PhD.
U.R. Khalmukhamedova.	Senior teacher of the Department of Computational and Applied Linguistics.
N. Sh. Inogamova.	Senior teacher of the Department of Computational and Applied Linguistics.
F.R. Topildieva.	Senior teacher of the Department of Computational and Applied Linguistics.
D.K. Khudoyberganova.	Senior teacher of the Department of Computational and Applied Linguistics, PhD.
Z.M. Kasimova.	Senior teacher of the Department of Computational and Applied Linguistics, PhD.
S.M. Polatova.	Senior teacher of the Department of Computational and Applied Linguistics, PhD.
G.G. Kholiyorova	Senior teacher of the Department of Computational and Applied Linguistics.
O. O. Yulbarsov.	Teacher of the Department of Computational and Applied Linguistics.
F.M. Aliyev.	Teacher of the Department of Computational and Applied Linguistics.
M.S. Jurakulova.	Teacher of the Department of Computational and Applied Linguistics.
Z.G. Kadirova.	Teacher of the Department of Computational and Applied Linguistics.
G. I. Kuvondikova.	Teacher of the Department of Computational and Applied Linguistics.
M.F. Abdunazarova.	Teacher of the Department of Computational and Applied Linguistics.
D. Kholbekova.	Teacher of the Department of Computational and Applied Linguistics.
D.A. Mirvaliyeva.	Teacher of the Department of Computational and Applied Linguistics.
M.H. Eshmamatova.	Teacher of the Department of Computational and Applied Linguistics.
Q.R. Ibodullayeva.	Teacher of the Department of Computational and Applied Linguistics.
U.M. Askarova.	Teacher of the Department of Computational and Applied Linguistics.
S.A. Abdurakhmonova.	Teacher of the Department of Computational and Applied Linguistics
N.Z. Urinboeva.	Master student of Computational and Applied Linguistics.
M. Khudoyberganova.	Master student of Computational and Applied Linguistics.

This proceedings of international scientific-practical conference including articles and theses deals with some directions of Computational Linguistics, including Corpus Linguistics, Computer Technologies in language education, Language and speech analysis in NLP (Linguistic analyzer: morphological, syntactic and semantic analysis, speech analysis and synthesis), Linguistic database and software of Machine Translation, Computer Lexicography: ontology, thesaurus, as well as Modern Linguistic Issues.

The proceedings of scientific articles and theses is intended for professors and teachers, doctoral students, independent researchers, masters and students who are conducting scientific research in such areas as Computational Linguistics, Computer Technologies, and Applied Linguistics.

The authors are responsible for the content and quality of the articles and theses included in the book.

RESPONSIBLE EDITOR:
DSc, prof. N.Abdurakhmonova

EDITORIAL BOARD:
PhD, prof. Z.Taxirov
PhD, dots. A.Iskandarova
PhD Z.Kasimova
PhD D.Fattaxova
PhD D.Xudoyberganova

This proceeding of international scientific - practical conference was recommended for publication by the Scientific and technical commission of the National University of Uzbekistan named after Mirzo Ulugbek (protocol № 3 dated March 27, 2024).

MUNDARIJA
*****
TABLE OF CONTENTS

I SHO‘BA
KORPUSLINGVISTIKASINING DOLZARB MASALALARI
*****
SECTION 1
CURRENT ISSUES OF CORPUS LINGUISTICS

Г. МАДИЕВА, К. ПІРМАНОВА, А. СЕРИКБАЕВА. ИННОВАЦИИ В ИЗУЧЕНИИ ПОЭЗИИ АБАЯ: РАЗРАБОТКА КАЗАХСКОГО ПОЭТИЧЕСКОГО ПОДКОРПУСА 14

E.G‘. XONNAZAROV. ZAMON MA’NOSINING TUR, YUR, YOT, O‘TIR FE’LLARIDA VOQELANISHINI KORPUSDA TEGLASH MASALASI 21

Д. ТОҚМЫРЗАЕВ, Е. ЖҰБАНОВА, А. ПІРМАНОВА. СОЗДАНИЕ ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКОЙ МОДЕЛИ КАЗАХСКОГО ТЕКСТА 26

S.R. BAXODIROV. MATN TOZALASH TIZIMI DASTURLARINING UMUMIY TAVSIFI. AFZALLIKLARI VA KAMCHILIKLARI 33

I.Z. BEKMURADOVA. SLENGLARNI AVTOMATIK TEGLASHDA SLENG BIRLIKLARI LUG‘ATINING AHAMIYATI.....39

A. SULTANOVA. QORAQALPOQ TA‘LIMIY KORPUSINI YARATISHNING NAZARIY ASOSLARI 42

M. ESHMAMATOVA. TURKIY TILLARDA KORPUSLARNING KENG KO‘LAMI VA ULARNING RIVOJLANISHI.....47

D.B. AXMEDOVA. O‘ZBEK TILI KORPUSLARIDA O‘Z QATLAM LEKSIKASI ETIMOLOGIYASINI SEMANTIK KENGAYTMA SIFATIDA IFODALASH MUAMMOLARI 52

G.I. KUVONDIKOVA, N. URINBOEVA. CREATION OF THE EDUCATIONAL CORPUS OF STUDENT SPEECH IS AN EXAMPLE OF WORLD EXPERIENCE.....55

K.CH. RAJABOVA. KORPUSGA ASOSLANUVCHI GRAMMATIK VA LEKSIKOGRAFIK TADQIQOTLAR 62

M. NORMAMATOVA, N. ABDURAXMONOVA. BOLALAR NUTQI KORPUSI SHAKLLANISH BOSQICHLARI.....67

A. VOSILJONOV. JAHON KORPUS LINGVISTIKASIDA DIALEKTAL KORPUSLAR.....71

M. SIROJIDDINOVA. KORPUSGA ASOSLANGAN MASHINA TARJIMASI UCHUN IKKI TILLI FRAZEOLOGIK LUG‘AT YARATISHNING LINGVISTIK AHAMIYATI 75

O.O. YULBARSOV, YU.S. ABDIRAHIMOVA. MATN TAHLILIDA TERMINOLOGIYA MASALASI 79

S. OTAMURODOVA. IDENTIFICATION OF SEMANTIC FEATURES OF PHRASAL VERBS IN A PARALLEL CORPUS.....83

А. АЙТБЕККЫЗЫ, А. ШОРМАКОВА. ОБЗОР И РАЗРАБОТКА ПРОГРАММЫ ПО СБОРУ ЭЛЕКТРОННЫХ РЕСУРСОВ НА КАЗАХСКОМ ЯЗЫКЕ ПО ЗАКОНОДАТЕЛЬСТВУ РЕСПУБЛИКИ КАЗАХСТАН 88

R.T. O‘RINBOYEVA, I.J. ASQAROV. O‘ZBEK KORPUSSHUNOSLIGIDA DIALEKTAL TA’MINOT YARATISH MASALALARI.....94

M. SULTONOVA. DEVELOPING PRE-SERVICE STUDENTS' CORPUS LITERACY SKILLS IN TEACHING ENGLISH 98

Г. ХИДИРОВА. ЛИНГВОМЕТОДИЧЕСКИЙ ПОТЕНЦИАЛ ПАРАЛЛЕЛЬНОГО КОРПУСА...101

R.T. MUFTILLAYEVA. NUROTA TUMAN LINGVISTIK TAMINOTI (TUZILISHI VA TAVSI..118

N. URINBOEVA, G.I. KUVONDIKOVA. CORPUS ANALYSIS OF SPEECH UNITS IN SOCIAL NETWORKS....122

P.V. ORIPOVA, G.I. TOIROVA. MINNATDORCHILIK – MULOQOT MANBAI SIFATIDA....126

E. NAZIROVA, N. ABDURAKHMONOVA, K. USMONOVA. CHALLENGES IN CORPUS LINGUISTICS AND MACHINE TRANSLATION 130

SH. GULYAMOVA, M.V. NURBOYEVA. FRAZEOLOGIK POLISEMIYA VA UNI KORPUS TIZIMIDA MODELLASHTIRISH MUAMMOLARI....134

D.A. FATTAXOVA, U.T. TURAYEVA. TURKCHA - O'ZBEKCHA PARALLEL KORPUSI UCHUN IBORALARNING SEMANTIK TAHLILI 139

N.B. MURADOVA. MUALLIFLIK KORPUSINING AHAMIYATI....144

D.O. O'ROQOVA. KORPUSDAN QIDIRUV TIZIMI SIFATIDA FOYDALANISH 148

K. VASLIDDINOVA. THE INSTRUMENTAL OBJECTIVE ROLE OF CORPUS IN RESEARCHES OF VARIOUS FIELDS 154

A. QARSHIYEV, S.A. KARIMOV, M.S. TURSUNOV. O'ZBEKCHA SO'ZLARNING GRAMMATIK TAHLIL QILISH ALGORITMLARI 159

M.F. XOMIDOVA. KORPUS LINGVISTIKASIDA INTERTEKSTUALLIKNING ROLI....164

R.T. O'RINBOYEVA, D.G. ALLABERDIYEVA. CHILDES KORPUSLAR TIZIMIDA O'ZBEK TILI UCHUN MA'LUMOTLAR BAZASI YARATISH....168

M. URAZALIYEVA. KOMPYUTER LINGVISTIKASIDA AUDIOKORPUS MASALALARI 172

G. ERGASHEVA, N. INOGAMOVA. O'ZBEK TILINING TA'LIMIY KORPUSI UCHUN BOLALAR ADABIYOTINI YARATISHNING DIDAKTIK ASOSLARI YUZASIDAN AYRIM MULOHAZALAR 175

II SHO'BA
TIL TA'LIMIDA KOMPYUTER TEXNOLOGIYALARI
*****
SECTION 2
COMPUTER TECHNOLOGIES IN LANGUAGE EDUCATION

А.Р. ГАТИАТУЛЛИН, Н. АБДУРАХМОНОВА. БАЗА ЗНАНИЙ ПОРТАЛА «ТЮРКСКАЯ МОРФЕМА » КАК ЛИНГВИСТИЧЕСКИЙ РЕСУРС ДЛЯ СОЗДАНИЯ ИНСТРУМЕНТОВ ОБУЧЕНИЯ ТЮРКСКИМ ЯЗЫКАМ....179

K. MAVLONOVA. O'QUVCHILARDA MEDIA SAVODXONLIK KO'NIKMASINI SHAKLLANTIRISH 187

A.A. JUMABOEV, A. SAITKHANOVA. COMPUTER TECHNOLOGIES IN LANGUAGE EDUCATION 194

D.A. FATTAXOVA, D. JUMANAZAROVA. O'ZBEK TILI MORFOLOGIK LUG'ATINING LINGVISTIK TA'MINOTI 198

J.B. BERDIYEV. BERT MODELINING QO'LLANILISHI VA VAZIFALARI 202

М. АХМЕДОВА, Д. ФАТТАХОВА. РОЛЬ И ФУНКЦИЯКОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ В ЯЗЫКОВОМ ОБРАЗОВАНИИ.....209

K.T. ODILOVA. ПОВЫШЕНИЕ ОБРАЗОВАТЕЛЬНЫХ РЕЗУЛЬТАТОВ С ПОМОЩЬЮ МУЛЬТИМОДАЛЬНЫХ ТЕКСТОВ: НА ПРИМЕРЕ УЗБЕКИСТАНА 214

SH. ASADOVA, H. ABDURAHMONOVA. XORIJIY TILLARNI O‘RGANISHDA KOMPYUTER TEXNOLOGIYALARINING O‘RNI.....218

Д. ЖУПАРБЕК. ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ ВЕБ-КВЕСТОВ ДЛЯ ОРГАНИЗАЦИИ САМОСТОЯТЕЛЬНОЙ РАБОТЫ НА РАЗЛИЧНЫХ УРОВНЯХ 223

M.G‘. HASANOVA, N. ABDURAXMONOVA. TIL TA’LIMIDA KOMPYUTER LINGVISTIKASI IMKONIYATLARIDAN FOYDALANISH: MUAMMO VA YECHIMLAR.....232

С. ИБРАГИМОВА, Г. МИРЗАЕВ. ЛИНГВИСТИЧЕСКИЙ ТРЕНАЖЕР ОБУЧАЮЩИЙ УЗБЕКСКОМУ ЯЗЫКУ КАК ИНОСТРАННОМУ НА БАЗЕ АНАЛИЗАТОРА РЕЧИ 236

A.N. AKHMEDOVA. THE IMPORTANCE OF CORPORA IN TRANSLATION TEACHING AN..240

III SHO‘BA
NLPDA TIL VA NUTQ TAHLILI
(LINGVISTIK ANALIZATOR: MORFOLOGIK, SINTAKTIK VA SEMANTIK TAHLIL, NUTQ ANALIZI VA SINTEZI)
*****
SECTION 3
LANGUAGE AND SPEECH ANALYSIS IN NLP (MORPHOLOGICAL, SYNTACTIC AND SEMANTIC ANALYSIS; SPEECH ANALYSIS AND SYNTHESIS)

L. BOIZOU, M. HOUSSOUBA. MAKING USEFUL TOOLS FOR LANGUAGE COMMUNITIES WITHOUT BIG CORPORA: A VERSATILE MORPHOLOGICAL ANALYSER AND SPELLCHECKER FOR SONGHAY BASED ON FREE SOFTWARE 243

E. NAZIROVA, K. USMONOVA. ADVANCED NLP AI TOOLS IMPORTANCE IN MACHINE TRANSLATION FROM UZBEK TO ENGLISH.....254

V. KARYUKIN, N. ABDURAKHMONOVA. THE DEVELOPMENT OF AN APPLICATION FOR COLLECTING AND FORMING CORPORA OF THE KAZAKH LANGUAGE 259

Э.Ш. НАЗИРОВА, К.А. УСМАНОВА. ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ЯЗЫКОВОМ ПЕРЕВОДЕ.....265

Z. ISMADIYOROV. SEMANTIK TEGLASHNING FE’L SO‘Z TURKUMIGA OID GURUHLAR 269

M.T. UZOQOVA. SINTAKTIK SINONIMIYA VA UNI ANIQLASH XUSUSIDA 272

SH. OTAJONOVA, N. INOGAMOVA. NATURAL LANGUAGE PROCESSING (NLP) APPLICATIONS IN LANGUAGE TEACHING.....277

M.S. JO‘RAQULOVA, Z.G. QODIROVA. STRUKTUR TAHLIL METODLARI ASOSIDA YARATILGAN SINTAKTIK NAZARIYALAR.....283

А. МИНДУБАЕВ. НЕЧЕТКАЯ ЛОГИКА НА БАЗЕ ЯЗЫКОВЫХ МОДЕЛЕЙ: ПОСТРОЕНИЕ ГИБРИДНОЙ СИСТЕМЫ КЛАССИФИКАЦИИ ДИАЛЕКТОВ И ОПРЕДЕЛЕНИЯ ДИАЛЕКТИЗМОВ 292

M. SUYUNOVA. NATURAL LANGUAGE PROCESSING AND ITS MAIN APPROACHES.....303

E. MALIKOV. BIR SO‘Z TURKUMIDAGI OMONIM SO‘ZLARNI ANIQLASH USULLARI 309

М.И. ИСМОИЛОВА. ЎЗБЕК ТИЛИДА ҚАРАШЛИЛИК/ТЕГИШЛИЛИК МАЪНОСИНИНГ СЎЗ СЕМАНТИКАСИ ТАЪСИРИДА ЯШИРИН ИФОДАЛАНИШИ УСУЛИ 316

SH. ABDISALOMOVA. LINGVISTIK CHEKLOVLAR – KOREFERENTLIKNI AVTOMATIK HAL ETISHNING ASOSI 323

A. OTEMISOV, SH. SHARBAEV. MORFOANALIZATOR BOSQICHI UCHUN QORAQALPOQ TILIDAGI FE'L SO'Z TURKUMINI FORMALLASHITISH MODULI 328

G.E. UMURZAKOVA. UNDOV SO'ZLAR PRESUPPOZITSIYA SIGNALIZATORI SIFATIDA ...333

Z.I. FAYZULLAYEVA, N.N. KARIMOV. NLPDA MATNLARNI UMUMIYLASHTIRISH VA LEKSIK TAHLIL 337

B. ELOV, A. ABDULLAYEV, N. XUDAYBERGANOV. O'ZBEK TILI KORPUSI MATNLARI ASOSIDA TIL MODELLARINI YARATISH 343

В.Б. БАРАХНИН, Д.Б. МЕНГЛИЕВ. АЛГОРИТМИЧЕСКИЙ ПОДХОД К АНАЛИЗУ ЭМОЦИОНАЛЬНОЙ ОКРАСКИ И ПОЭТИЧНОСТИ УЗБЕКСКИХ ТЕКСТОВ 353

U. YAXSHIBOYEV. SIYOSIY MATNLARNI O'RGANISHDA NUTQIY YONDASHUV 358

U. TULIEV, N. ABDURAKHMONOVA. METHOD OF DEVELOPING ARTIFICIAL PATTERNS TO ANALYZING SEMANTIC SIMILARITY OF DOCUMENTS.....361

IV SHO'BA
MASHINA TARJIMASINING LINGVISTIK VA DASTURIY TA'MINOTI
*****
SECTION 4
LINGUISTIC DATABASE AND SOFTWARE OF MACHINE TRANSLATION

R. ALAYEV, G. MAXMUDJONOVA. COUNTVECTORIZER YORDAMIDA SO'ZLAR STATISTIKASINI ANIQLASH.....365

N.S. SOTVOLDIYEVA. MASHINA TARJIMASIDA SINONIMLAR BILAN BOG'LIQ MUAMMOLAR 370

U. ABDURAXMONOV. INGLIZCHA-O'ZBEKCHA MASHINA TARJIMASI UCHUN FRAZEOLOGIK BIRLIKLAR BAZASINI YARATISH TAVSIFI.....375

M.S. SHARIPOV, O.O. SOBIROV, G.M. SHARIPOVA. MASHINALI TARJIMA TIZIMLARI UCHUN O'ZBEK TILI SO'ZLARINI LEMMALASH VA GAP STRUKTURALARINI TAHLIL QILISH.....378

M. XUDOYBERGANOVA. MASHINA TARJIMASIDA SINTAKTIK QOLIPLARNING QIYOSIY TAHLILI (QORAQALPOQ TILI MISOLIDA) 383

A.Z. OTEMISOV, A. ABDULLAEVA. THE ROLE OF LINGUISTICS IN MACHINE TRANSLATION SOFTWARE DEVELOPMENT 387

M.A. JO'RAQULOVA. PARALLEL KORPUS MASHINA TARJIMASINING O'RGANISH OBYEKTI SIFATIDA 391

R. YULDASHEVA. A LINGUISTIC STUDY OF COLLOCATION IN ENGLISH.....395

B. KHOLBEKOVA. THE ESSENCE OF THE CHARMING POETRY OF ZAHIRUDDIN MUHAMMAD BABUR 399

G. SHAMSIYEVA. MASHINA TARJIMASI UCHUN GRAMMATIK SATHDA MODELLASHTIRISH BOSQICHLARI 403

N. SHAMIEVA. BREAKING DOWN LANGUAGE BARRIERS: THE POWER AND PROGRESS OF MACHINE TRANSLATION 410

H. ZOKIROVA. TERMINOLOGY AS A LEARNING OBJECT IN MACHINE TRANSLATION....414

G.A. SHIKHNAZAROVA, D.R. GULYAMOVA, N.R. SAMANDAROVA. MATHEMATICAL MODELING OF MACHINE LEARNING ALGORITHMS 417

V SHO‘BA
KOMPYUTER LEKSIKOGRAFYASI: ONTOLOGIYA, TEZAURUS
*****

SECTION 5
COMPUTER LEXICOGRAPHY: ONTOLOGY, THESAURUS

N. ABDURAXMONOVA, U. ASKAROVA, YU. YULBARSOV. KORPUSNING LINGVISTIK VA TA’LIMIY AHAMIYATI.....420

G. PO‘LATOVA. O‘ZBEK TILI TEZAURUSI UCHUN SIFAT SO‘Z TURKUMI LINGVISTIK BAZASINI YARATISH 428

Z. UBAYDULLAYEVA. TEZAURUS TUSHUNCHASI VA ILMIY YONDASHUV 434

M. SAPAROVA. ISSUES, THEORIES, AND APPROACHES IN THE PRESENT-DAY LEXICOGRAPHIC SEMANTIC STUDIES.....437

Z.G. QODIROVA, M.S. JO‘RAQULOVA. TERMINOLOGIK ONTOLOGIK BILIMLAR BAZASINI MODELLASHTIRISH BOSQICHLARI 442

L. BABAJANOVA. BILINGVAL TEZAURUS YARATISHDA O ‘ZBEK VA INGLIZ TILLARINING VALENTLIK MUNOSABATI.....453

G‘.G‘. ABDUVAXOBOV. O‘ZBEK TILIDA ELEKTRON LUG‘ATLAR YARATISH ZARURATI 458

N. ABDURAXMONOVA, G. XOLIYOROVA. O‘ZBEK TILI TEZAURUSINI YARATISHDA STANDART ME’YORLAR.....463

G. HASANOVA. THE PROBLEMS OF REPRESENTING NEAR-SYNONYMS IN A COMPUTATIONAL LEXICON.....469

S.O. OYBEKOVA. TEZAURUSLAR VA ULARNI TAHLIL QILISH TAMOYILLARI 474

VI SHO‘BA
ZAMONAVIY TILSHUNOSLIK MASALALARI
*****

SECTION 6
ISSUES OF MODERN LINGUISTICS

N.I. ADIZOVA. SAYOHAT TURIZMIGA OID MATNLARDA MILLIY-MADANIY BIRLIKLARNING LEKSIK-PRAGMATIK XUSUSIYATLARI 478

D.B. JUMAYEVA, S.M. USMONOVA. BOSHLANG‘ICH SINF O‘QUVCHILARINI ORFOEPIK ME’YORLARGA O‘RGATISH VA UNING AMALIY AHAMIYATI 484

N.M. MIROBIDOVA. O‘ZBEK AYOLLAR NUTQIDA KONNOTATSIYA VA PRAGMATIK T...487

S.D. NURMONOVA. XALQONA QASAMLARDA FONATSION VOSITALAR 492

D.X. XOLBEKOVA, K. IBODULLAYEVA. TILSHUNOSLIKDA LINGVOSTATISTIK TAHLIL MUAMMOLARINING O‘RGANILISHI 495

SAJID V. VALAPPIL, N. ABDURAKHMONOVA. ISSUES OF MODERN LINGUISTICS: - INHIBITING LINGUISTIC CREATIVITY AMONG YOUTH ON SOCIAL MEDIA BY USING EMOJI AND INTERNET SLANG.....504

S.A. ABDURAXMONOVA. O‘ZBEK BOLALAR NUTQINING KORPUS ASOSIDA PSIXOLINGVISTIK TADQIQI.....509

I.E. UMIROV. SEMANTOGRAFIK TARJIMA YOZUVINING LEKSIK, GRAMMATIK VA STILISTIK XUSUSIYATLARI.....513

Н. АБЛАЕВА, И. ДЖУМАНИЯЗОВА. ВОЗМОЖНОСТИ КОРПУСНОЙ ЛИНГВИСТИКИ В ИЗУЧЕНИИ РУССКОГО СЛОВООБРАЗОВАНИЯ.....519

A.Q. ABDULLAYEV, B.B. ELOV, M.T. SAMATBOYEVA. O‘ZBEK TILIDAGI MATNLARDA ATOQLI OTLARNI AVTOMATIK ANIQLASH NE VA NER TUSHUNCHASI.....524

M.F. ABDUNAZAROVA. GENDEROLINGVISTIKA ZAMONAVIY TILSHUNOSLIK SOHASI..528

D. XUDAYBERGANOVA, Z. KASIMOVA. SOTSIOLINGVISTIKADA O‘RGANILADIGAN MASALALAR XUSUSIDA.....534

J. MAHMUDOVA. AN ANALYSIS OF PRISON ARGOT AND ITS TYPES.....539

K. IBODULLAYEVA, D. XOLBEKOVA. SCIENTIFIC DISCOURSE AS AN OBJECT OF STUDY OF SCIENTIFIC RESERCH.....542

A.R. ISKANDAROVA. IJTIMOIY MAQOM, IJTIMOIY ROL, LISONIY ISHTIMOIYLASHUV....547

A.R. ISKANDAROVA, G. JABBOROVA. TOHIR MALIKNING “ALVIDO BOLALIK” QISSASIDA QO‘LLANGAN SOTSIOLEKTLARNING SINTAKTIK KONSTRUKSIYALARI VA MA’NOVIY 551

F. KULTAYEVA. ERGONIMLARNI MA’NOVIY VA VAZIFAVIY KLASSIFIKATSIYALASH..555

S. BOYSARIYEVA. MORFEM LUG‘ATLAR KLASSIFIKATSIYASI VA TABIATI.....560

M.B. BOTIROVA. O‘ZBEK TILIDA REKLAMALARNING SINTAKTIK STRUKTURASI.....564

M.N. XUSANOVA. TARIXIY DIGNITONIM VA FALERONIMLARNING LISONIY TADQIQI...567

F.R. TOPILDIYEVA. FARIDA AFRO‘Z SHE’RIYATINING LINGVOPOETIK TADQIQI VA O‘ZIGA XOS JIHATLARI.....570

Z.T. TAXIROV. TIL, USLUBIYAT, NUTQ MADANIYATI VA ADABIY TAHRIR.....574

D.B. JUMAYEVA, N.O. NOROVA. BOSHLANG‘ICH SINFLARDA IMLO QOIDALARI ASOSIDA TOG’RI YOZISHNI SHAKLLANTIRISHGA O‘RGATISH.....581

N. RASULOV, A. AKBAROVA. O‘ZBEK TILIDAGI HARAKAT MAZMUNLI SLENGLAR.....585

S. PO‘LATOVA. IZOGLOSSA.....589

A.T. KAZAXBAEVA. T.JUMAMURATOV ASARLARI TILIDAGI PRAGMATIK BELGILAR.....592

D. SULAYMANOVA. QO‘SHMA GAPLARDA TRANSFORMATSIYA: NOMINATIV BIRLIK (CHEGARALI QURILMA)LARDA POG‘ONAVIYLIK.....597

M.A. ABJALOVA, G.SH. SAFAROVA, N. RAVSHANOVA. ZAMONAVIYLASHGAN VA “TIRILGAN” ESKI SO‘ZLAR.....603

U.R. HALMUKHAMEDOVA. THE SIGNIFICANCE OF TRANSLATION TEXTS IN TEACHING THE UZBEK LANGUAGE.....607

M.Q. ABDULLAYEVA. TILSHUNOSLIKDA MEDIAMATNNING O‘RNI.....611

I SHO‘BA
KORPUS LINGVISTIKASINING DOLZARB MASALALARI

ИННОВАЦИИ В ИЗУЧЕНИИ ПОЭЗИИ АБАЯ: РАЗРАБОТКА КАЗАХСКОГО ПОЭТИЧЕСКОГО ПОДКОРПУСА

Мадиева Гульмира
доктор филологических наук, профессор
Казахский национальный университет имени аль-Фараби.

Пірманова Күнсұлу
старший научный сотрудник, постдокторант
Казахский национальный университет имени аль-Фараби.

Серикбаева Айжан
старший научный сотрудник, постдокторант
Казахский национальный университет имени аль-Фараби.

Аннотация: Предлагаемое исследование, посвященное разработке казахского поэтического подкорпуса, актуально в контексте сохранения культурного наследия и применения современных технологий для его анализа. Работа фокусируется на творчестве известного казахского поэта Абая Кунанбаева с целью создания цифровой базы для морфологического и поэтического изучения его стихов. Особенностью проекта является отсутствие аналогов такого рода исследований, что подчеркивает его инновационный характер. Методологические основы проекта включают в себя разработку критериев разметки и адаптацию инструментов обработки естественного языка. Первый этап исследования охватывает разработку теоретической базы и первичный анализ текстов, что обеспечит качественный фундамент для последующего глубокого изучения и классификации данных. Проект имеет значительные перспективы для развития цифровой гуманитарной науки и обогащения культурного пространства Казахстана, открывая новые горизонты для изучения национальной литературы и культуры в цифровую эпоху.

Ключевые слова: цифровые гуманитарные науки, поэтический корпус, казахская поэзия, поэзия Абая, морфологический анализ, обработка естественного языка, культурное наследие.

Annotation: The study dedicated to the development of the Kazakh poetic corpus is pertinent in the context of cultural heritage preservation and the application of modern technologies for its analysis. The focus of the work is on the writings of the renowned Kazakh poet Abai Kunanbaev, creating a digital database for the morphological and poetic study of his verses. A distinctive feature of the project is the absence of direct analogs in such research, highlighting its innovative nature. The methodological foundations of the project include the development of markup criteria and the adaptation of natural language processing tools. The initial stage of the research encompasses the development of a theoretical base and preliminary text analysis, which

will provide a qualitative foundation for further in-depth study and data classification. The project holds significant prospects for the development of digital humanities and enriching the cultural space of Kazakhstan, opening new horizons for the study of national literature and culture in the digital age.

Keywords: digital humanities, poetic corpus, Kazakh poetry, Abai's poetry, morphological analysis, natural language processing, cultural heritage.

В современном мире, в котором цифровые технологии используются практически во всех сферах жизни, становится все более очевидной важность развития цифровых гуманитарных наук. Одним из направлений в этой области является создание и анализ поэтических корпусов, что позволяет не только сохранить культурное наследие, но и предоставить новые возможности для его изучения и интерпретации [Жубанов А., 2013: 25]. Разработка казахского поэтического подкорпуса, сосредоточенная на морфологической и поэтической разметке стихов Абая Кунанбаева, отражает современные тенденции в этой области.

Актуальность и новизна данной темы обусловлены несколькими факторами. Во-первых, творчество Абая Кунанбаева является средоточием, сердцевиной казахской литературы и культуры, однако комплексный цифровой анализ его творчества только начинается. Во-вторых, разработка такого корпуса позволит применять современные методы обработки естественного языка и лингвистического анализа, что откроет новые горизонты для исследования поэтики казахского языка и культуры.

Основанием для разработки данной темы служат исходные данные, полученные из предыдущих лингвистических исследований [Жубанов А., 2016: 159] казахского языка, а также существующие цифровые ресурсы казахской литературы. Важно отметить, что наблюдается недостаток комплексных работ, посвященных анализу стихотворений Абая с использованием современных технологий.

Необходимость проведения научно-исследовательской работы (НИР) обусловлена стремлением расширить понимание поэтической истории казахской культуры и литературы, а также обеспечить сохранение и популяризацию национального наследия в цифровую эпоху. Научно-технический уровень разработки предполагает создание цифровой базы данных, обладающей высоким уровнем доступности и функциональности для проведения морфологического и поэтического анализа.

В рамках подготовки проекта были проведены исследования, которые показали отсутствие прямых аналогов разрабатываемого подкорпуса, что подчеркивает его уникальность и инновационный потенциал. Кроме того, важным аспектом является метрологическое обеспечение НИР по проекту "Разработка казахского поэтического корпуса: морфологическая и поэтическая разметка стихов Абая", которое включает в себя разработку стандартов для морфологической и поэтической разметки текстов.

Связь данной работы с другими научно-исследовательскими работами выражается в ее вкладе в развитие цифровой лингвистики, а также в обогащении культурных и образовательных цифровых ресурсов Казахстана. Разработка казахского поэтического корпуса является не только актуальной задачей, но и важным шагом на пути сохранения и развития национального культурного наследия.

Цель проведенного исследования: разработка и проверка методологических основ для морфологической и поэтической разметки текстов, что включает в себя выбор соответствующих инструментов и технологий, а также определение методических подходов для анализа и классификации стихов Абая.

Задачи проведенного исследования:

- разработка критериев для морфологической и поэтической разметки стихов;
- выбор и адаптация инструментов обработки естественного языка для поэтического наследия Абая;
- проведение первичного анализа выбранных произведений Абая для оценки эффективности методов;
- оценка текущего состояния исследований в данной области для интеграции лучших практик.

Этот этап является фундаментальным для всей программы, так как заложенные здесь методологические основы определяют качество и точность всего последующего анализа. Он направлен на устранение начальных технических и методологических препятствий, что позволит эффективно перейти к следующим этапам исследования, включая расширенный анализ и классификацию данных, а также разработку пользовательского интерфейса для доступа к разрабатываемому подкорпусу.

Работа, выполненная на данном этапе, обеспечивает необходимую основу для дальнейшего углубленного исследования стихов Абая, что способствует достижению конечной цели программы – созданию полноценного, структурированного и легко доступного цифрового малого поэтического корпуса, который будет служить важным ресурсом для исследователей, студентов и всех, кто интересуется казахской литературой и культурой.

С помощью поэтического подкорпуса Абая станет возможным облегчить проведение различных лингвистических исследований поэтического творчества Абая, анализ структуры стихотворений Абая, изучение опыта использования языковых средств для создания образов и эмоционального воздействия, а также исследование развития казахской поэзии как исторического процесса.

1. Выбор направления исследований.

В рамках программы "Разработка казахского поэтического корпуса: морфологическая и поэтическая разметка стихов Абая" было выбрано направление исследований, основанное на современных методах обработки естественного языка и цифровой лингвистике. Основная цель – создание структурированного и аннотированного поэтического корпуса стихотворений Абая Кунанбаева, который будет способствовать глубокому пониманию его творчества и казахской поэтической традиции.

Методы решения задач включали использование алгоритмов машинного обучения для морфологической, поэтической разметки и анализа текстов. Сравнительная оценка с аналогами показала, что применение этих методов для казахского языка является новаторским подходом, способным обеспечить более глубокий и точный анализ.

2. Процесс теоретических и экспериментальных исследований.

Теоретические исследования включали анализ лингвистических особенностей казахского языка и поэтических структур в поэзии Абая: различные методы компьютерной

лингвистики, включая статистический анализ и методы машинного обучения, для обработки текстов и извлечения необходимой информации.

Экспериментальные работы охватывали разработку и тестирование алгоритмов для морфологической разметки, а также создание интерфейса для доступа к корпусу. Было проведено несколько циклов тестирования для оценки точности и эффективности разработанных методов.

3. Обобщение и оценка результатов исследований.

Полученные предварительные результаты показали, что технико-экономическая эффективность проекта проявляется в возможности использования корпуса в образовательных и научных целях, который является уникальным в своем роде благодаря специфики казахского языка и поэзии, что способствует повышению интереса к казахской культуре и языку. В то же время, была выявлена необходимость дополнительных исследований для дальнейшего улучшения алгоритмов морфологической разметки и расширения функциональности корпуса, включая возможности для интерактивного анализа и более глубокой статистической обработки данных.

В целом, поэтический подкорпус Абая ориентирован на то, чтобы сделать поэзию более доступной и понятной как для академического сообщества, так и для широкой публики, а также способствовать более глубокому пониманию и оценке поэтического наследия [http://web-corpora.net].

Морфологическая разметка стихов Абая включает в себя процесс присвоения каждому слову в тексте меток, обозначающих их морфологические характеристики. Для этого можно использовать следующие методы и условные обозначения:

Методы морфологической разметки:

Автоматическая разметка: Использование компьютерных программ, которые автоматически анализируют текст и присваивают метки на основе предварительно заданных правил и словарей. Такие программы могут быть обучены на специализированных корпусах казахского языка.

Ручная разметка: Анализ и разметка текста специалистами-лингвистами. Этот метод обеспечивает более высокую точность, особенно в случае сложных или неоднозначных языковых конструкций.

Гибридный подход: Сочетание автоматической и ручной разметки для увеличения точности и эффективности процесса.

Условные обозначения для морфологической разметки [Жубанов А., 2015: 37]:

Для казахского языка могут быть использованы следующие условные обозначения, общепринятые в знаковой системе национальных корпусов:

Существительное (С): N
Прилагательное (П): ADJ
Глагол (Г): V
Наречие (Нар): ADV
Местоимение (Мест): PRON
Числительное (Числ): NUM
Предлог (Предл): PREP
Союз (Союз): CONJ
Междометие (Межд): INTJ

Падеж (например, именительный, винительный): NOM, ACC и т.д.

Время (например, настоящее, прошедшее): PRES, PAST и т.д.

Число (единственное, множественное): SG, PL

Эти обозначения применяются к каждому слову или словоформе в стихотворении для указания их морфологических характеристик. Это позволяет проводить более детальный анализ текста и изучать его языковые особенности.

Создание моделей морфологической разметки для стихов Абая Кунанбайулы представляет собой сложный процесс, который включает несколько этапов:

1. Сбор и подготовка данных:

Репозиторий стихов Абая: централизованное хранилище всех доступных стихотворений Абая, которые собраны из различных источников. Репозиторий предоставляет доступ к этим данным для лингвистических исследований в разных целях с использованием поэтического подкорпуса Абая, разработки языковых моделей, обучения машинного обучения и других целей, связанных с изучением или анализом языка творчества Абая.

Предварительная обработка: чистка текста от технических ошибок (например, исправление опечаток, стандартизация орфографии).

2. Разработка морфологической разметки:

Определение тегов: создание списка морфологических тегов, применимых конкретно к поэтическому языку Абая (например, часть речи, падеж, число, время и т.д.).

Ручная разметка: начальная ручная разметка выборки текстов для создания обучающего набора данных.

3. Обучение модели:

Выбор алгоритма: Определение подходящего алгоритма машинного обучения (например, CRF, LSTM, Transformer).

Обучение: использование размеченных вручную данных для обучения модели.

Валидация и тестирование: проверка эффективности модели на отдельном наборе тестовых данных.

4. Оптимизация и масштабирование:

Итеративная оптимизация: настройка параметров модели и алгоритмов для улучшения точности.

Расширение данных: постепенное добавление новых данных для улучшения обобщающей способности модели.

Технические аспекты:

Использование НЛП библиотек: применение библиотек обработки естественного языка, таких как NLTK, spaCy или машинного обучения (TensorFlow, PyTorch) для реализации модели.

Работа с подкорпусом: управление и обработка больших текстовых данных с использованием соответствующих инструментов и технологий.

Создание такой модели, безусловно, требует значительных усилий и сотрудничества лингвистов, программистов и специалистов по обработке естественного языка, в нашем случае – поэтическое наследие Абая. Это многоступенчатый процесс, требующий постоянного обновления и оптимизации для обеспечения высокой точности и полезности в исследовательских и образовательных целях.

Для примера модели морфологической разметки стихотворной строки Абая "Желсіз түнде жарық ай" (переводится как "Яркая луна в безветренную ночь"), можно использовать следующие условные обозначения:

Желсіз (безветренный):
Часть речи: Прилагательное (ADJ)
Падеж: Именительный (NOM)
Число: Единственное (SG)
түнде (ночь):
Часть речи: Наречие (ADV)
Падеж: В предложном (LOC)
Число: Единственное (SG)
жарық (яркий):
Часть речи: Прилагательное (ADJ)
Падеж: Именительный (NOM)
Число: Единственное (SG)
ай (луна):
Часть речи: Существительное (N)
Падеж: Именительный (NOM)
Число: Единственное (SG)

Таким образом, морфологическая разметка этой строки выглядит следующим образом:

Желсіз (ADJ, NOM, SG)
түнде (ADV, LOC, SG)
жарық (ADJ, NOM, SG)
ай (N, NOM, SG)

Эта разметка позволяет провести детальный анализ морфологических особенностей поэтического языка, использованного в стихотворении Абая, и способствует глубокому пониманию текста как с лингвистической, так и с литературной точек зрения.

Благодарность: Исследование финансируется Комитетом науки Министерства науки и высшего образования Республики Казахстан (грант № BR21882334).

Список использованных источников:

1. Жұбанов А.Қ. Мәтіндер корпусы – лингвистер үшін ас күшті ақпараттық ресурс //«Фонетика ғылымдарының өзекті мәселелері: даму көздері мен болашағы» атты халықаралық ғылыми-теориялық конференция материалдары. – Алматы, 2012. – 186-193 бб.
2. Жубанов А.К., Кабетенов А.Т. Казахская компьютерная лингвистика в период независимости Республики Казахстан // Сборник: Проблемы современной прикладной лингвистики (ответственный - А.В. Зубов). – Минск: МГЛУ, 2013. декабря. – С. 23-30.
3. Жубанов А.К. Принципы автоматизации морфологической разметки текстов Национального корпуса казахского языка (НККЯ) // Материалы международной научно-практической конференции «Контрастивные исследования и прикладная лингвистика». Минск: МГЛУ, 2015. – 227 с.

4. Жұбанов А. Қазақ тілінің ұлттық корпусы – қуатты ақпараттық ресурс // Актуальные проблемы диалектологии языков народов России. XVI Всероссийская конференция. – Уфа, 1–4 июня 2016 г. – 157-161-бб.
5. http://web-corpora.net

ZAMON MA'NOSINING TUR, YUR, YOT, O'TIR FE'LLARIDA VOQELANISHINI KORPUSDA TEGLASH MASALASI

Xonnazarov Elyor G'ofurovich
Filologiya fanlari bo'yicha falsafa doktori (PhD),
ToshDO'TAU katta o'qituvchisi.

Annotatsiya: Mazkur maqolada o'zbek tilidagi davomiylikni ifodalovchi holat fe'llari – turmoq, yurmoq, yotmoq, o'tirmoq fe'llarining grammatik shakllar bilan zamon ma'nosini voqelantirishdagi o'ziga xosliklari haqida ma'lumot berilgan. Bu o'ziga xoslik keltirilgan fe'lllarning semantikasi bilan aloqador ekanligi tahlillar asosida ochib berilgan. Shuningdek, amaliy jihatdan zamon ma'nosining bu fe'llarda voqelanishini korpus matnlarda teglash yuzasidan modellar ishlab chiqilgan.

Kalit so'zlar: zamon, grammatik shakl, fe'l, korpus, annotatsiya, o'tgan zamon, hozirgi zamon.

Abstract: This article provides information about the peculiarities of the verbs of Uzbek language – turmoq, yurmoq, yotmoq, o'tirmoq – in expressing the meaning of tense with grammatical forms. Based on the analysis, it was revealed that this peculiarity is related to the semantics of the mentioned verbs. Also, from a practical point of view, models have been developed for tagging the meaning of time in these verbs in corpus texts.

Keywords: tense, grammatical form, verb, corpus, annotation, past tense, present tense.

Аннотация: В данной статье представлена информация об особенностях глаголов узбекского языка – турмок, юрмок, ётмок, утирмок – в выражении значения времени грамматическими формами. На основе анализа выявлено, что данная особенность связана с семантикой упомянутых глаголов. Также с практической точки зрения разработаны модели обозначения значения времени у этих глаголов в корпусных текстах.

Ключевые слова: время, грамматическая форма, глагол, корпус, аннотация, прошедшее время, настоящее время.

Fe'lning grammatik xususiyatlarini tadqiq qilgan olimlar tomonidan zamon ma'nosining ifodalanishida fe'l leksik ma'nosining ta'sirga ega ekanligi aytib kelinsa-da, aynan qaysi fe'llarning qanday ta'sirga ega ekanligi maxsus o'rganilmagan. Biroq hozirgi zamon doirasida *yotmoq, yurmoq, turmoq, o'tirmoq* fe'llarining -ib shakli bilan o'ziga xos tarzda mazmunan hozirgi zamonni ifodalashi masalasiga to'xtalingan. Davomiylikni ifodalovchi holat fe'llari (*turmoq, yurmoq, yotmoq, o'tirmoq)*ning mustaqil holat fe'li va ko'makchi fe'l vazifalarida -(i)b shakli bilan hozirgi zamon ma'nosini anglatishi, hozirgi zamonning -yap, -yotir, -yotib kabi sintetik shakllarining kelib chiqishida asos vazifasini bajarishi, shuningdek, bir-biridan ma'noviy farqlanishi kabi masalalar o'zbek tilshunosligida, xususan, A.Hojiyev ishlarida [Ҳожиев, 1962, 51-68; Ҳожиев, 1973, 156-157] batafsil yoritilgan. Shu sababli biz bu o'rinda –ib qo'shimchasi bilan shaklan o'tgan zamonda turgan mazkur fe'llarning aslida hozirgi zamonni ifodalashini annotatsiyalash jarayonida hisobga olish qanchalar ahamiyatli ekanligiga urg'u bermoqchimiz.

N.Abdurahmonova morfologik razmetka tegger yordamida amalga oshirilsa ham, bu dasturlarning aksariyati avtomatik razmetkadan keyingi tuzatishni talab qilishiga, jumladan, o‘zbek tili uchun aynan istisnoli vaziyatlar ko‘p bo‘lganligi sababli qayta tahrirlash zarurati yuzaga kelishiga e’tibor qaratadi [Abduraxmonova, 2021, 303]. Olimaning istisnoli vaziyatlar haqidagi fikrlari zamon shakllari annotatsiyasi uchun ham taalluqli bo‘lib, bu vaziyatlarni imkon qadar modellarda qamrab olish orqali natijani nisbatan yaxshilash mumkin. Biroq “cheklangan modellardagi grammatika nuqsonlardan xoli bo‘lmaydi”[Kalonova, 2012, 97]. Shunga qaramay, ba’zi istisnolardan tashqari matnning katta qismi uchun ishlashi sababli modellarning ahamiyati yuqori baholanadi.

Turmoq, yurmoq, yotmoq, o‘tirmoq fe’llarining -ib shakli bilan hozirgi zamonni anglatishi ularning leksik ma’nosi bilan bog‘liq bo‘lib, ular holat fe’li, ko‘makchi fe’l bo‘lganda davomiylik ma’nosini ifodalaydi. Harakat ma’nosini bildirganda esa -ib shakli bilan o‘tgan zamon ma’nosini anglatadi. Ushbu fe’llar -ib shakli bilan hozirgi zamon ma’nosini ifodalashda ikki xil holatda qo‘llaniladi:

1. Holat ma’nosi bilan mustaqil fe’l sifatida: *Bundan bir yil muqaddam tashkil etilgan Maktabgacha ta’lim vazirligi oldida juda ko‘p muammolar va ularni yechish vazifasi* ***turibdi.*** *(marifat.uz)*

2. Yetakchi fe’lga birikib ko‘makchi fe’l sifatida: *Umumiy hisobda 10 mingga yaqin maktab faoliyat ko‘rsatib* ***turibdi.*** *(marifat.uz)*

Ta’limiy korpus (uzschoolcorpara.uz) matnlarida mazkur fe’llarning -ib qo‘shimchasini olib shaxs-sonda tuslangan holatlaridan anglashiladigan zamon ma’nosini tahlil qildik. Tahlil natijalari quyidagi jadvalda aks etgan:

1-jadval

T/r	Fe’l shakli	O‘tgan zamon ma’nosida		Hozirgi zamon ma’nosida	
		Soni	Foizi	Soni	Foizi
1	yur+ib+shaxs-son	34	2%	1509	98%
2	tur+ib+shaxs-son	45	0,2%	24183	99,8%
3	o‘tir+ib+shaxs-son	4	0,5%	849	99,5%
4	yot+ib+shaxs-son	12	0,8%	1427	99,2%
Jami		95	0,3%	27968	99,7%

Jadvaldan ko‘rinadiki, keltirilgan fe’l shakllari bugungi kunda o‘tgan zamon ma’nosida juda kam ishlatiladi. Mavjud qo‘llanishlarda kontekstning ta’siri ham yaqqol seziladi. Masalan: *Xayriyat, Mastura operatsiyadan bardam* ***turibdi.*** *(A.Qahhor)*

turmoq fe’li -ib zamon shakli bilan qolgan uch fe’lga nisbatan ancha faol ishlatiladi. Uning o‘zbek tilining barcha uslublarida qo‘llanila olishi bu natijaning yuzaga kelishidagi asosiy sabab hisoblanadi. Shuningdek, *turmoq, yotmoq* fe’llari jonli va jonsiz subyektlar harakat-holatini ifodalay oladi. Biroq *yurmoq va o‘tirmoq* fe’llari esa faqat jonli subyektlar harakat-holatini anglata oladi. Qolaversa, *yurmoq* va *turmoq* fe’llarining son jihatdan -ib shakli bilan o‘tgan zamonda *o‘tirmoq va yotmoq* fe’llariga nisbatan ko‘proq qo‘llanilishi, bizningcha, fe’l semantikasiga ham aloqador hisoblanadi. Chunki *yurmoq, turmoq* fe’llarida holatga nisbatan harakat ma’nosi ustunlik qiladi (ayniqsa, *yurmoq* fe’lida bu yaqqol seziladi). Masalan: “*Telefonimda yurgan masofamni o‘lchaydigan dastur bor. Kecha qarasam, bir smenada 24 kilometr masofa* ***yuribman.*** *Biror marta ham o‘tirmadim”, – degan hamshira.* (daryo.uz) O‘*tirmoq, yotmoq* fe’llarida esa harakatga

nisbatan holat ma'nosi ustunlik qiladi. Shunga qaramay, kontekstga bog'liq holda, ba'zan ko'chma ma'noda qo'llanganda bu fe'llar ham o'tgan zamon ma'nosini voqelantiradi. Masalan: *Kecha tomoshabinlarimizdan biri: «Tasbehni ham o'ttiz uch marta o'giramiz, siz ham o'ttiz uch kun* ***o'tiribsiz****. Bunda ham bir hikmat bor. O'tgan kunlaringiz, butun hayotingiz go'zal sarhisob bo'libdi», deb yozibdi. (kun.uz).*

Tilshunos A.Hojiyev "Yotmoq, turmoq, yurmoq, o'tirmoq fe'llari harakat bildirganda boshqariluvchi so'z jo'nalish yoki chiqish kelishigi formasida bo'ladi. Holat bildirganda o'rin-payt kelishigi formasida bo'ladi" [Ҳожиев, 1962:52] deya kelishik shakllari orqali mazkur fe'llarning harakat yoki holat anglatayotganligini farqlaydi. Olim fikrini quyidagi misollar orqali asoslaydi:

divanga yotibdi (o'tgan zamon) – divanda yotibdi (hozirgi zamon)
stulga o'tiribdi (o'tgan zamon) – stulda o'tiribdi (hozirgi zamon)
joyidan turibdi (o'tgan zamon) – joyida turibdi (hozirgi zamon) [Ҳожиев, 1962:52]

Keltirilgan misollarda, haqiqatan ham, kelishik shakllari orqali harakat va holat ma'nosi, shu bilan bir qatorda zamon ma'nosi ham farqlangan. Ammo ta'limiy korpus matnlaridagi bu fe'llarning qo'llanishlarini tahlil qilganimizda har doim ham bu mezonga asoslanish to'g'ri natija beravermasligiga guvoh bo'ldik. Xususan, mazkur fe'llarning holat ma'nosini anglatishini faqat o'rin-payt kelishigi bilan chegaralash mumkin emas. Gapda bu kelishik shakli ishlatilmasa ham holat ma'nosi anglashilishi mumkin. Masalan: – *Shu bilganim o'zimga yetadi, chuchvaradan maqsad go'sht bilan xamir yeyish bo'lsa ikkovi ham* ***turibdi,*** *qo'shib yeyaveraman! (A.Qahhor)*

To'rtala fe'l ham hozirgi zamon ma'nosida o'z tarkibiga bo'lishsizlik shaklini qabul qilmaydi. Agar fe'l tarkibida bo'lishsizlik qo'shimchasi bo'lsa, demak, u o'tgan zamon ma'nosini anglatayotgan bo'ladi.

Masalan: *Kutib turibman* – hozirgi zamon ma'nosi; *Kutib turmabman* – o'tgan zamon ma'nosi.

Demak, bu fe'llarning -ib shakli bilan qo'llanilishida hozirgi zamon ma'nosi yetakchilik qiladi. Ayni shakldan o'tgan zamon ma'nosining anglashilishi juda kam hollarda bo'lib, bu xususida J. Abdullajonova shunday deydi: "Bu fe'l formalarining o'tgan zamon eshitilganlik ma'nosi og'zaki nutq uchun xarakterli bo'lib, badiiy adabiyotda uchramaydi"[Абдуллажонова, 1993:51]. Bular asosida xulosa qilish mumkinki, *turmoq, yurmoq, yotmoq, o'tirmoq* fe'llari holat fe'li ma'nosida -ib shakli bilan asosan hozirgi zamon ma'nosini anglatadi, ayrim hollarda ularda o'tgan zamon ma'nosining ifodalanishi esa kontekstdagi maxsus vositalar ta'sirida zamonning ko'chishi sifatida baholanadi. Masalan: *Ko'zlarimni ochsam, ToshMIda* ***yotibman****, ana azob… (kun.uz)* Bu misolda kontekstdagi *ko'zlarimni ochsam* birikmasi ta'sirida *yotibman* fe'li o'tgan zamon ma'nosini anglatgan.

E'tiborli jihati, bu fe'llar o'tgan zamonning boshqa shakllari -di, -gan kabilarni olganda ham fe'l semantikasiga bog'liq ravishda o'tgan zamondagi davomiylik ma'nosini ifodalaydi. Masalan: *Bir kuni besh-oltita ochiq xotin orasiga kirib qolib, ne vaqtgacha ko'nglim g'ash, ta'bim kir bo'lib* ***yurdim.*** *(A.Qahhor) Shu ma'noda, aytish mumkinki, umummilliy uyg'onishning har ikkalasi poydevorida ham tarix, etnografiya, manbashunoslik, san'atshunoslik, numizmatika, arxeologiya singari sohalar yetakchilar qatorida* ***turgan.*** *(xs.uz)*

yotmoq, turmoq, o'tirmoq, yurmoq fe'llarining -(i)b shakli bilan ham hozirgi zamonni, ham o'tgan zamonni voqelantira olishi avtomatik tahlil qilishda muammoni yuzaga keltiradi. Bu muammo korpus matnlaridagi fe'llarning semantik jihatdan harakat va holat ma'nosini bildirishiga

ko‘ra teglanganligi bilan bog‘liq holda yechiladi. *yotmoq, turmoq, o‘tirmoq, yurmoq* fe’llari harakat ma’nosini bildirganda -(i)b shakli bilan o‘tgan zamon ma’nosini, holat ma’nosini bildirganda esa hozirgi zamon ma’nosini anglatadi. Ushbu fe’llarning hozirgi zamonni ifodalashi semantikasida mavjud davomiylik semasiga aloqadorlikda yuzaga keladi. Shuningdek, bu fe’llar ko‘makchi fe’l sifatida qo‘llanganda ham holat ma’nosida bo‘ladi va bunda har doim hozirgi zamon ma’nosini ifodalaydi. Buni quyidagicha modellashtirish mumkin:

Agar *yotmoq, turmoq, o‘tirmoq, yurmoq* fe’llari harakat fe’liga mansub bo‘lsa, ularning tarkibidagi -ib shakli o‘tgan zamonni anglatadi.

F=Harakat fe’li={yot/tur/o‘tir/yur}+↓N+ib+Sh2 bo‘lsa, “ib=Z1(2,al2)”

Shuningdek, bu fe’llar tarkibida -ma bo‘lishsizlik shakli mavjud bo‘lsa, o‘tgan zamon ma’nosi anglashiladi.

F{yot/tur/o‘tir/yur}+↓N+BB2+b+Sh2 bo‘lsa, u holda “b=Z1(2,al1)”

Agar *yotmoq, turmoq, o‘tirmoq, yurmoq* fe’llari holat fe’liga mansub bo‘lsa, u holda ularga qo‘shilgan -ib zamon ko‘rsatkichi hozirgi zamonni anglatadi, ya’ni:

F=Holat fe’li={yot/tur/o‘tir/yur}+↓N+ib+Sh2 bo‘lsa, “ib=(Z2)”.

Agar *yotmoq, turmoq, o‘tirmoq, yurmoq* fe’llari ko‘makchi fe’l bo‘lsa, u holda ularga qo‘shilgan -ib zamon ko‘rsatkichi hozirgi zamonni anglatadi, ya’ni:

F=Ko‘makchi fe’l={yot/tur/o‘tir/yur}+↓N+ib+Sh2 bo‘lsa, “ib=Z2”.

Yot, tur, o‘tir, yur fe’llarining -ib shakli bilan hozirgi zamonni anglatishidagi yana bir o‘ziga xos jihat shundaki, zamon shaklidan oldin -ma bo‘lishsizlik qo‘shimchasi ishlatilmaydi. Agar -ma ishlatilsa, fe’ldan hozirgi zamon emas, o‘tgan zamon ma’nosi anglashiladi.

Yot+ib+shaxs-son, tur+ib+shaxs-son, o‘tir+ib+shaxs-son, yur+ib+shaxs-son shaklidagi fe’llarning korpus matnlaridagi qo‘llanilishida garchi hozirgi zamon ma’nosi ustuvor bo‘lsa-da, avtomatik annotatsiyalash jarayonida yuqori natijaga erishish uchun kam miqdorda bo‘lsa-da, bu shakldan anglashiladigan o‘tgan zamon ma’nosiga ham e’tibor qaratish kerak. Demak, korpus matnlaridagi fe’llar semantik jihatdan harakat va holat fe’llariga ajratilgandagina yuqoridagi modellar natija beradi va yuqoridagi fe’llarning -ib zamon shakli bilan o‘tgan yoki hozirgi zamon ma’nosini anglatayotganligini to‘liq farqlab bera oladi.

Umuman olganda, *yurmoq, turmoq, o‘tirmoq, yotmoq* holat fe’llari semantikasida davomiylik ma’nosi mavjudligi sababli yana qo‘shimcha davomli zamon shakllarini qabul qilmaydi. Bu fe’llarning -ib shakli bilan hozirgi zamonni anglatishi ham leksik ma’nosidagi davomiylik semasi ta’sirida yuzaga keladi. Avtomatik annotatsiyalashda dastur grammatik shaklni qoidaga ko‘ra tasnifda berilgan zamon ma’nosi bilan izohlaydi. Natijada annotatsiyada xatolik kuzatiladi. Zamonni ifodalovchi grammatik shakllarning korpus matnlarida anglatgan turli ma’nolarini statistik o‘rganish va tavsifda o‘rganish natijalariga tayanish shakl haqida nisbatan ishonchli ma’lumot beradi. Tahlillar zamonlarning ko‘chishi bilan bog‘liq misollardagi umumiyliklarni aniqlash va bu asosda ularning modellarini ishlab chiqish orqali bu muammoni qisman hal etish mumkinligini ko‘rsatdi. Omonim zamon grammatik shakllarining zamon ma’nosini saralashda ularning so‘zdagi pozitsiyasidan, polisemantik zamon shakllarining ma’nosini aniqlashda esa standart konstruksiyalar tarkibida uchrashidan kelib chiqib morfotaktik qoidalar asosida modellar tuziladi.

Foydalanilgan adabiyotlar ro‘yxati:

1. Abduraxmonova N. Kompyuter lingvistikasi. – Toshkent: Nodirabegim, 2021. – B. 303.
2. Абдуллажонова Ж. Феъл замон формаларининг услубий хусусиятлари. –Тошкент: Фан, 1993. – Б. 51.
3. Ҳожиев А. Ҳозирги замон феъллари // Феъл замонлари. – Тошкент: Фан, 1962. – Б. 51–68.
4. Ҳожиев А. Феъл. – Тошкент: Фан, 1973. – Б. 156-157.
5. Калонова Д. Ўзбек тилида компьютер луғатчилиги. // Ўзбек тили ва адабиёти. – 2012. – №3. – Б. 97.

СОЗДАНИЕ ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКОЙ МОДЕЛИ КАЗАХСКОГО ТЕКСТА

Тоқмырзаев Дархан
Программист
Казахский национальный унивкерситет имени аль-Фараби

Жұбанова Еңлік
Программист
Международный университет информационных технологий

Пірманова Айнұра
Математик-статист, PhD докторант
Казахский национальный унивкерситет имени аль-Фараби

Аннотация: В настоящее время лингвистическая статистика переходит от первоначального этапа описания языковых явлений к разработке теорий, способных количественно предсказывать языковые закономерности. Основная задача исследования заключается в анализе математического распределения языковых единиц в текстах, что является ключевым для моделирования языковых феноменов. В рамках данной работы было исследовано статистическое распределение основных частей речи в текстах на казахском языке, включая тысячи словоупотреблений из романа М. Ауэзова «Путь Абая» и газетных текстов. Используя методы математической статистики, такие как нормальное распределение, распределение Пуассона и распределение Шарлье, были проверены и подтверждены теоретические закономерности распределения данных частей речи. Эксперименты показали значительные особенности в распределении языковых элементов, что способствует дальнейшему пониманию структуры языка и его функционирования в различных контекстах.

Ключевые слова: лингвистическая статистика, математическое распределение, тексты на казахском языке, части речи, распределение Пуассона, Нормальное распределение, моделирование языка.

Annotation: Currently, linguistic statistics is moving from its initial stage of describing linguistic phenomena to developing theories capable of quantitatively predicting language patterns. The main task of the study is to analyze the mathematical distribution of linguistic units in texts, which is key to modeling language phenomena. This research investigated the statistical distribution of major parts of speech in texts in the Kazakh language, including thousands of word usages from M. Auezov's novel "The Path of Abai" and newspaper texts. Using methods of mathematical statistics such as normal distribution, Poisson distribution, and Charlier distribution, theoretical patterns of distribution of these parts of speech were examined and confirmed. The experiments revealed significant features in the distribution of linguistic elements, contributing to a deeper understanding of the structure of language and its functioning in various contexts.

Keywords: linguistic statistics, mathematical distribution, Kazakh language texts, parts of speech, Poisson distribution, Normal distribution, language modeling.

Лингвистическая статистика сейчас завершает свой первоначальный этап, а именно описание лингвистических явлений, и переходит к следующему, более сложному этапу. Это — проблема создания теории, которая могла бы предсказывать языковые закономерности в количественном измерении.

Таким образом, определение математического распределения языковых единиц в тексте становится одной из наиболее важных задач при создании модели языкового явления [Фрумкина Р., 1962: 124-133].

По данной проблеме были проведены исследования, касающиеся проверки соответствия статистического распределения частей речи в текстах на казахском языке некоторым теоретическим закономерностям. В качестве теоретических закономерностей распределения рассматривались «нормальное распределение», «распределение Пуассона» и «распределение Шарлье» в математической статистике. В качестве объектов исследования были выбраны тексты: 10 тысяч словоупотреблений из романа М. Ауэзова «Путь Абая» и 50 тысяч словоупотреблений из газетного текста. Было проверено статистическое распределение наиболее основных частей речи, таких как существительные, глаголы, прилагательные и наречия, в этих текстах. Важным условием исследования становится определение минимального объема текста, при котором не нарушается закономерность распределения конкретного лингвистического элемента (например, определенной части речи) при его использовании в языке.

Эксперимент, проведенный с такой целью, представлен следующим образом:

Разделение слов в тексте по частям речи и их соответствие условным обозначениям;

Разделение общего объема текста на микрочасти, состоящие из 25, 50, 100, 200 и 500 словоупотреблений, то есть на серийные выборочные части;

Соблюдение переменных мер: общий объем текста по каждому стилю – N; объем микрочасти (серии) – k; количество микрочастей на каждую серию – n;

Принятие обобщенной записи каждого случая – B =k (знак равенства условный). Здесь N=n. Например, если рассмотреть текст романа "Путь Абая", состоящий из 10000 словоупотреблений, разделенный на три части: N1=2500, N2=5000, N3=10000, и если каждый Ni разделить на микрочасти объемом k=25, 50, 100, 200, 500 словоупотреблений (n), их количество будет меняться. Например, если k=25, то: N=2500, то B=100 частей; N=5000, то B=200 частей; N=10000, то B=400 частей. Аналогично, если k=50, то B=50; B=100; B=200 или если k=100, то B=25; B=50; B=100. Текст газеты, выбранный для эксперимента, состоящий из 50000 словоупотреблений, также был разделен и исследован по серийным частям, как показано выше;

Для каждого варианта была подсчитана частота необходимой части речи в микрочастях, и было определено количество серий, в которых часть речи встречается с одинаковой частотой. На основе этих данных был создан следующий дискретный вариационный ряд распределения, состоящий из частоты и количества микрочастей:

x_i	x_1	x_2	x_3	…	x_s
n_i	n_1	n_2	n_3	…	n_s

Здесь xi (i=1,2,3,...,s) - частота определенной части речи, ni (i=1,2,3,...,s) - количество микрочастей, соответствующих каждой частоте xi, n = сумма всех ni. $\sum_{i=1}^{s} n_i$ =n.

По результатам проведенного эксперимента изучалась степень соответствия вариационных рядов распределения основных частей речи в текстах художественной литературы и газетных материалов с некоторыми теоретическими закономерностями. В качестве теоретических закономерностей были рассмотрены законы распределения Пуассона, Шарлье (типы А и В), а также нормальное и логарифмически-нормальное распределения.

В качестве примера создания дискретного и непрерывного вариационного ряда рассмотрим выборочную часть текста романа "Путь Абая", состоящую из N=10000 словоупотреблений. Предположим, что длина внутренних частей (микровыборок) или серийных выборок составляет k=100 словоупотреблений, что делает количество серий B =k=B =100. Далее, определим частоту использования существительных в каждой серийной части текста, и предположим, что минимальное количество употреблений – 19, а максимальное – 43. На основе этого интервала (19; 43) можно создать следующий дискретный вариационный ряд, увеличивая каждую частоту (число использований) на "1" [Джубанов А., 1987: 131]:

x_i	19	20	21	22	23	24	25	…	40	41	42	43	Сумма ni:
n_i	1	0	2	3	3	3	5	…	1	0	0	1	100

Для перехода от дискретного вариационного ряда к непрерывному вариационному ряду выбирается шаг равный Δx, и формируется интервальный ряд (xi; xi+Δx). Арифметическое среднее граничных значений интервала обозначается как xj′, а соответствующее количество nj′ для каждого интервала можно найти по формулам: xj′ = (xi + xi+Δx) / 2, (1) nj′ = количество элементов в интервале (xi; xi+Δx). (2) Здесь i=1,2,3,...,n — порядковый номер в дискретном ряду, j=1,2,3,...,m — порядковый номер в непрерывном ряду.

Если Δx=3, тогда непрерывный вариационный ряд может быть представлен следующим образом (пример в Таблице 1):

Таблица 1. Пример непрерывного вариационного ряда

Интервал (xi; xi+3)	Арифметическое среднее xj′	Количество nj′
(19; 22)	20.5	количество_1
(22; 25)	23.5	количество_2
(25; 28)	26.5	количество_3
...	...	...
(40; 43)	41.5	количество_n

Здесь количество_1, количество_2, ... количество_n — это значения nj′ для каждого интервала, вычисленные в соответствии с указанными формулами. Эти данные помогут проиллюстрировать как частоты определенной части речи распределяются в тексте, что существенно для анализа с помощью теоретических закономерностей, таких как законы распределения Пуассона и Шарлье, а также нормального и логарифмически-нормального распределений.

Таким образом, сформированные дискретные и непрерывные вариационные ряды помогают определить эмпирические закономерности распределения частот частей речи в тексте. Эти данные необходимы для использования в специализированных компьютерных программах, которые позволяют вычислять теоретические частоты и статистические параметры.

Давайте теперь рассмотрим формулы для двух типов теоретических законов распределения, которые мы обсуждали ранее.

Закон распределения Пуассона

Основным теоретическим законом в лингвистических исследованиях является закон распределения Пуассона. Этот закон часто используется в экспериментах, когда вероятность каждого отдельного события (p) очень мала, а количество экспериментов очень велико.

Можно предположить, что эти условия также выполняются для частот распределения частей речи в текстах на казахском языке.

Для использования закона Пуассона, как уже упоминалось выше, нам необходимо создать дискретный вариационный ряд на основе статистических данных о частях речи. В эксперименте, определяя частоту возникновения желаемого события, например, встречаемость существительных (x), "математическое ожидание" (a) или его арифметическое среднее и количество экспериментов (n), формула закона Пуассона записывается следующим образом [Смирнов Н., 1969: 280]:

$$n^T = \frac{a^x \cdot e^{-a}}{x!} \cdot n. \qquad (3)$$

В этом (3) уравнении формула арифметического среднего, которая заменяет "математическое ожидание", записывается следующим образом:

$$a \approx \bar{x} = \frac{1}{n} \cdot \sum n_i \cdot x_i' \ . \qquad (4)$$

Закон нормального распределения

Можно сказать, что закон нормального распределения широко используется в статистических исследованиях. Основная причина этого — простота графической формы этого закона и его близость по форме к многим «одногорбым» распределениям. Еще одно преимущество закона нормального распределения заключается в обширности опыта математическо-статистической обработки данных во время экспериментальных наблюдений. Одним из важнейших аспектов закона является возможность его

использования для определения объема части текста и статистической оценки (тестирования).

Переменная и непрерывная величина x принимает значения в интервале от минус бесконечности до плюс бесконечности. Распределение переменной "x" по нормальному закону выражается следующим образом:

$$n^T = \varphi(x) \cdot n = \frac{n \cdot \Delta x}{\sigma} \cdot \frac{1}{\sqrt{2\Pi}} \cdot \ell^{-\frac{(x-a)^2}{2\sigma^2}}. \qquad (5)$$

Здесь x — длина интервала, необходимая для перехода от дискретного вариационного ряда к непрерывному ряду, а — "математическое ожидание" (определяется уравнением 4, упомянутым выше). 2σ2 — это дисперсия, также называемая "центральным моментом второго порядка", и её математическая формула выглядит следующим образом:

$$\sigma^2 = \mu_2 = \frac{1}{n-1} \cdot \sum_i x_i \cdot (x_i - a)^2. \qquad (6)$$

Математическое ожидание (a) и дисперсия (σ^2) являются ключевыми статистическими параметрами, описывающими законы распределения. Например, с их помощью можно определить рассеянность значений переменной x и форму кривой распределения на графике. В практических исследованиях для определения степени соответствия полученных данных теоретическому "нормальному распределению" важно знать параметры "асимметрии" и "эксцесса":

Асимметрия (Skewness) - показывает степень асимметрии распределения данных относительно среднего значения. При нормальном распределении ассиметрия равна нулю. Положительное значение указывает, что распределение имеет длинный правый хвост, а отрицательное - что хвост длиннее слева.

Эксцесс (Kurtosis) - мера остроты пика распределения данных. Для нормального распределения эксцесс равен 3. Значения эксцесса, превышающие 3, указывают на более острый пик по сравнению с нормальным распределением, а значения меньше 3 - на более плоский.

Эти параметры помогают более точно оценить, насколько эмпирическое распределение данных соответствует нормальному распределению, что важно для многих статистических тестов и анализа данных.

$$A = \frac{\mu_3}{\sigma^3}, \qquad (7)$$

$$E = \frac{\mu_4}{\sigma^4} - 3. \qquad (8)$$

Здесь σ обозначает стандартное отклонение, 3μ3 и 4μ4 — это центральные моменты третьего и четвёртого порядков соответственно:

Стандартное отклонение (σ) — это мера разброса значений вокруг среднего (математического ожидания). Оно показывает, насколько широко значения варьируются от среднего значения распределения. Рассчитывается как квадратный корень из дисперсии.

Третий центральный момент (3μ3) — используется для оценки асимметрии распределения. Положительное значение 3μ3 указывает на асимметрию распределения с длинным правым хвостом (правосторонняя асимметрия), отрицательное значение — с длинным левым хвостом (левосторонняя асимметрия).

Четвёртый центральный момент (4μ4) — используется для оценки эксцесса распределения. Это мера того, насколько плоским или острым является пик распределения по сравнению с нормальным распределением. Если 4μ4 больше, чем у нормального распределения, пик будет более выраженным, а если меньше — более плоским.

Эти параметры важны для полного характеристического описания распределения, особенно в статистическом анализе, где точное понимание формы распределения может влиять на выводы исследования.

$$\mu_3 = \frac{1}{n} \cdot \sum_i n_i \cdot (x_i - a)^3 \ , \qquad (9)$$

$$\mu_4 = \frac{1}{n} \cdot \sum_i n_i \cdot (x_i - a)^4 \ . \qquad (10)$$

Для нормального распределения значения третьего центрального момента (асимметрия, A) и четвёртого центрального момента (эксцесс, E) равны нулю. Поэтому, если значения асимметрии (A) и эксцесса (E) для анализируемого распределения также близки к нулю, можно предположить, что эмпирическое распределение приближается к нормальному. Наоборот, чем больше отклонения значений A и E от нуля, тем меньше соответствие между экспериментальным распределением и нормальным законом распределения.

Для более глубокого изучения теоретических законов распределения, таких как Шарлье типов A и B, а также логарифмически нормального распределения, можно обратиться к следующей литературе:

Статистические методы анализа и обработки данных - этот источник может предоставить подробные объяснения и примеры использования различных типов распределений, включая их применение в реальных исследованиях.

Теория вероятностей и математическая статистика - книга, где обсуждаются основы теории вероятностей, включая различные типы распределений и их свойства.

Прикладная статистика и основы метрологии - источник, который может быть полезен для понимания применения статистических методов в научных и инженерных задачах, включая анализ распределений [Романовский В., 1963: 94], [Хальд Л., 1956: 61].

Эти ресурсы помогут получить более глубокое понимание различных распределений и методов их анализа в контексте прикладной статистики и данных.

Список использованных источников:

1. Джубанов А.Х. Квантитативная структура казахского текста (опыт лингвистического анализа на ЭВМ). Алма-Ата, 1987. С. 147.
2. Романовский В.И. Математическая статистика. Ташкент, 1961. Кн. 1. 637 с.; 1963. Кн. 2. 794 с.
3. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностной и математической статистики. М., 1969. 511с.
4. Фрумкина Р.М. О законах распределения слов и классов слов // Структурно-типологические исследования. М., 1962. С. 124-133.
5. Хальд Л. Математическая статистика с техническим приложением. М., 1956. 644 с.

MATN TOZALASH TIZIMI DASTURLARINING UMUMIY TAVSIFI. AFZALLIKLARI VA KAMCHILIKLARI

Baxodirov Sanjarbek Rahmatali o‘g‘li
ToshDO‘TAU, kompyuter lingvistikasi mutaxassisligi magistranti.
E-mail: sanjarbahodirov9901@gmail.com

Annotatsiya: Matnni tozalash dasturi matn ma’lumotlarini samarali tozalash va tartibga solish uchun mo‘ljallangan kuchli vositadir. Murakkab algoritmlari va interfeysi bilan u keraksiz belgilarni olib tashlash, ortiqcha bo‘shliqlarni yo‘qatish, matnni formatlash, takroriy jumlalarni bartaraf etishga qodir. Ushbu dasturlar katta hajmdagi matn ma’lumotlarini osonlik bilan boshqarishi mumkin. Bu marketing, tadqiqot va ma’lumotlarni tahlil qilish kabi turli sohalarda foydalanish uchun juda qulay hisoblanadi. Matnni tozalash dasturi ko‘p tarmoqli vosita bo‘lib, uni shaxsiy imtiyozlar va talablarga mos ravishda sozlash mumkin. U ommaviy ishlov berish, moslashtirilgan filtrlar va tozalangan matn ma’lumotlarini turli formatlarda saqlash va jo‘natish funksiyalarini o‘z ichiga olgan bir qator xususiyatlarni taklif etadi. Matnni tozalash dasturi tizimining qulay dizayni va mustahkam funksionalligi matnni qayta ishlash jarayonini soddalashtiradi va har bir foydalanuvchiga qiynalmasdan foydalanish imkoniyatini yaratadi. Bundan tashqari, ushbu dasturlarni avtomatlashtirish orqali foydalanuvchilarning qimmatli vaqtlari tejaladi hamda hujjatlarida izchillik va aniqlik ta’minlanadi.

Kalit so‘zlar: vaqtni tejash, avtomatlashtirish, html teglari, URL manzillari, agressiv tozalash, ma’lumotlarni yo‘qalishi, ommaviy ishlov berish.

Abstract: This is a powerful tool for effectively cleaning and organizing text data. It can remove unnecessary markers and eliminate excessive gaps, format the text, and remove repetitive sentences. These tools can easily manage large amounts of text data. Marketing is very useful in various fields such as research and data analysis. Text cleaning software is a versatile tool that can be customized to personal preferences and requirements. It offers public use, customized filters, and functions for storing and sending cleaned and formatted text data in various formats. The user-friendly design and robust functionality of the Text Refinement Program simplifies the process of revising text and provides every user with an effortless experience. In addition, automating these programs saves users valuable time and ensures precision and clarity in their documents.

Keywords: time saving, automation, html tags, URLs, aggressive cleaning, data loss, batch processing.

Аннотация: Это мощный инструмент для эффективной очистки и организации текстовых данных. Он умеет удалять ненужные маркеры и устранять лишние пробелы, форматировать текст и удалять повторяющиеся предложения. Эти инструменты могут легко управлять большими объемами текстовых данных. Маркетинг очень полезен в различных областях, таких как исследования и анализ данных. Программное обеспечение для очистки текста — это универсальный инструмент, который можно настроить в соответствии с личными предпочтениями и требованиями. Он предлагает общедоступное использование, настраиваемые фильтры и функции для хранения и отправки очищенных и отформатированных текстовых данных в различных форматах. Удобный дизайн и

надежная функциональность программы улучшения текста упрощают процесс редактирования текста и предоставляют каждому пользователю удобство работы. Кроме того, автоматизация этих программ экономит пользователям драгоценное время и обеспечивает точность и ясность их документов.

Ключевые слова: экономия времени, автоматизация, html-теги, URL-адреса, агрессивная очистка, потеря данных, пакетная обработка.

KIRISH

Ma’lumotlarni tahlil qilish va tabiiy tilni qayta ishlashning jadal rivojlanayotgan davrda matn ma’lumotlarining sifati, undan olinayotgan tushunchalarning aniqligiga bo‘lgan talab oshib bormoqda. Matn tozalash dasturlari keraksiz belgilar, nomuhim so‘zlarni olib tashlash va nomuvofiqliklarni bartaraf etish orqali matn ma’lumotlarini tahlil qilish uchun tayyorlaydi. Ushbu dasturlar matnni qayta ishlash vazifalarining sezilarli darajada oshirish mumkin bo‘lgan qator afzalliklarni taqdim etadi. Biroq, har qanday texnologiya singari, ushbu dasturlar ham ma’lum cheklovlar va foydalanuvchilar bilishi kerak bo‘lgan kamchiliklarga ega.

Matnni tozalash dasturining keng tarqalgan turlaridan biri imlo tekshiruvi bo‘lib, u berilgan matndagi imlo xatolarni aniqlash va tuzatish uchun mo‘ljallangan.Yana bir turi grammatik tekshiruv bo‘lib, foydalanuvchilarga yozishdagi grammatik xatolarni aniqlash va tuzatishga yordam beradi. Boshqa matnni tozalash dasturlari takroriy jumlalar, so‘zlarni o‘chirish, matnni formatlash yoki ahamiyatsiz va ortiqcha ma’lumotlarni aniqlash va o‘chirishga qaratilgan bo‘lishi mumkin.

Matnni tozalash dasturlarining ishlatishning asosiy afzalliklaridan biri shundaki, ular matnli ma’lumotlarning umumiy sifatini yaxshilashga yordam beradi, ularni aniqroq, izchil va tahlil qilishni osonlashtiradi. Ushbu dasturlar, shuningdek, foydalanuvchilarning matndagi xatolarni aniqlash va tuzatish jarayonini avtomatlashtirish orqali ularning vaqt va qo‘l mehnatini tejash imkonini beradi. Biroq,

matn tozalash dasturlari ham ba’zi kamchiliklarga ega. Ushbu dasturlar har doim ham berilgan matndagi barcha xatolarni, ayniqsa, o‘zga tildagi so‘zlar, lahjalar qatnashganda kutilgan natijani bermasligi mumkin. Bundan tashqari, ba’zi matnni tozalash dasturlari aniqlash va tuzatish mumkin bo‘lgan xatolar turlari bo‘yicha cheklovlarga ega bo‘lishi mumkin.

ASOSIY QISM

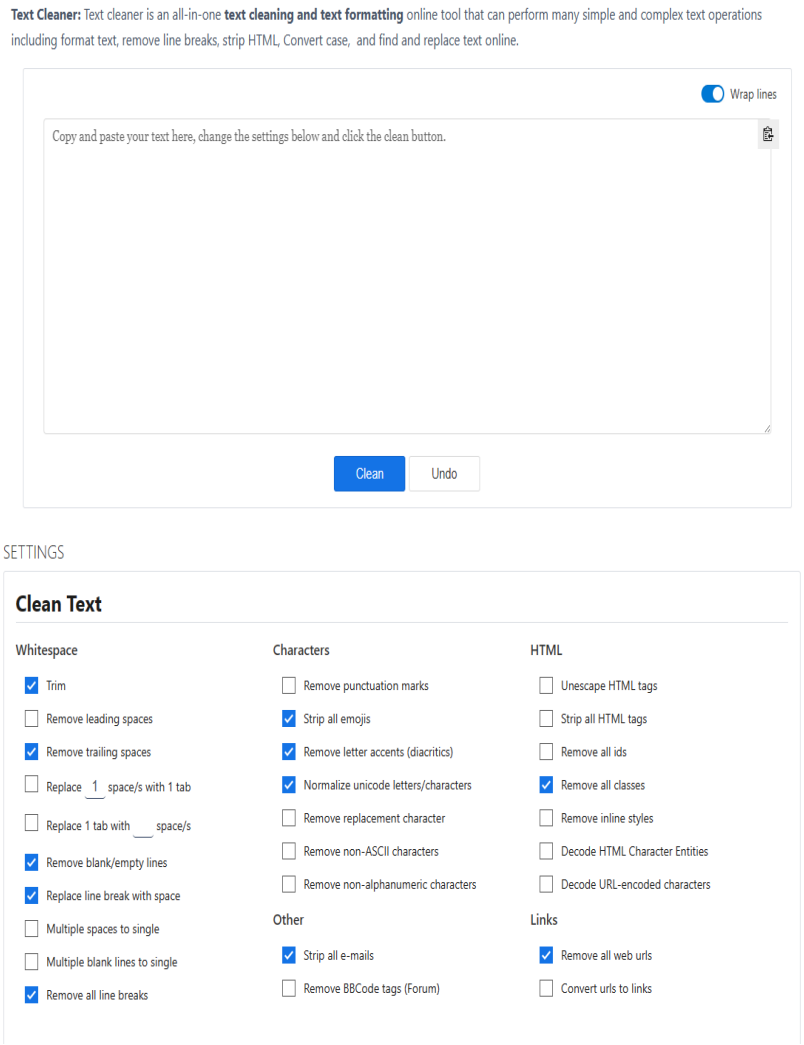
Matnni tozalash dasturlari katta hajmdagi tuzilmagan ma’lumotlar bilan ishlaydigan har bir inson uchun muhim vosita sanaladi. Ushbu dasturlar matnli ma’lumotlarni tahlil qilish, qayta ishlash va tozalash uchun mo‘ljallangan. Matnni tozalash dasturlarining bir necha turlari mavjud bo‘lib, ularning har biri o‘ziga xos xususiyat va imkoniyatlarga ega.

1.1 Text cleaner: Text Formatter, Formatting online. Text cleaner yoki Clean text – bu juda murakkab matnlarga ishlov berish jarayonida bajarishi mumkin bo‘lgan matnni tozalash va formatlash uchun mo‘ljallangan onlayn vosita.

Ushbu dastur ortiqcha bo‘shliqlar va keraksiz belgilarni olib tashlashi mumkin. Shuningdek, u harflarni hajmini o‘zgartirishi, tipogrfiya qo‘shtirnoqlarini o‘zgartirishi, takroriy satrlar,paragraflar va so‘zlarni o‘chirishi, qalin va kursivli Unicode harflarini oddiy harflarga aylantirishi, tinish belgilari va orasidagi masofani tuzatishi, harf urg‘ularini olib tashlashi, belgilarning kodlarini dekodlashi, HTML teglari va URL larni olib tashlashi mumkin. Bu dastur matnni onlayn formatlashda juda moslashuvchan sanalib, shaxsiy ehtiyojdan kelib chiqib

sozlamalarni o‘zgartirish mumkin. Ushbu vositani eng katta afzalliklaridan biri, shaxsiy konfiguratsiyangiz avtomatik ravishda brauzeringizda saqlanadi, shuning uchun keying safar yana tashrif buyurganingizda sozlamalarni to‘liq o‘zgartirishingiz shart emas.

Ushbu yordamchi dasturning asosiy maqsadi, formatlangan matnni formatsizlantirish va matn protsessorlari, veb sahifalar, PDF-fayllar, mijozlar qisqacha ma’lumotlari va elektron pochtalardan to‘g‘ridan-to‘g‘ri ko‘chirilgan matnlarda ko‘pincha mavjud bo‘lgan barcha ma’nosiz belgilarni olib tashlashdir. Ushbu veb-ilova tadqiqot, ishlab chiqish yoki tijorat maqsadlarida har qanday jismoniy shaxs, kompaniya, ofis yoki tashkilot uchun bepul foydalanish mumkin. Mazkur veb ilovani 1-rasmda ko‘rishingiz mumkin:

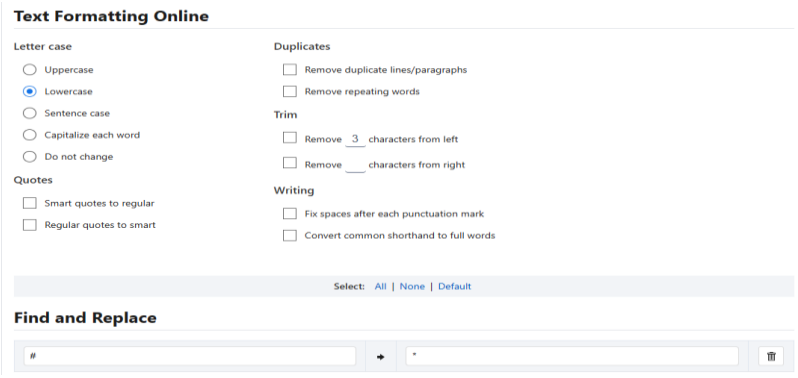


1-rasm. Veb ilovaning ko‘rinishi

Qanday ishlatiladi?

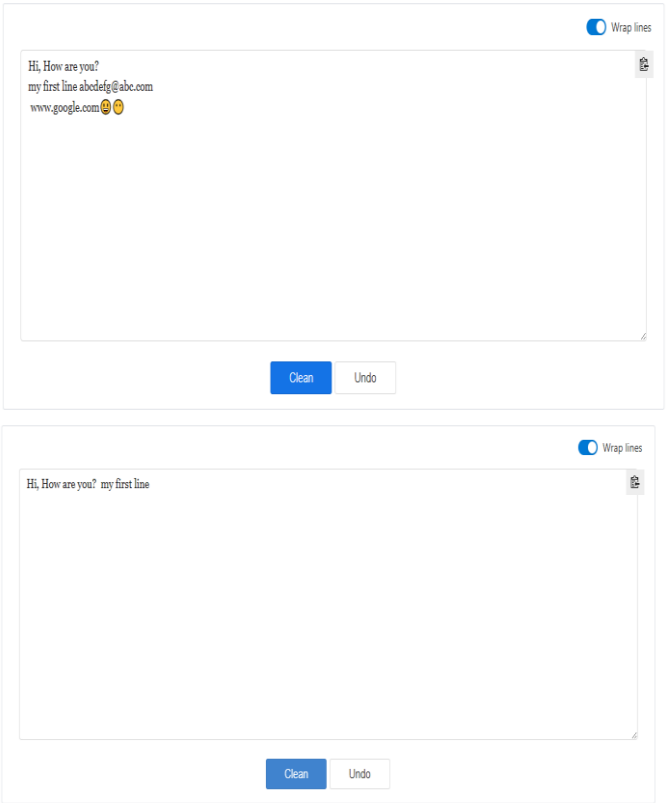
Matnni kiritish maydoniga nusxa ko‘chirasiz va joylashtirasiz, katakchalarni belgilash yoki belgini olib tashlash orqali 1-rasmdagi sozlamalarni sozlash va tozalash (clean) tugmasini bosasiz. Matningizni tozalangan versiyasi natijalar oynasida paydo bo‘lishi kerak. Agarda siz o‘ylagan natija chiqmasa, xavotir olish shart emas, siz orqaga qaytib, kiritish yorlig‘ini bosishingiz mumkin. Bekor qilish(Undo) tugmasi yordamida tozalashda xatolik yuz bersa, qayta ortga qaytishingiz va matnni oldingi ko‘rinishga keltira olasiz. Bundan tashqari, yuqorida

aytilganidek matn formatlarini o‘zgartirish, takroriy jumlalarni xoxishimizga qarab o‘zgartirsak bo‘ladi. Quyida 2-rasmda ko‘rishingiz mumkin.



2-rasm. Matn formatlash.

“Find and Replace” funksiyasi yordamida esa, birinchi jadvalga qidirayotgan belgi yoki so‘zni kiritishimiz va uni o‘zimizga kerakli belgiga alishtirishimiz mumkin. Bu orqali katta hajmli matnlarda yuzaga kelgan xatoliklarni bir soniyada hal etsak bo‘ladi. Quyidagi 3-rasmda **“strip all emojis”, “strip al e-mails”, “remove all web urls”, “replace line breaks with space”** katakchalarini belgilash orqali qisqa ko‘chirilgan matnni tozalab ko‘ramiz. [Zia A, 2024]



3-rasm. Matn tozalash jarayoni

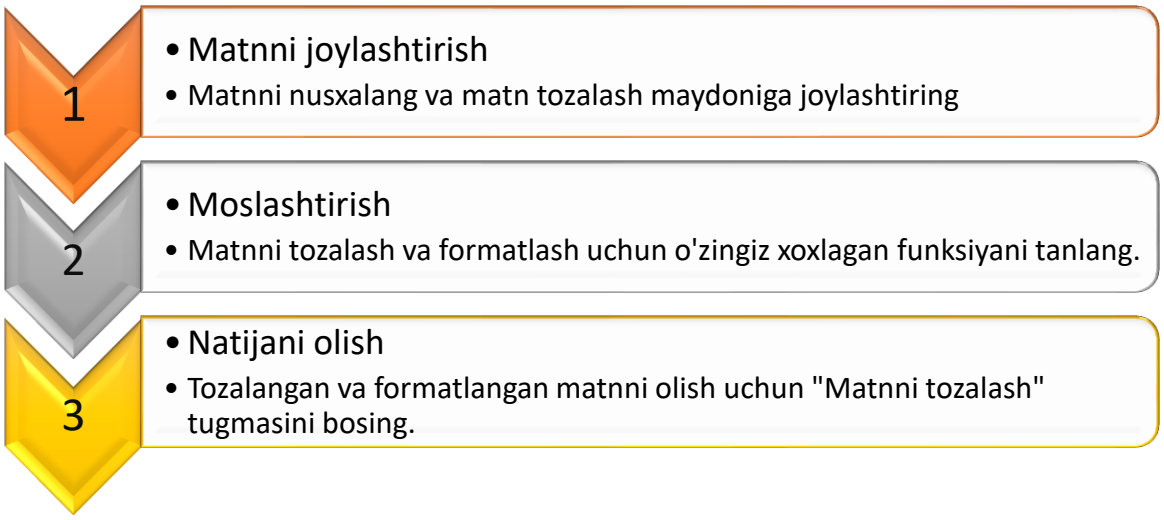
Mazkur dasturni ishlatish davomida hech qanday ishingizga halal beradigan reklamalar uchramaydi, tizimni ishlashini qiyinlashtirmaydi, veb-brauzerdan to‘g‘ridan-to‘g‘ri foydalanishingiz mumkin, hech qanday dasturni o‘rnatish shart emas. Matnni bir necha soniya ichida tozalab beradi.

Asosiy kamchiliklaridan biri, matnni faqat nusxa ko‘chirish asosida kiritish mumkin xolos, fayl(docx,pdf) shunga o‘xshash ko‘rinishlarda joylashtirib bo‘lmaydi. Bu dastur 2021-yilda tuzib chiqilgan va so‘ngi o‘zgartirish 2022-yilda kiritilgan.

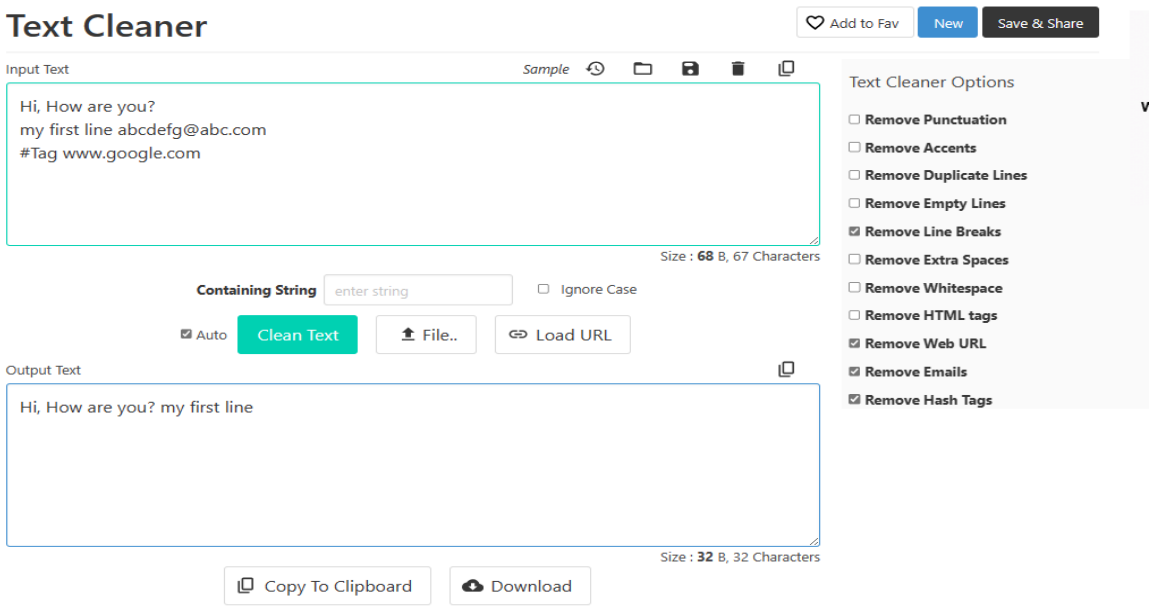
Dastur manzili: https://ravana69.github.io/textcleaner

1.2 Code beautify. Text Cleaner. CodeBeautify’s Text Cleaner – bu matnni tez, oson tozalash va formatlashda yordam beradigan veb ilova. Ushbu dastur matndan tinish belgilari, harf urg‘ulari, takroriy qatorlar, bo‘sh qatorlar, qo‘shimcha bo‘shliqlar, HTML teglari, veb URL manzillari, elektron pochta xabarlari va hash teglarini olib tashlashda yordam beradi. Mazkur ilova matnga ommaviy ishlov berishda yoki ma’lum bir yo‘nalishda ishlov berishda (masalan, faqat URL manzilini tozalash) qulay hisoblanadi. Shuningdek, eng katta afzalliklaridan biri matnni fayl ko‘rinishida hamda nusxa ko‘chirish asosida amalga oshirsa bo‘ladi.

Matnni tozalash vositasidan qanday foydalanish kerak:



Quyidagi 4-rasmda matn tozalash jarayonini ko‘rishingiz mumkin:



4-rasm. Matn tozalash jarayoni

“Text cleaner” matn formatlashni soddalashtiradi, vaqtni tejab, izchillikni taminlasa-da, ba’zi bir kamchiliklarni ham kuzatish mumkin, Misol uchun, matnga asoslangan fayllar bilan

cheklangan, murakkab formatlashda ishlamasligi mumkin va internetga ulanishni talab qiladi. Bundan tashqari, ilovadan foydalanish davomida turli xil reklamalar ish faoliyatiga ta’sir o‘tkazadi [Williams.2024]

Mazkur dastur 2024-yilda tuzib chiqilgan.

Bog‘lanish uchun: support@codebeautify.org

Manzil: https://codebeautify.org/text-cleaner

XULOSA

Xulosa qilib aytadigan bo‘lsak, matnni tozalash tizimlari turli tahliliy jarayonlarda qo‘llaniladigan ma’lumotlarning aniqligi va ishonchliligini ta’minlashda hal qiluvchi rol o‘ynaydi. Ushbu maqolada ko‘rib chiqilgan dasturlar, masalan, emojilar, tinish belgilari, HTML tehlari, URL manzillari, e-mail va boshqa elementlarni olib tashlash, stemming va lemmatizatsiya va imlo tekshiruvi matnli ma’lumotlar sifatini yaxshilashga yordam beradigan matnni tozalash tizimlarining muhim tarkibiy qismlaridir. Matnni tozalash tizimlaridan foydalanishning asosiy afzalliklaridan biri keraksiz va ahamiyatsiz belgilarni olib tashlash orqali ma’lumotlarni tahlil qilish samaradorligini oshirish qobiliyatidir. Bu yaxshi natijalarga va aniqroq tushunchalarga olib keladi. Bundan tashqari, matnni tozalash tizimlari matn ma’lumotlarining umumiy o‘qilishi va tushunarliligini yaxshilashga yordam beradi, bu esa foydalanuvchilarga ma’lumotni tahlil qilishni osonlashtiradi.

Ammo shuni ta’kidlash kerakki, matnni tozalash tizimlari ham kamchiliklardan xoli emas. Kamchiliklarning ba’zilari tozalash jarayonida ma’lumotlarning potentsial yo‘qolishi, shuningdek, natijalarning to‘g‘riligini ta’minlash uchun muayyan holatlarda qo‘lda aralashuv zarurligini o‘z ichiga oladi.

.

Foydalanilgan adabiyotlar:

1. Chandra R. Text Cleaning in Python: Effective Data Cleaning Tutorial, 2023. Text Cleaning in Python: Effective Data Cleaning Tutorial – Kanaries
2. Otten N.V. Top 20 Essential Text Cleaning Techniques [Practical How to Guide in Python] / Data Science, 2023. https://spotintelligence.com/2023/09/18/top-20-essential-text-cleaning-techniques-practical-how-to-guide-in-python
3. Pankaj. How to Remove Spaces from a String in Python / DigitalOcean, 2022. How To Remove Spaces from a String In Python | DigitalOcean
4. Roepke B. How to Clean Text Like a Boss for NLP in Python / Data Knows All, 2024. https://dataknowsall.com/blog/textcleaning
5. Williams A. Remove emails from Text in Python (methods+examples) / PYtutorial, 2023. https://pytutorial.com/remove-emails-from-text-in-python-methods-examples
6. Yeung J.A. How to remove emojis from string in Python /The Web Dev, 2022. https://thewebdev.info/2022/04/16/how-to-remove-emojis-from-a-string-in-python
7. Zia A. How to Remove Whitespaces in Python Strings? /jQuery-AZ /2024. https://www.jquery-az.com/how-to-remove-whitespaces-in-strings-of-python

SLENGLARNI AVTOMATIK TEGLASHDA SLENG BIRLIKLARI LUG‘ATINING AHAMIYATI

Bekmuradova Iroda Zokir qizi
Alisher Navoiy nomidagi
Toshkent davlat o‘zbek tili va adabiyoti universiteti
kompyuter lingvistikasi mutaxassisligi magistranti.
E-mail: irodabekmuradova9@gmail.com

Annotatsiya: Ushbu maqolada sleng tushunchasi, sleng qatnashgan matnlarni teglashda sleng lug‘atlarining ahamiyati hamda bugungi kunda bunday lug‘atlariga bo‘lgan ehtiyojning ortishi kabi masalalar tahlil qilinadi. Sleng lug‘atlarini tuzish, slenglar izohi va boshqa xususiyatarini korpusda matnga teg sifatida biriktirish jarayoniga to‘xtalib o‘tiladi.
Kalit so‘zlar: sleng, noadabiy leksika, korpus, teg, sleng lug‘ati.

Abstract: This article discusses the concept of slang, the issues such as the importance of slang dictionaries and the increasing need for such dictionaries today are analyzed in tagging texts with slang. It focuses on the process of creating slang dictionaries, attaching slang annotations and other features to the text as tags in the corpus.
Keywords: slang, non-literary lexicon, corpus, tag, slang dictionary.

Аннотация: В этой статье анализируется концепция сленга, важность сленговых словарей для разметки текстов сленгом, а также растущая потребность в таких словарях сегодня. Основное внимание уделяется процессу создания сленговых словарей, прикреплению сленговых аннотаций и других функций к тексту в виде тегов в корпусе.
Ключевые слова: сленг, нелитературная лексика, корпус, тег, сленговый словарь.

Bugungi o‘zbek so‘zlashuv uslubida slenglar ko‘p qo‘llanilishi bilan xarakterlidir. Slenglar sotsiolingvistik birlik sifatida ko‘p qo‘llanadigan noadabiy leksik qatlamga oid bo‘lib, doimiy o‘rganishni talab etadigan so‘zlardir, ular adabiy me’yor doirasiga kirmasa-da, yozma nutqda kamdan-kam ishlatilsa ham, kundalik hayotda ko‘p qo‘llaniladi. Shu sababli, so‘zlashuv nutqining muhim va ajralmas qismi hisoblanadi. Slenglar til foydalanuvchilarning barcha toifasi uchun birday tushunarli bo‘lolmaydi, chunki, bunday birliklar ko‘p hollarda qisqalik hamda aytilmoqchi bo‘lgan fikrni boshqalardan yashirish maqsadida ishlatiladi.

O‘zbek tilshunosligida sleng birliklar hali yetarli darajada tadqiq etilmagan, shuning uchun mavzuga doir keng sotsial tadqiqotlar, slenglar lug‘atini yaratish kabi masalalar yangi ilmiy izlanishlarga muhtoj. Slengni ishlatish juda oson bo‘lganidek, bir qarashda tushunish ham soddaga o‘xshab ko‘rinadi. Ammo eng oddiy narsalarni aniqlash, albatta, muhokama qilish eng qiyin, chunki sleng so‘zlashuv nutqiga xos bo‘lganligi hamda uni qo‘llashdan maqsad ilmiy qonunlar, grammatik qoidalar va falsafiy nazariyalarga asoslanmagani, qulaylik uchun ishlatilishi sababli bu jarayonni izohlash biroz murakkab. U qanday paydo bo‘lsa, shundayligicha so‘zlashuv nutqida yashaydi. [Eric Partridge, 1933:20]

O‘zbek tilshunosligidagi sotsiolingvistik tadqiqotlarda slenglar ko‘proq yoshlarning nutqida uchraydigan o‘ziga xos yashirin ma’noga ega bo‘lgan leksik birliklar sifatida e’tirof

etiladi. Shu jihatdan qaraganda bugungi kunda yoshlar o‘rtasidagi muloqot jarayonida keng qo‘llaniladigan ba’zi chet so‘zlari, ko‘chma ma’noda qo‘llanilib, ommalashtirilgan ayrim o‘zbekcha so‘zlar, iboralar va o‘xshatishlar slenglarni tashkil etadi.

Odatda, slenglar, ayniqsa, ularning dastlabki shakllanishi davrida, keng ommaga tushunarsiz bo‘ladi. Til korpusida sleng birliklarni ajratib tahlil qilish, ularga avtomatik teg qo‘yishning ahamiyati shundaki, bunday birliklar qatnashgan matnning u yoki bu tahlili amalga oshirilganda slenglarni oddiy leksik birliklar kabi tahlil qilib ketish, avtomatik tarjima jarayonida so‘zni to‘g‘ridan-to‘g‘ri belgilangan tilga tarjima qilib qo‘yish kabi xatoliklarni to‘xtatishga erishish mumkin. Slenglarni teglash bunday birliklarga izoh berish orqali amalga oshiriladi. Korpusda birliklarni teglash muayyan maqsadda to‘plangan matnlar majmui, til birliklari yig‘indisi, tabiiy tildagi elektron shaklda saqlanadigan yozma yoki og‘zaki, kompyuterlashtirilgan qidiruv tizimiga dasturiy ta’minot asosida joylashtirilgan matnlar jamlanmasi orqali amalga oshiriladi. Undagi matnlar tilni ifoda etish, tilning rivojlanish bosqichlari, janrlari, uslublari, hududiy va ijtimoiy variantlarining barcha xilma-xilligini aks ettiradi. [Hamroyeva Sh, 2020:52]

So‘zlarni teglashning lug‘atga asoslangan yondashuvi quyidagi bosqichlarni o’z ichiga oladi:



1-rasm. Slenglarni avtomatik teglashning lug‘atga asoslangan usuli bosqichlari

Matndagi birliklarni teglashning lug‘atga asoslangan usuli matn ma’lumotlariga teglar yoki toifalarni belgilash uchun oldindan yig‘ilgan so‘zlar to‘plami (lug‘atlar)dan foydalanishni o‘z ichiga oladi. Bunda qilinishi kerak bo‘lgan eng birinchi ish mavjud sleng birliklarni to‘plab, lug‘at shakllantirishdir. Bunda lug‘atdagi so‘zlarning muayyan toifalari bilan bog‘liq so‘z va iboralar ro‘yxati yaratiladi. Bunday lug‘at qo‘lda tuzilishi yoki “text-mining” usuli yordamida yaratilishi mumkin.

Sleng	**So‘z turkumi**	**Izohi**	**Misol**	**Sinonimi**
Strelka	Ot	*Uchrashuv*	1. Strelka belgiladi. [og‘zaki] 2.Strelkani soat 5 ga, “Anhor” kafesiga urdim. [og'zaki] 3.Namuncha xursandsan?! Strelkang bormi deyman [og‘zaki] 4.Turishingni qara, shu ahvolda strelkaga chiqasanmi? [og‘zaki] 5.Strelka restoranda. [og‘zaki]	–

			6.Strelkami deyman - dedi Qamariddin ko'zini ma'noli qisib. [T.Malik, “Alvido, bolalik”] 7. Shoshyapsanmi, strelkang nechchida? [og‘zaki] 8.Kecha strelkada boshqacha eding? [og‘zaki] 9.Strelkam bor. [filmdan] 10.Strelkadan chiqib to'g'ri uyga keldim. [og‘zaki]	
Qovun	Ot	*Biror narsani eplolmaydi gan, bir aytgandan tushunmaydigan*	1.Qovunmisan, nima balo?! [og‘zaki] 2.G'irt qovun bola. [og‘zaki] 3. Qovunsan-da, o‘rtoq [yoshlar suhbatidan] 4. O‘zingni qovunga solma, hammasini tushuntirib turibsan-ku [og‘zaki] 5. Voy, Sabriya, qovunsan, qovun! [ijt.tarmoq]	Chuchvara, guvala, chumbul
Nol besh	Son	*Bo'yi past, pakana*	1. “Nol besh” kelyapti. [Yoshlar suhbatidan] 2. O‘zi “nol besh-u”, osilgan dorini qarasin . [Og‘zaki] 3. “Nol besh”lar “bir”larga qarshi! [og‘zaki] 4. Shu “nol besh” bolaning nimasi yoqadi senga, hayronman?! [ijt.tarmoq] 5. O‘rtadagi “nol besh”ni bir yoqlik qil. [og‘zaki] 6. O‘zi “nol besh” bo‘lsa ham mard ekan. [og‘zaki] 7. O‘sha “nol besh”ga ayt, meni tinch qo'ysin. [ijt.tarmoq] 8. Ko‘p og‘ir yuk ko‘tarma, bunaqada “nol besh” bo‘lib qolasan. [og‘zaki suhbatdan]	–
Uxlatmoq	Fe’l	*Aldab ketmoq, chuv tushirmoq*	1. MIBning soxta xodimi fuqaroni 31 000 dollarga “uxlatdi”. 2. Akam seni “uxlatib” ketibdi. [og‘zaki suhbatdan] 3. Shuncha pulga “uxlabsan”, jo‘ra! [ijt.tarmoq] 4. Ana endi aniq uxlading [og‘zaki suhbatdan]	Kiydirib ketmoq
Kiydirmoq	Fe’l	Boplamoq	1. To‘g‘ri, til bilgan el bilar, ammo ona tilimiz darajasini eplab ko‘tarolmayotgan bir vaziyatimizda bu xorijiy tillar targ‘ibotini “kiydirib” qo‘yish qanchalik ma’qul?! [xs.uz]	–
Kiydirib ketmoq	Fe’l	Aldab ketmoq, chuv tushirmoq	1. “Qo‘shnisi kiydirib ketsayam yo‘lli to‘svolish qattan chiqdi o‘zi?” deb yozadi Bemalolxo‘ja. [bugun.uz] 2. – Nima bo'ldi? – Kiydirib ketdi [filmdan]	Uxlatmoq
Guvala	ot	Hech narsani tushunmaydigan, biror ishni eplolmaydi gan	1. – Sen bolani onang, aslida, insofsiz qilib tuqqan. Bolaligingda ham bizlarga xiyonat qilib turarding. Shoyusuf guvalaning bog‘i esingdami? Cho‘tir, biz boqqa tushganimizda bu nima qilardi, a? [saviya.uz]	Qovun, chuchvara, chumbul

2-rasm. Sleng birliklari lug‘atida so‘zlar izohi

Matn ma’lumotlarini lug‘atga asoslangan holda teglash uchun kiritilgan matn lug‘atdagi so‘z yoki iboralarning mavjudligi uchun tekshiruvdan o‘tkaziladi. Agar moslik topilsa, matnga

mos teg tayinlanadi. Slenglarni teglashda lug‘atga asoslangan yondashuv boshqa usullarga qaraganda nisbatan oson va muayyan kalit so‘zlar kerakli teglarni yuqori darajada ko‘rsatadigan matn ma’lumotlarining ayrim turlari uchun samarali bo‘lishi mumkin, ammo lug‘atga asoslangan usul o‘ziga xos tarzda ayrim kamchiliklarga ega. Ulardan biri shundaki, tahlil natijasi lug‘atning sifati va to‘liqligiga bog‘liq. Agar lug‘at barcha tegishli atamalarni o‘z ichiga olmasa, teglash natijalari samarasi kamayishi mumkin. Bundan tashqari, bu usul noaniq yoki kontekstga bog‘liq atamalarni teglashda ba’zi mos kelmaydigan tomonlarga ega. Lug‘atga asoslangan usulning ishlashini yaxshilash uchun kontekstdan xabardor lug‘atlardan foydalanish, teglarni bashorat qilish uchun mashinani o‘rganish modellarini kiritish yoki bir nechta lug‘atlarni birlashtirish kabi usullardan foydalanish mumkin.

Umuman olganda, slenglarni avtomatik teglashda sleng lug‘atlarning ahamiyati katta bo‘lib, unga asoslangan yondashuv matn ma’lumotlarini tahlil qilish uchun foydali sanaladi.

Foydalanilgan adabiyotlar:

1. Berdialiyev A. O‘zbek sinxron sotsiolingvistikasi. – Toshkent, 2019.
2. Hamroyeva Sh. Korpus lingvistikasi atamalarining qisqacha izohli lug‘ati. – Globe edits, 2020. B.52.
3. Partridge E. Slang Today and Yesterday. – London: Routledge and Kegan Paul, 1979. – P.20.
4. Беликов В. И., Крысин Л. П. Социолингвистика. – Москва, 2001. С.161.
5. Демская-Кульчицкая О.М., Семеренко В.Р., Ющенко Р.А. Методы автоматической разметки текстов национального корпуса языка. // Компьютерная математика, 2005. С.70-75.

Sultanova Aydin
Ajiniyoz nomidagi Nukus DPI
O‘zbek tili kafedrasi dotsenti, p.f.f.d (PhD).

Annotatsiya: Ushbu maqolada ta’limiy korpus, uning ahamiyati, qoraqalpoq ta’limiy korpusini yaratishning nazariy asoslari haqida so‘z yuritiladi. Shuningdek, qoraqalpoq ta’limiy korpusining ta’lim samaradorligidagi ahamiyati haqida ma’lumot beriladi.
Kalit so‘zlar: korpus, til korpusi, matn korpusi, ta’limiy korpus, milliy korpus, elektron kutubxona, elektron o‘quv lug‘atlar.

Аннотация: В данной статье речь пойдет об учебном корпусе, его значении, теоретических основах создания Каракалпакского учебного корпуса. Также пойдет речь о значении каракалпакского учебного корпуса и эффективности обучения.
Ключевые слова: корпус, языковой корпус, текстовый корпус, образовательный корпус, Национальный корпус, электронная библиотека, электронные учебные словари.

Annotation: This article will focus on the educational building, its significance, and the theoretical foundations of the creation of the Karakalpak educational building. We will also talk about the importance of the Karakalpak educational building and the effectiveness of training.
Keywords: Corpus, language Corpus, text corpus, educational Corpus, National Corpus, electronic library, e-learning dictionaries.

“Korpus” atamasi Lotincha “corpus” so‘zidan olingan bo‘lib, “tana, vujud” degan ma’noni anglatadi. Hozirgi kunda bu tushuncha lingvistik tahlil uchun olingan ma’lum bir til, dialekt yoki tilning boshqa kichik matnlar to‘plamlarini ifodalaydi. Bu atama quyidagi ma’nolarni ham anglatadi: 1. Har qanday matnni o‘zida ifoda etishi mumkin; 2. Kompyuter matnining asosi bo‘lishi mumkin; 3. Til yoki xilma-xillikni ifodalash uchun namuna sifatida olingan kompyuter matnining cheklangan to‘plami bo‘lishi mumkin [McEnery va Vilson., 1996:218]. Korpus o‘zaro lingvistik munosabatga ega bo‘lgan turli xil tillarni qamrab oluvchi matnlarning katta namunalarini o‘z ichiga qamrab olishi mumkin.

Dunyo tilshunosligida korpus sohasidagi maqsadli tadqiqotlar XX asrning 40-yillarida Blumfild, Friz va Bondjerslar tomonidan boshlangan; N.Frensis va G.Kuchera ilk marta korpus tuzish prinsiplarini ishlab chiqqan. Rus tilshunosligida V.P.Zaxarov, A.B.Kutuzov, Ye.V.Nedoshivina, V.V.Rikov, V.A.Plungyanlar korpus, uning turlari, korpus tuzish va teglash tamoyillari borasida tadqiqot olib borishgan.

O‘zbek tilshunosligida kompyuter lingvistikasi, matnga leksikografik ishlov berish va lingvostatistik tahlil etish borasida muayyan tadqiqotlar amalga oshirilgan. Korpus yaratish bo‘yicha ilmiy-tadqiqot ishlarini olib borgan olimlar sifatida N.Abdurahmonova, Sh.Hamroyeva, G.Ergasheva, J.Jumaboyevalarni alohida e’tirof etish lozim. Professor N.Abdurahmonova: «Korpus matnlar jamlanmasi – bu mashina o‘qiy oladigan formatda matnlarni linvistik analiz qilish, tabiiy tilni qayta ishlashga mo‘ljallangan bazadir. Korpus lingvistikasini kompyuter lingvistikasining asosiy yo‘nalishi sifatida e’tirof etish mumkin, chunki matnni qayta ishlovchi matnlar korpusga asoslanadi», - deb ta’rif beradi. [N.Abdurahmonova., 2022:141.] Darhaqiqat,

korpus lingvistikasi bugungi kunda zamonaviy ta'lim jarayoni uchun eng zarur manba bo'lib xizmat qiladi. Bugungi axborotlashgan davrda, zarur ma'lumotlarni tez va oson topishda, albatta, korpuslarning ahamiyati katta.

Korpus kompyuterda foydalanish uchun, tadqiqotlarda va amaliyotda qo'llash uchun, manba sifatida keltirilayotgan til vakili sifatida, inson tomonidan qayta ishlashga va cheksiz ma'lumotlarni shakllantirish va taqdim etish uchun qulay hisoblanadi [Dash., 2005:35]. Foydalanuvchi hech qanday lingvistik tushuntirishlarsiz izchil va uzluksiz ma'lumot olishini xohlaydi va korpus aynan mana shu ehtiyojni qondirishi zarur. Zamonaviy texnologiyalar bizga kompyuterda korpus yaratish va uni saqlash imkonini beradi, shunda biz istalgan vaqtda kerakli ma'lumotni oson topib olishimiz mumkin. Korpusni tashkil qilgan tarkibiy qismlar haqidagi tegishli hujjatlar matndan alohida saqlanishi lozim. Bunday yondashuv bizga korpusdan foydalanish jarayonida yengillik beradi.

O'qituvchi uchun korpus tengsiz xazina, undan har bir o'quv mashg'uloti uchun serqirra va mazmunli o'quv materiali tayyorlash mumkin bo'ladi. Yuqorida aytib o'tganimizdek, korpusda xilma-xil matnlar jamlanib, undan o'qituvchi istalgan shaklda foydalana oladi. Korpus turli janr va uslubdagi matnlarni qamrab olganligi sababli har qanday talabni qondirish imkoniga ega [Abdurakhmonova, N. 2021,2022]. Korpus qamrovi juda keng, salmoqdorlik darajasi yuqori bo'lgan tizimlashtirilgan kutubxonadir. Foydalanishga qulay, vaqtni ancha tejaydi. Elektron qidiruv tizimi jihatidan boshqa dasturlardan farq qiladi. Korpus bo'yicha qidiruv foydalanuvchiga belgilangan so'zning turli kontekstdagi barcha shakldagi ko'rinishini topib beradi. Lug'atda qaysi o'rinda ekanligini, uning variantlarini aniq ko'rsatadi. Qidirilayotgan so'zning birikish imkoniyatiga ega bo'lgan so'zlar qatorini, denotativ hamda konnotativ ma'nolarini belgilay oladi. So'zdan foydalanish chastotasi yoki statistikasini adibning yozgan asarida tasvirlaydi.

Bugungi kunda ta'lim sohasiga nazar soladigan bo'lsak, "Zamonaviy o'qituvchi" o'z kasbining mohir ustasi, nutqi ravon, o'quvchilarni faniga qiziqtira oluvchi bo'lishi kerak. Buning uchun albatta elektron manbalar, elektron kutubxona, elektron lug'at, multimedia materiallari, audio va video manbalarga murojaat qiladi. Bularning barchasi "Ta'limiy korpus" tushunchasini beradi.

Nutq o'stirish, o'quvchilarda o'zbek tilida to'g'ri va ravon so'zlash, nutq madaniyatini shakllantirish uchun so'z va grammatik vositalar boyligini hosil qila borish lozim. Shunda nutq o'stirish tushunchasi keng qamrovda qabul qilinadi[Abdurakhmonova, N. 2021,2022]. Zero, o'zbek tili darslarida grammatik vositalarni o'rganishga nisbatan kommunikativ yondashuv joriy etilayotgan ekan, ta'lim jarayoni tinglovchiga qaratilmay tuziladigan gaplar, hikoyalar tarzidagi «nutq o'stirish»nigina emas, balki muayyan vaziyatlarda namoyon etiladigan muloqot madaniyatini ham, axborot olish madaniyatini ham o'z ichiga olishi kerak. Aks holda, Ye.N.Solyovova to'g'ri ta'kidlaganidek, hatto ona tilida so'zlari va uslubiy xususiyatlarini o'rniga qo'yib aytilgan gaplar shaklan nutq sanalishi mumkin, lekin ular mohiyatan tinglovchiga qaratilgan nutqni ifoda etmasligi ham mumkin [90;10-b.]. Ta'lim qoraqalpoq tilidagi maktablarning o'zbek tili darslarida o'quvchilarning mustaqil fikrlash ko'nikmalarini, ijodiy matn yaratish qobiliyatlarini rivojlantirish masalalari bir zum bo'lsa-da, o'qituvchining diqqat-e'tiboridan chetda qolmaydi. Bunda «topshiriq shartining o'quvchini izlash, qidirish, noma'lumni izlab topishga undashi bilan» xarakterlanishiga e'tibor beriladi [84;12-b.]. O'zbek tilini o'rgatishda qiyosiy metodddan foydalanish yaxshi samara beradi. Bunda, albatta, lug'atlarga yuzlanamiz. Biz «O'zbekcha – qoraqalpoqcha, qoraqalpoqcha – o'zbekcha lingvistik

terminlarning maktab lug‘ati» ni [A.Sultanova., 2019:96.]yaratdik. Bugungi kunda esa mana shu kabi lug‘atlarni elektron shaklda yaratish vaqti keldi. Elektron lug‘at imkoniyatlari kengaydi, ilmiy-nazariy asoslari yaratildi, amaliyotda qo‘llash mumkin bo‘lgan ilk namunalari qo‘llanila boshladi. Fandagi bu yangilanishlar axborot texnologiyalarini tilshunoslikka tatbiq etish bilan bog‘liq istiqbolli ilmiy yo‘nalishlar paydo bo‘lishiga yo‘l ochdi. Bu esa korpus, korpus lingvistikasi, uning shakllanishi, taraqqiyoti, bugungi holati va korpus tuzishning umumiy tamoyillarini o‘rganish zaruratini hamda loyihaning mavzusi dolzarbligini belgilaydi. Takomillashib borayotgan kompyuter lingvistikasi yo‘nalishida avtomatik tarjima sifatini yaxshilash, tilni lingvistik modellashtirish, har bir tilga oid so‘zlar lug‘atini yaratish hamda muayyan tilning ko‘p asrlik milliy-madaniy merosdan foydalanish imkonini oshirish maqsadida ularni elektronlashtirish jahon tilshunosligida dolzarb masalaga aylandi. Tilshunoslikda, xususan, kompyuter lingvistikasi sohasida korpus yaratish, mavjud korpuslar hajmini kengaytirish, matnni avtomatik qayta ishlaydigan dasturlarni ishlab chiqish kabilar yechimini kutayotgan muhim masalalardan biri bo‘lib turibdi. O‘zbek tilining ta’limiy korpusini yaratish xorijiy tajribalardan kelib chiqib, bosqichma-bosqich ma’lumotlar shakllantirilishiga qaratilgan va o‘zbek adabiy tilining bugungi so‘zligini qamrab olgan elektron darslik, ko‘p tilli so‘zlashgichlar va o‘zbek tilining boshqa tillarda ekvivalenti bo‘lmagan, tarjima qilinmaydigan leksik birliklari – milliy-madaniy so‘zlarning izohi berilgan glossariy, o‘zbekcha to‘g‘ri talaffuz ko‘nikmalarini shakllantirishga qaratilgan audio va video materiallar hamda mobil ilova qamrab olingan multimedia mahsulotlari majmuasini yaratish ko‘zda tutilgan. Bu esa respublikamizning iqtisodiy o‘sishi va ijtimoiy rivojlanishini ta’minlaydigan ilmiy va texnologik zaxiralarni shakllantirishga xizmat qiladi.Ta’limiy korpus - o‘zbek tilini davlat tili va ikkinchi til, shuningdek xorijiy til sifatida chuqurroq o‘rganish imkonini beradi.

O‘zbek tili bo‘yicha bir qator ta’limiy korpuslar yaratilgan, lekin shu kungacha qoraqalpoq tilining ta’limiy korpusi ishlab chiqilmagan. Bu esa biz filologlar zimmasiga katta vazifalar yuklaydi. Chunki qoraqalpoq tilining o‘ziga xos xususiyatlarini namoyon etish, bu tilda yaratilgan boy badiiy merosni yosh avlodga yetkazib berishning eng samarali usullaridan biri uning ta’limiy korpusini yaratish hisoblanadi. Shuningdek, bu orqali o‘zbek va qoraqalpoq tili mutaxassislari uchun tarjima ishlarini oson bajarishlariga yordam beradi.

Milliy korpusda bosh maqsad jonli tilni ko‘rsatish bo‘lgani sababli unda turli uslub va janrdagi matnlar korpus birligi sifatida ishlatiladi. Ta’limiy korpusdagi matnlar til ta’limi, o‘quvchilarni ma’naviy-axloqiy tarbiyalashga qaratilgan matnlardan iborat bo‘ladi [Abdurakhmonova, N. 2021,2022]. Ta’limiy korpusga ko‘proq maqol, matal, naql, ertak, rivoyat janrlari kiritiladi. Asl maqsad – ona tilini o‘rgatish bilan birga, tarbiyaviy maqsad ham asosiy planda bo‘ladi. Ta’limiy korpusni Milliy korpusdan farqlab turuvchi yana bir jihat shuki, undagi qidiruv tizimi, video-audio, multimedia materiallari ko‘proq til ta’limi – tilni o‘rgatishga qaratiladi.

Xulosa o‘rnida, qoraqalpoq tili ta’limiy korpusining muhimligi shundaki, o‘zbek va qoraqalpoq tili filolog mutaxassislari, shuningdek, tarjimonlik bilan shug‘ullanuvchi shaxslar bu tillarning boy imkoniyatlaridan, adabiy merosidan, lug‘at boyligidan tez va oson foydalanish imkoniga ega bo‘lishadi, shuningdek yoshlarning lug‘at boyligi ortib, yozma savodxonligi takomillashib boradi.Ta’limiy korpusdan ilmiy xodimlar, oliy va o‘rta maxsus o‘quv yurtlari o‘qituvchi va talabalari, xalq ta’limi tizimi sohasida faoliyat yurituvchi o‘qituvchilar, barcha maktab yoshidagi o‘quvchilar, shuningdek, xorijda yashovchi vatandoshlarimiz, xorijlik

o‘qituvchi va talabalar, aholining o‘zbek va qoraqalpoq tiliga qiziquvchi barcha qatlamlari o‘z faoliyatlarida ununmli foydalanishlari mumkin.

Foydalanilgan adabiyotlar ro‘yxati:

1. O‘zbekiston milliy ensiklopediyasi. 2000-2005.
2. A.Qirg‘izboyev. Korpus milliy til xazinasi. O‘zbek milliy va ta’limiy korpuslarini yaratishning nazariy va amaliy maalalari mavzusidagi xalqaro ilmiy – amaliy konferensiya materiallari. . –Toshkent: 2021.
3. Raupova L. O‘zbek milliy va ta’limiy korpuslarini yaratishning nazariy hamda amaliy masalalari, Xalqaro ilmiy-amaliy konferensiya, –Toshkent: 2021.
4. O‘zbek tili milliy korpusi tizimi.info@navoiy-uni.uz
5. Ma’ripov J. Antroposentrizm–tilshunoslikning zamonaviy yonalishi sifatida //Zamonaviy dunyoda innovatsion tadqiqotlar: Nazariya va amaliyot. –T.: 2022.
6. Абдурахмонова, Н., & Абдувахобов, Г. (2021). О ‘QUV LUG ‘ATINI TUZISHNING NAZARIY METODOLOGIK ASOSLARI. *МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА*, *4*(6).
7. Abdurakhmonova, N. (2021). Formal-Functional Models of The Uzbek Electron Corpus. *ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies*, *10*(8), 59-66.
8. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
9. Абдурахмонова, Н., & Абдувахобов, Г. (2021). О ‘QUV LUG ‘ATINI TUZISHNING NAZARIY METODOLOGIK ASOSLARI. *МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА*, *4*(6).
10. Abdurakhmonova, N., Shakirovich, I. A., & O‘G‘Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. *Science and Education*, *3*(9), 146-156.
11. Tolibovna A. K. et al. Functions of Allusion and Allusion as a Marker of Intertextuality and Precedence //European Multidisciplinary Journal of Modern Science. –T.: 2022. –S. 485-487.
12. Ma’ripov J. K. A brief information about tenses //O‘zbekiston respublikasi oliy va o‘rta. –S. 464.
13. Tolibovna A. K. et al. Features of Anthropocentric Study of Sacred Texts //JournalNX. –T. 8. –No. 1. –S.
14. Sultanova A.M. O‘zbekcha – qoraqalpoqcha, qoraqalpoqcha – o‘zbekcha lingvistik terminlarning maktab lug‘ati. – T.: 2019.

TURKIY TILLARDA KORPUSLARNING KENG KO‘LAMI VA ULARNING RIVOJLANISHI

Eshmamatova Mahliyo
O‘zbekiston Milliy universiteti
O‘zbek filologiyasi Kompyuter lingvistikasi
va amaliy tilshunoslik kafedrasi stajyor-o‘qituvchisi.
E-mail: meshmamatova21@gmail.com

Annotatsiya: Umumiy korpus yozma yoki og‘zaki matnlarni o‘z ichiga olishi mumkin yoki ikkala ommaviy axborot vositalaridan matnlarni saralab oladi. Yaqinda biz yozish texnologiyalari va avtomatlashtirilgan transkripsiya dasturiy ta’minotining takomillashuvi tufayli og‘zaki korpuslar sonining ko‘payishiga guvoh bo‘ldik. Tilshunoslar til o‘zgarishini va korpus ma’lumotlaridagi til o‘zgarishini kuzatish uchun, birinchi navbatda, ma’lum bir nuqtada yoki vaqt oralig‘ida tildan foydalanishning “oniy tasvirini” taqdim etish maqsadida sinxron korpusini yaratdilar. Monitor korpusi bu yerda taqdim etilgan boshqalardan yangi materiallar qo‘shilishi bilan doimiy ravishda o‘sib borishi (dinamik) ma’nosida farq qiladi. Tilning o‘z vaqtida lingvistik hisobi uchun tuzilgan korpuslar odatda tekshirilayotgan tilning turli davrlarida tildan foydalanishni ifodalovchi matnlarni o‘z ichiga oladi.

Kalit so‘z: korpus,metama’lumot,umumiy korpus, NLP, lug‘at.

Abstract: A general corpus may include written or spoken texts, or a selection of texts from both media. Recently, we have seen an increase in the number of spoken corpora due to improvements in writing technologies and automated transcription software. linguists have created synchronic corpora to track language change and language change in corpus data, primarily to provide a "snapshot" of language use at a specific point or time period. The monitor case differs from the others presented here in the sense that it is constantly growing (dynamic) with the addition of new materials. Corpora constructed for a linguistic account of a language in time usually contain texts representing language use in different periods of the language under investigation.

Keywords: corpus, metadata, general corpus, NLP, dictionary.

Аннотация: Общий корпус может включать письменные или устные тексты или подборку текстов с обоих носителей. В последнее время мы наблюдаем увеличение количества устных корпусов благодаря усовершенствованиям в технологиях письма и программном обеспечении для автоматической транскрипции. лингвисты создали синхронные корпуса для отслеживания изменений языка и языковых изменений в данных корпуса, в первую очередь для того, чтобы предоставить «моментальный снимок» использования языка в определенный момент или период времени. Корпус монитора отличается от других представленных здесь тем, что он постоянно растет (динамично) с добавлением новых материалов. Корпуса, созданные для лингвистического описания языка во времени, обычно содержат тексты, представляющие использование языка в различные периоды исследуемого языка.

Ключевые слова: корпус, метаданные, общий корпус, НЛП, словарь.

Turk tilshunosligi tadqiqotlarida korpusdan foydalanish, yaqinda paydo bo‘lgan sohalardan biridir. Ushbu kechikishning asosiy sabablaridan biri 2000 - yillarning boshlariga qadar hech qanday lingvistik korpus mavjud emasligidir. Jarayonda ishtirok etayotgan ko‘pchilikka ma’lumki, korpus qurish ancha mehnat va ko‘p vaqt talab qiladigan faoliyat bo‘lib, u qat’iy institutsional yordamni talab qiladi. Korpusni qurishda birinchi navbatda e’tiborga olinadigan ba’zilari umumiy dizaynga taalluqlidir: masalan, kiritilgan matn turlari, matnlar soni, alohida matnlarni tanlash, matnlar ichidan matn namunalarini tanlash va matn namunalari uzunligi. Ularning har biri ongli ravishda namuna olish qarorini o‘z ichiga oladi.

• **Umumiy korpus:** Umumiy korpusni qurishda harakatlantiruvchi kuch teng va nomoyon bo‘lishi mumkin bo‘lgan tildan foydalanishning mos yozuvlar korpusini yaratishdir. Umumiy korpus yozma yoki og‘zaki matnlarni o‘z ichiga olishi mumkin yoki ikkala ommaviy axborot vositalaridan matnlarni saralab oladi. Asosiy maqsad - turli janrlar, sohalar va turlardagi matnlarni teng tarzda taqdim etish, natijada xulosalar shuni ko‘rsatadiki korpus ma’lumotlarining miqdoriy va sifat tahlillaridan olingan ma’lumotlar tildan foydalanish uchun to‘g‘ri keladi. Britaniya Milliy Korpusi (BNC) 100 million so‘zni o‘z ichiga olgan va 4048 yozma matn va o‘n million so‘z transkripsiya qilingan og‘zaki ma’lumotlarni o‘z ichiga olgan zamonaviy ingliz tilining shunday umumiy ma’lumot korpusidir[Abdurakhmonova, N. 2021,2022]. Respublika va mintaqaviy gazeta va jurnallar, ommabop va akademik kitoblar, universitet insholari, elektron pochta namunalari, nashr etilmagan xatlar va turli yoshdagi, muassasalar va kitobxonlarning hisobotlari. BNC muvaffaqiyati vakili va muvozanatli umumiy korpus sifatida boshqalarni uning asosiy dizayn tamoyillarini, jumladan Amerika Milliy Korpusini, Koreya Milliy Korpusini, Polsha Milliy Korpusini va yaqinda Turkiya Milliy Korpusini qabul qilishga olib keldi.

• **Ixtisoslashgan korpus:** Nisbatan kichik o‘lchamli va janr yoki soha bo‘yicha ixtisoslashgan, bu turdagi korpuslar ko‘proq xilma - xil va ko‘proq mavjud. Ixtisoslashgan korpus yaratishdagi hozirgi uslublar asosan professional va akademik sohalarda kuzatiladi. Bunday ixtisoslashgan korpusning ba’zi turlari orasida Professional Spoken American English Corpus (CPSA) va Michigan Corpus of Academic Spoken English (MICASE) kiradi. Ixtisoslashgan korpus kattaroq umumiy korpusdan tegishli matn ma’lumotlarini olish orqali ham yaratilishi mumkin.

• **Yozma korpus:** Jigarrang korpus nafaqat birinchi korpus, balki ayni paytda ingliz tilining zamonaviy davrdagi birinchi yozma korpusidir. Korpus ma’lumotlarini tashkil etuvchi matnlar yozma ommaviy axborot vositalaridan to‘plangan.

• **Og‘zaki korpus:** Umumiy yoki yozma korpus bilan solishtirganda, tilning og‘zaki korpusini qurish va izohlash qiyinroq. Yaqinda biz yozish texnologiyalari va avtomatlashtirilgan transkripsiya dasturiy ta’minotining takomillashuvi tufayli og’zaki korpuslar sonining ko‘payishiga guvoh bo‘ldik. Og‘zaki ingliz tili uchun korpuslar 1960-yillarning oxirida yaratilgan masalan, London - Lund Corpus (LLC) (Greenbaum and Svartvik 1990), keyin esa boshqalar, jumladan Lancaster/IBM Spoken English Corpus (SEC), Kembrij va Nottingem. Ingliz tilidagi nutqlar korpusi (CANCODE) (Karter va MakKarti 2004). Turk tilining yagona mavjud va lingvistik jihatdan ishonchli yangi avlod so‘zlashuv korpusi Og‘zaki Turk korpusidir (Ruhi et al.2010). Turk Milliy Korpusida (TNC) (Aksan et al. 2012) ham mavjud bir million so‘zning og‘zaki komponenti BNC dizayn tamoyillariga rioya qilishning aksi sifatida.

• **Sinxron korpus:** tilshunoslar til o‘zgarishini va korpus ma’lumotlaridagi til o‘zgarishini kuzatish uchun, birinchi navbatda, ma’lum bir nuqtada yoki vaqt oralig‘ida tildan foydalanishning

“oniy tasvirini” taqdim etish maqsadida sinxron korpusini yaratdilar. Bunday korpuslarda barcha matnlargacha bo‘lgan vaqt oralig‘idan tanlanishi kerak, sinxron ravishda mavjud bo‘lgan til turlarini hisobga olish. Ingliz tilining xalqaro korpusi (ICE) Buyuk Britaniya, AQSh, Avstraliya, Kanada va ingliz tili birinchi til bo‘lgan boshqa mamlakatlarda so‘zlashuvchi ingliz tilini sinxron tahlil qilish uchun qurilgan (Greenbaum 1991). U bir millionlik yigirma korpusdan iborat.

• **Diaxronik korpus:** Tilning o‘z vaqtida lingvistik hisobi uchun tuzilgan korpuslar odatda tekshirilayotgan tilning turli davrlarida tildan foydalanishni ifodalovchi matnlarni o‘z ichiga oladi. Ovoz yozish texnologiyalarining so‘nggi tarixini hisobga olsak, diaxronik korpus vaqt o‘tishi bilan yozma tilni ifodalaydi, masalan, Xelsinki diaxronik korpusi inglizcha matnlar (Rissanen va boshq. 1991) Learner Corpora: Til sinfida korpusdan foydalanish, o‘qitish va o‘rganish kontekstlarida o‘z o‘rnini topdi. Masalan, International Corpus of Learner English (ICLE) (Granger 2003) va uning subkorpusi sifatida Turk Xalqaro O‘rganuvchi Corpus of English (TICLE) (Kilimci and Can 2009) so‘nggi yillarda o‘qitish kontekstida tadqiqot manbai bo‘lib kelgan.

• **Monitor korpusi:** Monitor korpusi bu yerda taqdim etilgan boshqalardan yangi materiallar qo‘shilishi bilan doimiy ravishda o‘sib borishi (dinamik) ma’nosida farq qiladi. Bank of English (BoE) va Corpus of Contemporary American English (COCA) (Devis 2008) ingliz tili uchun ushbu turdagi taniqli korpuslardir. Korpus, birinchi navbatda, tilni haqiqiy ma’lumotlar asosida empirik tarzda o‘rganish uchun tildan foydalanishni muvozanatli tarzda ifodalash uchun tuzilgan. Tilshunoslik tahlillarida korpusning roli va funksiyasiga qarab, turli nuqtai nazardan qarash mumkin, tadqiqot savollari bo‘yicha. Lüdeling and Kytö lingvistik tahlillarda korpusdan foydalanishni uchta asosiy maqsadda umumlashtiradi:

1) **empirik yordam**
2) **chastotali ma’lumot**
3) **metama’lumot**

Korpus qidiruv tizimi tadqiqotchilarga yordam beradi. Ularning savollariga mos bo‘lgan haqiqiy tildan foydalanish uchun ma’lumotlar topish, ya’ni ular hozirda muqobil emas, balki haqiqiy tildan foydalanish uchun ma’lumotlar keltirish mumkin[Abdurakhmonova, N. 2021,2022]. Tilning tuzilishi va ishlatilishiga oid dalillarni korpusdan topish, lingvistik tahlilning ma’lum darajasi bilan chegaralanib qolmaydi, balki tovushdan shaklgacha va funktsiyaga qadar barcha darajalarda ishlaydi. Korpusdagi ma’lumotlar saralanadi va izohlanadi, shuning uchun namuna olishning aniq turini ta’minlaydi. Bu gipotezalarni empirik tarzda tasdiqlaydi. Haqiqiy til namunalari jamlanmasi sifatida korpus so‘rovi tildan foydalanishning ilgari ko‘zda tutilmagan iqtiboslarini qaytaradi. Korpusdan olingan iqtiboslar shunchaki ma’lum bir lingvistik ko‘rinishni ifodalamaydi, balki miqdoriy ma’lumotni ham beradi.

Bugungi kunda turkiy tilda kamida uch xil turdagi korpus mavjudligini aytib o‘tishimiz mumkin:

1) katta o‘lchamli umumiy tilshunoslik korpusi

2) aniq tadqiqot savollarini o‘rganish uchun tuzilgan va faqat quruvchilar uchun mo‘ljallangan kichik o‘lchamli ixtisoslashtirilgan korpuslar

3) hech qanday til mezonlarisiz qurilgan NLP korpuslari yaratilgan va foydalanuvchilar uchun taqdim etilgan, ong balki turli ilovalar uchun ishlab chiqilgan algoritmlarni sinab ko‘rish vositalari sifatida

Turk tilini boshqa tillarga nisbatan yaxshi o‘rganilgan til deb ayta olmaymiz, uning tarixi va grammatikasi yaxshi hujjatlashtirilgan. Hozirgi turkshunoslikdagi aksariyat tilshunoslik asarlari

nutq tahlili, pragmatika yoki sintaksis kabi oz sonli sohalarga qaratilgan. Biz kamdan - kam hollarda semantika yoki leksikologiya yoki boshqa sohalardagi ishlarni topamiz, chunki ular boyitilgan ma'lumotlar to'plamini talab qiladi. Jumladan, yangi tashkil etilgan Respublikaning dastlabki yillarida leksikaning turkiylashuvidan kelib chiqqan holda, o'sha davrdagi mavjud shevalardan lug'at tuzishga qaratilgan. Lug'at XIII asrdan boshlab 160 ga yaqin turli tarixiy matnlardan turkiy tilga oid lug'aviy birikmalarni jamlagan. Juda kam nazariy tadqiqotlardan tashqari, deyarli barcha lingvistik tahlillar empirik va ma'lumotlarga asoslangan. Turk tilshunosligidagi tipik tadqiqot savolni tahlil qilishda " ma'lumotlar bazasi " yoki "ma'lumotlar to'plami" ni jamlaydi. Aytishimiz mumkinki, deyarli barcha foydalanishga asoslangan empirik tadqiqotlarda qo'llaniladigan juda kichik o'lchamli maxsus korpuslar mavjud. Biroq, ular o'zlarining shakli va hajmi bilan qattiq cheklangan, ular boshqa tadqiqotchilar uchun mavjud emas va ma'lumotlar muayyan muammo bilan to'plangan. Turk tilidagi hisoblash tilshunosligi bo'yicha ish turkiy korpus lingvistikasiga qaraganda uzoqroq tarixga ega. Turk tilida korpus tadqiqotining dastlabki boshlanishi NLP tadqiqotlari va hisoblash tilshunosligi tahlillari bilan bog'liq edi. Hisoblash tilshunosligida va NLPda keng ko'lamli korpuslar "amaliy" maqsadlar uchun qurilgan. Turkiyada hisoblash tilshunosligining keng qamrovli tarixi hali yozilmagan; ammo, bu sohada oldingi ishlar haqida ba'zi – ba'zi joylarda havolalar bor. Lingvistik tahlil uchun ma'lum bo'lgan birinchi elektron korpus Koksal (1976) tomonidan "avtomatik morfologik tahlil" uchun yaratilgan. Koksal o'z algoritmini kundalik gazetalardan tasodifiy tanlangan 1534 so'zli matn namunasi korpusida sinab ko'rdi va baholadi. Koksalning ishi turkiy tillarning boy morfologiyasi va mumkin bo'lgan morfema birikmalarini tan oladi, shuningdek, kelajakdagi asosiy vazifalarga ishora qiladi, qo'llashning potentsial sohalariga e'tibor qaratadi, avtomatlashtirilgan til tahlillari uchun kattaroq korpus yaratishga undaydi. Zamonaviy turk tilini ifodalash uchun ishlab chiqilgan va tuzilgan birinchi elektron lingvistik korpus.

Foydalanilgan adabiyotlar:

1. Leech, G. 100 million words of English: The British National Corpus (BNC). Language research, 1992.
2. Xiao, Z., McEnery, A. Two approaches to genre analysis: Three genres in modern American English. Journal of English Linguistics, 33(1), 2005. 62-82.
3. Gomez, I. F. Interaction in academic spoken English: the use of 'I'and 'you'in the MICASE. Information technology in languages for specific purposes: Issues and prospects, 2006. 35-51.
4. Rissanen, M., Ihalainen, O., Kytö, M. The Helsinki Corpus of English Texts. Diachronic and Dialectal. Helsinki:1991.
5. Granger, S. The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. Tesol Quarterly, *37*(3), 2003538-546.
6. Kilimci, A. Negotiation of meaning in L2 academic writing. In Proceedings of 1st international conference on literature, languages and linguistics. Athens:2009.
7. Davis, M. The Corpus of Contemporary American English (COCA): One Billion Words, 1990–2019.
8. Абдурахмонова, Н., & Абдувахобов, Г. (2021). O 'quv lug 'atini tuzishning nazariy metodologik asoslari. *Международный журнал искусство слова*, *4*(6).

9. Abdurakhmonova, N. (2021). Formal-Functional Models of The Uzbek Electron Corpus. *ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies*, *10*(8), 59-66.
10. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
11. Abdurakhmonova, N., Shakirovich, I. A., & O‘G‘Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. *Science and Education*, *3*(9), 146-156.
12. Kytö, M., & Lüdeling, A. (Eds.). Corpus linguistics: an international handbook. De Gruyter Mouton.:2009.
13. Köksal, F. J coupling measurements using Carr—Purcell spin-echo techniques. Chemical Physics Letters, *44*(1), 1976187-189.
14. Aksan, M., & Aksan, Y. Лингвистические корпуса: взгляд с турецкого языка. В обработке турецкого естественного языка.Шпрингер, Чам: 2018. (стр. 291-315)
15. Мирзахалов Дж., Бабу А., Атаман Д., Кариев С., Тайерс Ф., Абдурауфов О.,Челлаппан С. Масштабное исследование машинного перевода на тюркских языках, 2021.

O‘ZBEK TILI KORPUSLARIDA O‘Z QATLAM LEKSIKASI ETIMOLOGIYASINI SEMANTIK KENGAYTMA SIFATIDA IFODALASH MUAMMOLARI

Axmedova Dildora Bahodirovna
BuxDU doktaranti (DSc).

Annotatsiya: Maqolada etimologik lug‘atning tuzilishi, unda o‘zbek tilining o‘z qatlamiga mansub bir ma’noli hamda omonim so‘zlarning etimologik talqini hamda ularni semantik kengaytma sifatida qayta ishlash imkoniyatlari borasida fikr yuritilgan. Lug‘at muallifining ta’kidlashicha, hozirgi o‘zbek adabiy tilidagi so‘zlarga tovush tomonidan etimologik talqin berishda qadimgi turkiy tildagi unlilarga qattiq-yumshoqlik jihatidan xos bu belgisi o‘zbek tilining ayrim shevalarida o‘z kuchini qisman saqlagan bo‘lsa ham, adabiy tilda inobatga olinmaydi.

Kalit so‘zlar: etimologik lug‘at, semantik kengaytma, til korpuslari, lug‘at maqolasi, qadimgi turkiy til, omonimlar, tag ma’no.

Annotation: The article discusses the structure of the etymological dictionary, the etymological interpretation of synonyms and homonyms belonging to its own layer of the Uzbek language, and the possibilities of processing them as a semantic extension. The author of the dictionary states that in the etymological interpretation of the words in the modern Uzbek literary language by sound, this characteristic of the vowels of the ancient Turkic language in terms of hardness and softness has partially preserved its power in some dialects of the Uzbek language. is not taken into account in the literary language.

Keywords: etymological dictionary, semantic extension, language corpora, dictionary article, ancient Turkic language, homonyms, tag meaning.

Аннотация: В статье рассматривается структура этимологического словаря, этимологическая интерпретация синонимов и омонимов, принадлежащих к собственному пласту узбекского языка, а также возможности их обработки как семантического расширения. Автор словаря констатирует, что при этимологической интерпретации слов в современном узбекском литературном языке по звуку эта характеристика гласных древнетюркского языка по твердости и мягкости частично сохранила свою силу в некоторых диалектах узбекского языка. язык. не учитывается в литературном языке.

Ключевые слова: этимологический словарь, семантическое расширение, корпус языка, словарная статья, древнетюркский язык, омонимы, значение тега.

O‘zbek tilining etimologik lug‘atidan keng jamoatchilik foydalanishini ko‘zda tutib tilshunoslik terminlari deyarli ishlatilmaganligi, sharhlar ommabop yo‘sinda, ixcham tarzda bayon qilinganligi ta’kidlanadi. Bu esa semantik kengaytma izohlarini ishlab chiqishda bizga qulaylik beradi. Chunki til korpusi yoki boshqa axborot qidiruv vositalaridan keng jamoatchilik foydalanadi. Bu esa semantik kengaytma uslubini soddalashtirishni taqozo etadi.

Lug‘atda qo‘shimcha qo‘shish bilan yangi so‘z yuzaga kelsa, «yasalgan» so‘zi, grammatik shakl yuzaga kelsa, «hosil qilingan» so‘zi ishlatilgan. Asos bilan qo‘shimcha orasiga qo‘shuv alomati, bular bilan hosila orasiga barobar alomati qo‘yildi. So‘z qismlari orasidagi munosabatni ko‘rsatishda qavslardan foydalanildi. Tovush jihatidan o‘zgargan so‘zning oldingi va keyingi

shakli o‘tish belgisi (>) yordamida ko‘rsatilgan. Biz ham semantik kengaytmada xuddi shu belgilarni saqlab qolishni ma’qullaymiz.

So‘zlarga etimologik talqin, odatda, ikki jihatdan beriladi: 1) mazmun jihatdan; 2) shakl (tovush) jihatdan. Mazmun jihatidan talqin berish tub deb qaraluvchi so‘zlarda yasama deb qaraluvchi so‘zlarga nisbatan qiyin. Etimologik lug‘at maqolasining tuzilishi quyidagicha:

AVAYLA – «ehtiyot qil» – *Seni avaylab o‘stirdim.* Bu fe’l asli ayrim turkiy tillarda, shu jumladan, qirg‘iz tilida mustaqil ishlatilib, «ehtiyotkorlik» ma’nosini anglatuvchi abay otidan ***-la*** qo‘shimchasi bilan yasalgan; keiynchalik o‘zbek tilida ikki unli oralig‘idagi ***b*** undoshi ***v*** undoshiga, ***a*** unlilari esa ***ä*** unlilariga almashgan: ***abay+la = abayla > äbäylä.*** [Раҳматуллаев Ш., 2000: 20]

Yoki boshqa bir izohga, omonim so‘zlarning berilishiga e’tibor qaratamiz:

ADASH I «ismi bir xil (kishilar)». *Ikkimiz adashmiz.* Bu so‘z qadimgi turkiy tildagi «ism» ma’nosini anglatuvchi a:t so‘zining at shaklidan *-dash* qo‘shimchasi bilan yasalgan (ESTY, 1, 204), keyinchalik **t** undoshi **d** undoshining ta’sirida **d** undoshiga almashgan, so‘ngra **dd** undoshlaridan biri talaffuz qilinmay qo‘ygan (TRS, 15); o‘zbek tilida ***a*** unlilari ***ä*** unlilariga almashgan: ***at-dash = atdash > addash > adash > ädäsh.***

ADASH II: **ADASH BO‘L-** «oyoq kiyimining poyi boshqa bir oyoq kiyimining bir poyi bilan almashib qol-». *Kechagi yig‘inda bir poy kalishim adash bo‘lib qolibdi. Bu so‘z* ***ädäsh-*** fe’lining eski o‘zbek tilidagi ism eshi bo‘lib, mustaqil ishlatilmaydi, ***ädäsh bol-*** qo‘shma so‘zi tarkibida keladi.

ADASH- «to‘g‘ri yo‘lni yo‘qot-», «yanglish-». *Qorong‘i tushganidan keyin adashib qoldim.* Bu so‘z asli qadimgi turkiy tildagi «to‘g‘ri yo‘lni yo‘qot-» ma’nosini anglatuvchi **a:z-** fe’lining **a:d-** shaklidan «biroz» ma’nosini ifodalovchi **–(a)sh** qo‘shimchasi bilan hosil qilingan (ESTY, I: 95), keyinroq **a:** unlisining cho‘ziqlik belgisi yo‘qolgan (KRS, 22); bu so‘z dastlab «to‘g‘ri yo‘ldan biroz chetga chiq-» ma’nosini anglatgan. O‘zbek tilida **a** unlilari **ä** unlilariga almashgan: ***(a:z- > a:d-) +ash = a:dash- > adash- > ädäsh.*** [Раҳматуллаев Ш., 2000: 20-21]

So‘z tilning tarixiy taraqqiyoti davomida mazmunan o‘zgarmay, tovush jihatidan o‘zgargan bo‘lishi mumkin. Lug‘at maqolalarida ana shunday tovush o‘zgarishlari qadimgi turkiy tildan hozirgi o‘zbek tiliga qadar qanday voqe bo‘lgani bayon qilingan. [Раҳматуллаев Ш., 2000: 9]

Qadimgi turkiy tilda, masalan, «ich tomonga tort-» ma’nosini anglatgan op- (o‘p-) fe’lidan «bo‘sa ol-» ma’nosini anglatgan öp- (o‘p-), «o‘roq bilan kes» ma’nosini anglatgan or- (o‘r-) fe’lidan «sahn bo‘ylab harakatlan» ma’nosini anglatgan ör- (o‘r-) fe’li, «havo bo‘ylab harakatlan» ma’nosini anglatgan uch-fe’lidan «ikki bilan to‘rt oralig‘idagi butun miqdor son» ma’nosini anglatgan uch so‘zi, «maysa ma’nosini anglatgan ot (o‘t) otidan «bir joydan ikkinchi joyga harakatlan-» ma’nosini anglatgan öt- fe’li unli tovushining yumshoqligi bilan farqlangan. Bulardan anglashiladiki, semantik kengaytmada tovushlar bilan ma’no farqlanadigan har qanday holat aks etishi kerak.

Shuningdek, Sh.Rahmatullayev quyidagi masalaga e’tiborimizni qaratadi: «Qadimgi turkiy tilda (shuningdek, hozirgi turkiy tillarning ko‘pchiligida) unlilarning qisqa-cho‘ziqligi ham so‘zlarni o‘zaro farqlashga xizmat qilgan... qadimgi turkiy tilda «asp (ish hayvoni)» ma’nosini anglatuvchi at (ot) so‘zi bilan «ism» ma’nosini anglatuvchi a:t so‘zi o‘zaro unli tovushning qisqa/cho‘ziqligi bilan farqlangan; «kovla» ma’nosini anglatgan och- fe’lidan, «ochiqqan»

ma’nosini anglatuvchi a:ch sifati va «ochiq holatga keltir-» ma’nosini anglatuvchi ach- fe’lidan unli tovushining cho‘ziqligi bilan farqlangan” [Раҳматуллаев Ш., 2000: 11].

Qadimgi turkiy tilga xos hodisalardan biri – tag ma’nosi bir xil asosning ikki so‘z sifatida ishlatilishi. Bunday asosning ham fe’l, ham ot bo‘lib, qisman ham ot, ham sifat bo‘lib ma’no tarmoqlanishi hodisasini qadimgi turkiy tilga nisbatan semantik sinkretizm deb nomlab, omonimlardan farqlash to‘g‘ri [Севортян Э.В., 2001: 40). Asli omonimlar deb turli tovush o‘zgarishlari natijasida, so‘z o‘zlashtirish oqibatida tasodifan shaklan bir xil bo‘lib qolgan so‘zlarga aytiladi. Qadimgi turkiy tildagi hodisa esa tamoman boshqacha tabiatga ega bo‘lib, ikki har xil turkum ma’nosining bir asos bilan anglatilishiga teng. Bu hodisani shartli ravishdagina semantik omonim deb yuritish mumkin. Masalan, qadimgi turkiy tilda «kovla-» ma’nosini ham, «chuqur» ma’nosini ham or so‘zi anglatgan. Keyinchalik ko‘pchilik turkiy tillarda, shu jumladan, o‘zbek tilida ham or- fe’li iste’moldan chiqqan, or oti esa saqlangan. Shuningdek, «alanga ol-» ma’nosi ot- so‘zi bilan, «alanga» ma’nosi o:t so‘zi bilan anglatilgan; keyinchilik ot- fe’li iste’moldan chiqib ketgan, hozirgi o‘zbek tilida ot otigina mavjud. Yuqoridagi juftliklarda tag ma’no bir xil, ikki so‘z ana shu tag ma’noning go‘yo tarmoqlanishi natijasida yuzaga kelgan; bundan qat’i nazar, birinchi juftlikda unlilar orasida farq ko‘rinmasa, ikkinchi juftlikda unlilar qisqa-cho‘ziqligi jihatidan farqlangan. Xullas, bu yerdagi hodisa zamirida ma’no taraqqiyoti yotadi. Bunday ma’no taraqqiyoti natijasida bir turkum so‘zidan boshqa bir turkum so‘zi o‘sib chiqadi. Bunday ma’no taraqqiyoti qadimgi turkiy tildan to hozirgi kungacha davom etib keladi. Etimologik lug‘atda ana shunday jarayonni ham to‘g‘ri aks ettirishga harakat qilingan.

Foydalanilgan adabiyotlar:

1. Axmedova D.B. Semantik kengaytma tushunchasi va uning til korpusidagi ahamiyati // Ilm sarchashmalari. Urganch, 2022. №9. – B. 66-69.
2. Баҳриддинова Б.М. Ўзбекистонда ўқув луғатчилиги: лингвистик асослари, тарихи ва истиқболлари: Филол. фан. доктори (DSc) дисс. автореф.– Самарқанд. 2020.
3. Ҳамроева Ш. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
4. Нарходжаева Х.Ш. Ўзбек тилида жараён англатувчи терминларнинг лингвистик тадқиқи: Филология фанлари бўйича фалсафа доктори (PhD) дисс., – Тошкент, 2007. – 154 б.
5. Ўзбек тилининг изоҳли луғати. 5 жилдли. Учинчи жилд. Н – Тартибли. – Тошкент: “Ўзбекистон миллий энциклопедияси” Давлат илмий нашриёти, 2007. 608-б.
6. Шимчук Э.Г. Русская лексикография: учеб. пособие для студ. филол. фак. высш. учеб. заведений / Э.Г.Шимчук. - М.: Издательский центр «Академия», 2009.
7. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011.

CREATION OF THE EDUCATIONAL CORPUS OF STUDENT SPEECH IS AN EXAMPLE OF WORLD EXPERIENCE

Kuvondikova Gavkhar Isomiddinovna
Teacher of the National University of Uzbekistan.

Urinboeva Nazokat
Master's degree student
of the National University of Uzbekistan.

Abstract: This article presents the importance of the educational corpora of students' speech and the analysis of corpora samples created in this regard. Extensive work has been done on Arab, Brazilian and Lithuanian student corpora. Examples of them include the development of criteria for teaching the language as a mother tongue, as a second and foreign language, and the creation of statistics on the use of academic words by students. Adaptation of texts or spoken speech samples (audio or multimedia files) attached to the corpus content for the purpose of the user, as well as the important aspects of following the rules of providing metadata in order to make the corpus easier to use are expressed in the article.

Keywords: Educational corpus, Arab student corpus, Brazilian student corpus, Lithuanian corpus

Аннотация: В данной статье представлена значимость учебного корпуса речи студентов и проведен анализ образцов корпусов, созданных, в связи с этим. Обширная работа была проделана над студенческими корпусами арабских, бразильских и литовских стран. К их примерам можно отнести разработку критериев обучения языку как родному, как второму и иностранному, а также создание статистики использования академических слов учащимися. Адаптация текстов или образцов устной речи (аудио или мультимедийных файлов), прикрепленных к контенту корпуса, для целей пользователя, а также важные аспекты соблюдения правил предоставления метаданных с целью облегчения использования корпуса выражены в статья.

Ключевые слова: образовательный корпус, арабский студенческий корпус, бразильский студенческий корпус, литовский корпус.

The growing public interest in the use of corpora as a resource in the study of language and linguistics led to the creation of TaLC in 1994 as a result of a discussion among members of ICAME (the international computer archive of modern and medieval English). . The first conference on educational corpora was held in Lancaster in 1994, and it was noted that the use of computer text corpora in research was well established and that they were widely used for teaching purposes, corpora data in the preparation of teaching materials, and foreign languages and discussions on the direct creation of corpora by students in linguistic education are on the agenda.

Creation of the first educational corpora Sylviane Granger and her scientific team created the "International corpus of learner English" (ICLE - International corpus of learner English). The purpose of this corpus is to create methods for correcting common errors of language learners by classifying them [D. Stewart: 2004].

The following characteristics of English language learners were taken into account when creating this educational corpus [G. Sylviane; 199]: educational environment, age, mother tongue, stage of education, nature of tasks.

Language learners were selected for this corpus within the following criteria:

1) type of learners;

2) age: adults

3) level of education: high level (advanced);

4) task: essays

The educational international corpus of the English language is based on the students of English as a second language of French, German, Chinese, Japanese, and Czech nationalities.

The international educational corpus of the English language consists of several linguistic instruments.

• The ICE Markup Assistant is an automatic system that simplifies the process based on a set of standard symbols for this instrumental layout;

• TAGSELECT – the Ice tag Selection system (tagging system) is an automatic sorting system of word group tags generated using the TOSCA tagger for each word in the corpus;

• ICEMARK – Used for syntactically parsing the tagged text before the TOSCA parser. Also, one of its other functions is the ability to change the tags of word groups, edit the errors of the words found in the text, and see untagged sentences.

• ICECUP – Corpus Utility is functionally important for analysis, concordance and search of this instrumental corpus.

Student corpus research is a relatively young but high-performing field of corpus linguistics, and the world experience is rich in extensive and diverse research in this field. One of these is the ALC project, a collection of written and spoken materials collected from Arabic language learners in Saudi Arabia. The aim of this project is to serve as a source of information for the study and teaching of Arabic as a second language and for further linguistic research.

Prior to the creation of the ALC corpus, several studies on student corpora were conducted in Saudi Arabia. These are: Pilot Arabic Learner Corpus (Abuhakema et al., 2008), Corpus of L2 (Hassan and Daud, 2011). The Pilot Arabic Learner Corpus (PALC) is a corpus of approximately 9,000 words of Arabic texts created by American English speakers learning Arabic. Some of the student texts were written while the students were studying Arabic in the United States. This is considered to be an intermediate level and consists of 3818 words. The second part, about those who went to study in Arab countries, was written in Arabic, and it was a high level and made up 4741 words. Corpus of L2 is a corpus of Malaysian Arabic language learners and consists of students' written texts. This corpus is mainly designed to check the incorrect use of Arabic compounds. It contains approximately 240,000 words created by students from 60 universities. 97% of them are Malaysians and were collected in the first and second year of Arabic studies. Corpus materials were developed on the topic of descriptive and comparative essays.

Based on data from both corpora, a further project ALC was created about the student corpora [Abdurakhmonova, N. 2019,2021,2022]. This corpus was more detailed and comprehensive than the corpus created before. First of all, special attention was paid to the body design. In this case, the ALC is intended to consist of a number of small cases. The design of the body is divided according to the following criteria:

non-native speakers of Arabic	students whose native language is Arabic
Men	Women
Pre-university students	university students
oral materials	written materials

The information in the ALC corpus is both oral and written, both of which have two stages. For the first stage of the oral part, the conditions of "describing a photograph" and for the second part of the "interview" have been set. For the first stage of the written part, the topic "Holiday Adventures" was chosen. Students wrote an essay in the classroom for 40 minutes. At the next stage, the students will discuss the content of the discussion "Should students choose education based on their interest?" Or the bright developing areas of the economy?" wrote an essay on the topic. Only the students had to write the essay for two days as homework, not in the classroom or in the auditorium. During the essay, students are given the opportunity to use dictionaries and grammatical information. ALC metadata is classified as:

Reader metadata Text metadata

1. Age 1. Place of registration
2. Gender 2. Year of writing
3. Nationality 3. City of registration
4. Mother tongue 4. The time is fixed
5. The number of languages you can speak 5. The time is not set
6. Number of years studying Arabic 6. Use of literature
7. Number of years spent in Arab countries 7. Use of grammar books
8. General level of education 8. Use of monolingual dictionary
9. Level of education 9. Use of other resources
10. Year/Semester 10. Using a bilingual dictionary
11. Educational institution 12. Text volume

In Saudi Arabia, except for pre-school educational institutions, all other educational institutions are taught in single-sex classes, that is, men and women do not mix. For this reason, special attention is paid to the gender component in the research. To work with women, only female specialists were included. Therefore, the part devoted to information about women's speech was 20%. Especially the oral part of the corpus is laborious and time-consuming work, which made up only 10% of the ALC.

The distribution of words in 4 groups that make up the ALC criterion was as follows.

The level of the students' mother tongue	**The level of education of the students**	**The gender of the students**	**The form of the material**
National 50% 100,000 words	Pre-university students 70% 140,000 words	Male 80% 160,000 words	Written speech 90% 180,000 words
Non-national 50% 100,000 words	University students 30% 60,000 words	A woman 20% 40,000 words	Oral speech 10% 20,000 words

The ALC corpus includes 1,585 texts (written and spoken), 282,732 words developed by 942 students. The corpus differs from other corpora in that it is based on metadata. According to

the metadata, the age of the students is from 16 to 42, and the students of 67 different nationalities participated. Text metadata is classified as follows. In ALC, two genres constituted narrative (67% of ALC content) and discussion (%). According to where the texts were written: students wrote their texts in class (62% of ALC data) at home (31%), but all audio recordings were produced in class (7%). The average length of texts is 178 words33

Another major study of corpora of student speech was conducted in Brazil. Researcher Larisa Goulart Da Silva, a professor at the University of Warsaw, focused her research on how Brazilian students use academic words and how corpus linguistic tools work based on AWL, a pre-existing list of academic words. Also, this study can be a source for creating a list of academic vocabulary used by Brazilian students. For this, the criteria for receiving essays from Brazilian students studying in Britain has been developed [Abdurakhmonova, N. 2019,2021,2022]. Essays were collected in two stages, the essay topic in the first stage was descriptive, and in the second stage, essays were discussed. Larisa Goulart Da Silva initially tried to find answers to the following questions in the students' written works.

a) What is the vocabulary of essays written by Brazilian students?

b) How does it compare with the vocabulary profile of other academic British corpora?

c) What words do Brazilian students use in the Academic Word List (AWL)?

d) What are the differences between Brazilian students' use of academic words and those represented in the British Academic Written English (BAWE) corpus?

The texts collected in the corpus are classified based on the following groups:

- Life Sciences,
- Social studies,
- Arts and humanities,
- Physical sciences.

In the section of groups, we can see the level of students' use of academic words in the table below.

T|r	Corpus	Number of running words	AWL
1	Art	875.000	9.30%
2	Science	875.000	9.10%
3	Commerce	875.000	12%
4	Law	875.000	9.40%
5	Linguistics	2.031	12.60%
6	Applied linguistics	5.137	17.60%
7	Sociology	2.084	12.44%
8	Social psychology	2.059	14.38%
9	History	2.036	14.49%
10	Devolepment	2.023	12.26%
11	Medicine	2.024	6.72%
12	Zoology	2.026	7.31%
13	Anatomy	11.356	8.60%

The researcher worked in cooperation with the "Languages Without Borders" (SwB) program. The program covers 10,740 Brazilian students from 87 British universities. These students came from 131 Brazilian universities representing different regions of Brazil. According to information provided on the official SwB website, of these 10,740 students, 774 are doctoral

students, 571 are full-time PhD students, 8,864 are visiting undergraduate students, and 531 are postdoctoral fellows. The researcher recruited only undergraduate students for the corpus. The results of the study show that Brazilian students use academic vocabulary at the same level as British students, but there are some peculiarities in the use of academic vocabulary by Brazilian students. Given these unique characteristics, some pedagogical recommendations for teaching EAP to Brazilian students are presented below. Nevertheless, the recommendations given here mainly take into account the "Languages Without Borders" program. For example, BAWE students used 43 prefixed word forms, while Brazilian students used only one prefixed word form, which means that Brazilian students used less suffixes and prefixes than BAWE students. Because of this, it has been hypothesized that prefixes such as "under" and "over" may not be comfortable for Brazilian students' speech because they are not derived from Latin. However, further research is needed to confirm this hypothesis.

Student corpora are important in students' second language acquisition. In Lithuania, a corpus of EFL students was created in 2011. Students of the Faculty of Philology of the English-speaking Vilnius University in Lithuania participated in the study[Abdurakhmonova, N. 2019,2021,2022].The corpus focused on examining the phraseology of student speech and various aspects of academic written and spoken English. Materials in EFL are classified under the following groups:

Local - Non-local

Women – Men

The oral part of the research begins with an informal discussion about university life, hobbies, travel or the future. Then interviewees are asked to choose one of three topics: "The country that surprised you", "Your favorite movie to watch", "The game you liked or disliked". Each interview ends with a short story based on a picture. Interviews averaged 200 words. In the written part, students wrote an argumentative essay for 40 minutes. Research shows that there are common mistakes for different groups of students. For example, excessive use of connectives and incorrect use of lexical expressions in the written text, some deviations in the use of grammatical patterns. It has been proven that the reason for uncertainty and hesitation in students' oral speech or time-consuming attempts is that the speech is subject to grammatical rules, or that the students could not speak without a sequence of grammatical formulas.

The educational corpus of the Uzbek language was created within the framework of the practical project "Creation of the educational corpus of the Uzbek language" numbered AM-FZ-201908172, its initial presentation was held on April 23, 2021, and its base is currently being enriched and improved. Today in this case

1) morphoanalyzer (automatic morphological analysis);

2) dividing the word (form) into syllables;

3) provide comment(s);

4) show antonyms;

5) synonymizer (a program to present synonyms to the word entered in the search);

6) presentation of homonyms;

7) giving the pronunciation (paronym) of the word with its explanation;

8) providing information from the "Native Language Encyclopaedia" related to the searched word;

9) display phrases in which the searched word is included;

10) it has the ability to display a range of ratings for various features.

At the moment, the volume of linguistic data of the educational corpus of the Uzbek language has the following indicator:

№	The name of the base unit	Count
1	Books	184
2	Internet texts	4 78 908
3	Phrases are tagged words	36 897
4	Words	5600
5	Contexts	5562
6	Number of words annotated	4575
7	Phrases	1384
8	Word order	413
9	Synonymous words	993
10	Antonym words	870
12	Homonyms	1636
13	Concepts in "Encyclopedic Dictionary of Mother Tongue"	1636

Various studies are being carried out to enrich the educational corpus, to enrich its base with more useful and necessary information for the student. The most popular of these studies is the study of students' speech and corpora. Student corpora research is a relatively young but highly developed field of corpus linguistics.

In conclusion, it can be said that using the positive results and experiences achieved in the educational corpora of student speech studied above, the creation of corpora of student speech in the educational corpus of the Uzbek language and the formation of methodical developments on bub are among the priority tasks in the future.

Foydalanilgan adabiyotlar:

1. Abdurakhmonova N, Urdishev K. Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL). 2019;6(1-2019):131-7.
2. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. InProceedings of the International Conference on Language Technologies for All (LT4All) 2019.
3. Abdullah Alfaifi, Eric Atw, Hedeya Ibrahim. Arabic Learner Corpus (ALC) v2. A New Written and Spoken Corpus of Arabic Learner2014 Saudia Arabia
4. D. Stewart, S. Bernardini, G. Aston Introduction: Ten years of TaLC / Corpora and language learners, Vol 17, John Benjamins Publishing Company Amsterdam/Philadelphia, 2004, - P. 3
5. Larisa Goulart Da Silva Academic vocabulary: a corpus linguistics study on how Brazilian students write academic English. 2016-a. Brazilia
6. Abdurakhmonova, N., & Urdishev, K. (2019). Corpus based teaching Uzbek as a foreign language. *Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL)*, *6*(1-2019), 131-7.
7. Абдурахмонова, Н., & Абдувахобов, Г. (2021). O ‘quv lug ‘atini tuzishning nazariy metodologik asoslari. *Международный журнал искусство слова*, *4*(6).

8. Abdurakhmonova, N. (2021). Formal-Functional Models of The Uzbek Electron Corpus. *ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies*, *10*(8), 59-66.
9. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
10. Abdurakhmonova, N., Shakirovich, I. A., & O‘G‘Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. *Science and Education*, *3*(9), 146-156.
11. Granger, S. (2003a). Error-tagged learner corpora and CALL: A promising synergy. CALICO Journal, 20(3): 465-480.
12. Granger, S. (2003b). The international corpus of learner English: A new resource for foreign language learning and teaching and second language acquisition research. TESOL Quarterly, 37(3): 538-546.
13. Hammarberg, B. (2010). Introduction to the ASU corpus, a longitudinal oral and written text corpus of adult learners’ Swedish with a corresponding part from native Swedes. Stockholm University:Department of Linguistics.
14. Thompson, P. (2005). Spoken language corpora. In M. Wynne (Ed.), Developing Linguistic Corpora: a Guide to Good Practice (pp. 59-70). Oxford: Oxbow Books.

KORPUSGA ASOSLANUVCHI GRAMMATIK VA LEKSIKOGRAFIK TADQIQOTLAR

Rajabova Karomat Chorshanbi qizi
O‘zMU, Kompyuter lingvistikasi mutaxassisligi magistranti.

Annotatsiya: Ushbu maqolada korpusga asoslanuvchi grammatik va leksikografik tadqiqotlar, korpus lingvistikasi oldida turgan asosiy vazifalar, lingvistik masalalar o‘z aksini topgan.Leksikografiya, ideografik lug‘atlarning korpus lingvistikasidagi o‘rni va ahamiyati misollar orqali ko‘rsatilib, ular haqida mulohazalar bildirilgan.

Kalit so‘zlar: korpus lingvistikasi, leksikografiya, ideografik lug‘at, ideografik teglash, tezaurus

Annotation: Grammatical and lexicographic studies based on the corpus, the main tasks and linguistic issues facing the corpus linguistics are reflected in this article.

Keywords: corpus linguistics, lexicography, ideographic dictionary, ideographic tagging, thesaurus

Аннотация: В статье отражены грамматические и лексикографические исследования на основе корпуса, основные задачи и лингвистические проблемы, стоящие перед корпусной лингвистикой.

Ключевые слова: корпусная лингвистика, лексикография, идеографический словарь, идеографическая разметка, тезаурус

Bugungi kunda til korpuslarida turli xil tadqiqotlar olib borilmoqda. Ular orasidan leksikografik tadqiqotlarni alohida ko‘rsatish o‘rinlidir. Buyuk lug‘atshunoslardan biri V.I.Dal shunday yozadi: “Алфавитный способ расположения слов крайне туп и сух. Самые близкие и сродные речения разносятся далеко врозь и томятся тут и там в одиночестве; всякая живая связь речи разорвана и утрачена, читать такой словарь нет сил, на десятом слове ум притупеет и голова вскружится, потому что ум наш требует во всем какой нибудь разумной связи, постепенности и последовательности” [С.Баранов.1995]. Ya’ni, V.I.Dalning fikricha, so‘zning lug‘atda joylashish tamoyili lug‘atning o‘quvchi uchun foydalanish qulayligi bilan belgilanadi.

Til ongimizda sistema sifatida yashashini bilamiz. N.V. Krushevskiy so‘zlarni lug‘atda joylashtirish masalasiga to‘xtalar ekan: “Hech qachon tilning asosiy xarakterini unutmaslik lozim. So‘z–narsaning belgisi. So‘z, narsa haqidagi tasavvur assotsiatsiya haqidagi qonun asosida juftlik hosil qilgan. Shunday ekan, so‘zlar bizning ongimizga ham o‘zi anglatadigan jismlar guruhi sifatida kirishi kerak” [Крушевский Н. 1998: 296], – degan edi. Leksikani ideografik nuqtayi nazardan o‘rganishning asosiy maqsadi so‘zning turli ma’nolarini bus-butun, yaxlitligicha o‘rganish, ularning o‘z o‘rnida qo‘llash va qo‘llanishini ta’minlashdir. So‘zlarni semantik maydonga yig‘ib o‘rganish ularning turli xususiyati, leksik ma’nosining emotsional ta’siri, boshqa jihatini kuzatib, leksik ma’nosini o‘zaro qiyoslab o‘rganishda muhim omil sifatida xizmat qiladi [Abdurakhmonova, N. 2021,2022]. Semantik maydonning aniq chegarasi mavjud emas, bu tabiiy hol, chunki bir so‘z turli pozitsiyada turlicha talqin qilinishi mumkin. Masalan, **olma** *mevalar nomi*

semantik maydoni va **olma** – *daraxtlar nomi* semantik maydoni yoki *o‘simliklar dunyosi* mavzuviy qurshovi tarkibida hamishtirok etishi mumkin.

Har bir tilning leksikasi o‘sha tilda gapiruvchi uchun faol, nofaol holatda bo‘ladi. Faol qatlamga kundalik so‘zlashuv so‘zlari kiradi, nutq jarayonida u so‘zlarni miyamizdan qidirib o‘tirishga hojat yo‘q, bu so‘zni aytishga ehtiyoj tug‘ilgan zahotiyoq talaffuz etamiz. Nofaol so‘z esa biz biladigan, lekin kundalik nutqimizda kam ishlatadigan so‘zlarimizdir. Faol so‘zda ma’no, talaffuz orasidagi munosabat ikki tomonlama: M.T. Nofaol so‘z orasidagi aloqa bir tomonlama, chunki uni o‘qiymiz yoki biror kishi nutqidan eshitamiz, uning nutqimizda faollashuvi tez sodir bo‘lmaydi, uni eslash kerak [Abdurakhmonova, N. 2021,2022]. Demak, T→M. Bundan kelib chiqadiki, eshitganimiz, o‘qiganimizni tushunishga yordam beruvchi lug‘at (alifbo lug‘ati) foydali, so‘zni butun mohiyati, sistemasi bilan ko‘z oldimizga keltiruvchi lug‘at (ideografik lug‘at) esa ma’lum bir til imkoniyatidan bemalol, oson foydalanish imkoniyatini beradi. Shuning uchun bunday lug‘at ham muhim axborot banki sirasiga kiradi. Yu.N. Karaulov ideografik lug‘atga shunday ta’rif beradi: “Barcha tashkil qiluvchi birliklari orasida kuchli semantik munosabat yaqqol ko‘rinuvchi lug‘atni tezaurus deb ataymiz” [Караулов Ю.1990]. Oddiy izohli lug‘atdan biror so‘zni qidirmoqchi bo‘lsak, uning moddiy tomoni(harflar birikmasi)ga e’tibor beramiz. Bunday lug‘atga kalit bitta: alifbo. Yu.N.Karaulov alifbo lug‘ati, tezaurus haqida shunday misol keltiradi: “D.N.Ushakov lug‘atida ***dissoniruyushiy*** so‘zi *dissonirovat, dissotsiatsiya* so‘zlari orasida keladi. Bu so‘z musiqa sohasiga oid, bu sohaga tegishli barcha so‘zlar haqida bilimga ega bo‘lish uchun butun lug‘atni o‘qib chiqish lozim. X.Kasares tezaurusi195da esa bu so‘z “Musiqa” maqolasi ichida quyidagi sifatlar qatorida keladi: “... *мелодичный, гомофонический, консонирующий, созвучный, аккордийный, гармонический, диатонический, полухроматический, энгармонический, полифалический, характерный, фугированный...”*. Bu maqola 508 so‘zni qamrab olgan. Qisqacha aytganda, butun lug‘atni emas, undagi atigi 1-2 betni varaqlasak, kerakli so‘z topiladi. Demak, tezaurusda sistema, semantika markazda turadi. Kalit so‘zlar esa alifbo tartibida bo‘ladi. Ideografik lug‘atning yana bir ahamiyatli tomoni shundaki, u turli tillar leksikasini qiyosiy o‘rganishda muhim ahamiyatga ega, tarjima lug‘ati tuzishda qo‘l keladi. Chunki ma’lum tildagi narsa bir so‘z bilan atalsa, boshqa tilda bir necha so‘z bilan izohlanuvchi birlikka to‘g‘ri kelishi mumkin. Yevropada ideografik lug‘atga qiziqish uyg‘otgan “Ingliz so‘z, iboralari tezaurusi” (“Тезаурус английских слов и выражений”) muallifi P.M.Rodjet o‘z lug‘ati so‘z boshisida aytadi: “Bog‘li matn tuzayotgan yozuvchi yoki xohlagan kishi nima demoqchiligini yaxshi bila turib, ayni paytda o‘rinli ibora yoki so‘zni topolmasa, boy, kerakli so‘z tez topiladigan, to‘g‘ri sistemalashtirilgan, guruhlarga bo‘lingan bir kitobdan yordam oladi. Bir so‘z bilan aytganda, tezaurusga murojaat etadi... Biz aytmoqchi bo‘lgan fikrimizni qanchalik aniq tasavvur qilmaylik, ba’zan shunday bir vaziyatga tushib qolamizki, bizga lug‘at boyligimizdagi so‘z yetishmay qoladi. O‘z fikrimizni aniq ifodalash uchun bitta so‘z o‘rniga bir necha ijobiy, salbiy, kuchli, kuchsiz, ba’zan esa keraksiz so‘zlarni ishlatib yuboramiz. Ideografik lug‘atning eng muhim xususiyati shundaki, u foydalanuvchiga bir fikrning barcha qirralarini ochib beruvchi so‘z, iboralarning boy to‘dasini tortiq qiladi”. Demak, ideografik lug‘at maqolalari odatdagidek alifbo tartibida emas, balki ma’no (bosh so‘z, so‘z birikmaning leksik ma’nosi) asosida tartib berilgan lug‘atdir. Alifbo tartibli lug‘atdan aynan bir so‘z ma’nosini bilish mumkin bo‘lsa, ideografik lug‘at biror tushuncha haqida, bir ma’noni qanday so‘zlar bilan ifodalash mumkinligi borasida ma’lumot beradi[Abdurakhmonova, N. 2021,2022].

Leksikografiya ideografik lug‘atlar tuzish borasida katta yutuqlarga erishdi, uning ideografik tezaurus, analogik lug‘at, o‘quvtematik lug‘at kabi turlarini yaratdi. Leksikografiya tarixida ingliz, fransuz, nemis, ispan tillari materiali asosida tuzilgan lug‘atlar anchagina. Bu siraga rus
leksikograflari K.Plots, K.Babov, A.Vorgulev, L.Rozkovsova, Z.Ganusova, V.V.Morkovkin, Yu.N.Karaulov, nemis leksikografi G.O.Fogt, ingliz leksikograflari Dj.Roudel, P.M.Rodjetlar ishlarini kiritish mumkin. Ideografik lug‘atning ahamiyati korpusda ideografik razmetkaga ehtiyoj borligini ko‘rsatadi. Muallif asari mavzularini tasnif qilish natijasida birliklar orasidagi mantiqiy aloqadorlik asosida birlashgan so‘zlar to‘plamini aks ettiruvchi ideografik lug‘at paydo bo‘ladi. A.P.Chexov nasriy, dramatik asarlarining mavzuviy razmetkasi xususiyatlari E.V.Surovseva tomonidan alohida tadqiq etilgan. Bunday mavzuviy razmetkani amalga oshirish uchun uning 583ta nasriy, 17 ta dramatik asari matni tanlangan. Tahlilga olingan matnlar uchun muallifning mukammal asarlari to‘plami (Чехов А.П. Полное собрание сочинений и писем. В 30 томах. М.: Наука, 1974-1983.) asos qilib olingan. Ideografik teglash uchun avtomatik algoritm – maxsus dastur bo‘lmaganligi sababli teglash qo‘lda amalga oshiriladi. Chunki har bir muallif asari mavzu doirasi turlicha bo‘ladi. Avval muallifning barcha asarlari matni tahlil qilinib, keyin umumiy xulosaga kelinadi. Mavzuviy teg matn mavzusi asosida leksik-semantik guruhlar shakllantirishni maqsad qilib belgilaydi. Matn tanlanib, undan ma’lum bir mavzu atrofida birlashgan so‘zlar belgilanadi. So‘zlarni belgilashda matnni qayta aytib berish uchun butun matn mazmunini to‘liq ifodalashga xizmat qiluvchi so‘z, iboralar ajratiladi. Aynan shu matndan boshqa bir leksik-semantik guruhga (LSG) mansub so‘zlar alohida rang bilan belgilanadi. Masalan, E.V.Surovseva A.P.Chexov asarlari ideografik razmetkasi uchun yozuvchining “Semiz va oriq” hikoyasini obyekt sifatida tanlaydi. Birinchi bosqichda hikoyadan *uchrashuv*, ikkinchi bosqichda *bolalik, xotira* mazmuniy guruhiga kiruvchi so‘zlarni alohida ranglar bilan belgilaydi, *do‘stlik* mazmuniy guruhiga kiruvchi so‘zlarni kursivda ajratadi. Shunday tahlildan keyin ko‘rinadiki, LSG bir xil emas. Bu guruhlar *holat, mavhum tushuncha, qahramon tipi* yoki *voqea joyi* kabi guruhlar bo‘lishi mumkin. Demak, bitta hikoyada bir necha LSGga kiruvchi so‘zlar guruhi aniqlandi. Faqat bitta LSGga kiruvchi so‘zlardan iborat asar (hikoya) kam uchraydi. Masalan, “Bahorni kutish”da faqatgina *uchrashuv,* “Odamning qisqacha anotomiyasi”da *ilm,* “Parcha”da *provinsiya* kabi LSGlar mavjud. Ba’zi hikoyalarida esa o‘nlab LSG aniqlangan. Asarda hikoyachi, qahramonda xayol surish, fikr yuritish, tahlil qilishga moyillik bo‘lsa, o‘sha asarni mavzuviy razmetkalash oson kechadi. Barcha asarlar tahlil qilingandan so‘ng mavzular ikkiga: makro, mikromavzuga ajratilib, ba’zi mavzular bir vaqtning o‘zida ikki makromavzuga tegishli bo‘lib qolishi mumkin. Deylik, *hayit* LSGi mazmuniga ko‘ra ham *din,* ham *vaqt* makromavzusiga kiritilishi mumkin. E.V. Surovsevaning fikricha, korpusda bu tipdagi razmetkaning mavjudligi rus tili (umuman olganda, barcha tilni)ni xorijiy til sifatida o‘qitishning samaradorligini oshirishda, korpus materialidan ta’limiy maqsadda foydalanishda ahamiyatli. Chunki mavzulashtirilgan matn xorijiy til ta’limi jarayonida o‘quv materiali tayyorlashda axborot manbai vazifasini bajaradi. Hatto bu usulda razmetkalangan korpus yordamida jahon mualliflik kartinasini kuzatish qulayligi ham paydo bo‘ladi. Ideografik razmetkali korpus yordamida bir muallif asarlarini yaxlit holda g‘oyaviy-badiiy tahlil qilishda, u yoki bu mavzuni qaysi usul va vosita bilan ifodalaganligini o‘rganishda ham asqotadi. Masalan, A. Qahhorning inson, ma’naviyat mavzusi ko‘tarilgan barcha asarlarini bir prizmadan o‘tkazish orqali u yoki bu til ifoda vositalarini qo‘llash mahoratini ham o‘rganish

mumkin. Shuningdek, stilistik, lingvo-poetik, kognitiv, pragmatik, lingvomadaniy, konseptologik, diskursiv nuqtai nazardan o‘rganishda qo‘l keladi.

Xulosa o‘rnida shuni aytish mumkinki, har qanday yuksalayotgan fan o‘z amaliyotida kompyuterlar taraqqiyotining yutuqlaridan foydalanadi, lekin ushbu fanlar (korpus lingvistikasi, kompyuter lingvistikasi, kompyuter matematikasi, kompyuter fizikasi) kompyutershunoslikning bir qismi sifatida qabul qilinishi joiz emas. Mazkur fanlar tahlil ob’yektini tadqiq etishda kompyuterdan vosita sifatida foydalanadi, xolos. Shuni aytish mumkinki, korpus lingvistikasi kompyuter lingvistikasining alohida bir qismi yoki bo’limi emas. Ular yo‘nalish hamda tadqiqot ob’yekti, maqsad va muammolari turlicha bo‘lgan mustaqil soha hisoblanadi. Har bir tilning leksikasi o‘sha tilda gapiruvchi uchun faol, nofaol holatda bo‘ladi[Abdurakhmonova, N. 2021,2022]. Faol qatlamga kundalik so‘zlashuv so‘zlari kiradi, nutq jarayonida u so‘zlarni miyamizdan qidirib o‘tirishga hojat yo‘q, bu so‘zni aytishga ehtiyoj tug‘ilgan zahotiyoq talaffuz etamiz. Nofaol so‘z esa biz biladigan, lekin kundalik nutqimizda kam ishlatadigan so‘zlarimizdir. Faol so‘zda ma’no, talaffuz orasidagi munosabat ikki tomonlama: M↔T. Nofaol so‘z orasidagi aloqa bir tomonlama, chunki uni o‘qiymiz yoki biror kishi nutqidan eshitamiz, uning nutqimizda faollashuvi tez sodir bo‘lmaydi, uni eslash kerak. Demak, T→M. Bundan kelib chiqadiki, eshitganimiz, o‘qiganimizni tushunishga yordam beruvchi lug‘at (alifbo lug‘ati) foydali, so‘zni butun mohiyati, sistemasi bilan ko‘z oldimizga keltiruvchi lug‘at (ideografik lug‘at) esa ma’lum bir til imkoniyatidan bemalol, oson foydalanish imkoniyatini beradi. Shuning uchun bunday lug‘at ham muhim axborot banki sirasiga kiradi. Yu.N. Karaulov ideografik lug‘atga shunday ta’rif beradi [Караулов Ю.1990]: “Barcha tashkil qiluvchi birliklari orasida kuchli semantik munosabat yaqqol ko‘rinuvchi lug‘atni tezaurus deb ataymiz”. Oddiy izohli lug‘atdan biror so‘zni qidirmoqchi bo‘lsak, uning moddiy tomoni(harflar birikmasi)ga e’tibor beramiz. Kompyuter texnologiyalarining rivoji natijasida zamonaviy tilshunoslikda korpus va kompyuter lingvistikasi terminlarini semantik tahlil, korpus lingvistikasi, kritik diskurs tahlil, shuningdek, deskriptiv, qiyosiy va komponent tahlil metodlari asosida tadqiq etish tobora dolzarb ahamiyat kasb etib bormoqda. Tilshunoslik nuqtai nazaridan yondashsak, kompyuter lingvistikasining paydo bo’lishi bugungi kunda fanning bir qancha muammolariga yechim topish imkoniyatini beradi.

Foydalanilgan adabiyotlar:

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ,2013. – 161 с.
2. Идеографический словарь русского языка. (Сост. О.С.Баранов.) – М., 1995. – 4166 статей.
3. Информационные технологии в лингвистике: учеб. Пособие / Л.Ю.Щипицина. – М.: ФЛИНТА: Наука, 2013. –128 с.
4. Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. – М.: Прометей, 1990.
5. Структурно-функциональная компьютерная модель татарских морфем. – Казань: Фэн, 2003. – 220 с.
6. Абдурахмонова, Н., & Абдувахобов, Г. (2021). O ‘quv lug ‘atini tuzishning nazariy metodologik asoslari. *Международный журнал искусство слова*, *4*(6).

7. Abdurakhmonova, N. (2021). Formal-Functional Models of The Uzbek Electron Corpus. *ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies*, *10*(8), 59-66.
8. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
9. Abdurakhmonova, N., Shakirovich, I. A., & O‘G‘Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. *Science and Education*, *3*(9), 146-156.
10. Информационные технологии в лингвистике: учеб. Пособие / Л.Ю.Щипицина. – М.: ФЛИНТА: Наука, 2013. –128 с.
11. Ҳамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
12. Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011
13. Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.
14. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистикасослари. Монография. – Германия: GlobeEdit, 2020. – 250 б
15. Шаров С.А. Представительный корпусрусского языка в контексте мирового опыта // https//lamb.viniti.ru
16. www.ruscorpora.ru
17. www.uzbekcorpus.uz
18. www.philol.msu.ru/~lex/corpus
19. http://www.philol.msu.ru/~lex/corpus
20. Ataboev N.B. Chomsky's Critics to Corpus and the Problem of Expert Speakers' Linguistic Intuition // Филология масалалари. – Тошкент. (4)/2015.– Б. 108-111.
21. Атабоев Н.Б. Корпус тилшунослиги шаклланишининг диахроник тадқиқи // Бердақ номидаги Қорақалпоқ давлат университетининг Ахборот-номаси. – Нукус. №4/ 2018. – Б. 147-150.

BOLALAR NUTQI KORPUSI SHAKLLANISH BOSQICHLARI

Normamatova Mashhura
O‘zbekiston Milliy Universiteti 2-bosqich talabasi.
E-mail: mashhuranormamatova0@gmail.com

Ilmiy rahbar: **Abduraxmonova Nilufar**
Mirzo Ulug‘bek nomidagi O‘zbekiston milliy universiteti
Professori, filologiya fanlari doktori.
E-mail: n.abduraxmonova@nuu.uz

Annotatsiya: Ushbu maqolada bolalar nutqi korpuslarining mohiyati hamda bolalar nutqi korpuslarining shakllanish bosqichlari borasida fikr yuritiladi. Bugungidek jadal rivojlanib borayotgan dunyoda korpus har qanday sohada muhim ahamiyat kasb etib borayotgan universal manbadir. Korpuslar rivojlanish tarixiga nazar tashlaydigan bo‘lsak, dastlabki korpuslar tilshunoslik sohasida vujudga kelgan. Bugungi kunda esa, korpuslar nafaqat tilshunoslikda balki, tibbiyot, pedagogika, psixologiya yo‘nalishlarida ham o‘z o‘rnini egallashda davom etmoqda. Shuningdek, korpuslar til o‘qitish sohasida ham muhim ahamiyatga molik hodisa bo‘ldi. Bu borada bolalar nutqi korpuslari, ayniqsa, katta rol o‘ynamoqda. Maqolada bolalar nutqi korpuslaridan ayrimlariga qisqacha to‘xtalib o‘tildi. Keltirilgan korpuslar korpus lingvistikasining bolalar nutqi korpusi yo‘nalishidagi eng muvaffaqiyatli ishlar sirasiga kiradi.

Kalit so‘zlar: bolalar nutqi korpusi, nutqni avtomatik aniqlash, CHILDES, Brayan MakVinni, TaklBank, Demut Sesoto korpusi, Ketrin Demuth.

Abstract: This article examines the nature of children's speech corpora and the stages of formation of children's speech corpora. In today's rapidly developing world, the body is a universal resource that is becoming increasingly important in any field. If you look at the history of the development of corpora, the first corpora were created in the field of linguistics. Today, corpora continue to occupy their place not only in linguistics, but also in medicine, pedagogy and psychology. In addition, corpora have become an important phenomenon in the field of language teaching. Children's speech corpora play a particularly important role in this regard. The article briefly touches on some corpora of children's speech. The given corpora are among the most successful works of corpus linguistics in the field of corpora of children's speech.

Keywords: child speech corpus, automatic speech recognition, CHILDES, Brian McWhinney, TalkBank, Demuth Sesotho Corpus, Catherine Demuth.

Аннотация: В данной статье рассматривается природа детских речевых корпусов и этапы формирования детских речевых корпусов. В современном быстро развивающемся мире корпус является универсальным ресурсом, приобретающим все большее значение в любой сфере. Если взглянуть на историю развития корпусов, то первые корпуса были созданы в области лингвистики. Сегодня корпусы продолжают занимать свое место не только в лингвистике, но и в медицине, педагогике и психологии. Кроме того, корпуса стали важным явлением в сфере преподавания языка. Особенно важную роль в этом отношении играют детские речевые корпуса. В статье кратко затронуты некоторые корпуса детской

речи. Приведенные корпуса относятся к числу наиболее успешных работ корпусной лингвистики в области корпусов детской речи.

Ключевые слова: корпус детской речи, автоматическое распознавание речи, CHILDES, Брайан Маквинни, TalkBank, Demuth Sesotho Corpus, Кэтрин Демут,

So‘ngi bir necha o‘n yilliklar davomida korpus lingvistikasi doirasida samarali ishlar qilindi. Bu davr mobaynida olimlar tomonidan turli korpuslar ishlab chiqildi. Biroq, korpus lingvistikasi borasidagi ko‘pgina tadqiqotlar kattalar nutqiga asoslangan edi. Bolalar nutqi kattalar nutqiga nisbatan artikulyar, akustik, jismoniy va lingvistik xususiyatlariga ko‘ra murakkab ekanligi bilan ancha murakkab jarayonlarni taqozo etardi. Ehtimol shu bois, bir muncha vaqt bolalar nutqi bilan ishlash e‘tibordan chetda qoldi. Bu esa ushbu soha tadqiqot ishlarining eng e’tiborli nuqtasi ekanini anglatadi.

Nutq korpuslari borasidagi ishlarni o‘rganish jarayonida tadqiqotchida “bolalar nutqi korpusining mohiyati nimadan iborat?” degan savol tug‘ilishi tabiiy. Dastlab ushbu savolga atroflicha javob berishga harakat qilamiz.

Bir vaqtlar elektron ta’lim kelajak ta’limida nuhum rol o‘ynaydi deb qaralgan bo‘lsa-da, u an’anaviy ta’lim tizimida tez sur’atlarda o‘zgartirib yuborishi kutilmagan edi. COVID-19 pandemeyasi natijasida onlayn o‘qitish dasturlari, til ilovalari, video konferensiyalar, virtual repetitorlik muhim o‘qitish vositalariga aylandi. Barcha o‘rganuvchilar, jumladan, eshitish imkoniyati cheklangan bo‘lgan o‘rganuvchilar uchun real vaqt rejimida taglavhalar (subtitr)ni taqdim etish uchun nutqni aniqlash texnologiyalariga yanada ehtiyoj kuchaydi. Nutqni aniqlash texnologiyasi qo‘l yozuvi yoki imlo bilan bog‘liq qiyinchiliklarga duch kelgan bolalar uchun ham qulaylik yaratadi. Disleksiya yoki boshqa o‘rganishda nuqsonlari bo‘lgan bolalar qo‘lda yozish o‘rniga, o‘z ovozlaridan foydalanib yozish uchun nutqdan matnga texnologiyasidan foydalanishi mumkin. Bolalar – multimedia o‘yinlari, o‘quv dasturlari va o‘quv materiallari nuqtai nazaridan avtomatik nutqni aniqlash texnologiyalaridan eng samarali foydalanadigan qatlamdir. Mana shu jihatlarni inobatga olgan holda bolalar nutqi korpusi korpus lingvistikasining muhim yo‘nalishlaridan biri desak mubolag‘a bo‘lmaydi.

Korpus lingvistikasi tarixiga nazar tashlaydigan bo‘lsak, bolalar nutqi korpusi mavzusi nisbatan yangi masalalardan hisoblanadi. Shunga qaramasdan, nutqni avtomatik aniqlash texnologiyasiga asoslangan holda ko‘plab bolalar nutqi korpuslari ishlab chiqildi. Quyidagi jadvalda ulardan eng muvaffaqiyatlilarini ko‘rishingiz mumkin.

	Korpus nomi	Muallif	Tillar	Ma’ru-zachilar	Davo-miyligi	Yosh oralig‘i	Sana
1	CHILDES	Brayan MakVinni, Ketrin Snou	25 ta til			turlicha	1984
2	Demut Sesoto korpusi	Ketrin Demuth	Sesoto	4	98 soat	2-4	1992
3	TIDIGITS	R.Gary Leonard	Ingliz	326 (101 bola)		6-15	1993
4	CMU Kids Corpus	Maxine Eskinaze	Ingliz	76		6-11	1997
5	CU bolalar o‘qish va taklif qilingan nutq korpusi	Andreas Hagen	Ingliz	663		K – G5	2001
6	CU Story Corpus	Andreas Hagen	Ingliz	106	40 soat	G3–G5	2003
7	TABLO	Abe Kazemzadeh	Ingliz	256	40 soat	K-G4	2005

8	Shvetsiya NICE Korpusi	Linda Bell	Shvet	5580		8-15	2005
9	PF-STAR bolalar nutqi korpusi	Martin Rassel	Ingliz	158	14,5 soat	4-14	2006
10	Providence Corpus	Ketrin Demuth	Ingliz	6	363 soat	1-3	2006
11	CSLU Kids' Speech Corpus	Khaldoun Shobaki	Ingliz	1100		K-G10	2007
12	Lion korpusi	Ketrin Demuth	Fransuz	4	185 soat	1-3	2007
13	CHIEDE	Marta Garrote	Ispan	59	8 soat		2008
14	CASS_CHILD	Jun Gao	Mandarin	23		1-4	2012
15	CFSC	R.M.Paskual	Filippin	57	8 soat	6-11	2012
16	CALL-SLT	Manny Rayner	Nemis				2014
17	SingaKids – Mandarin	Nensi F.Chen	Mandarin	255	125 soat	7-12	2016
18	Boulder Learning-MyST Corpus	Cole va boshqalar	Ingliz	1371	393 soat	G3-G5	2019

Bolalar nutqi korpusi doirasida yaratilgan dastlabki loyiha – CHILDES korpusi edi. CHILDES (Child Language Data Exchange System) 1984-yilda Karnegi Mellon universitetining psixologiya bo‘limi doktorlari Brayan MakVinni va Ketrin Snoular tomonidan yaratilgan. Dastlab, ushbu dasturning maqsadi bolalar tilini o‘zlashtirish transkriptlari va yozuvlarini tadqiqotchilarga bepul, ommaviy axborot sifatida taqdim qilishdan iborat edi. Keyinchalik esa bu dastur korpus sifatida shakllantirildi. Uning eng dastlabki transkriptlari 1960-yillarga to‘g‘ri keladi. 1990-yillarning boshlarida CHILDESda topilgan ma’lumotlardan osonlikcha foydalanish imkoniyati keng tarqalgani tufayli, undan foydalangan holda bolalar tilini o‘zlashtirish bo‘yicha yaratilgan tadqiqotlar soni sezilarli darajada oshdi.

CHILDES transkriptlarida turli tillar va yoshdagilar mavjud. Transkriptlarning aksariyati ssenariylashtirilmagan o‘zaro suhbatlardan iborat. Ttranskripsiyalar suhbat transkriptlarini ishlab chiqarish uchun standartlashtirilgan formatni ta’minlovchi CHAT transkripsiya formatida kodlangan. Bu tizim har qanday til o‘rganuvchilar: bolalar, ikkinchi tilni o‘rganuvchilar va tiklanuvchi afaziklar bilan suhbatlarni transkripsiya qilish uchun ishlatilishi mumkin. Diskurs darajasidagi transkripsiyadan tashqari, CHAT tizimi fonologik va morfologik tahlil imkoniyatlariga ham ega. CHILDES asosan yosh bolalarning tilini tahlil qilish uchun ishlatiladi va kattalarning bolalar nutqiga munosabatiga qaratilgan.

Ayni paytda korpusda 30 ta tildagi 230 ta korpus mavjud bo‘lib, u nisbatan kattaroq TalkBank korpusining tarkibiy qismiga aylantirildi. Loyihada ma’lumotlar bazasidan foydalangan 4000 nafar bolalar tili tadqiqotchilari va talabalar ro‘yxati mavjud hamda ushbu materiallardan foydalanish asosida chop etilgan 3000 dan ortiq maqolalar qayd etilgan. Loyiha, shuningdek, tildan foydalanishning turli darajalari, jumladan, leksika, sintaksis, morfologiya, fonologiya, nutq va hikoyani o‘rganish uchun foydali bo‘lgan kompyuter dasturlari to‘plamini yaratdi.

Bolalar nutqi korpusi borasida olib borilgan loyihalardan ahamiyatlisi sifatida Demut Sesoto korpusini ham keltirish mumkin. Ushbu korpus 1980-82-yillarda Afrikaning janubidagi Lesoto davlatida Ketrin Demuth tomonidan tuzilgan. Ushbu korpusni yaratishdan maqsad past xarajatli jarayon (ya’ni, professional bo‘lmagan ko‘ngillilar va norasmiy yozib olish muhitlari yordamida yozilgan bir nechta ovozlardan foydalangan holda yuqori sifatli nutqni avtomatik aniqlash tizimini joriy etishni tekshirish edi. Ma’lumotlar Moxotlong tumanidagi 550 kishidan iborat kichik Lesoto tog‘ qishlog‘ida to‘olangan. Natijada lug‘aviy fe’llar, ½ million morfemalar va taxminan 13,250nta iborani o‘z ichiga olgan 98 soatlik nutq korpusi paydo bo‘ldi. Ushbu

materiallar qayta ishlangach CHILDES formatiga o‘tkazildi va korpusning tarkibiy qismiga qo‘shildi.

Bugungi kungacha yaratilgan va yaratilayotgan bolalar nutqi korpuslarining har biri nafaqat ta’lim, balki tibbiyot sohasidagi ko‘plab muammolarga ham o‘ziga xos yechimlar taqdim etmoqda. Bu jihatdan ularning har biri o‘ziga xosdir. Biroq, ularning yaratilishiga asos bo‘lgan CHILDES korpusi bugungi kunda ham yangi korpuslar uchun o‘ziga xos qolip hamda boy ma’lumotlar bazasi vazifasini o‘tamoqda desak, mubolag‘a bo‘lmaydi.

Foydalanilgan adabiyotlar:

1. Kodner, Jordan. 2019. Estimating child linguistic experience from
2. historical corpora. Glossa: a journal of general linguistics 4(1):
   122. 1–14. DOI: https://doi.org/10.5334/gjgl.926
3. Abdurakhmonova, N. (2021). Formal-Functional Models of The Uzbek Electron Corpus. *ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies*, *10*(8), 59-66.
4. Sharipbay, A., Shirinova, R., Abdurakhmonova, N., & Qodirova, Z. (2023). Ontological knowledge-based models representing medicine terminology. *Advances in Social Sciences and Economics*, *1*(1), 4.
5. Gatiatullin, A., Suleymanov, D., Prokopyev, N., & Abdurakhmonova, N. (2022, November). “Turkic Morpheme”: From the Portal to the Linguistic Platform. In *World Conference Intelligent System for Industrial Automation* (pp. 181-189). Cham: Springer Nature Switzerland.
6. Dusmurod, N., Nilufar, N., Alimova, S., Abdurakhmonova, N., & Karabayeva, D. (2019). Social issues of rational nutrition.
7. http://icame.uib.no/ij35/Hanhong_LI_and_Alex_C_Fang.pdf
8. https://childes.talkbank.org/access/Other/Sesotho/Demuth.html
9. https://childes.talkbank.org/access/Eng-NA/0docs/MacWhinney1990.pdf
10. https://osf.io/zqb6c/wiki/home/

JAHON KORPUS LINGVISTIKASIDA DIALEKTAL KORPUSLAR

Vosiljonov Azizbek
Farg‘ona davlat universiteti
Ijtimoiy-gumanitar fanlar kafedrasi o‘qituvchisi.

Annotatsiya: Ushbu maqolada Britaniya ingliz dialektlarining Xelsinki korpusi (HD) hamda Chexiya milliy korpusining chex dialektlari korpusi misolida dialekt korpuslar konsepsiyasi zamonaviy til fanining yutuqlari va dialektologlar uchun mavjud bo‘lgan materiallarning tabiati bilan bog‘liq bo‘lgan xalq nutqini o‘rganish bo‘yicha maktabning asosiy yo‘nalishlarini hisobga olgan holda ishlab chiqilganligi yoritilgan.
Kalit so‘zlar: dialektologiya, dialekt korpus, intervyu, axborotchilar, geografik joylashuv, SED, yunikod.

Annotation: In this paper, using the examples of the Helsinki Corpus of British English Dialects (HD) and the Czech National Corpus of the Czech Dialects, the concept of dialect corpora relates to the progress of modern linguistics and the nature of the materials available to dialectologists in the school of vernacular studies. It is explained that it was developed taking into account the main directions.
Keywords: dialectology, dialect corpus, interview, informants, geographic location, SED, unicode.

Аннотация: В данной статье на примере Хельсинского корпуса британских английских диалектов (HD) и Чешского национального корпуса чешских диалектов концепция диалектных корпусов представляет собой исследование народной речи в связи с достижениями современной лингвистики и характер материала, доступного диалектологам, поясняется, что он был разработан с учетом основных направлений школы.
Ключевые слова: диалектология, диалектный корпус, интервью, информаторы, географическое положение, СЭД, юникод.

Milliy til va uning shevalarini informatsion dunyoga olib chiqish, uni elektronlashtirish nafaqat tilshunoslik, balki shevashunos tadqiqotchilar oldida turgan muhim masaladir. Muloqotning turli sohalarida milliy tillarning ishlashini ifodalovchi korpuslar bilan bir qatorda, milliy til doirasida ajralib turadigan alohida lingvistik jamoalardagi muloqotni taqlid qiluvchi korpuslar ham kerak. Bu tipdagi eng muhim til shakllanishlari shevalardir. Dialekt matnlarining mashina fondini yaratish zarurligi haqidagi g‘oyalar dastavval 1980-yillarda A.S.Gerd va V.E.Goldin tomonidan ilgari surildi[Goldin, 1990; 17]. Hozirgi vaqtda dunyo tilshunosligida dialektal nutqning alohida elementlarini ifodalovchi bir qator korpuslar mavjud: dialektal matnlarning xorijiy korpusi (masalan, ingliz dialektlarining Xelsinki korpusi, Kirkning Shimoliy Irlandiya transkripsiyalangan nutq korpusi (NITCS), IViE (ingliz tilidagi intonatsion o‘zgarishlar) korpusi, BBC Voices); rus tili milliy korpusining (NCRL) dialekt subkorpusi; Pustosha – Moskva viloyati, Shaturskiy tumani, shu jumladan matnlar – dialektdagi ma'ruzachilar nutqining namunalari shular jumlasidandir.

Britaniya ingliz dialektlarining Xelsinki korpusi (HD) asosan Sharqiy Angliya va Janubi-G‘arbiy hududlardan, Lankashirdan kichik to‘plamga ega bo‘lgan orfografik tarzda yozib olingan audio yozuvlar to‘plamidir.

Korpusning maqsadi dialektologiya, sotsiolingvistika, nutq tahlili, morfologiya, sintaksis va fonologiya sohalarida lingvistik tadqiqotlar uchun material berishdir. Korpus shuningdek, aloqa etnografiyasi, mahalliy urf-odatlar va tarix kabi tilga oid bo‘lmagan, ko‘p tarmoqli tadqiqotlar uchun material beradi. Korpusning asosiy ma’lumotlari og‘zaki dialekt nutqining audio yozuvlaridir Hajmi 1 008 641 so‘zdan iborat bo‘lib, jami 187 ta fayldan tashkil topgan. Birinchi bosqich 2006-yilda yakunlandi, ikkinchi bosqichi esa korpus materiallarini ilgari nashr etilmagan korpus ma’lumotlari orqali kengaytirish rejalari bilan davom etmoqda [Abdurakhmonova, N. 2021,2022]. Yozuvlar 1970 va 1980-yillarda Finlyandiya aspirantlari tomonidan qilingan, bu ishda professor Garold Orton va boshqa Lids olimlari maslahat bergan. Korpusda og‘zaki dialekt nutqining orfografik transkripsiyalari mavjud.

Namuna olish jarayonining tamoyillari ma’lum darajada mahalliy va ma’lumot beruvchilarni tanlash uchun *Lids ingliz dialektlari tadqiqoti (SED)* mezonlariga asoslangan. Ma’lumot to‘plash usuli sifatida, asosan, so‘rovnomadan foydalanish, ma’lumot beruvchining fikrlash oqimini to‘xtatib qo‘yish va nutqning uzoq cho‘zilgan qismini yozib olish usullari orqali ifoda etilgan. Ma’lumot beruvchilarning asosiy qismi qishloq aholisining “dala ishchilari” hisoblanadi va ularning yosh va jins kategoriyalarida ham turli xil farqlanishlar mavjud. Jumladan, boshqa dialektal korpuslardan farqli ravishda erkak informatorlar ayol informatorlarga nisbatan ko‘pchilikni tashkil qiladi. Yosh munosabati esa, asosan, nafaqa yoshidagi shaxslar hisoblanadi. Ma’lumot to‘plash birinchi navbatda tekshiriladigan joylar tarmog‘ini o‘rnatish va keyin ushbu joylarda tegishli ma’lumot beruvchilarni topishdan boshlanadi. Qishloq pochta bo‘limlari ma’lumot beruvchilarni topishda katta yordam beradi, chunki ular potentsial ma’lumot beruvchi bo‘lgan nafaqaxo‘rlar ro‘yxatini saqlaydilar. Do‘konlar ham yaxshi ma’lumot manbai sanaladi, chunki do‘kondorlar odatda qishloq aholisini yaxshi bilishadi. Bundan tashqari, qishloq hayotiga qiziquvchi mahalliy aholi bilan uchrashish ma’lumot to‘plashda katta yordam berdi, chunki ular dala ishchisiga potentsial ma’lumot beruvchilar haqida qo‘shimcha ma’lumot berishlari va hatto o‘zlari ham suhbatdosh sifatida yordam berishlari mumkin[Abdurakhmonova, N. 2021,2022]. Ular axborot beruvchilar nutqidagi notanish dialekt shakllarini farqlashda va izohlab berishda yordam bergan.

Har bir intervyu Britaniya ingliz dialektlarining Xelsinki korpusi (HD) da moslashtirilgan beshta parametr kodlari bilan ta’minlangan . Ular quyidagilar:

B: HD fayl kodi (<B DICAM13>)

N: tuman, qishloq, ma’lumot beruvchi ma’lumoti (|N_CAM_LANDBEACH_SJ)

Y: ma’lumot beruvchi(lar)ning yoshi(lar)i (<Y 86>)

X: ma’lumot beruvchi(lar)ning jinsi (<X WOMEN>)

H: ma’lumot beruvchi(lar)ning kasbi(lar)i (<H HOUSEWIFE>)

Asosiy parametr kodlaridan keyin quyidagi tafsilotlarni ham ko‘rish mumkin:

- raqamli arxiv kodi (Dig. CAM32A)
- tasma uzunligi daqiqalarda (47:37 min)
- dala ishchisining bosh harflari bilan yozilgan yil yoki sana (Rec 1974, AO tomonidan)
- transkriptorning bosh harflari (AO)
- jami va olib tashlangan so‘zlar soni (bizning WC 6,075, tozalangan WC 5,796)

● arxiv sahifalari soni (41 arxiv sahifasi)
● yakuniy kompyuter qo‘lyozmasi sanasi (CMS 9.1.2001)

Umuman olganda, Britaniya ingliz dialektlarining Xelsinki korpusi (HD) Britaniya ingliz tili dialektlarini ilmiy, ta’limiy maqsadlarda o‘rganishda maqbul qo‘llanma hisoblanadi.

Chexiya milliy korpusining chex dialektlari korpusi. 2017-yilda Chexiya Milliy Korpusi instituti o‘zining “Og‘zaki korpus” bo‘limi tomonidan tayyorlangan Chexiya milliy korpusining yangi ixtisoslashtirilgan chex dialektlari korpusini taqdim etdi[Goláňová H. Waclawičová M. 2018; 79]. Korpus Chexiya Respublikasi hududida mavjud bo‘lgan an'anaviy mintaqaviy dialektlarni qamrab oladi va taqdim etadi. Uning birinchi ommaviy versiyasida korpusning o‘lchami taxminan 100 000 so‘zdan iborat, lekin ko‘proq ma’lumotlar doimiy ravishda to‘plangan va kengaytirilgan versiyalari kelajakda nashr etiladi.

Auditoriya nuqtai nazaridan, dialekt korpusi til mutaxassislari (dialektologlar, boshqa tilshunoslar va tegishli sohalardagi tadqiqotchilar), shuningdek, keng ommaning havaskorlarini qamrab olishni maqsad qilgan. Shuningdek, u barcha ta’lim darajalarida o‘quv manbasi bo‘lib xizmat qiladi. Dialekt korpusi Chexiya Respublikasining barcha an’anaviy dialekt hududlarida qilingan nutq yozuvlari va ularning transkriptlaridan iborat. Joriy versiyada umumiy uzunligi 13 soat bo‘lgan 324 ta yozuv mavjud bo‘lib, ular 178 ta noyob dinamikni o‘z ichiga oladi. Ma’ruzachilar soniga dialektologik yozuvlar nisbatan qisqa (taxminan 1–6 daqiqa, odatda 2 daqiqa) bo‘lishi ta’sir qiladi, chunki ular asosan bitta mavzu bo‘yicha bir kishining hisobiga qaratilgan. Hududiy qamrov nuqtai nazaridan korpusda Chexiya Respublikasining barcha dialekt mintaqalari, shu jumladan Polshadagi chex tilida so‘zlashuvchilarning yozuvlari mavjud. Hozircha u Chexiya, Moraviya va Sileziya chegaralaridan olingan yozuvlarni o‘z ichiga olmaydi, ammo ular an'anaviy dialekt mintaqalariga tegishli emas. Ikkinchi jahon urushidan keyin aholining ko‘p ko‘chirilishi sababli, bu asosan nemis tilida so‘zlashuvchi hududlarda an'anaviy dialekt substrati yo‘q[Abdurakhmonova, N. 2021,2022].

Korpusda [http://www.korpus.cz] to‘plangan yozuvlar turli manbalardan olingan, buning natijasida ular juda uzoq vaqtni qamrab oladi. Ular eski va yangi bo‘lgan ikki vaqt qatlamiga bo‘lingan va bu ma’lumotlardan qidiruvlarni yoki korpus bilan bog‘liq boshqa operatsiyalarni cheklash uchun foydalanish mumkin. Qadimgi qatlam 1950-yillarning oxiridan 1980-yillargacha yozilgan yozuvlardan iborat. Ushbu materialning bir qismi Chexiya Fanlar akademiyasining Chex tili institutining dialektologiya bo‘limi tomonidan to‘plangan va Chexiya lingvistik atlasiga qo‘shimcha sifatida nashr etilgan; qolganlari shaxsiy to‘plamlardan olingan bo‘lib, ular asosan ilgari nashr etilgan. Yangi qatlam esa 1990-yillardan hozirgi kungacha bo‘lgan yozuvlarni qamrab oladi. Ushbu yangi qatlam turli universitetlardagi tadqiqot faoliyati davomida, mustaqil shaxslar tomonidan, eng muhimi, Chexiya Milliy Korpusi institutining o‘zi tomonidan olib borilgan izlanishlarning yozuvlarini o‘z ichiga oladi.

Iloji boricha ko‘proq asl sheva xususiyatlarini yozib olish uchun ma’lumotchi sifatida eng qadimgi avlod vakillarigina ishtirok etgan. Axborotchilar qishloqdagi mahalliy aholi orasidan tanlab olingan, ular aholining o‘troq qatlamiga mansub bo‘lib, asosan shu yerda butun umrini o‘tkazgan, qishloq xo‘jaligi turmush tarziga yoki ma’lum bir hunarga bog‘langan. Ularning barchasi 60+ yosh guruhiga to‘g‘ri keladi, Yozib olish bilan bog‘liq metama’lumotlar yozib olish joyi haqida batafsil ma’lumotni o‘z ichiga oladi, masalan, mahalliy joy turi (shahar, qishloq) va uning o‘lchami, geografik joylashuvi - mamlakat, mintaqa (Bogemiya, Moraviya, Sileziya), dialekt guruhi *(skupina),* kichik guruh *(podskupina),* bo‘linish *(úsek)* va tur (tip) kabi. Aloqa

holatining qo‘shimcha xususiyatlariga yozuv manbai, yozib olish sanasi, vaqt qatlamining a’zoligi, asosiy mavzu, nutq turi (monolog, dialog va ularning kombinatsiyasi), ma’ruzachilar soni, va tadqiqotchining mavjudligiva hokazolar kiradi. Korpusda informatorning jinsi, yoshi, ma’lumoti, bolalik davridagi yashash joyi va eng uzoq yashash joyidan tortib, ma’ruzachining eng uzoq professional kasbigacha bo‘lgan bir qator metama’lumotlar kuzatib boriladi. Ushbu ma’lumotlarning barchasi korpus interfeysida ko‘rsatilishi va chastota statistikasini olish uchun ishlatilishi mumkin[Abdurakhmonova, N. 2021,2022]. Agar bizni shimoli-sharqiy Bogemiyadagi dialekt ovozlari darajasidagi hodisalar qiziqtirsa, biz matnning qidiruvni ***cheklash*** funksiyasidan foydalanib, boshqa hududlar qidiruvini cheklashimiz orqali biz istagan mintaqadagi dialektni topishimiz mumkin yoki ushbu dialekt mintaqasi asosida o‘z doimiy subkorpusimizni yaratishimiz mumkin. Matn tinish belgilari yozma chex tilining standart qoidalariga amal qiladi, lekin jumlalar bosh harf bilan boshlanmaydi.

Shunday qilib dialektal korpus yaratish jarayonida og‘zaki folklorlik xususiyatlariga ega bo‘lmagan matnlar, og‘zaki folklorlik xususiyatlariga ega bo‘lgan matnlar, yozma folklorlik xususiyatlariga ega bo‘lmagan matnlar va yozma folklorlik xususiyatlariga ega bo‘lgan matnlarning barchasini to‘laqonli kiritish dialektal bazaning shakllanishi uchun o‘ziga xos bosqich bo‘lib xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Goldin. Mashina fondi, 1986-1990.
2. Абдурахмонова, Н., & Абдувахобов, Г. (2021). O ‘quv lug ‘atini tuzishning nazariy metodologik asoslari. *Международный журнал искусство слова*, *4*(6).
3. Abdurakhmonova, N. (2021). Formal-Functional Models of The Uzbek Electron Corpus. *ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies*, *10*(8), 59-66.
4. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
5. Абдурахмонова, Н., & Абдувахобов, Г. (2021). O ‘QUV LUG ‘ATINI TUZISHNING NAZARIY METODOLOGIK ASOSLARI. *МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА*, *4*(6).
6. Abdurakhmonova, N., Shakirovich, I. A., & O‘G‘Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. *Science and Education*, *3*(9), 146-156.
7. Mahmudov, M.Ə. Kompüter dilçiliyi / M.Ə. Mahmudov. – Bakı: Elm və təhsil, – 2013. – 352 s
8. Goláňová, H. Waclawičová, M. Co je v ČNK nového Ix (Zprávy z českého národního korpusu). Korpus – gramatika – axiologie, 2018 (17), pages 78–82
9. https://varieng.helsinki.fi/CoRD/corpora/Dialekts/
10. http://www.korpus.cz

KORPUSGA ASOSLANGAN MASHINA TARJIMASI UCHUN IKKI TILLI FRAZEOLOGIK LUG‘AT YARATISHNING LINGVISTIK AHAMIYATI (BADIIY ASAR MISOLIDA)

Sirojiddinova Muharram
O’zbekiston Milliy universiteti Kompyuter lingvistika mutaxassisligi magistranti.

Annotatsiya: Ushbu tezisda kompyuter texnologiyasi va tarjimashunoslik sohalarining integratsiyalashuvi natijasida mashina tarjimasi sohasining paydo bo‘lishi, ushbu sohaning yo‘nalishlari va korpusga asoslangan mashina tarjimasining muammolari, ya’ni badiiy asarlardagi frazeologik birliklarni boshqa tilga tarjima qilishdagi nomuvofiqliklar hamda bu muammolarni hal qilishda badiiy asarlarning ikki tilli yoki ko‘p tilli parallel korpuslarini yaratish va ular asosidagi frazeologik birliklar lug‘atini yaratishning lingvistik ahamiyati yoritilgan.

Kalit so‘zlar: frazeologik lug‘at, qoidaga asoslangan mashina tarjimasi (MT), korpus asosidagi mashina tarjimasi (Corpus-based MT), gibrid mashina tarjimasi (Hybrid machine translation (HMT).

Abstract: In this thesis, the emergence of the field of machine translation as a result of the integration of the fields of computer technology and translation studies, the directions of this field and the problems of corpus-based machine translation, that is, inconsistencies in the translation of phraseological units in works of art into another language, and the use of works of art in solving these problems the linguistic importance of creating bilingual or multilingual parallel corpora and creating a dictionary of phraseological units based on them is highlighted.

Keywords: phraseological dictionary, rule-based machine translation (MT), corpus-based machine translation (Corpus-based MT), hybrid machine translation (Hybrid machine translation (HMT).

Аннотация: В данной диссертации рассматривается возникновение области машинного перевода в результате интеграции областей компьютерных технологий и переводоведения, направления этой области и проблемы корпусного машинного перевода, то есть несоответствия в перевода фразеологизмов в произведениях искусства на другой язык и использования произведений искусства при решении этих задач подчеркивается лингвистическое значение создания двуязычных или многоязычных параллельных корпусов и создания на их основе словаря фразеологизмов.

Ключевые слова: фразеологический словарь, машинный перевод на основе правил (MT), корпусный машинный перевод (Corpus-based MT), гибридный машинный перевод (Hybrid Machine Translation (HMT).

Bugungi kunda kompyuter texnologiyasi, avtomatik tarjima, mashina tarjimasi, google tarjimon, amazon tarjimon, microsoft tarjimon, DeepL, parallel korpus kabi atamalarning kundalik iste’molda tez-tez takrorlanishi, nafaqat tarjimashunoslik, balki mashina tarjimasi (XX asr 50-yillaridan kompyuter va tilshunoslik sohalarinining integratsiyasidir) sohalarida ko‘plab ishlar amalga oshirilganligidan dalolat beradi. Ko‘pchilikka ma’lumki, bu boradagi birinchi tajriba

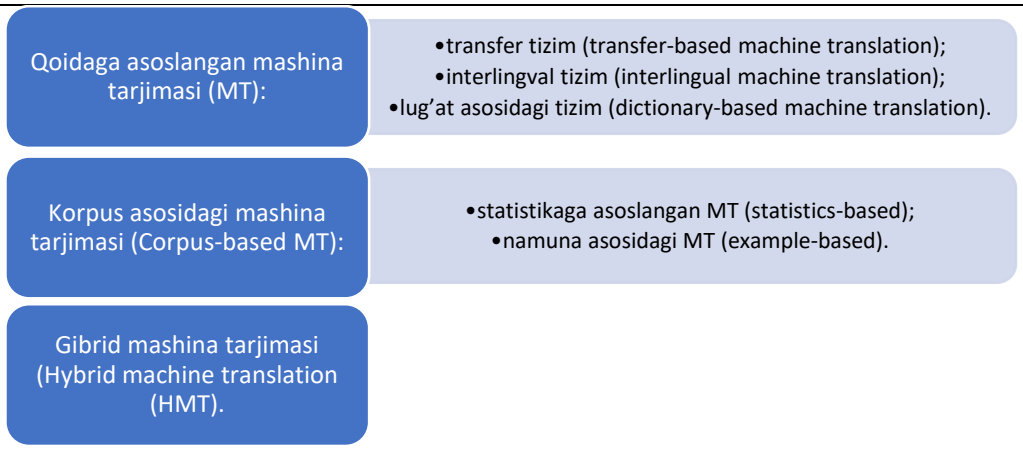
1954-yil AQSHda Jorjtaun universitetida o‘tkazilgan bo‘lib, 1960-yilda esa mustaqil fan sifatida shakllangan [Rahimov A. 2011: 160].

Jahonda mashina tarjimasi sohasini rivojlantirish uchun ko‘plab olimlar ilmiy izlanishlar olib borgan. N.D. Andreev, I.A. Melchuk, I.I. Revzin, V.Yu. Rozensveyg, Yu.N. Marchuk, R.G. Piotrovskiy, Yu.A. Motorin, K.B. Bektaev, A.N. Belyaev, I.K. Belskiy, A.V. Zubov, G.E. Miram, L.L. Nelyubin, V.I. Perebiynos, V.A. Chijakovskiy, Ye.A.Shingarev, G.G.Belonogov, R.G.Kotov, Babushkina N.V, Z.Shalyapina, O.Yu. Mansurova, A.S. Panina, A.A. Xoroshilov (Rossiya); A.But, R. Richans, J. Xatchins, J.Allen, P. Braun (AQSh); M. Nagao (Yaponiya); A.Vaxer (Estoniya); Frantsiyada J.Astrouni; R.Sinha, A.Jain (Hindiston); B.Bleyzer, U.Shvol, A.Storrer (Germaniya) singari tadqiqotchilarning ishlari shular jumlasidandir [N.Abduraxmonova. 2012]. Shu bilan birgalikda, D.O.Dobrovolskiy, Yu. Tao, A.A.Kokoeva va V.Zaxarov kabi olimlar parallel korpusning tuzilishi, tarkibi va imkoniyatlari bo’yicha tadqiqotlar olib borib, ko‘plab asarlar yaratishdi. Ularning fikricha, mashina tarjimasi sohasida parallel korpuslar yuaratish muhim ahamiyatga ega, chunki unga asl matn va tarjima matn biteks ko‘rinishida kiritiladi. Bu esa matn tajrimasida konteksga qarab matn mazmunini aniqlash va boshqa tilga tarjima qilishga semantik muvofiqini topish hamda parallel matnlarning semantik munosabatiga oid ko‘plab kerakli ma’lumotlarni o‘z ichiga olish imkonini beradi.

Shuningdek, respublikamizda ham ushbu sohani rivojlantirishga qaratilgan jiddiy tadqiqotlar amalga oshirilgan hamda chet tildagi matnlarni o‘zbek tiliga va yoki buning aksini tarjima qiluvchi dasturning lingvistik ta’minotini yaratishga ham alohida e’tibor qaratilgan. Ushbu sohaga dasturchi va matematiklar ham e’tibor qaratib muayyan tadqiqotlar olib borganligi eslatish ham bu sohaning dolzarbligini anglatadi [Хакимов М. 2008: 297].

Tarjimashunoslikda frazeologik birliklarga oid masalalar G‘.Salomov,Q.Musaev, Yu.Po‘latov, S.Abidxodjaeva, G.K.Odilova,I.G‘afurov, O.Mo‘minov, N.M.Qambarov, M.Xolbekov, Sh.Safarov kabi olimlarning tadqiqotlarida atroflicha o‘rganilgan. Zamonaviy tilshunoslikda frazeologik birliklarni boshqa tillar bilan qiyosiy o‘rganib, ularning milliy-madaniy xususiyatlarini tadqiq qilish keng tus oldi. Bu jihatdan L.Kulieva, K.Musaev, I.Stepanova, E.Arsenteva, I.N.Isabekov singari tilshunos olimlarning ishlari bir-biriga qardosh bo‘lmagan ikki til doirasida olib borilgan tadqiqotlar yuzasidan ahamiyatlidir.

Bugungi kunda mashina tarjimasining quyidagi yo‘nalishlari mavjuddir. Bu yo‘nalishlarni aniqlashda mashina tarjimasini lingvistik aspektda o‘rgangan Jon Xatchinsning xizmatlari kattadir. Olim mashina tarjimasi tizimlarida lingvistik tahlil bosqichlarini quyidagicha tasniflaydi [N.Abduraxmonova. 2012]:



Hozirda bu yo‘nlishlardan eng keng tarqalagani sifatida korpusga asoslangan mashina tarjimasini keltirish mumkin. Bunda parallel korpusga kiritilgan bitekslarning (inglizcha – o‘zbekcha matn) har biri avtomatik tahlil qilinadi va tarjima uchun muayyan xulosalarga kelinadi[Abdurakhmonova, N. 2021,2022]. Mohir tarjimonlar yordamida tajima qilingan parallel matnlardagi so‘zlar, frazalar konteksga qarab ma’nolari aniqlanadi.

Tilshunosligimizda o‘zbek tili frazeologiyasini qarindosh bo‘lmagan ingliz tili frazeologiyasi bilan solishtirish ikkala tilning o‘ziga xos xususiyatlarini yaxshiroq anglashga, tillararo frazeologik birliklarning mohiyatini aniqlashga imkon beradi. O‘zbek va ingliz tillarining frazeologik birliklari o‘rtasidagi o‘xshashlik va farqlar nimada, ular tilda, asosan, qanday namoyon bo‘ladi; ular qanday intralingvistik va ekstralingvistik omillar bilan belgilanadi, frazeologik birliklarning tillararo ekvivalentligi darajasi qanday; qardosh bo‘lmagan tillarning frazeologiyasini qanday moslash mumkin degan savollarga javob topish nafaqat tilshunosligimizning, balki mashina tarjimasi sohasining istiqbollarini ham belgilaydi[Abdurakhmonova, N. 2021,2022].

A.V.Kunin frazeologizmlarni quyidagicha ta’riflagan: «frazeologik birlik – qisman yoki to‘laligicha ko‘chma ma’noli so‘zlarning barqaror birikmasi» A.Kunin frazeologik birliklardagi turg‘unlik belgisini ularning leksikasida ishlatilishi, semantikasi, morfologik, sintaktik tuzilishida mavjudligi orqali sharhlaydi [Nurmuhammedov Y. 2022: 90]. Bundan ko‘rinib turibdiki, har bir tilning frazeologiyasi o‘sha tilning madaniyati, yashash tarzi, geografik joylashuvi, odatlariga bog‘liq bo‘lganligi uchun ularni qardosh bo‘lmagan boshqa tilga tarjima qilganda o‘zining anglashilgan ma’nosiga putur yetdi. Agar bunday birliklar berilgan asarlarni boshqa tilga tarjima qilsak, asarning mazmun-mohiyati, g‘oyaviy ahamiyati va tarjima sifatiga zarar yetadi. Chunonchi, inglizlarda “A little bird told me” frazema mavjud bo‘lib, agar ma’lumotni oldindan eshitgan yoki bilgan, lekin kimdan eshitgani aytishni xohlamasangiz, shu iboradan foydalanishingiz mumkin. Bu ibora o‘zbek tilidagi “ teshik quloq eshitadi” iborasiga semantik jihatdan mos keladi. Ushbu ibora asarda Did you hear that Ann is going to marry? Where did you find this information? A little bird told me tarzida qo‘llanilgan bo‘lib, mashina tarjimasi orqali tarjima qilinganda, Enn turmushga chiqmoqchi ekanligini eshitdingizmi? Bu ma’lumotni qayerdan topdingiz? Kichkina qush aytdi kabi tarjima qilinib, ma’no jihatdan uslubiy g‘alizlikni keltirib chiqaradi. Shuningdek, “Aching heart” idiomasi o’zbek tilida “ko’ngli yarim yoki yuragi yarim” iborasiga to’g’ri keladi. Lekin tarjima dasturlar orqali tarjima qilinsa, “yaralangan yurak yoki og’riyotgan yurak” deb tarjima qilinadi. Bu tarjima esa hamma konteks mazmuniga ham muvoqil kelavermaydi. Yoki “Able to take (can take)” iborasi so‘z o‘zbek tilida semantik jihatdan “Tosh kelsa kemirmoq, suv kelsa simirmoq” iborasiga to‘g‘ri keladi. Eng achinarlisi, bu ibora tarjima dasturlar orqali tarjima qilinsa, “olish mumkin” deb tarjima qilinadi. Matn tarjimasida bunday xatoliklarni ko‘p bora takrorlanishi asl matn mazmunini keskin o’zgartiradi. Shunday ekan, badiiy asarlar asosidagi frazemalar lug‘atini yaratib, parallel korpusga biteks ko‘rinishida kirtilsa va doimiy boyitib borilsa, tarjima jarayonida har qanday badiiy matnning asliyatidagi ma’nosini saqlab qolish imkoniyati yaratiladi

Xulosa

Ta’kidlash joizki, bugungi kunda mashina tarjimasi sohasida juda ko‘plab tadqiqotlar amalga oshirildiki, buning natijasi o‘laroq, so‘zma-so‘z tarjima usulidan voz kechildi, tarjima matn ancha mahalliylashtirildi, ilmiy va rasmiy matnlarning bir tildan boshqa tildagi tarjimalari mukammallashtirildi. Lekin har bir tilning o‘ziga xos lingvistik xususiyatlari, leksik va frazeologikbirliklari hamda so‘zlarining ma’no nozikliklari boshqa tilga to‘laqonli mos kelmasligi

sababli badiiy matnlarni boshqa tilga tarjima qilinganda, asl matnning nafaqat g‘oyasi, balki mazmuni ham keskin o‘zgarishga uchramoqda. Shu sababli, badiiy asarlarning ikki tilli biteks matnlarini hamda ular asosidagi frazeologik birliklar lug‘atini yaratish va ushbu sohaga tatbiq qilish ushbu sohani yanada rivojlantirish va tarjima sifatini yaxshilashga xizmat qiladi. Bu esa, o’z navbatida, yuqorida ko‘zda tutilgan muammolarning qisman hal qililishiga olib keladi.

Adabiyotlar ro‘yhati:

1. Abduraxmonova N. Inglizcha matnlarni o‘zbek tiliga tarjima qilish dasturining lingvistik ta’minoti (sodda gaplar misolida).2018.
2. Abdurahmonova N. Mashina tarjimasi lingvistik asoslari. Qo‘llanma. T.: Akademnashr, 2012.
3. Абдурахмонова, Н., & Абдувахобов, Г. (2021). O ‘quv lug ‘atini tuzishning nazariy metodologik asoslari. *Международный журнал искусство слова*, *4*(6).
4. Abdurakhmonova, N. (2021). Formal-Functional Models of The Uzbek Electron Corpus. *ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies*, *10*(8), 59-66.
5. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
6. Abdurakhmonova, N., Shakirovich, I. A., & O‘G‘Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. *Science and Education*, *3*(9), 146-156.
7. Nurmuhammedov Y. Frazeologiyaning dolzarb masalalari. O‘quv-uslubiy qo‘llanma. – S.: SamDCHTI, 2022. – 90 b.
8. Rahimov A. Kompyuter lingistikasi asoslari. – T.: Akademnashr, 2011. – 160b.
9. Хакимов М.Х. Расширяемый входной язык математического моделирования естественного языка для многоязычной ситуации машинного перевода // ЎзМУ хабарлари, 2009. – № 1. – С.75-80;

MATN TAHLILIDA TERMINOLOGIYA MASALASI

Yulbarsov Ochilbek Obidjon o‘g‘li
O‘zMU o‘qituvchisi.

Abdirahimova Yulduz Sobir qizi
O‘zMU 4-bosqich talabasi.

Annotatsiya: Ushbu tezisda terminologik lug‘at tuzish jarayonida, ya’ni terminografiyada amalga oshiriladigan tahlil turlari hamda matn tahlilida terminalogiya masalalari yoritilgan. Tahlilning bosqichlari haqida fikr yuritiladi. Terminografik jihatdan ahamiyatli bo‘lgan eng muhim lingvistik xususiyatlarga to‘xtalib o‘tilgan.

Kalit so‘zlar: termin, terminologiya, terminologik lug‘at, terminografiya, lingvistik va kontseptual, onomasiologik yondashuv, kontseptual ma'lumot.

Abstract: In this thesis, the types of analysis performed in the process of creating a terminological dictionary, that is, in terminography, as well as the issues of terminology in text analysis are highlighted. Also, the stages of the analysis are considered. The most important linguistic features, which are terminographically significant, have been touched upon.

Keywords: term, terminology, terminological dictionary, terminography, linguistic and conceptual, onomasiological approach, conceptual information.

Аннотация: В данной тезис освещены виды анализа, выполняемые в процессе создания терминологического словаря, то есть в терминографии, а также вопросы терминологии при анализе текста. Рассмотрены этапы анализа. Затронуты важнейшие лингвистические особенности, имеющие терминографическое значение.

Ключевые слова: термин, терминология, терминологический словарь, терминография, лингвоконцептуальный, ономасиологический подход, концептуальная информация.

Hozirgi kunda yangi texnika vositalari qaysi sohaga ko‘proq kirib kelgan bo‘lsa, shu yo‘nalishga doir terminlarni o‘rganishga ehtiyoj oshib bormoqda. Termin umumadabiy so‘zlardan tubdan farq qiladi va ma’lum bir fan sohasiga mansub so‘z va so‘z birikmalarini o‘z ichiga oladi. Terminlarning qo‘llanish, tarqalish ko‘lami muayyan terminologik sistema bilan cheklangan bo‘lib, ular insonning faoliyati doirasidagi aniq uslubda harakat qiladi, voqelanadi [Dadaboyev H. 2020:6].

Terminologiyasi ma’lum darajada barqarorlashgan barcha tillarda terminlar doimiy tartibga solib turiladi. Tartibga solish terminologik me’yorlar asosida amalga oshiriladi. Terminologiyaning rivojida fan sohalariga oid maxsus lug‘atlarni nashr qilib turish ham muhim ahamiyatga ega. Lug‘atga kiritilishidan oldin har bir termin to‘plami turli xil vositalar va usullardan foydalangan holda lug‘at turiga mos keladigan turli nuqtai nazardan tahlil qilinishi kerak. Ta’rifga ko‘ra, terminologik lug‘at terminlarni, ya’ni lingvistik birliklarni o‘z ichiga oladi va ular tegishli tushunchalar bilan moslashtiriladi. Shuning uchun bunday birliklarning har qanday

tahlili ikki bosqichdagi o‘rganishni talab qiladi: lingvistik va kontseptual[Abdurakhmonova, N. 2019, 2021,2023].

Lingvistik tahlil atamani to‘g‘ri aniqlash va ishlatish imkonini beradigan barcha muhim jihatlar bilan bog‘liq. Terminlarning lingvistik harakati umumiy til so‘zlaridan juda farq qilishi va kontekstga qarab o‘zgarishi mumkin. Shuning uchun bunday ma‘lumotlar umumiy emas, balki ixtisoslashgan til korpuslaridan olinishi kerak. Terminografik jihatdan ahamiyatli bo‘lgan eng muhim lingvistik xususiyatlar:

– orfografik/fonetik variantlar;
– transliteratsiya/transkripsiya shakllari;
– morfologik birliklar (masalan, yunon yoki lotincha ma’noli atama elementlari);
– nutq xususiyatlarining bir qismi;
– sintagmatik xususiyat (birikma, ibora);
– uslub;
– chet tilidagi ekvivalentlar;
– terminning holati: neologizm, standartlashtirilgan atama, eskirgan atama, muallif atamalari va boshqalar.

Konseptual (nazariy) tahlil terminni o‘rganishning eng muhim elementidir, chunki u termin ifodalaydigan bilimlar miqdorini aks ettirishga qaratilgan. U atamaning terminologik paradigmasini, ya’ni ma’lum bir lisoniy birlikning (termin bilan ifodalangan) boshqa birliklar (terminlar) bilan semantik yoki ontologik munosabatlari to‘plamini qayta qurishga qaratilgan[Abdurakhmonova, N. 2019, 2021,2023]. Tahlil jarayonida eng ko‘p uchraydigan munosabatlar qatoriga giponimiya-giperonimiya, meronimiya-holonimiya va boshqa assotsiativ munosabatlar kiradi [Lukszyn J. 200:22]. Terminologiyaga an’anaviy yondashuv terminlarni tahlil qilishda onomasiologik yondashuvni talab qiladi, ya’ni tushuncha va uning xususiyatlarini aniqlash hamda nomlashda tildan tashqari voqelikni hisobga olish jarayoni. Bu atamalarni matn muhitida (korpus) o‘rganishni talab qiladigan eng yangi yondashuvdir [Temmerman R.2000:41]. Semantik aloqalar endi ixtisoslashtirilgan matn korpusidan o‘rtacha 92 foiz aniqlik bilan avtomatik ravishda olinishi mumkin. Bunday turdagi avtomatik vositalar mavjud bo‘lmaganda yoki aniqlik darajasi qoniqarli bo‘lmasa, mavjud ontologiyalar, semantik tarmoqlar yoki tezaurus kabi boshqa vositalardan foydalanish mumkin. Raqamlashtirilgan semantik tarmoqlarga Prinston universitetida ishlab chiqilgan WordNet va tegishli milliy tarmoqlarini misol qilib ko‘rsatish mumkin. Avtomatik yoki qo‘lda bo‘lishidan qat’i nazar, bunday tahlillar semasiologik-onomasiologik usulda amalga oshiriladi, ya’ni tegishli ma’lumotlar maxsus korpuslardan olinadi va keyin semantik tahlil qilinadi, ehtimol boshqa ma’lumotnomalar yordamida bo‘lishi ham mumkin[Abdurakhmonova, N. 2019, 2021,2023]. Bundan tashqari, har qanday terminologik leksikaning kognitiv qiymati har xil turdagi terminologik birliklarga tegishli oldindan belgilangan qiymatlardan foydalangan holda miqdoriy o‘lchovlar bilan vaqtincha baholanishi mumkin. Shunga qaramay, matn korpusining ahamiyatini e’tiborsiz qoldirib bo‘lmaydi. Olingan va aniqlangan semantik munosabatlar lug‘atda keltirilgan terminologiyani semantik sohalarga ajratish, lug‘atning tizimli tuzilishini ishlab chiqish, bilim xaritalarini tuzish va xususan, turli tizimlarning terminologiyalarini qarama-qarshi qo‘yishga yordam beradi. Konseptual ma’lumotlar lug‘atda aniq shaklda, qo‘shimcha semantik metatil yoki alohida lug‘at ilovasida tavsiflovchi taqdimot orqali yoki tizim orqali yashirin tarzda taqdim etilishi mumkin.

Bundan tashqari terminologiya har doim o‘zgarib turuvchi soha hisoblanadi. Olimlar fan va texnikaga oid terminlarni standartlashtirishga oid fikrlarni beradilar. Muayyan terminning iste’molga kirib kelishida uning yaratuvchisi yoki ijtimoiy hodisalari ta’sirida ba’zi terminlarning o‘rniga boshqa birining qo‘llanish holatlarini inkor etib bo‘lmaydi. Bunda terminologik lug‘atlarni tuzish haqidagi terminografiya sohasi ahamiyatlidir. Terminografiya vazifalariga: 1) chegaralangan leksemalarning har bir lingvistik sathlar bo‘yicha formal va modellashtirilgan tavsifini berish; 2) matn-matn munosabati o‘rtasida turgan umumiy va farqli lisoniy paradigmalarni aniqlash; 3) avtomatlashgan terminologik bazani shakllantirish; 4) ilmiy matnlarga oid leksik, morfologik va sintaktik xususiyatlarni ikki til bo‘yicha solishtirish va ularni o‘xshash hamda farqli jihatlari bo‘yicha kategoriyalarga ajratish; 5) ikki tilli ilmiy matnlarda qo‘llanadigan leksemalarning miqdori bo‘yicha sinonimik, omonimik va polisemantik tizimini tuzish orqali ularning dominantini bazaga kiritish kabilar kiradi [N.Abduraxmonova. 2018:53].

Xulosa o‘rnida shuni aytish mumkinki, terminologik korpus, terminologik ma’lumot bazalari va lug‘atlar yaratishda lingvistik va kontseptual tahlil qilish muhim bosqichlardan sanaladi.

Foydalangan adabiyotlar:

1. Dadaboyev H. O‘zbek terminologiyasi. O‘quv qo‘llanma –Toshkent: 2020.
2. Lukszyn. J., Zmarzer. W. Teoretyczne podstawy terminologii. Warszawa: Katedra Języków Specjalistycznych UW. 2006. pp. 7-22.
3. Temmerman. R. Towards new ways of terminology descript: The sociocognitive approach. Amsterdam: 2000. pp. 41.
4. N. Abduraxmonova. Mashina tarjimasining lingvistik ta’minoti. (Monografiya). – Toshkent, 2018.
5. Aripov, M., Norov, A., & Abdurakhmonova, N. (2019). Composing Modification of the Uzbek Phonetic Alphabet Based on International Phonetic Alphabet. International Journal of Advanced Manufacturing Technology, 6, 10731-10735.
6. Abdurakhmonova, N., & Urdishev, K. (2019). Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL), 6(1-2019), 131-7.
7. Mengliev, D. B., Abdurakhmonova, N., Hayitbayeva, D., & Barakhnin, V. B. (2023, November). Automating the transition from dialectal to literary forms in Uzbek language texts: an algorithmic perspective. In 2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE) (pp. 1440-1443). IEEE.
8. San’atbek, M., Mirsaid, A., & Nilufar, A. (2018). Modeling WordNet type thesaurus for Uzbek language semantic dictionary. International Journal of Systems Engineering, 2(1), 26.
9. Abdurakhmonova, N. (2021). Formal-Functional Models of The Uzbek Electron Corpus. ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies, 10(8), 59-66.
10. Aripov, M., Sharipbay, A., Abdurakhmonova, N., Razakhova B.: Ontology of grammar rules as example of noun of Uzbek and Kazakh languages. In: Abstract of the VI International Conference “Modern Problems of Applied Mathematics and Information Technology Al-Khorezmiy-2018.–Tashkent: 2018. pp. 37–38.

11. Abduraxmonova N. Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences). Aftoref, 2018.
12. Abdurakhmonova N., Tuliyev U. Morphological analysis by finite state transducer for Uzbek-English machine translation. Foreign Philology: Language. Literature, Education, 2018 (3):68.
13. Abdurakhmonova N., Urdishev K. Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL), 2019.
14. https://uz.m.wikipedia.org

IDENTIFICATION OF SEMANTIC FEATURES OF PHRASAL VERBS IN A PARALLEL CORPUS

Otamurodova Salomat
Toshkent davlat agrar universiteti
Tillar kafedrasi o‘qituvchisi.
E-mail: soloahmedovna@gmail.com

Annotation: The scientific, technological and information revolutions of the 20th century, which covered all areas of human knowledge and practice, could not but affect linguistics. We are talking not only about new forms of recording, such as audio or video recording and electronic data storage, but about phenomena that fundamentally change our views and approaches to the functioning of language. In recent years, corpus linguistics has become widespread - a branch of applied linguistics whose goal is the creation and maintenance (using machine media and appropriate access tools) of large arrays of language data.

Keywords: corpus linguistics, parallel corpora, phrasal verbs, semantic structure, phraseological unit.

Аннотация: Научно-техническая и информационная революции XX века, охватившие все области человеческого знания и практики, не могли не затронуть языкознание. Речь идет не только о новых формах записи, таких как аудио- или видеозапись и электронное хранение данных, но и о явлениях, которые коренным образом меняют наши взгляды и подходы к функционированию языка. В последние годы широкое распространение получила корпусная лингвистика — раздел прикладной лингвистики, целью которого является создание и поддержание (с использованием машинных носителей и соответствующих средств доступа) больших массивов языковых данных.

Ключевые слова: корпусная лингвистика, параллельные корпуса, фразовые глаголы, семантическая структура, фразеологизм.

Introduction

Today, corpus linguistics is actively and successfully used in various fields, such as applied linguistics, philology, stylistics, methods of teaching foreign languages, theory and practice of translation, criminal linguistics, content analysis, discourse analysis, genre analysis.

The role of text linguistics1 is determined by the desire to deeply study the connections of language with various aspects of human activity realized through text. Research conducted both in our country and abroad makes it possible to carry out detailed developments of the essential characteristics of the general linguistics of the text. The question of the boundaries of text linguistics is debatable, namely: does it include pragmatics, functional semantics and syntax, the currently actively developing rhetoric, or does it intersect with them as it intersects with poetics, psycholinguistics and communication theory, entering as an integral part of the theory of text. When creating and developing electronic corpora of languages of the world, the consistent research of NLP and language technologies has become an urgent task [Abdurakhmonova N.2021:202].

Parallel corpora

The phrase "multilingual corpora" also refers to "parallel corpora," which are collections of materials that have been translated into many languages and are aligned across those languages. The primary applications of parallel corpora in research are in translation studies and contrastive linguistics. They also have a significant influence on the creation of translation resources for language learners and on translation training. Beginning in 2016, the Multilingual Student Translation project seeks to integrate learner corpus research with corpus-based translation studies by creating a new, sizable multilingual student translation corpus.

Parallel corpora are divided into monolingual, bilingual and multilingual. In monolingual corpora, dialects and language variants are contrasted. For example, such varieties of English as English as a native language and English as a foreign language remained outside the scope of scientific interest until the advent of new technologies, which made it possible to involve a significantly larger number of compared speech products in contrastive analysis [Abdurakhmonova, N. 2021,2022].

Bilingual and multilingual corpora can be divided into two main types:

1) corpora representing a set of original texts written in any source language, and texts-translations of these source texts into one or more other languages;

2) corpora that combine texts from the same thematic area, independently written in two or more languages; Such corpora help in working with terminology and are often used by translators.

These kinds of corpora are used for comparative language studies (lexicology, grammar, stylistics, translation studies, etc.) and for the development of efficient translation techniques, such as machine translation and the creation of bilingual and multilingual terminological dictionaries.

Electronic dictionaries are excellent resources for researching words and phraseology. Although they often don't offer the same contexts for individual search items as full-text corpora do, the information retrieved can be utilized for additional searches inside historical corpora. Long-term efforts that span a great deal of historical time are large-scale dictionaries, which attempt to cover the vocabulary's evolution in a language. The Oxford English Dictionary Online (OED Online) for English, Der digitale Grimm for German, and Svenska Akademiens wordbook for Swedish are a few examples of these kinds of undertakings.

Large-scale text collections have showed potential, especially for researchers studying literature, although smaller corpora can also be helpful study tools. For literary and socio-historical study, whole text collections like the Innsbruck Computer Archive of Machine-Readable English Texts and the Corpus of Middle English Prose and Verse are invaluable resources. Legal and social historians as well as historical linguists might benefit from electronic editions such as the An Electronic Text Edition of Depositions 1560-1760 (ETED), which include transcriptions of early official documents.

The current stage of research and development in the field of machine translation and natural language processing language requires the use of new methods of representing linguistic reality, reflecting the complex the nature of natural languages and, at the same time, combining the capabilities of already existing approaches. The formation of corpora of parallel texts makes it possible to use machine learning in addition to traditional methods based on heuristic rules compiled by a human expert [Abdurakhmonova, N. 2021,2022].

However, in order to be able to automatically extract rules from parallel tests, it is necessary develop fairly subtle methods for comparing linguistic objects, with the main basis there must be

their semantic characteristics. Until now, semantic alignment of parallel texts has been considered mainly from the point of view terms of constructing lexical-semantic models, in our case we are talking about the semantics of structural components, i.e. ways of configuring linguistic objects (in semiotic linguistics they received name of structural signs [1]). These studies are motivated by the fact that the applied methods for automatically aligning parallel texts word by word or sentence by sentence lead to the emergence of a large number of noises that can only be eliminated manually at the stage post-editing. Phrasal verbs for researchers in the field linguistics skills are attractive element. This linguistic unit is dedicated to quite a large number of works of various linguistic whist. The increased interest in these units explains is determined not only by their prevalence in the linguistic environment (12 thousand according to dictionary entries only), but an abnormal fullness in the field of polysemy, synonymy, antonymy and in other lexical-semantic ski and thematic groups. It should also be taken into account that, since the 1960s, the frequency of use phrasal verbs increase significantly in scientific literature, journalism and academic texts the style. Subsequently, according to the law of analogy, in this productive language model, more and more new ones are included elements. In particular, verbal derivatives, image- bathrooms from nouns or adjectives in terms of conversion, are one of the leading word forms vocal trends at the present stage of development language. Determining the essence and importance of phrasal verbs is one of the modern tasks of the modern English phraseology.

According to Shansky, phraseological units are linguistic units consisting of two or more accentuated components, repeating, of a speech nature, having content and structure in a fixed (i.e., constant) sense. Considering phraseology as a branch of linguistics, A.F. Alefirenko considers its main task to be the mechanisms for coordinating phraseological components (syntactic, lexical or morphological) with non-free units of words, and the phraseological whole is one of these components, will lead to unification [Abdurakhmonova, N. 2021,2022]. He speaks.

A distinctive feature of phraseological units as part of the entire lexical-semantic system of any language is the stability and integrity of their meanings, which allows them to be repeated in speech in the form of semantically indivisible units. Many phrases serve as the basis for new formations due to a revision of the meaning of the words that make up the phraseological unit (A.B. Kunin. 2005).

In her guide to phrasal verbs, Jane Povey devotes an entire chapter to the use of particles and their semantic meaning, takes into account grammatical structure and properties of the verb, but semantic relationships between the components of the phrasal verb there remain inconsistencies followed. We selected for our study
phrasal verbs with particles showing direction of movement and related to compositional(compositional). The division of phrasal verbs into compositional and non-compositional (non compositional), in which the meaning of the verb and postposition completely or partially changes, and therefore deduce the meaning of the verb from the meaning of its components does not work out, is traditional.

Along with the three main verbs of state, the microsystem of broad-meaning verbs includes words with a complicated semantic structure, which is formed by adding a dynamic, active component to the stative semantics of *be* or *have*. Thus, the dynamic correlates of *have* are the verbs *take* and *give*, denoting the action of including an object in a sphere, the center of which is the subject intake, and in give a recipient (participant in the situation; perceiver) that is different

from the subject; in the latter case, the transition of an object into the sphere of the recipient may mean the exclusion of the object from the sphere of the subject [Abdurakhmonova, N. 2021,2022].

Both verbs, having become broad-meaning, base their semantics on a significantly expanded representation of the subject, under which actions are also included (*take a look, give a look)*. Although both of these analytical lexemes are denotatively close to each other, which is reflected in their common Uzbek translation of *look*, the broad-meaning verbs in their composition retain the semantic opposite - *give a look* characterizes a *look* as a way of transmitting information about the state of the subject, while *take a look* means the search for visual information by the subject.

Object of study

The object of the study are literally characteristics of English phrasal verbs. At present time, a lot of academics focus on phraseology as a science. One of the concepts is related to this notion that phraseological units are cultural phenomena. As a result, phraseology is now a subject of sociolinguists. One of the best definitions of phraseological units is as follows: it is a stable, coherent combination of words with partially or fully figurative meaning [Anna Jansone].

The term “phraseology” is derived from the Greek words phrases-speech and logos-teaching. One-third part of colloquial speech cover phraseological units. Although phraseological units have different names (phraseologisms, phrases, stable combinations of words, phraseological expressions), they still have one main feature, namely constancy. For this reason, phraseological units are stable combinations of words that cannot be understood from the point of view of their grammatical structure in the vocabulary and, most importantly, from the point of view of their understandable component in spoken language [Abdurakhmonova, N. 2021,2022].

Information was selected using documents to collect sentences containing phrasal verbs during data collection. During the data collection process, the novel was read carefully and attention was paid to its sentences. After this, the types and meanings of phrasal verbs were analyzed and reflected in a parallel corpus.

The study of the semantics of phrasal verbs in a parallel corpus serves as a resource for the future creation of linguistic resources of synonymy, antonymy, hyperanonymy or machine translation. If the authors worked with the literature in this direction, now this information is stored in the form of a database and is used as a linguistic resource.

Conclusion

Some of these aspectual meanings are inherent in the basic semantics of certain subgroups of verbs. So, for example, the meanings of incompleteness or repetition can be expressed using lexical derivation (word production), in particular prefixation (oversimplify, outdo, underestimate).

Thus, from the point of view of the most general division of verbs into subclasses, while relying on the grammatical structure of the English language, we can distinguish two aspectual subclasses of phrasal verbs in the English language

The basis of the above division is the relationship of verbal semantics to the idea of procedural limitation, i.e. to any boundary, beyond which the process expressed by the verb or its implied semantics cannot objectively exist. For example, the verb arrive expresses an action that obviously can develop before arriving at the destination. Having reached it, the action stops.

The verb start denotes a transition from some initial state to a subsequent action, thus denoting the boundaries between the two. In contrast to these two verbal units, the verb move

expresses a process that is in itself alien to any idea of limitation (initiative, discontinuous or ultimate).

References:

1. Abdurakhmonova N. Computer models of electronic corpus of the Uzbek language(monograph)/Tashkent: Muharrir, 2021.
2. Alefirenko N.F., Linguoculturology: the value-semantic space of language, 2009.
3. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). Uzbek electronic corpus as a tool for linguistic analysis. In *Компьютерная обработка тюркских языков. TURKLANG 2022* (pp. 231-240).
4. Abduraxmonova, N. Z. Q., & Urazaliyeva, M. Y. (2022). O ‘zbek tili elektron korpusida (http://uzbekcorpus. uz/) og ‘zaki matnlar korpusini yaratishning nazariy va amaliy masalalari. *Academic research in educational sciences*, *3*(3), 644-650.
5. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, *11*(19), 9117.
6. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In *of the International Conference on Language Technologies for All (LT4All)*.
7. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).
8. Shaskiy N.M., Phrazeology of Modern Russian language, 1996.
9. Англо-русский фразеологический словарь: Ок. 20000 фразеол. единиц. – 4 –е изд., перераб. и доп... –М.: Рус. яз., 1984
10. Джейн Поуви Английские фразовые глаголы и их употребление, 1990.

ОБЗОР И РАЗРАБОТКА ПРОГРАММЫ ПО СБОРУ ЭЛЕКТРОННЫХ РЕСУРСОВ НА КАЗАХСКОМ ЯЗЫКЕ ПО ЗАКОНОДАТЕЛЬСТВУ РЕСПУБЛИКИ КАЗАХСТАН

Айтбеккызы Аружан
КазНУ имени аль-Фараби.

Шормакова Асем
КазНУ имени аль-Фараби.

Аннотация: Работа посвящена обзору программ по сбору электронных ресурсов. В условиях стремительной цифровизации средств обучения, культурных архивов и коммуникаций, вопрос эффективного сбора, хранения и использования языковых данных становится ключевым для сохранения культурного наследия и поддержания языкового разнообразия. В исследовании проведен литературный обзор, охватывающий актуальные аспекты электронных ресурсов, языковых данных и программ сбора данных. Целью работы является разработка программы, специфичной для казахского языка. В процессе исследования решаются задачи изучения законодательной базы, анализа существующих программ и разработки универсальной системы сбора данных на казахском языке.
Ключевые слова: электронные ресурсы, казахский язык, краулинг.

Annotation: The work is devoted to the development of a program for collecting electronic resources. With the rapid digitalization of teaching media, cultural archives and communications, the issue of effective collection, storage and use of language data becomes key to the preservation of cultural heritage and the maintenance of linguistic diversity. The study conducted a literature review covering relevant aspects of electronic resources, language data and data collection programs. The goal of the work is to develop a program specific to the Kazakh language. In the process of research, the tasks of studying the legislative framework, analyzing existing programs and developing a universal data collection system in the Kazakh language are solved.
Keywords: electronic resources, Kazakh language, crawling.

Введение

Растущий спрос на данные на казахском языке в сфере научных исследований, образования, культуры и бизнес-аналитики. Разработка не только удовлетворит потребности в доступе к информации на национальном языке, но и соответствует законодательству о защите данных и языковом многообразии. Законодательство Республики Казахстан отражает стремление к гармоничному сочетанию традиций и современности, а также поддержанию и развитию казахского языка в цифровой эпохе. Программа сбора электронных ресурсов призвана создать эффективный механизм для агрегации, хранения и доступа к разнообразным материалам на казахском языке, включая литературные произведения, исследования, языковые ресурсы и другие материалы. Таким образом, создание программы по сбору электронных ресурсов на казахском языке отражает стремление к сохранению, развитию и цифровой адаптации своего уникального языкового и культурного наследия в эпоху современных технологий.

Литературный обзор

В современном информационном обществе электронные ресурсы играют ключевую роль в сохранении и развитии языков и культур. В связи с этим, разработка программ по сбору электронных ресурсов казахского языка в соответствии с законодательством Республики Казахстан становится необходимой задачей. Целью данного литературного обзора является рассмотрение актуальных исследований и публикаций, касающихся этой темы, с целью обоснования необходимости разработки подобной программы. Статья "Аналитическая обработка текстовых ресурсов и документов на казахском языке" авторов Рахимовой и других (2019) посвящена разработке интеллектуальных поисковых систем, способных искать и извлекать новую информацию из текстовых данных на казахском языке в области образования. Авторы обосновывают актуальность данной темы в связи с растущим объемом данных в цифровой форме, предоставляющих возможность доступа к различным источникам электронных документов. Применение интеллектуальных поисковых систем позволит удовлетворить информационные потребности пользователей, и поэтому разработка информационно-аналитических поисковых систем, позволяющих работать с данными на казахском языке, является актуальной.

Цель исследования заключается в разработке эффективных алгоритмов и моделей для интеллектуальных поисковых систем на основе современных технологий в области информационного поиска и обработки естественного языка. Авторы стремятся создать системы, способные проводить анализ текстовых ресурсов на казахском языке, выявлять новые знания и улучшать качество образования в соответствии с потребностями современного информационного общества. Таким образом, данная работа имеет важное значение для развития интеллектуальных поисковых систем в контексте использования казахского языка и образования.

Работа Рахимова и Сатыбалдиева (2020) посвящена разработке системы автоматического сбора и обработки открытых данных на казахском языке из сети интернет, что имеет практическую значимость в задачах сбора и анализа текстов. Введение работы обосновывает актуальность выбранной темы, проводится обзор существующих подходов и формулируются задачи исследования. Одной из центральных задач является сбор и первичная обработка текстовых данных с последующим анализом. Авторы подчеркивают, что сбор данных является первоочередной задачей из-за неструктурированности открытых данных в сети интернет, требующей дополнительной обработки [1-2].

В работе представлена система обработки веб-страниц казахско-язычных порталов, а также приведено практическое применение данного подхода на реальных данных открытых ресурсов с помощью созданной системы. В рамках исследования представлен подход индексирования документов с использованием признаков. Ожидается, что разработанная система поможет структурировать открытые данные с ресурсов интернета на казахском языке и провести анализ собранных данных.

Методы сбора электронных ресурсов

1.Веб-краулинг (Web Crawling) [3]:

Веб-краулинг — это процесс автоматического сканирования веб-страниц для извлечения информации.

Веб-кроулинг, также известный как веб-сканирование или веб-спайдинг, — это автоматическое переползание (навигация) по веб-страницам в Интернете с целью сбора

информации. Краулеры или веб-пауки — это программные роботы, которые просматривают веб-страницы и собирают с них данные для различных целей, таких как индексация содержимого для поисковых систем, анализ структуры веб-страниц, извлечение информации и другие цели.

Процесс веб-ползания начинается с предоставления краулеру URL-адреса веб-страницы, которую необходимо посетить. Затем краулер загружает содержимое страницы и анализирует его, извлекая ссылки на другие страницы. Затем краулер переходит по этим ссылкам и продолжает процесс переползания, повторяя его до тех пор, пока не переполнит весь набор страниц или не будут выполнены другие условия остановки.

Веб-краулеры широко используются для различных целей, в том числе:

1. индексирование содержимого для поисковых систем, чтобы пользователи могли находить информацию в Интернете.

2. сбор данных для анализа, исследования или отслеживания онлайн-активности.

3. анализ структуры веб-сайтов для оптимизации производительности или улучшения качества работы пользователей.

4. проверка ссылок на веб-сайтах или обнаружение неработающих ссылок.

5. сканирование контента в целях безопасности или обнаружения угроз.

На языке программирования Python можно использовать библиотеки, такие как Scrapy или BeautifulSoup, для написания краулеров, которые могут автоматически переходить по ссылкам и извлекать данные с веб-страниц.

2.Использование API:

API (Application Programming Interface) - это интерфейс, позволяющий различным программным приложениям взаимодействовать друг с другом. API определяет методы и правила взаимодействия между различными программными компонентами, позволяя им обмениваться данными и выполнять определенные функции.

Основными особенностями API являются:

- Структурированные запросы и ответы: API определяет структуру запросов, которые могут быть отправлены программному обеспечению, и формат ответов, получаемых в ответ на эти запросы.
- Набор методов и функций: API предоставляет набор методов или функций, которые могут быть вызваны для выполнения определенных действий или операций. Эти методы и функции обычно документируются, чтобы разработчики могли понять их работу и использование.
- Стандартизация взаимодействия: использование API стандартизирует способ взаимодействия приложений друг с другом, облегчая интеграцию и обмен данными между различными системами.

Примерами API являются веб-интерфейсы, обеспечивающие доступ к данным или услугам через Интернет, API операционных систем, обеспечивающие управление ресурсами компьютера, API приложений, обеспечивающие доступ к функциям приложения, и т. д.

Многие онлайн-сервисы предоставляют API для доступа к своим данным и функциональности.

На языке Python можно использовать библиотеки для работы с API, например, requests для выполнения HTTP-запросов. Более подробную информацию можно увидеть в таблице 1.

Таблица1. Особенности методов сбора данных

Можно еще перечислить следующие методы сбора:

3.Веб-скрапинг [4]:

Веб- скрапинг — это процесс автоматического сбора информации с веб-страниц для извлечения данных. Веб-скрапинг выполняется с помощью программных средств, называемых веб-очистителями, которые получают доступ к веб-страницам, загружают их содержимое и извлекают нужные данные.

Основными компонентами веб-срапинг являются:

1. Загрузка веб-страницы: веб-скребки загружают HTML-код веб-страницы с помощью HTTP-запросов.

2. Анализ HTML: после загрузки страницы веб-скребки анализируют HTML-код страницы для поиска релевантных данных. Это может включать поиск определенных тегов (пример, или <p>) или атрибутов (пример, класс или идентификатор).

3. Извлечение данных: найденные HTML-элементы анализируются и извлекаются данные, соответствующие указанным критериям.

4. Обработка данных и резервное копирование: извлеченные данные могут быть обработаны и сохранены в нужном формате, например, в базе данных, CSV или JSON-файле.

Приложения для веб-анализа могут быть самыми разными - от автоматического сбора цен на товары в интернет-магазинах до сбора заголовков новостей с различных новостных сайтов. Однако важно отметить, что веб-анализ должен выполняться в соответствии с руководящими принципами использования Интернета и законами о защите данных. На некоторых веб-сайтах действуют правила, запрещающие несанкционированное удаление их содержимого, поэтому перед выполнением веб-очистки важно убедиться, что правила и политика этого веб-сайта не противоречат этим правилам [5-6].

Использование библиотек для анализа HTML-кода веб-страниц и извлечения нужных данных.

Метод сбора данных	Преимущества	Недостатки
Веб-краулинг	Автоматизация процесса	учет правовых и этических норм
Использование API	Доступ к данным через официальные источники	Ограниченный доступ
Веб-скрапинг	Гибкость и адаптивность	Нарушение правил использования данных
Скачивание и архивирование	Удобство хранения и передачи	Ограниченный доступ к определенным источникам

Используя и анализируя методы сбора анализа, программа включает в себя элементы веб-краулинга (для загрузки и отображения веб-страницы) и веб-скрапинга (для анализа HTML-кода и извлечения информации, такой как текстовые блоки из тегов). Более подробнее показана на рисунке 1.



Рисунок 1. Общая схема программы по сбору данных.

Начало программы выглядит так (Рисунок 2)

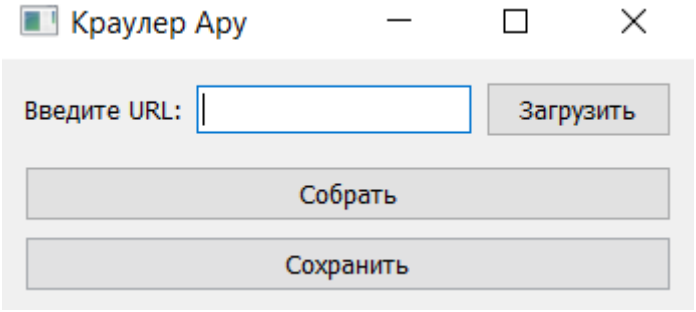


Рисунок 2. Начало работы программы

После нажатия кнопки “Собрать” программа собирает интернет-ресурсы с указанных источников.

Вывод

В заключении надо сказать, что автоматический сбор данных для любого языка очень необходим. Количество данных и объем нужен большой так как для обработки требуется большая статистика. Поэтому если владеть современными нынешними инструмента сбора данных, то можно облегчить задачу. И данная предложенная программа актуальна для юридических документов на казахском языке.

Использованная литература:

1. Shormakova, Zh. Zhumanov, B. Abduali, D. Rakhimova and D. Amirova Analytical Processing of Textual Resources and Documents in the Kazakh Language // Journal of Engineering and Applied Sciences 14 (20): 7714-7721, 2019
2. Д.Р. Рахимова, А.Р. Сатыбалдиев АЛГОРИТМ СБОРА ТЕКСТОВЫХ ДАННЫХ НА КАЗАХСКОМ ЯЗЫКЕ // «Физика-математика ғылымдары» сериясы, №2(70), 2020
3. Okogwu, Flora Ifeoma, "Understanding Electronic Resources Collection Development Practices Through Selected Theories" (2021). // Library Philosophy and Practice (e-journal). 5704.
4. Краулер, интернет ресурс (ru-brightdata.com)
5. Владимир Дронов, Николай Прохоренок // Python 3 и PyQt 5. Разработка приложений (2016)

6. Abdurakhmonova, N., & Ismailov, A. S. (2022). APPLYING WEB CRAWLER TECHNOLOGIES FOR COMPILING PARALLEL CORPORA AS ONE STAGE OF NATURAL LANGUAGE PROCESSING. In СОВРЕМЕННАЯ ФИЛОЛОГИЯ. СОЦИАЛЬНАЯ И НАЦИОНАЛЬНАЯ ВАРИАТИВНОСТЬ ЯЗЫКА И ЛИТЕРАТУРЫ (pp. 22-27).
7. https://primeminister.kz/decisions

O‘ZBEK KORPUSSHUNOSLIGIDA DIALEKTAL TA’MINOT YARATISH MASALALARI

O‘rinboyeva Ro‘za Toliboyevna
Mirzo Ulug‘bek nomidagi
O‘zbekiston Milliy universiteti magistranti.
E-mail: orinboyevaroza@gmail.com

Asqarov Ilhomjon Jamshid o‘g‘li
Mirzo Ulug‘bek nomidagi
O‘zbekiston Milliy universiteti magistranti.
E-mail: asqarovilhomjon21@gmail.com

Annotatsiya: Dialektal korpus yaratish masalasi soha vakillari oldidagi muhim vazifa hisoblanib, tilimizning barcha ko‘rinishlarini sof holda saqlab qolish imkonini beradi. Muayyan hudud shevasining ma’lumotlar bazasini yaratish hamda uni algoritmlash mutaxassislardan zaruriy bilim va ko‘nikmalarni talab etadi. Ushbu maqola bu vazifaning dolzarbligini yoritib berishga qaratilgan.

Kalit so‘zlar: korpus, dialekt, transkripsiya, sheva, ta’minot, fonema.

Abstract: The issue of creating a dialectal corpus is considered an important task for the representatives of the field, which allows us to preserve all aspects of our language in its pure form. Creation of a database of a specific regional dialect and its algorithmization requires the necessary knowledge and skills of specialists. This article aims to clarify the relevance of this task.

Keywords: corpus, dialect, transcription, dialect, supply, phoneme.

Аннотация: Вопрос создания диалектного корпуса считается важной задачей для представителей отрасли, позволяющей сохранить все проявления нашего языка в чистом виде. Создание базы данных конкретного регионального диалекта и ее алгоритмизация требуют необходимых знаний и навыков специалистов. Данная статья призвана прояснить актуальность данной задачи.

Ключевые слова: корпус, диалект, транскрипция, диалект, снабжение, фонема.

Kirish.

Dunyo miqyosida kompyuter lingvistikasi fani doirasida ahamiyatga molik ishlar amalga oshirilmoqda. Ular orasida eng muhimlaridan biri hisoblangan korpusshunoslik o‘z mutaxassislaridan amaliy natijalarni kutmoqda. Ko‘plab sohalar uchun boy manba bo‘la oladigan til korpuslarini yaratish va ularni takomillashtirish bugungi ta’lim jarayonining sifatini oshirish, o‘rganuvchilarga til o‘qitishning yangi uslub va tamoyillarini tatbiq etish, jamiyatning barcha qatlamlari uchun til o‘rganishni osonlashtirishda samarali vositalardan biridir.

O‘zbek tili korpusining yaratilishi til imkoniyatlarini umumlashtirish va ularni foydalanuvchilarga namoyon etishdagi asosiy omil vazifasini o‘tadi. Bu, o‘z navbatida, foydalanuvchilarga o‘zbek tilidagi sheva so‘zlar bilan tanishish, ularning muloqot jarayonidagi semantik ma’nolarini aniqlash hamda geografik jihatdan tarqalish hududlarini o‘rganish

imkoniyatini beradi[Abdurakhmonova, N. 2021,2022].. Korpusga yuklanadigan audio materiallar orqali foydalanuvchi muayyan shevaga oid so‘zning talaffuzi va uning qo‘llanish doirasi bilan tanishadi. O‘zbek shevalarining o‘ziga xos xususiyatlari bir necha maqolalarda, tadqiqot ishlarida yoritib berilgan, ammo ular korpus sifatida shakllantirilmagan. Bu esa mavzuning dolzarbligini belgilaydi. Ushbu maqolada o‘zbek tili korpusi uchun dialektal ta’minot yaratish masalalari xususida so‘z yuritiladi.

Adabiyotlar tahlili va metodlar.

Dialektal korpus yaratish bo‘yicha jahon miqyosida ahamiyatga molik ishlar amalga oshirilgan. Hozirda olimlar ularning funksiyalarini yanada takomillashtirish ustida ish olib borishmoqda. Ko‘pgina tillardagi dialektal korpuslar foydalanuvchilarga havola etilgan. Ularning yaratilishi haqida turli xil tezis, maqola, dissertatsiya ishlari orqali boxabar bo‘lishimiz mumkin. Xususan, rus tili dialektal korpusi haqida A.B.Letuchiy [Летучий А.Б., 2003-2005], I.B.Kachinskaya, D.V.Sichinava [Качинская И.Б., Сичинава Д.В., 2015], tatar tili dialektal korpusi bo‘yicha Liliya Yansaeva [Лилия Я.А., 2008], ingliz tili dialektal korpusi bo‘yicha Lieselotte Anderwald, Susanne Wagner [Anderwald L., Wagner S., 2007] kabilar atroflicha fikr yuritib o‘tgan.

Lieselotte Anderwald va Susanne Wagner “Corpus linguistics and dialectology” [Anderwald L., Wagner S.] nomli maqolasida hozirda ishlayotgan dialektal korpuslarni tahlilga tortib, ulardagi materiallarni bir necha toifalarga ajratadi. Uning fikricha, dialektal ta’minot uchun ko‘proq sotsiolingvistik materiallar to‘planadi. Muallif ularning eng yaxshisi sifatida SED yozuvlarini ko‘rsatadi. O.Y. Kryuchkova, V.E. Goldinlar dialektal matnlarni ilmiy tahlil qilish uchun nafaqat leksik, grammatik, balki fonetik hodisalarni ham imkon qadar to‘liq aks ettirishni ta’kidlaydi [Kryuchkova O.Y., Goldin V.E., 2011]. Bunda birinchi maqsad orfografik yozishga yaqin bo‘lgan transkripsiyaga mos keladi.

O‘zbek tili uchun dialektal korpus yaratishda quyidagi bosqichlarni amalga oshirish lozim:

1. Dialektal materiallarni to‘plash: har bir dialektning o‘ziga xos xususiyatlari bo‘lib, ularni o‘zida aks ettirgan matnlar to‘planadi.
2. Analiz qilish: to‘plangan materiallar tahlil qilinadi, umumiy va farqli jihatlar aniqlanadi.
3. Ta’minot yaratish: tahlil natijalariga asosan har bir dialektning xususiyatlari so‘zlar, iboralar, frazalar va shakllar ko‘rinishida bo‘ladi.
4. Sinovdan o‘tkazish: yaratilgan dialektal ta’minot sinovdan o‘tkaziladi, mavjud kamchiliklar bartaraf etiladi.
5. Ta’minotni integratsiyalash: tayyor bo‘lgan ta’minot platforma, dastur yoki internet saytlariga yuklanadi.

O‘zbek dialektal korpusini yaratish doirasida ushbu vazifalarning qay tartibda bajarilishi N.Z.Abduraxmonova [Abduraxmonova N., 2001], M.A.Xolova [Xolova M., 2020], Z.X.Rustamova [Rustamova Z., 2022] singari olimalarning ishlarida batafsil yoritib berilgan. Ular o‘zbek tili dialektal korpusining mohiyati, xususiyatlari haqida bayon etishgan. Dialektolog olimimiz Samixon Ashirboyev [Ashirboyev S., 2021] o‘zining darslik qo‘llanmasida shevalarni qanday tasniflash, ularni transkripsiyada qanday ifodalash yo‘l-yo‘riqlarini batafsil yoritib bergan. Qolaversa, N.Murodova [Murodova N., 2000], M.Saidova [Saidova M., 1995], M.Q.Begaliyev [Begaliev M.Q., 2007], .Z.Y.Ibragimova [Ibragimova Z.Y., 2009]lar shevalarni hududiy kesimda o‘rganib, ularni o‘z ishlarida nazariy jihatdan aks ettirganlar.

Dialektal ta’minot yaratishning amaliy ahamiyati.

Dialektal ta’minot yaratish korpusshunoslikda keng imkoniyatlarni ochib beradi. Uning amaliy ahamiyati bir nechta muammolarga yechim topa olishi bilan belgilanadi:

1. Sayt yoki dastur orqali foydalanuvchilarni ularning o‘z lahja, dialekti yangiliklaridan boxabar qiladi;
2. Yaratilgan dialektal ta’minot korpusga yuklanib keyingi tadqiqotlar uchun asos bo‘lib xizmat qiladi;
3. Dialektal materiallar va uning audiomatnlaridan o‘zbek nutq sintezatori uchun manba sifatida foydalanish mumkin;
4. Foydalanuvchilar (asosan, dialektologlar) ga dilaektal materiallarni transkripsiyasi bilan birga taqdim etadi;
5. Aloqador bo‘lgan so‘zlarni va iboralarini aniqlash: dialektal ta’minot shevalar va lahjalar o‘rtasidagi o‘zaro munosabatlarni taqqoslash imkoniyatini beradi;
6. Regional foydalanuvchilarga qaratilgan mahsulotlar va xizmatlar: agar kompaniya yoki tashkilot biron bir hududga xizmat ko‘rsatmoqchi bo‘lsa, dialektal ta’minot bu hududning dialektidagi so‘z va ifodalarni tushunish hamda foydalanuvchilarni qiziqtirishga yordam beradi.

Dialektologiyada olib boriladigan har qanday amaliy tadqiqot ishi dialektal ta’minotni talab qiladi. Bu, o‘z navbatida, masalaning amaliy ahamiyatini belgilab beradi.

Xulosa.

O‘zbek tili shevalari nazariy jihatdan yetarlicha o‘rganilgan. Ularning fonetik xususiyatlari muayyan qonun-qoidalarga tushirilgan. Amaliy jihatdan dialektal manbalarni bir korpus ostida jamlash soha doirasidagi navbatdagi yutuq hisoblanadi. Dialektal ta’minot yaratish bu boradagi dastlabki bosqich bo‘lib, kompyuter lingvistlari oldidagi muhim vazifalardan biridir.

Foydalanilgan adabiyotlar:

1. Летучий А.Б. Корпус диалектных текстов: задачи и проблемы // Национальный корпус русского языка: 2003-2005. М.: Индрик, 2005.
2. Качинская И.Б., Сичинава Д.В. Диалектный подкорпус сегодня// Тр. ин-та русского языка им. В.В.Виноградова РАН. 2015.
3. Лилия Я.А. Диалектный синтаксис татарского языка. Диссертации на соискание ученой степени кандидата филологических наук. Казань, 2008.
4. Anderwald L., Wagner S. The Freiburg English Dialect Corpus: Applying Corpus-Linguistic Research Tools to the Analysis of Dialect Data. Beal, J.C., Corrigan, K.P., Moisl, H.L. (eds) Creating and Digitizing Language Corpora. Palgrave Macmillan, London. https://doi.org/10.1057/9780230223936_3
5. Abduraxmonova N.Z. O‘zbek tili elektron korpusining kompyuter modellari. Monografiya. GlobeEdit, 2021.
6. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). Uzbek electronic corpus as a tool for linguistic analysis. In *Компьютерная обработка тюркских языков. TURKLANG 2022* (pp. 231-240).
7. Abduraxmonova, N. Z. Q., & Urazaliyeva, M. Y. (2022). O ‘zbek tili elektron korpusida (http://uzbekcorpus. uz/) og ‘zaki matnlar korpusini yaratishning nazariy va amaliy masalalari. *Academic research in educational sciences*, *3*(3), 644-650.

8. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, *11*(19), 9117.
9. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In *of the International Conference on Language Technologies for All (LT4All)*.
10. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).
11. Ashirboyev S. O‘zbek dialektologiyasi (darslik). T.: Nodirabegim, 2021.

DEVELOPING PRE-SERVICE STUDENTS' CORPUS LITERACY SKILLS IN TEACHING ENGLISH

Sultonova Munojat
PhD researcher of the Fergana state university.

Annotation: This article explores the importance of developing pre-service philology students' corpus literacy skills through various methods. The authors discuss the benefits of incorporating corpus linguistics into the curriculum, such as improving students' ability to analyze language patterns and trends. They also highlight different approaches to teaching corpus literacy, including hands-on activities and technology-based tools. Overall, this research emphasizes the value of equipping future philologists with the necessary skills to navigate and utilize corpora effectively in their academic and professional endeavors.

Keywords: pre-service, students, corpus linguistics, teaching, literacy, methods

Аннотация: В данной статье исследуется важность развития навыков корпусной грамотности у студентов-филологов с помощью различных методов. Авторы обсуждают преимущества включения корпусной лингвистики в учебную программу, например, улучшение способности учащихся анализировать языковые модели и тенденции. Они также освещают различные подходы к обучению корпусной грамотности, включая практические занятия и технологические инструменты. В целом, это исследование подчеркивает ценность предоставления будущим филологам необходимых навыков для навигации и эффективного использования корпусов в их академических и профессиональных начинаниях.

Ключевые слова: предварительная подготовка, студенты, корпусная лингвистика, преподавание, грамотность, методы.

Introduction

Corpus linguistics has become an essential tool for language analysis and research, providing valuable insights into language patterns and usage. As such, it is crucial for pre-service students in the field of philology to develop corpus literacy skills to enhance their understanding of language structure and usage. This article explores the various methods that can be employed to develop pre-service students' corpus literacy skills, ultimately preparing them for successful careers in philology.

Incorporating Corpus Linguistics into the Curriculum

One effective method for developing pre-service students' corpus literacy skills is through the integration of corpus linguistics into the curriculum. By incorporating corpus-based activities and assignments into coursework, students can gain hands-on experience with analyzing language data and identifying linguistic patterns. For example, students may be tasked with conducting corpus searches to investigate specific language phenomena or trends, allowing them to develop critical thinking and analytical skills.

Bardovi-Harlig and Mossman discuss the importance of using corpus-based materials in teaching pragmatic routines. They emphasize the benefits of incorporating authentic language data

into language learning materials to help students develop a better understanding of how language is used in real-life situations. Their research highlights the value of integrating corpus linguistics into materials development for language learning, ultimately enhancing students' ability to communicate effectively in various contexts [Bardovi-Harlig and Mossman., 2016: 250].

According to Aijmer, the use of corpora in language teaching, encompassing theoretical, pedagogical, and technological aspects are discussed in research. Findings from the research are explored how corpora can be used in different language teaching contexts, including vocabulary acquisition, grammar teaching, pronunciation training, and learner autonomy. Besides, it is examined the impact of corpus-based approaches on learner language use, motivation, and confidence. Covers the technology and software used to access and analyze corpora, including corpus query tools and concordances [Aijmer., 2009].

Campoy et al., also provides a comprehensive overview of corpus-based approaches to English language teaching (ELT). The research covers both theoretical and practical aspects, with a focus on the use of corpora in language teaching materials and classroom activities. As theoretical foundations they discussed the nature of corpora and their relevance to ELT, including the principles of corpus linguistics and the different types of corpora available. As a result, they explored the benefits and challenges of using corpora in language teaching, and provided guidance on how to select and use corpora effectively. As for pedagogical applications, they examined how corpora can be used to inform and enhance language teaching practices in various areas, such as vocabulary teaching, grammar instruction, pronunciation training, and writing instruction. They provide practical examples and case studies of corpus-based language teaching activities and materials [Campoy.,2010].

Callies argues for the importance of integrating corpus literacy into language teacher education programs. Corpus literacy refers to the ability to use corpora (collections of authentic language data) to inform language teaching and learning. As benefits of Corpus Literacy for Language Teachers, the research enables teachers to access and analyze authentic language data to inform their teaching practices, helps teachers develop a deeper understanding of language use and variation, supports teachers in creating more effective and engaging language teaching materials and fosters critical thinking and research skills in teachers. However, there are some challenges of Integrating Corpus Literacy into Teacher Education as it requires specialized training and knowledge of corpus linguistics. Moreover, it can be time-consuming to incorporate into already crowded teacher education programs. To overcome such challenges, it is recommended that to introduce corpus concepts and tools early in teacher education programs by providing hands-on training in corpus analysis and query tools, corpus-based activities into language teaching methodology courses with the collaboration between language teacher educators and corpus linguists. Callies concludes that integrating corpus literacy into language teacher education is essential for preparing teachers to meet the demands of 21st-century language teaching. By developing corpus literacy, teachers can enhance their teaching practices, create more effective materials, and foster learner autonomy [2019: 245].

Another approach to developing pre-service students' corpus literacy skills is through the use of technology-based tools. Online corpora platforms, such as the Corpus of Contemporary American English (COCA) or the British National Corpus (BNC), provide students with access to vast amounts of linguistic data for analysis (Baker et al., 2006). By familiarizing students with

these tools and teaching them how to effectively navigate and interpret corpus data, educators can help students develop proficiency in corpus analysis.

In addition to incorporating corpus linguistics into the curriculum and utilizing technology-based tools, hands-on activities can also be beneficial for developing pre-service students' corpus literacy skills [Sultonova M., 2023]. For example, educators may design interactive exercises that require students to analyze corpus data and draw conclusions about language usage. These activities not only reinforce theoretical concepts but also provide practical experience in working with corpora.

Developing pre-service students' corpus literacy skills is essential for preparing them for successful careers in philology. By incorporating corpus linguistics into the curriculum, utilizing technology-based tools, and designing hands-on activities, educators can help students develop the necessary skills to analyze language data effectively [Sultonova M., 2023]. Ultimately, these methods will empower pre-service students to navigate and utilize corpora proficiently in their academic and professional endeavors.

References:

1. Aijmer K. (Ed.). (2009). Corpora and language teaching. Gothenburg: John Benjamins Publishing.
2. Bardovi-Harlig, K., Mossman, S. Corpus-based materials development for teaching and learning pragmatic routines. In Tomlinson, B. (Ed.) SLA research and materials development for language learning. London: 2016. pp. 250-278.
3. Callies, M. Integrating corpus literacy into language teacher education. In S. Götz, J. Mukherjee (Eds.), Learner Corpora and Language Teaching. Philadephia: 2019. pp. 245–263.
4. M. Campoy, B.Cubillo, B. Belles-Fortuno, M.Gea-Valor (Eds.). Corpus-based approaches to ELT. London: 2010.
5. M. Sultonova. Ingliz tilini o’qitishda korporadan foydalanishni deduktiv va induktiv metodlari. Research and Implementation. 2023
6. M. Sultonova. Incorporating corpus literacy skills into teaching english teacher training. Research and Implementation, 2023.
7. M. Sultonova. Developing efl teachers’ sociolinguistic competence through online exchange programs in uzbekistan. Евразийский журнал академических исследований, 3(4 Special Issue), 202389–91. https://in-academy.uz/index.php/ejar/article/view/14475

ЛИНГВОМЕТОДИЧЕСКИЙ ПОТЕНЦИАЛ ПАРАЛЛЕЛЬНОГО КОРПУСА

Хидирова Гулнора
кандидат филологических наук, доцент
Университет Маъмуна.

Аннотация: В статье рассматриваются структура и назначение параллельного корпуса. Под корпусом понимается совокупность устных и письменных текстов, хранящихся в базе данных компьютера, также в корпусе представлено подробное описание точного времени написания собранных материалов, стиль текстов. Корпус может определять диапазон слов, денотативные и коннотативные значения искомого слова, описывает частоту или статистику использования того и иного слова в языке писателя. Параллельный корпус описывается как новый вид лингвистических ресурсов, которые все больше используются при обучении языкам. На основе данных корпусов выявляются несовершенства современных учебников. В параллельном корпусе собраны тексты оригинала и его переводов. Параллельные корпуса служат обучающими данными для статистических систем машинного перевода. С точки зрения переводоведения важной задачей параллельного корпуса является обнаружение переводческих изменений. Основная цель создания параллельного корпуса – разработка эффективных и адекватных методов перевода. Посредством параллельного корпуса достигается улучшение качества перевода.

Ключевые слова: корпусная лингвистика, текст, стиль, параллельный, структура, сопоставительный, преподавание языка, переводоведение, машинный перевод, словари, процесс автоматической нормализации текстов, семантика, статистические технологии.

Annotation: The article discusses the purpose and functions of the parallel corpus. A corpus is understood as a collection of oral and written texts stored in a computer database; the corpus also provides a detailed description of the exact time of writing of the collected materials and the style of the texts. Modern methodological literature actively uses examples and descriptions from parallel corpora when teaching a language; increasingly, the corpus is used to analyze and identify the shortcomings of existing textbooks in teaching foreign languages. The corpus can determine the range of words, the denotative and connotative meanings of the searched word, and describes the frequency or statistics of the use of a particular word in the writer's language. The parallel corpus is described as a new type of linguistic resource. Parallel corpora serve as training data for statistical machine translation systems. The parallel corpus contains the texts of the original and its translations. From the point of view of translation studies, an important task of a parallel corpus is the detection of translation changes. The main goal of creating a parallel corpus is to develop effective and adequate translation methods. By using a parallel corpus, translation quality is improved.

Keywords: corpus linguistics, text, style, parallel, structure, comparative, language teaching, translation studies, machine translation, dictionaries, process of automatic text normalization, semantics, statistical technologies.

Введение. Одной из глобальных проблем XXI века является сохранение национальной идентичности естественных языков. Актуальной задачей мировой науки

стало последовательное проведение исследований нейролингвистического программирования и технологий в создании и развитии электронных языковых корпусов. Научные и практические исследования в области иностранной корпусной лингвистики доказали, что корпус является желательным и необходимым основанием не только для ученых лингвистов, занимающихся работой со словом, но и для развития и сохранения нации.

И в Узбекистане одной из важнейших задач, стоящих сегодня перед прикладной лингвистикой, стало создание Национального корпуса узбекского языка. Помимо научно-исследовательских задач, корпус даст импульс в достижении множества экстралингвистических целей, в частности:

- повышение престижа узбекского языка в обществе;

- создание Национального корпуса узбекского языка, содержащего всю научную, теоретическую и практическую информацию на узбекском языке;

- популяризация узбекского языка в мировой информационной сети Интернет и обеспечение его достойного места на международном уровне;

- создание узбекских приложений программных продуктов, компьютерного обучения узбекскому языку, работы по созданию компьютерных программ для редактирования текстов на узбекском языке [Указ Президента - №5850; 2019].

На настоящий момент это направление в тюркологии также активно разрабатывается, отметим труды лишь некоторых ученых, например, вопросы корпусного представления турецкого языка исследовали Дениз, Аксан, Зейрек, Кемал Офлазер, Умут Озге Булар; уйгурского языка - Юсуп Аибаидулла, Ким Тенг Луа; башкирского языка - З.А. Сиразитдинов, Л.А. Бускунбаева; хакасского языка - А.В. Шеймович, А. В. Дыбо, В. С. Мальцева, Э. В. Султрекова; татарского языка - Ж. Сулейманов, О. Невзорова, Б. Хакимов; тувинского языка - А.Я. Салчак, Мира В. Бавуу-Сюрюн, Чодураа М. Монгуш [Абдурахмонова Н.; 2020].

Концептология Национального корпуса узбекского языка разрабатывается коллективами ученых Государственного университета узбекского языка и литературы, Узбекского национального университета, Ташкентского университета информационных технологий, Самаркандского, Андижанского, Термезского госуниверситетов.

Корпус — это электронная коллекция текстов, которая формируется на основе различных аспектов. Различают самые разнообразные классификации корпусов: по видам базы данных (устная, письменная), языку текстов (русский, немецкий, турецкий, параллелизму переводов текстов (двуязычный, трехъязычный), стилю (художественный, научный…), доступности базы (открытая, закрытая), географическому фактору (относящийся к одному государству/…) и др.

Таким образом, под корпусом понимается совокупность устных и письменных текстов, хранящихся в базе данных компьютера. Также в корпусе представлено подробное описание точного времени написания собранных материалов, стиль текстов. Пользователь может ссылаться на тексты, представленные в художественном, научном, официальном или публицистическом стилях, в зависимости от целей и задач. Это особенно полезно при обучении языку. Ввиду того, что корпус представляет собой систематизированную библиотеку с очень широким охватом и высокой степенью достоверности, учитель в школе может получить действенную поддержку в оперативном составлении заданий учащимся на

этапе закрепления знаний. Он прост в использовании, значительно экономит время [Абдурахмонова Н., Урдишев К., 2019].

Современная западная методическая литература уже активно использует примеры и описания из параллельных корпусов при обучении языку. Примеры из корпуса дают студентам практический материал, с которым они столкнутся при использовании языка в реальных ситуациях межкультурной коммуникации. Нужно отметить, что основным недостатком многих учебников являются изобретенные примеры, а их описания строятся на интуиции или субъективных выводах [Кокорева А. А., 2013]. В работе с корпусом этот момент исключается совершенно, т.к. он выдает примеры из реальных текстов. В связи с этим, все чаще корпус используется для анализа и выявления недостатков существующих материалов в преподавании иностранных языков.

Поиск по корпусу позволяет пользователю найти все формы указанного слова в различных контекстах, что выгодно отличает электронную поисковую систему от других. Наглядно показывает варианты слова, какое место они занимают в словаре. Он может определять диапазон слов, денотативные и коннотативные значения искомого слова, которые могут быть спряжены. Также описывает частоту или статистику использования того и иного слова в языке писателя.

Основная часть. Направления корпусной лингвистики, в том числе проекты электронных корпусов текстов, активно развиваются и имеют значительный прикладной потенциал в методике обучения иностранным языкам и переводу, а также в компьютерной лингвистике. Вопросам обучения языку и переводу на базе параллельных корпусов уделено большое внимание в работах таких ученых как М.Барлоу, Маккенери, Бэйкера, Уилсона, Дэнилсон и Райдингс, Заннети, Аренберга, Бланка, Брауна, Черча и Гейла, Девиса, Фостера и многих других.

Использование электронного корпуса текстов на иностранном языке при переводе художественных текстов позволяет найти способ устранения ошибок, заключающихся в употреблении таких конструкций и принятии таких решений, которые не являются очевидным выбором носителя языка. К ним относятся проблемы лексико-грамматической сочетаемости, орфографические, пунктуационные трудности и трудности, связанные с выбором оптимальной грамматической конструкции [Груздев Д. Ю., 2013].

К началу 21 века в мире было создано множество двуязычных/многоязычных корпусов, к примеру: EUROPARL; CHEMNITZ GERMAN-ENGLISH TRANSLATION CORPUS; KACENKA; Lancaster's ITU, Johansson and Hofland; HKUST; Erjavec et al.; "Agenda 21"; Kraaij; Resnik et al., Tadic и др.

В настоящее время действуют или находятся на стадии разработки корпуса многих языков мира: Национальный корпус русского языка [НКРЯ, 2003] - представительная коллекция текстов на русском языке общим объемом более 2 млрд слов, оснащенная лингвистической разметкой и инструментами поиска [https://ruscorpora.ru]. В состав НКРЯ вошли 38 корпусов. Некоторые из корпусов включают подкорпусы, в частности, Параллельный корпус содержит 22 параллельных подкорпуса общим объемом 165 млн слов.

TUD - Национальный корпус турецкого языка [https://www.tnc.org.tr/tr/] версии 3.0 представляет собой исчерпывающий, сбалансированный и представительный корпус-

справочник общего назначения объемом 50 миллионов слов, охватывающий период в 24 года (1990–2013 гг.)

С 2012 года разрабатывается Алматинский корпус казахского языка [АККЯ] – один из возможных вариантов Национального корпуса казахского языка [http://web-corpora.net/KazakhCorpus]. С 2019 года разрабатывается Национальный корпус таджикского языка (НКТЯ). Корпус создан в результате совместной работы таджикских и российских специалистов. На настоящем ресурсе [https://tajik-corpus.org/] находится письменный корпус литературного таджикского языка объемом 40,8 млн словоупотреблений. Каждая разобранная словоформа включает грамматическую информацию и перевод на русский и английский язык.

Проводятся теоретические и практические исследования по формированию веб-корпуса узбекского языка: Ташкентский государственный университет узбекского языка и литературы, Национальный университет Узбекистана, Ташкентский университет информационных технологий, филиал Самаркандского государственного университета ТАТУ. В 2020-2021 годах в Ташкентском государственном университете узбекского языка и литературы реализован прикладной проект «Создание учебного корпуса узбекского языка».

В 2021 году под руководством профессора Национального университета - Абдурахмоновой Н.З. запущен проект по созданию Электронного корпуса узбекского языка [https://uzbekcorpus.uz]. Современное состояние «Узбекского электронного корпуса» представляет собой комплекс лингвистических словарей, веб-страниц, морфологической базы данных узбекского языка, учебной и научной литературы, официальных и художественных текстов различных жанров. Корпус содержит набор устных и письменных текстов, программное обеспечение с возможностью управления и использования данных [https://uzbekcorpus.uz].

Под руководством профессора Самаркандского государственного университета Суюн Каримова проводится прикладное исследование «Проектирование национального корпуса узбекского языка и разработка программного комплекса». Основным результатом исследования станет программный комплекс, разработанный для создания национального корпуса узбекского языка. В 2022 году программное обеспечение было протестировано и готово к использованию в рамках созданного корпуса текстов эпоса «Алпомиш».

Рассмотрим параллельный корпус как новый вид лингвистических ресурсов. Автономная часть электронного корпуса способна вместить в себя огромное количество необходимой информации. В направлении машинного перевода существуют многоязычные корпуса, специально отформатированные для параллельного сравнения, называемые систематизированными параллельными корпусами.

Спектр использования параллельного корпуса достаточно широк: в обычном и машинном переводе, в исследовании проблем теории и практики перевода, лингвистики, вычислительной лингвистики. С начала 1980-х годов корпуса перевода использовались в вычислительной лингвистике для машинного перевода, а также для извлечения терминов, определения значения слов и др. [Абдурахмонова Н., 2016]. Вопросы построения, состава и возможностей параллельного корпуса исследуются в работах Д.О.Добровольского, Ю.Тао, В.Захарова, А.А.Кокоревой, Е.П.Сосниной.

Таким образом, параллельный корпус представляет собой коллекцию текстов-оригиналов вместе с их переводами на другой язык или языки. Параллельный корпус рассматривается как база для проведения эмпирических исследований в лексикографии и теории перевода, и как один из элементов в методике обучения иностранному языку. В переводоведении и контрастивной лингвистике параллельные корпуса занимают центральное место. Многие параллельные корпуса доступны через простые в использовании конкордансы, что значительно облегчает изучение межъязыковых явлений.

Далее опишем сам процесс создания электронного параллельного корпуса, состоящий из переводов на русский язык художественных произведений узбекских писателей. Разработка параллельного корпуса на основе узбеко-русского языковой пары, наряду с возможностью проведения на его основе исследований самых разных направлений, представляет собой новый научный метод исследования, осуществляемый непосредственно на представительном материале художественных переводов с узбекского языка на другие языки мира (метод может быть применен как для прямого, так и для обратного направлений языковых пар). Такой подход к решению переводческих проблем называется корпусным.

Параллельный корпус разрабатывается в целях создания электронной платформы по исследованию перевода и обучению переводу. В связи с этим общий алгоритм создания параллельного корпуса будет следующим: определение содержания коллекции → сбор оригинальных текстов и их переводов → установление программного инструмента → статистика корпусных данных → работа с данными корпуса. Создание корпуса тесно связано с его назначением, поэтому на основе корпуса будет возможным проведение лингвистического анализа художественного перевода произведений. Таким образом решается задача – как нужно структурировать параллельный корпус, предназначенный для анализа переводов, для создания переводных словарей и используемый в создании новых переводов. В связи с этим приоритетной задачей является определение необходимых параметров и характеристик, какими следует наделить корпус электронных текстов, позволяющие улучшить качество перевода художественного дискурса на русский язык.

Для этой цели мы предприняли попытку проанализировать опыт составителей национальных и других представительных корпусов текстов различных языков с тем, чтобы выработать методические рекомендации по составлению и использованию параллельных корпусов. Особое внимание к процедуре поиска информации в корпусе объясняется тем, что для ее извлечения из такого корпуса необходимо специальное программное обеспечение, т.н. корпус-менеджеры. На основе нескольких составленных корпусов будут выработаны рекомендации по поиску информации в них с помощью таких программ [Груздев Д. Ю., 2016]. Для разработки параллельного корпуса в качестве образца выбраны 3 действующих электронных корпуса: Национальный корпус русского языка, Национальный корпус японского языка и Национальный корпус турецкого языка.

Параллельный корпус предназначен для анализа, исследования переводов, поэтому, как упоминалось выше, создаётся целый комплекс сложных задач.

В процессе создания параллельного корпуса будет определена общая структура параллельного корпуса, дана характеристика отобранных текстов, описаны метаданные и поиск словоизменительных парадигм, таким образом, решается задача создания электронного корпуса текстов.

Во-вторых, создаётся программа - конкорданс (программное обеспечение) параллельных текстов. Для согласования и создания конкордансов параллельных текстов стоит задача выбора, установки и адаптации одного из вариантов программы-конкордансер ParaConc для исследуемой пары языков (узбекский-русский). Данная задача нуждается в дополнительном анализе для принятия решения о целесообразности использования веб-приложений вместо программ-конкордансеров, а также использования Интернета вместо корпуса текста, что в определенной степени позволит освободить переводчиков от необходимости составлять тематические корпусы [Груздев Д. Ю., 2016].

Разрабатывается сайт корпуса, через веб-интерфейс которого реализуется поиск морфологических форм, поиск с добавлением элементов метаданных и поиск стилистических единиц (н-р, реалии, экзотизмы, языковая картина писателя).

Далее также намечается разработка программы предварительной обработки текста, автоматическая лемматизация слов в текстах в обоих (русскоязычной и узбекоязычной) частях корпуса. Поиск в корпусе осуществляется посредством запроса по буквам или в виде набора символов (выбора параметров из меню) разной степени сложности. Электронный формат корпуса обусловливает обязательное наличие специальной разметки - лингвистической или метаязыковой информация о выбранных единицах корпуса: текст, предложение, текстоформа, звук.

Создание корпуса начинается с процедуры обработки текстов. Проведение предварительной обработки состоит из следующих этапов: 1) заполнение текстовых таблиц. Импорт данных таблицы в параллельный корпус; 2) выравнивание. На первом этапе выравнивание корпуса выполняется автоматически с использованием Paraconc. На втором этапе ошибки выравнивания исправляются в ручном режиме. Затем производится загрузка, которая осуществляется с помощью следующих действий: file – export – export corpus files. Выравненные тексты загружаются автоматически на платформе Paraconc. Для того чтобы файл открывался, загруженный файл должен быть сохранён вместе с микрософтом Paraconc.

Особое внимание к процедуре поиска информации в корпусе объясняется тем, что для ее извлечения из такого корпуса необходимо специальное программное обеспечение, т.н. корпус-менеджеры. На основе составленного корпуса будут выработаны рекомендации по поиску информации в них с помощью таких программ.

На основе анализа возможностей других электронных ресурсов, в основе которых лежат корпусы текстов, например, программы накопительного перевода, был выявлен ряд присущих им потенциальных недостатков, а именно – 1) использование переводных текстов и 2) навязывание переводчику единственно возможного варианта перевода. Ввиду отсутствия таких ограничений у тематических корпусов текстов мы предполагаем, что они дают возможность придать процессу перевода творческий характер. Они также являются более надежным источником фоновой информации. Все это в комплексе должно способствовать повышению качества переводимых на русский язык художественных текстов в целом. Для проверки данного предположения требуется проведение лингвистического эксперимента.

Поиск в корпусе осуществляется на основе языка регулярных выражений (regular expressions). Морфологическая нормализация для узбекского языка не производится. Так как целью создания корпуса является исследование переводов, то основной режим поиска

- поиск на платформе Paraconc. Кроме того, создаётся платформа поиска в Интернете. Там будет реализован поиск по лексическим единицам/словоформам с добавлением элементов метаданных (retrieval intersection). Например, мы ищем контексты для слова «мусофирлар» в текстах, по автору и конкретному переводчику.

Генерация частотного словника. Материал корпуса – художественный текст и, естественно, в оригиналах и переводах много слов и выражений, отражающих идиостиль писателя. Их перевод является приоритетным, поэтому создание словника или базы данных языка писателя - одна из функций параллельного корпуса. На первом этапе мы определяем частотность слов из текстов на узбекском языке и их переводов в текстах на русском языке и выравниваем их. Выровненные слова и обороты составляют новый текст (в формате txt). На втором этапе мы импортируем выровненные тексты по строчке с помощью инструментов (eclipse и java) в базу данных (database) и выполняем генерацию частотного словника по данным корпуса [Захаров В., Yuan Tao., 2015]. Словник составит основу Толково-переводного словаря произведений писателя.

Далее разрабатывается сайт корпуса, через веб-интерфейс которого реализуются поиск морфологических форм, поиск с добавлением элементов метаданных и поиск частотных слов [Захаров В., Yuan Tao., 2015]. Метаданные корпуса включают следующую информацию: язык, тип, автор, переводчик, время издания и название. В метаданных мы указываем два языка – узбекский и русский, три типа: лингвистика, литературоведение и переводоведение, два вида времени издания - год публикации оригинала и перевода, название на двух языках – узбекском и русском [Захаров В., Yuan Tao., 2015]. Таким образом, электронный корпус даст возможность изучать стандартные лексические и синтаксические соответствия между двумя языками, анализировать путем сопоставления различных переводов принципы, лежащие в основе той или иной переводческой стратегии [Маник С. А., 2019].

Параллельный корпус (ПК) будет состоять из двух частей:

1. Переводной корпус текстов на узбекском языке (например, все произведения А.Кадыри) и их переводов на русский язык и

2. Сопоставимый корпус (СК) из тематических текстов на русском языке. Тематические тексты представят художественные произведения русскоязычных писателей Узбекистана, например, таких как - С. Бородин, М. Шевердин, В. Ян, Я. Ильясов, А. Удалов, Б. Пармузин.

Совпадение стиля текстов параллельного и сравнительного корпусов обеспечивает соспоставимость особенностей переводного и оригинального русского языка.

Переводной русский язык сопоставляется не только с оригинальным узбекским языком в переводном корпусе, но и соригинальным русским языком в сравнительном корпусе. Оригинал и перевод каждого произведения переводного корпуса создают свой подкорпус, и оригинал каждого произведения сравнительного корпуса создает соответствующий подкорпус. Таким образом, сопоставляются и количествено анализируются не только сами корпуса, но и пары подкорпусов. Отметим, что учет баланса количества слов между подкорпусами повышает надежность исследования.

Структуру корпуса узбекского и русского языков можно представить в виде следующей схемы:



Как отмечалось выше, в рамках параллельного корпуса возможно проведение корпусных исследований, то есть исследование языка с помощью корпусных методов. Данный подход основывается на использовании электронных корпусов текстов - огромных массивов естественных текстов на иностранном языке, собранных на магнитном носителе и должным образом упорядоченных и размеченных для более быстрого поиска интересующей лингвистической информации. Корпусный метод основан на статистической обработке языка, т.е. в корпусных исследованиях используются количественные методы. Параллельные корпуса выявляют закономерности в тексте оригинала и текста перевода, так как одинаковая коммуникативная направленность в двух разноязычных текстах является объектом сопоставительных корпусов.

Следование корпусному подходу при решении переводческих проблем представляет возможности повышения качества перевода. В настоящего времени исследователи корпусной лингвистики решают задачу составления корпусов текстов и выработки методологии работы с ними для проведения дальнейших лингвистических исследований и использования данного ресурса в качестве одного из способов решения проблемы улучшения перевода. Исследования в русле корпусного подхода к решению переводческих проблем стали привлекать внимание широкого круга ученых лишь в недалеком прошлом, к примеру международная конференция по данному вопросу состоялась в России в 2004 году. Хотя методология проведения микроисследования с помощью корпуса текстов в процессе перевода не нова: неэлектронные корпуса текстов применялись переводчиками в работе и раньше, но очень редко, так как их громоздкость и малая репрезентативность вызывали критику и не позволяли оперативно использовать их в реальной переводческой деятельности [Владимов Н. В., 2005].

Также вторым назначением создаваемой платформы является использование корпуса для обучения переводу. В корпусной методике обучения языку и переводу задействован грамматико-переводной метод. Обучение языку осуществляется в параллельных электронных корпусах текстов с использованием программ-конкордансов параллельных текстов.

Известно, что на практике перевод ориентируется на возможность постредактирования, также исследователь сравнивает и оценивает различные стратегии и интерпретации текста. Много времени тратится на обращение к словарям. Такой расход времени при использовании лингвистических компьютерных технологий и электронных параллельных корпусов ощутимо сокращается, также при изучении приемов и способов перевода корпус предоставляет образцы профессионального перевода.

Параллельный корпус художественных переводов с узбекского языка на русский позволит решать различные лингвистические, переводческие и образовательные задачи. Например,

1. Исследование принципов перевода на русский язык на основе анализа перевода частей речи, грамматических категорий (залог, переходность глагола), основе анализа перевода предложений;

2. Определение основной переводной единицы на основе корпуса;

3. Практическое исследование перевода на основе корпуса (перевод художественного дискурса, правила трансформации простых и сложных предложений).

На платформе корпуса возможно осуществление оценки переводных текстов, исследование природы и универсальности переводного языка, изучение идиостиля писателя. Вследствие чего в рамках проекта будет возможной разработка толково-переводной словарь художественного языка писателя.

На самом деле создание корпусов имеет большое значение, прежде всего, для лексикографии. Известно, что время находит свое отражение в словаре языка, которое постоянно «растет». Зеркалом, которое отчетливо показывает богатство языка, является его словарь. Объем словарного содержания считается фактором, повышающим его ценность.

В преподавании языка такие корпуса также являются богатым источником материалов. Кроме того, параллельные корпуса служат обучающими данными для статистических систем машинного перевода.

В параллельном корпусе собраны тексты оригинала и его переводов. В переводоведении основное внимание уделяется выявлению особенностей, отличающих переводы от оригинальных текстов. Эти изменения могут быть индивидуальными для конкретной переводческой задачи или языковой пары, но также в соответствии с широкими лингвистическими особенностями перевода они могут указывать на общие черты, которые отличают текст перевода. Очевидно, что эти исследования являют собой наглядный эмпирический метод определения особенностей корпусных переводов, и с 1990-х годов применены многими учеными, к примеру, Beyker (1993; 1996), Johansson и Ebeling (1996), Hansen (2003); Teich (2003); Mauranen и Kujamäki (2004) и Hansen-Schirra, Neumann и Steiner (2012). Кроме того, параллельные корпуса используются в качестве справочного материала при обучении переводу и в сфере профессионального перевода, поскольку они обеспечивают быстрый и интерактивный доступ к переводческим решениям (например, к памяти переводов).

С точки зрения переводоведения важной задачей параллельного корпуса является обнаружение переводческих изменений. Как в переводоведении, так и в контрастивной лингвистике в последнее время многоязычные корпуса стали использоваться для изучения переводческих явлений, т.е. переводческих изменений или особенностей перевода, а также контрастивных различий между ними.

Мировая лингвистическая ситуация явилась побудительной причиной в формировании приоритетной задачи узбекского языкознания – это создание параллельных корпусов. Процесс автоматической нормализации текстов связан с морфологическим анализом, при котором используются методы токенизации, лемматизации и стемминга.

Токенизация – разделение речевых единиц естественного языка в соответствии с грамматическим значением, определение формы слов. Лемматизация – приведение словоформ к словарной (начальной) форме, т.е. в соответствии с их представлением в словаре, определяется основа и окончание, таким образом приводится к форме морфем. Стемминг – определение корня слова.

Токенизация	Лемматизация	Стемминг
лошадку	лошадк - а	лошадь
женского	женск - ий	жен

При использовании параллельных текстов сначала необходимо идентифицировать текстовые сегменты (фразы или предложения), а также адаптировать и скоординировать тип и жанр параллельных текстов. Алгоритм данной процедуры таков:

1 шаг. В целях лингвистического обеспечения параллельных корпусов сбор и сортировка коллекции англо-узбекских, узбеко-английских параллельных текстов.

2 шаг. Анализ репрезентативности текстов.

3 шаг. Сегментация слов, сочетаний и предложений в соответствии с уровнем сопоставления.

4 шаг. Создание переводных словарей, соответствующих текстам.

5 шаг. По завершению процесса сегментации сохранить все параллельные тексты в памяти перевода.

При создании автоматического анализатора текста сначала создается электронный морфологический словарь. По итогам морфологического анализа выдаются статистические данные о частеречной принадлежности слов. Так создается большая логико-семантическая база лексем. Морфологический анализ также важен для синтаксического анализа. Параллельный корпус служит связующим звеном для выявления глубинной семантики лексических единиц языков [Абдурахманова Н., Тулиев У., 2018].

Сравнительный корпус содержит важные для параллельного корпуса ключевые слова. Известно, что ключевые слова используются часто и передают основную смысловую нагрузку. Для выделенной части текста в отдельную группу собираются единицы с самостоятельным значением (т.е. знаменательные части речи отделяются от служебных), затем создается таблица единиц, классифицированных как ключевые слова. Выделение ключевых слов в тексте параллельные корпуса намного ускоряет работу.

Заключение. Параллельный корпус считается важной реальностью для широко распространённого межкультурного общения современной эпохи. Посредством параллельных корпусов возможно выявление универсалий для сопоставляемых языковых сред и культур, а также специфических ментальных характеристик языков, реалий и лакунарных единиц. Корпус параллельных текстов служит также для развития автоматического перевода, обеспечивает формирование компьютерной лексикографии. С

помощью корпуса параллельных текстов разрабатываются программы согласования и возможно создание различных специализированных словарей.

В настоящее время корпус представляет собой действенный инструмент, экономящий время и усилия. Проводятся теоретические и практические исследования по формированию … «корпусного обучения языку, parsing на основе теории зависимостей, технологии FST в морфологическом анализе корпуса, авторского корпуса, программного обеспечения и лингвистического обучения в создании национального корпуса, морфологического и семантического анализаторов корпуса, по созданию параллельных корпусов машинного перевода на основе нейротехнологии, учебного корпуса узбекского языка» [Абдурахманова Н., Тулиев У., 2018]. Применение параллельного корпуса эффективно не только в языкознании, но также и в переводоведении, двуязычной лексикографии и в областях, где необходимо сравнивать языки. Одним из преимуществ электронного корпуса является возможность посредством их наблюдать изменения слов, появление историзмов, неологизмов, расширение и сужение значения, появление в языке новых фразеологизмов. В языковых корпусах, представляющих собой электронную базу, постоянно вводится и размещается новая информация, в них находит свое отражение процесс периодического обогащения языка. Это радикально сокращает объем человеческого труда, так как требуется много усилий и средств для включения в словари печатной формы все те изменения, происходящие со временем в языке. А корпус дает возможность решить эту проблему в короткие сроки и без материальных затрат.

Ожидаемым результатом применения параллельных корпусов является полная автоматизация текста, который может быть использован не только для ручного (человеческого) перевода, но и для машинного. В настоящее время мы находимся в процессе создания технологии машинного перевода, использующей статистический способ. Создание памяти переводов, являющейся лингвистической основой для статистических технологий перевода (таких как Google), способствует развитию нейронной технологии машинного перевода. Еще одним важным аспектом параллельных корпусов является то, что они препятствуют появлению переводческих ошибок. До сих пор наблюдается в Интернете проблема неверного перевода, когда появляются слова, не имеющие никакого отношения к переводимому текста. Основная цель создания параллельного корпуса – разработка эффективных и адекватных методов перевода. Посредством параллельного корпуса достигается улучшение качества перевода.

Параллельный корпус представляет собой набор электронных текстов на языке оригинала (узбекский язык) и его перевод (русский язык). Параллельный корпус сохраняет оригинальную последовательность текста и служит:

- во-первых, для сопоставления различных языковых структур, фраз и слов в определенном контексте, т.е. для сравнительного анализа структур двух языков;
- во-вторых, в области переводов: для поиска эквивалентов оригинала в других языках;
- в-третьих, при обучении движков (систем) машинного перевода при изучении языка и при составлении словарей.

В статье рассмотрен процесс разработки параллельного корпуса узбекского и русского языков и возможных на его основе направлений исследования. Этот параллельный и компаративный корпус узбекского и русского языков предназначен для анализа переводов. Основными задачами проекта являются разработка дизайна корпуса,

формирование коллекции текстов, проведение предварительной обработки, создание поисковой системы и метаданных. В результате исследования создается параллельный корпусный веб-сервис. Разрабатывается программа-конкордансер (программное обеспечение) параллельных текстов. В рамках апробации и функционирования параллельного корпуса возможно издание толково-переводной словарь художественного языка писателя, а также новых переводов на русский язык художественных произведений.

Продуктом, результатом разработки является веб-сервис «Параллельный корпус» - идентифицируемая уникальным URL-адресом программная система со стандартизированными интерфейсами. Отметим, что данный веб-сервис является продуктом совместной научно-исследовательской деятельности лингвистов и программистов.

Особо актуальным будет использование веб-сервиса в качестве корпусного метода в научных исследованиях по художественному переводу. Также будет значимым применение продукта в качестве нового, эффективного инструмента обучения переводу студентов и молодых специалистов. Третьей важной областью приложения веб-сервиса будем сам образовательный процесс в вузах, посредством которого можно будет обучать иностранному языку. Таким образом, обозначаются два основных направления в наличии спроса на результаты разработки: во-первых, на основе параллельного корпуса, применив корпусный метод, станет возможным проведение научных исследований в сферах художественного перевода, сопоставительного языкознания, корпусной лингвистики, лексикографии. Во-вторых, параллельный корпус станет образовательным инструментом или новой методикой в процессе обучения иностранному языку и в преподавании теории и практики перевода.

На основе технологии создания веб-сервиса возможна разработка других оригинальных программных систем, имеющих приложение не только в лингвистике и переводоведении, но и в других направлениях общественно-гуманитарной сферы. Станет возможным открытие современных информационных центров узбекского языка в филиалах зарубежных вузов. Также отметим, что параллельный корпус художественных текстов войдет составной частью Национального корпуса узбекского языка.

Список использованной литературы:

1. Абдурахманова Н. Komputer lingvistikasi [учебник / Globe edit, 2020, 395 b.
2. Абдурахманова Н. Основы автоматического морфологического анализа для машинного перевода. - Известия Киргизского государственного технического университета. 2016; 2 [38]:12-7.
3. Abdurakhmonova, N., & Tuliyev, U. (2018). Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. *Literature, Education*, *3*, 68.
4. Abdurakhmonova, N., & Urdishev, K. (2019). Corpus based teaching Uzbek as a foreign language. *Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL)*, *6*(1-2019), 131-7.
5. Владимов Н. В. Корпусный подход к решению переводческих проблем: на материале письменных переводов с русского языка на английский: диссертация ... к.ф.н.: - Москва, 2005. - 198 с.

6. Груздев Д. Ю. Электронный корпус текстов как эффективный инструмент переводчика: диссертация ... к.ф.н.: Москва. 2013.
7. Захаров В., Yuan Tao. Создание и использование параллельного корпуса русского и китайского языков. - 2015.
8. Кокорева А. А. Корпус параллельных текстов в обучении иностранному языку // Вестник ТГУ. 2013. №2 (118).
9. Маник С. А. Параллельный корпус в практике перевода общественно-политических текстов (с английского на русский) // СИСП. 2019. №4.
10. Многоязычный корпус: лингвистические требования и технические перспективы. Семинар. Великобритания. Ланкастер - 2003.
11. Соснина Е.П. Параллельные корпусы в обучении языку и переводу. http://spr.fld.mrsu.ru/2015/09/e-p-sosnina-parallelnye-korpusy-v-obuchenii-yazyku-i-perevodu/
12. Указ Президента Республики Узбекистан «О мерах по кардинальному повышению роли и авторитета узбекского языка в качестве государственного языка», mУП - №5850. 2019.
13. https://uzbekcorpus.uz

MATNNING DISKURSIV TAHLILI

Mirvaliyeva Dildora Anvarovna
Mirzo Ulug‘bek nomidagi O‘zbekiston
Milliy universiteti stajyor o‘qituvchisi.
E-mail: dildoramirvalieva07@gmail.com

Annotatsiya: Mazkur maqola diskurs tushunchasi, uning kelib chiqishi va o‘rganishga bo‘lgan turlicha yondashuvlar hamda matnni diskursiv tahlil qilishning ahamiyati bayonotini o‘z ichiga oladi.

Kalit so’zlar. Diskurs, intuitiv, tahlil, yondashuv, matn, muloqot, fraza, adresat.

Abstract: This article includes a statement of the concept of discourse, its origins and different approaches to its study, and the importance of discursive text analysis.

Keywords: Discourse, intuition, analysis, approach, text, communication, phrase, addressee.

Аннотация: В данной статье изложено понятие дискурса, его истоки и различные подходы к его изучению, а также важность дискурсивного анализа текста.

Ключевые слова: Дискурс, интуиция, анализ, подход, текст, общение, фраза, адресат.

Zamonaviy tilshunoslikda hozirgi kunga qadar diskurs masalasining turli ko‘rinishlari (siyosiy, ijtimoiy, pedagogik, diniy, didaktik, xususiy, ommaviy) o‘rganilmoqda va ko‘pchilik tilshunoslar diskursni nutqiy faoliyat jarayoni hamda so‘zlashish uslubi kabi mazmunlarda talqin etishadi.

Diskurs- hissiy, bevosita, intuitiv, ya’ni muhokama orqali vosita bilan hosil qilinadigan mantiqiy dalildir. V.G. Borbotko: “Diskurs ham matndir, lekin u tilning kommunikativ birliklaridan - jumlalardan va ularning uzluksiz ichki semantik aloqada bo‘lgan kattaroq birliklarga birlashmasidan iborat bo‘lib, uni integral shakllanish sifatida qabul qilish imkonini beradi” degan ta’rifni beradi. “Haqiqiy vaziyatlarda so‘zlovchilarning tildan haqiqiy foydalanishini o‘rganish” , deb yozgan T. A. van Deyk “Diskurs tahlili qo‘llanmasi” da. Matnning tuzilish xususiyatlarini qamrab olishi va konteks tahlili jarayoni yanada ochib berishini alohida ta’kidlab o‘tadi.

*Diskurs [*Temirova F., 2022: 272*] – matndan ko’ra kengroq tushuncha bo‘lib, bir vaqtning o‘zida ham tilga xos jarayondan, hamda shu jarayonning natijasi bo‘lgan matndan iborat bo‘ladi. Diskurs so‘zi zamirida nutqning barcha kognitiv-kommunikativ funksiyalarini tushunish yotadi. Izlanishlardan ko‘rinib turibdiki, diskurs ham jarayon, ham matndir. Matnning turg‘un, tayyor mahsulot, diskursning esa kechayotgan nutqiy muloqot jarayoni sifatida talqin qilinishi ularning keskin farqlanishiga sabab bo‘ladi. Lingvistik adabiyotlarda “diskurs” atamasining qat’iy bir ma’nosi yo‘q, u ifodalaydigan hodisalar diapazoni juda keng, ya’ni “matnning qismi” dan yaxlit “nutq”gacha bo‘lgan hodisalarni ifodalash uchun ishlatiladi*[Abdurakhmonova, N. 2019,2021,2022].

Tilni pragmatik tadqiq etishga bag‘ishlangan ishlarda [Миловидов В., 1998: 39] diskurs atamasi sakkiz xil: so‘z muqobili, frazalardan o‘lchami bo‘yicha ortadigan birlik, fikrning

adresatga ta’siri, suhbat, so‘zlovchi pozitsiyasi, lisoniy birliklardan foydalanish, fikrning cheklangan turi va matn hosil bo‘lish shartlarini tadqiq etishga mo‘ljallangan nazariy qurilmalar ma’nolarida qo‘llangan.

Diskurs tahlili - bu kontekstda tildan ma’no olish uchun ishlatiladigan sifatli tahlil usuli hisoblanadi. Diskursiv tahlil odatda o‘zaro bog‘liq bo‘lgan ikkita usulda aniqlanadi [Austin A., 2023: 134]. Birinchidan, u real muloqotning lingvistik hodisalarini tadqiq qiladi. Ikkinchidan, u tilning shaklini emas, balki uning asosiy funksiyalarini hisobga oladi. Bu ikki jihat ikki xil kitobda: Maykl Stubbs o‘zining “Diskurs tahlili” asarida tahlilni lingvistik pragmatikaga havola qilsa, Jon Braun “chiziqlar orasidagi” tilni o‘rganishga harakat qiladi.

Shuning bilan birga, o‘zbek tilshunosligida ham diskurs borasida lingvistik tadqiqotlar tilshunos Sh.Safarov tomonidan amalga oshirilgan. U matn va diskurs muammolari to‘g‘risida quyidagilarni bayon etadi: “Agarda matn va diskursning har ikkalasi ham inson lisoniy faoliyatining natijasi bo‘lsa, ularni faqatgina zohiriy – formal ko‘rsatkichga asosan “og‘zaki” va “yozma” sifatlari bilan farqlash imkoniga gumonim bor”.

Diskursiv [Taxirov Z., 2021:214] (mantiqda mulohazaga asoslangan, mantiqiy qismlar ketma-ketl qatori) tahlil ma’noni aniqlash uchun korpus yoki ma’lumotlar to‘plamida taqdim etilgan tildan foydalanadi. Ushbu ma’lumotlar to‘plami intervyular yoki guruh muhokamasi stenogrammalarini o‘z ichiga olishib, nutqni tahlil qilishning ba’zi shakllari tilning o‘ziga xos xususiyatlariga (masalan, tovushlar yoki grammatikaga), boshqa shakllar bu tilning maqsadlariga erishish uchun qanday ishlatilishiga e’tibor beradi. Wodak va Krzyzanowski ta’kidlaganidek: “Diskuriv tahlili muammoga yo‘naltirilgan ijtimoiy tadqiqotlar uchun umumiy asos yaratadi”.

Diskurs tahlilining maqsadi tilning funksiyalarini va turli kontekstlarda ma’no qanday tuzilganligini, nutqning ijtimoiy, madaniy, siyosiy va tarixiy sharoitlarini o‘z ichiga olganligini o‘rganishdir [Kamolova A., 2024:191].

Matnni diskurs tahlil qilishda quyidagi bosqichlar muhim ahamiyatga ega:

-bir matnning diskurs tahlili uchun birinchi bosqich mazmun va kontekstni tushunish bo‘lib, bunda matnning qanday maqsadda yozilganligini, qaysi muhitda yoki texnologiyada yaratilganligini aniqlash, mazmunning asosiy tushunchalar va matndagi asosiy mavzuni anglash diskurs tahlili uchun boshlang‘ich nuqta vazifasini bajaradi.

-diskursiv tahlil jarayonida matndagi muhim so‘zlar, atamalarning tavsifi va ularga oid ta’riflar ko‘zdan kechirilib, ba’zi so‘zlar yoki atamalar belgilangan ma’noni o‘zgartirish uchun ishlatilishi mumkinligini o‘rganish tahlilning muhim qismi sanaladi.

-matnning muhim qismlarini aniqlash, ularning ta’kidlanishi yoki tavsifi diskurs tahlida muallifning maqsadi, fikrlaridagi tashqi va ichki bog‘liqlar tahlil qilinishi hamda bu tahlil matndagi fikr, xulosa, mazmunning ustuvliklarini aniqlashni o‘z ichiga oladi.

-matndagi fikr va nazariyalarni o‘rganish va baholash muallifning matnda qanday muammo yoki tanqidiy qarashlarga e’tibor qilganligini, fikrlarini isbotlash usullarini hamda matndagi muhim mantiqiy o‘zgarishlarni tushunishni jamlaydi.

-matndagi natijalarni ko‘rsatish va baholashda diskurs tahlilining natijalari, muallifning maqsadi, mazmun-mohiyati hamda matnni o‘rganuvchilar uchun qanday ma’noni qamrab oladiganligi baholanadi.

Diskurs tahlili ko‘p qadamli jarayon bo‘lib, matnni tushunish, o‘rganish va tahlil qilishning ko‘p tomonlama o‘zgarishlarini o‘z ichiga oladi[Abdurakhmonova, N. 2019,2021,2022].

Faoliyat doirasini ko‘rib chiqadigan bo‘lsak, qonunchilik, ish yuritishga (ishga aloqador yozishmalar, huquqiy munosabatlar sohasi) oid hujjatlarning diskurs tadqiqoti ma’muriy faoliyat sohasida tahlil va tushunchalar bilan bog‘liq bo‘lib, ish yurituvchi organlar tomonidan ishlatiladigan hujjatlarni o‘rganish, kommunikatsiya jarayonlarini tahlil qilish va ish yuritish protsessidagi muammolarni aniqlash bilan bog‘liq[Abdurakhmonova, N. 2019,2021,2022].

Shuningdek, yozma nutqning rasmiy uslubida, asosan, quyidagi munosabatlar doirasidagi hujjatlar tuziladi:

1. Huquqiy munosabatlarga oid: qonun, fuqarolik va jinoyat aktlari, nizom, shartnoma va boshqalar.

2. Idoraviy-ma’muriy shaklga oid: dalolatnoma, buyruq va farmoyishlar, turli ish qog‘ozlari (ariza, tavsiyanoma, tilxat, ma’lumotnoma kabi).

3. Diplomatik munosabatlarga doir: bayonot, nota, bitim, memorandum va boshqalar.

Ish yuritishga oid hujjatlarning diskurs tadqiqida mamuriy hujjatlarning matnlari tahlil qilinadi, ularning ma’noni o‘rganish, muammo yuzaga kelishi, so‘zlar va ifodalar orqali munozaralarda o‘tishi kabi asosiy choralarga e’tibor qaratiladi. Tadqiqot maqsadi hujjatlarning tuzilishini, ularning kommunikativ vazifalarini va muhitlarini, hujjatlarning o‘zgarishlarga olib kelishi, protokollar, qonunchilik va boshqaruvni ta’minlash jarayonlarini tushunishdir.

Tahlil natijalari ish yuritish tizimi va uning faoliyati bilan bog‘liq qarorlarni olish, boshqaruv organlari va ishlab chiqarish tashkilotlari uchun qarorlarni o‘rganishda yordam berishi mumkin. Bundan tashqari, bu tadqiqot o‘rganilgan ma’lumotlar asosida boshqaruv tizimini yanada mustahkamlash va uning faoliyatini mukammallashtirish uchun strategiyalarni ishlab chiqishga ham yordam berishi mumkin.

Foydalanilgan adabiyotlar ro‘yxati:

1. A.Austin. Discursive analysis: concept and role in modern linguistics.2024
2. В. Миловидов. Текст, контекст, интертекст: Введение в проблему сравнительного литературоведения. – Тверь: ТвГУ, 1998, с. 39.
3. Z. Taxirov. O‘zbek tilining amaliy stilistikasi [Matn]: darslik. - Toshkent: Tafakkur avlodi, 2021.
4. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). Uzbek electronic corpus as a tool for linguistic analysis. In *Компьютерная обработка тюркских языков. TURKLANG 2022* (pp. 231-240).
5. Abduraxmonova, N. Z. Q., & Urazaliyeva, M. Y. (2022). O ‘zbek tili elektron korpusida (http://uzbekcorpus. uz/) og ‘zaki matnlar korpusini yaratishning nazariy va amaliy masalalari. *Academic research in educational sciences*, *3*(3), 644-650.
6. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, *11*(19), 9117.
7. F. Temirova. Badiiy diskurs va diskursiv tahlil. FarDU ilmiy xabarlari**,** 2022**.**
8. M.Saidova, M.Axrorova. Zamonaviy tilshunoslikda diskurs tushunchasi va uning tahlili. Miasto Przyszłości Kielce, 2023.
9. Sh.Bobojonova. Diskursiv tahlil va uning akademik matnlarda ifodalanishi. O‘zbekiston Milliy universiteti xabarlari, 2022. [1/12].
10. O.Ablakulova. Diskurs tushunchasi va uning tadqiqi. Academic Research in Educational Sciences, 2021.

11. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In *of the International Conference on Language Technologies for All (LT4All)*.
12. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).
13. https://muhaz.org/6-mavzu-ish-yuritish-tili-va-uslubi-hujjat-turlari-va-xususiya.html

NUROTA TUMAN LINGVISTIK TAMINOTI (TUZILISHI VA TAVSIFI)

Muftillayeva Ruxsora Toshmuhammad qizi
Alisher Navoiy nomidagi Toshkent davlat
o‘zbek tili va adabiyoti universiteti magistranti.
E-mail: ruxsoramuftillayeva851@gmail.com

Annotatsiya: Dialektal korpusning lingvistik ta’minoti muayyan dialekt yoki tilning mintaqaviy xilma-xilligi, lingvistik xususiyatlarini ifodalovchi matnlar yoki yozuvlar to‘plamini ishlab chiqish va tavsiflashni anglatadi. Ushbu maqola dialektal korpusdagi lingvistik xususiyatlarni tizimli ravishda belgilash va tahlil qilishni o‘z ichiga oladi. Lingvistik annotatsiyalar dialektning fonetik, fonologik, morfologik, sintaktik va semantik xususiyatlari haqidagi ma’lumotlarni, shuningdek, so‘zlashuvchi demografiyasi va tildan foydalanish shakllari kabi sotsiolingvistik omillar haqidagi ma'lumotlarni beradi. Lingvistik ta’minotga ega dialektal korpusning tuzilishi va tavsifi muayyan dialektning o‘ziga xos lingvistik xususiyatlari haqida qimmatli fikrlarni berishi mumkin. Bundan tashqari, dialektal korpusning lingvistik ta’minoti tilni saqlash, tilni jonlantirish tashabbuslari va til o‘rgatish, o‘rganish uchun resurslarni ishlab chiqishga yordam beradi.

Kalit so‘zlar: Leksik, semantik, grammatik, fonetik, shevashunoslik, lug‘at

Annotation: Linguistic maintenance of a dialectal corpus refers to the development and description of a set of texts or records that represent the regional variety, linguistic features of a particular dialect or language. This article includes a systematic definition and analysis of linguistic features in the dialectal corpus. Linguistic annotations can include information about phonetic, phonological, morphological, syntactic, and semantic features of a dialect, as well as information about sociolinguistic factors such as speaker demographics and patterns of language use. The construction and description of a dialectal corpus with linguistic support can provide valuable insights into the unique linguistic features of a particular dialect. In addition, the linguistic support of a dialectal corpus can aid in the development of language preservation, language revitalization initiatives, and resources for language teaching and learning.

Keywords: Lexical, semantic, grammatical, phonetic, linguistics, dictionary

Аннотация: Лингвистическое поддержание диалектного корпуса подразумевает разработку и описание набора текстов или записей, которые представляют региональное разнообразие, лингвистические особенности конкретного диалекта или языка. Данная статья включает систематическое определение и анализ языковых особенностей диалектного корпуса. Лингвистические аннотации могут включать информацию о фонетических, фонологических, морфологических, синтаксических и семантических особенностях диалекта, а также информацию о социолингвистических факторах, таких как демографические характеристики говорящего и модели использования языка. Построение и описание диалектного корпуса с лингвистической поддержкой может дать ценную информацию об уникальных лингвистических особенностях конкретного диалекта. Кроме того, лингвистическая поддержка диалектного корпуса может помочь в развитии

инициатив по сохранению языка, его возрождению и ресурсов для преподавания и изучения языка.

Ключевые слова: Лексический, семантический, грамматический, фонетический, лингвистический, словарь.

Milliy til muayyan bir millat vakillari uchun tarixan umumiy bo‘lgan, milliy madaniyatning ko‘zgusi sifatida ularni ruhan birlashtiradigan vosita bo‘lib, u adabiy til bilan bir qatorda, xalq shevalarini… o‘z ichiga oladi [Mahmudov N, Sobirov A, Odilov Y, Ziyodullayev G: 2017; 7]. Vaqt o‘tishi bilan shevalar iste’moldan chiqib ketishi, yoki adabiy til o‘rnini egallashi mumkin. Shevalar tarixan qanchalik barqaror bo‘lar ekan, adabiy til uchun shevashunos tadqiqotchilar tomonidan keltirilgan lingvistik nuqtai nazarlarda bir xulosaga kelish qiyin kechadi. Shevalar adabiy tilning asosiy boyish manbai hisoblanadi. Dialektologiya (yunon. Dialektos – sheva so‘zidan) tilshunoslikning bir sohasi bo‘lib, u biror tilning mavjud dialektlarini, ya’ni mahalliy lahja va shevalarni o‘rganadi. Shevashunoslik – sheva (fran. So‘zlashish, odat, til), shunos (fran. bilish, o‘rganish) degan ma’nolarini bildiradi. Shevashunoslik mahalliy shevalar va ularning lahjalarini o‘rganish haqidagi fan hisoblanib, o‘zbek shevashunosligi O‘zbekistonda mavjud dialekt, lahja va shevalarni o‘rganadi [Xolova M, 2022; 12]. Jahon tilshunosligida ham dialektologiya sohasida ancha ishlar amalga oshirilgan. Turkologiyada Mahmud Qoshg‘ariy o‘zining „Devoni lug‘otit turk’’ asaridan shevalarni o‘rganishni boshlagan. O‘zbek tilshunosligida shevalarning fonetik, leksik, grammatik, xususiyatlariga oid T. Yo‘ldoshev [Yo‘ldoshev T, 1986; 34], N. Rajabov [Rajabov N, 1990; 30], S.Ibragimov [Ibragimov S, 1959; 56] kabi olimlar izlanishlar olib borgan [Xolova M, 2022; 12]. O‘zbek tili fonetik va leksik jihatdan turli xil. Bu o‘zbek tili shevalarining ham boyligi hisoblanadi. Yudaxin o‘zbek shevalarini tojik shevalar bilan munosabatiga ko‘ra tasniflaydi va ularni 5 guruhga ajratadi. (1-jadval)

1	Toshkent
2	Farg‘ona
3	Qipchoq
4	Xiva
5	Shimoliy o‘zbek

Qipchoq lahjasiga taalluqli shevalar juda katta territoriyaga tarqalgan. O‘zbekistonning barcha viloyatlarida mazkur lahja vakillarining borligi ham bu fikrning to‘g‘ri ekanligiga dalildir. Haqiqatan ham qipchoq lahjasining vakillari Toshkent viloyatining Ohangaron vodiysidagina emas, hatto Fargona, Andijon viloyatlarida ham yashaydilar. Sirdaryo, Samarqand, Buxoro, Qashqadaryo, Surxondaryo va Xorazm viloyatlari aholisining ko‘pchilik qismi ham qipchoq lahjasining vakillaridir. Navoiy viloyati shevalari ham qipchoq lahjasiga kiradi. Bugina emas o‘zbek tili qipchoq lahjasining shevalari Qozog'iston, Qirg‘iziston, Tojikiston respublikalari territoriyasida ham bor [Ashirboyev S, 2013; 31]. Hozirgi vaqtda o‘zbek shevalarida uchraydigan ayrim so‘zlar faqat keksa avlod tilida uchrayotgani, bu so‘zlarni yosh avlod vakillari tushunmasligi bu so‘zlarning iste’moldan chiqib ketishiga olib kelmoqda, buning sababi esa turmush tarzining zamonaviylashuvi bilan izohlanmoqda. Shuning uchun ham Nurota tumani shevasi tadqiqida tashqi axborot vositalaridan qisman uzilgan mobil aloqa vositalaridan foydalanish ko‘nikmasiga

ega bo‘lmagan, savodxonlik darajasi o‘rtacha bo‘lgan keksa avlod vakili va shahar zonalaridan hududiy jihatdan uzoq bo‘lgan qishloq va ovullarda yashovchilarning so‘zlashuviga urg‘u berildi. Hozirga qadar tadqiqotlarni kuzatgan holda aytish mumkinki, aksariyat tadqiqotlar shevalarning mavzuviy bo‘linuvchanligi boshqa turkiy til oilalariga qiyoshlash orqali [Raximov S, 1985; 18] olib borilgan.

B.I.Sinsius [Цинциус В, 1979; 4] shevaga oid lug‘at ma’lumotlarini qayta ishlashda mavzuviy guruhlashning bir necha bosqichini taklif qiladi [Xolova M, 2022; 26]. Bu boradagi K.M.Musayev [Musayev K, 1984; 25] hamda M.Xolova [Xolova M, 2022; 26] fikrlariga asoslanib Nurota shevasiga oid so‘zlarni quyidagicha tasniflaymiz. (2-jadval)

Fonetik dialektizmlar
Leksik dialektizmlar
Semantik dialektizmlar
Grammatik dialektizmlar

Fonetik dialektizmlar. Dialektizmning bu turiga adabiy til me’yorlaridan fonetik tarkibiga ko‘ra farqlanuvchi, ammo semantik qurilishiga ko‘ra bir ma’noda qo‘llanuvchi so‘zlar kiradi [Xolova M, 2022; 27]. M. tuvaq (tuvoq) ~ tovoq (ad.t): Muslima buvi *tuvoq*ni qozon chetiga joylashtirib qo‘ydi-da… [Sh.Xolmirzayev. „ O‘n sakkizga kirmagan kim bor’’, 27-b]

Nurota tumani aholisi „j’’ tovushini qo‘shib gapiradi, ammo bugungi kunda yosh avlod vakillari bunday gapirmaslikka harakat qilishadi. Bu holat ko‘proq tog‘li hududlarga yaqin bo‘lgan qishloqlarda (Chuya, Oqqula, Jarma, Yangibino, Sovuq buloq) qishloqlarida saqlanib qolgan.

Leksik dialektizmlar. Bu tur ma’lum bir shevaga oid bo‘lib, adabiy tilda kuzatilmaydigan, ammo o‘zbek tilining izohli lug‘atiga kiritilgan so‘zlar kiradi. M. og‘ariq – paykal egatlariga yoki umuman paykalga shoxariqdan suv olib taratish uchun xizmat qiladigan muvaqqat ariq. M.: Nasim bug‘doy poyada uqariq qazishdan erta qaytdi [T.Murod. „Qo‘shiq’’. 189- b.]

Semantik dialektizmlar. Dialektizmlarning bu turiga ma’lum bir shevaga xos so‘zlarning umumiy fonetik tarkibiga ega, semantik nomutanosib so‘zlar kiritiladi. M.: tön (to‘n) so‘zi Nurota shevasida to‘yda kuyov kiyadigan qalin paxta solib toshlar bilan bezatilgan kiyimni bildirsa, Boysun shevasida qishda kiyiladigan qalin paxta solib qavilgan qishki kiyimni bildiradi, Xakas tili Qizil shevasida tön so‘zining qishki kiyimdan tashqari, ko‘ylak, kiyim ma’nolarini ham anglatadi. Sätil (satil) (Nurota sh.) – asosan suyuqliklarni tashish uchun mo‘ljallangan sirli, plastmassa yoki temirdan ishlangan idish, Sätil (satil) (Namangan sh.) – sirlangan, ikki tomonidan ushlaydigan qulog‘i bor metall idish.

Grammatik dialektizmlar. Dialektizmlarning ushbu turiga shevaning umumiy semantik ma’noligi saqlanib, grammatik qurilishi farqlanuvchi so‘zlar kiritiladi. R.P.Abdina [Abdina R, 2018; 104] turkiy tilga oid shevalarda morfologik dialektizmlarning juda kam kuzatilishini qayd etadi. M.: ad.t. мойдырых / Xakas qizil sheva. Mānxap; ad.t. чыырынды ~ čiirïndï, чыырбас~čiirïnbas / Boysun shevasi.

Yuqoridagi fikrlardan kelib chiqib, o‘zbek tili „j’’ lovchi shevalarda ham morfologik qurilishi farqlanuvchi shevalarning deyarli kuzatilmasligi, adabiy til va shevalar orasida morfologik qo‘shimchalar bo‘yicha qisman farqlanishi, asarlarda o‘z aksini topgan grammatik

dialektizmlar Respublika hududidan qat'iy nazar barcha uchun tushunarli bo'lishi hamda qo'shimchalarning fonetik jihati orqali farqlanish holati kuzatilishini aytib o'tish joizdir [Xolova M, 2022; 27].

Xulosa qilib aytadigan bo'lsak, har qanday korpus resursining uzoq muddat saqlanishi va foydalanish uchun lingvistik ta'minot yordami juda zarur. Bu ayniqsa, dialektal korpus uchun muhim. Korpusdan foydalanuvchilar tomonidan tushunarsiz bo'lishi mumkin bo'lgan ko'plab omillar, jumladan, lug'aviy noaniqlik yoki morfologik noaniqlik, notanish sintaksis, nutq tuzilishi, standartdan sezilarli darajada farq qiluvchi fonologik shakllar tufayli yuzaga kelishi mumkin. Tilning ma'lum bir lahjasini yoki mintaqaviy xilma-xilligini ifodalovchi matnlar yoki yozuvlar to'plamini tuzish va tavsiflash orqali dialekt korpusining lingvistik bazasi tuziladi. Dialekt korpusidagi til xususiyatlarini tizimli ravishda izohlash va tahlil qilish dialektning o'ziga xos xususiyatlari haqida qimmatli tushunchalar beradi. Dialektal korpusni lingvistik ta'minot bilan hujjatlashtirish va tahlil qilish orqali biz til xilma-xilligining boyligi va murakkabligini yaxshiroq tushunishimiz mumkin, shu bilan birga turli til an'analarini doimiy ravishda o'rganish va tahlil qilishimiz mumkin.

Foydalanilgan adabiyotlar:

1. Ashirboyev S. O'zbek Dialektologiyasi. – Toshkent: 2013. -B. 31.
2. URL: https://uz.wikipedia.org/wiki/Adabiy_til
3. https://ru.wikipedia.org/wiki/диалектология
4. Абдина Р.П., Белоглазов П.Е., Каскаракова З.Е., Кызласов А.С. Принципы лексикотематической дифференциации диалектной лексики хакасского языка\ Мир науки, культуры, образования, 2018. №1 (68). – С. 104-120
5. 10-sinf ona tili darsligi. Mahmudov N.M.,.Sobirov A.Sh, Odilov Y.R., Ziyodullayeva G.Sh. – Toshkent: Davlat milliy ensiklopediyasi, 2017. –B. 7.
6. Ражабов Н. Ўзбек шеваларида феълнинг морфологик тузилиши. – Тошкент: Фан, 1990. – Б. 30.
7. Мусаев К.М. Лексикология тюркских языков. – Москва. Наука, 1984. – С. 25.
8. Раҳимов С. Сурхондарѐ вилояти ўзбек шевалари. – Тошкент: Фан, 1985. – Б. 18.
9. Ибрагимов С. Фарғона шевалари касб-ҳунар лексикаси. – Тошкент: Фан, 1959. – Б. 56.
10. Цинциус В.И. Проблемы сравнительно-исторического изучения лексики алтайских языков. – Ленинград: Наука. 1979. – С. 4
11. Холова М. ўзбек миллий шевалари корпусини тузишнинг лингвистик асослари (Бойсун тумани "ж"ловчишевалари мисолида):Фил. фан. бўйича фалсафа доктори (PhD)...дисс. Автореф. – Термиз, 2022.
12. Йўлдошев Т. Тожикистондаги ўзбек шевалари морфологияси (феъл). – Тошкент: Фан, 1986. – Б. 34

CORPUS ANALYSIS OF SPEECH UNITS IN SOCIAL NETWORKS

Urinboeva Nazokat
Master's degree student
of the National University of Uzbekistan.
E-mail: nnazokat71@gmail.com

Kuvondikova Gavhar Isomiddinovna
Teacher of the National University of Uzbekistan.

Abstract: This article describes the use of an electronic database to study and verify the analysis of speech units in social media with the help of a corpus, the identification of marked speech changes, their specific characteristics and the reasons for their changes.
Keywords: NLP, API, Twitter, Reddit, Facebook, LinkedIn.

Аннотация: В данной статье описано использование электронной базы данных для изучения и проверки анализа речевых единиц в социальных сетях с помощью корпуса, выявления выраженных речевых изменений, их специфических особенностей и причин их изменений.
Ключевые слова: NLP, API, Twitter, Reddit, Facebook, LinkedIn.

Introduction

Corpus analysis of speech units in social networks means identifying speech changes in Internet networks, analyzing them, determining the causes of the specified language changes, and creating a set of data about their specific characteristics. Texts on the Internet, posts on social media, blogs, forums, news and other data are considered the main sources, using corpora to study the prevalence of speech changes, their conversation topics, what thoughts or trends to pay attention to, or other important data can be obtained.

Literature analysis and methods

Many scientists have been studying changes in social processes through corpus analysis for several decades. However, the specific starting point may vary depending on the context and the particular line of research. The 1970s and 1980s saw the founding of corpus linguistics, with major contributions by scholars such as John Sinclair and Randolph Quirk, who focused more on linguistic description and analysis, and later on the study of social processes through corpus analysis. . In the 1990s, corpus linguistics began to expand its scope to include sociolinguistic inquiries. Scholars such as Douglas Bieber and Susan Conrad began using corpus methods to study language change in different social contexts. From the 2000s to the present, the use of corpus analysis in the study of changes in social processes has grown significantly. Advances in technology and the availability of large digital corpora have facilitated more sophisticated analyses.

Dan Thomas [1], David Bamman [4], Jacob Eisenstein [5], Cynthia Van Hee [6], David Juergens [7] have conducted research using social media discourse analysis, including Twitter

data. spent Their work explored linguistic and sociolinguistic aspects of online communication, including language change, sentiment analysis, and social dynamics.

Also, in a corpus-based analysis of changes in social processes, Timothy Graham [8], Nathan Mathias [9], Bernhard Rieder, Scott Hale, Christoph A. Hafner, and other researchers together examine various social phenomena, including political discourse, community formation, and language use. conducted research using Reddit data to explore topics related to models, online communities, collective action, and the impact of technology on social processes. Scientists David Lazer, Michael S. Bernstein [10] conducted a study using Facebook data to study social phenomena, such as information sharing, communication patterns and topics related to social networks on the platform, political communication, and collective behavior in communities. They used Facebook as a corpus to analyze social processes, online behavior, and changes in communication patterns within the platform [Abdurakhmonova, N. 2019,2021,2022].

In particular, the study of processes has developed over time, with the contribution of sciences and scientists. It remains an active area of research with constant changes in methodology and theory. The process of analyzing speech changes in social networks with the help of a corpus is carried out in certain stages.

First, select a suitable dataset for analysis and this dataset will include data from social networks such as Twitter, Facebook, Reddit, blogs, news portals and other sources. Data can be retrieved using automated or manual downloads, APIs (Application Programming Interfaces) or browser extensions to retrieve the required data from the collection. The received information is placed in a corpus or database, transferred to a suitable format and stored. It is also possible to use automated algorithms or data mining methods to detect language changes from the data in the corpus.

At the same time, a detailed analysis of selected speech changes can be done through language learning, distribution, exchange of ideas, use of words and expressions, comparison of actions on words, and other factors. The results of the analysis are then made available to data subjects, end users, or employees who have access to the data. Results can be presented through graphs, statistics, presentations and other visualizations. A detailed explanation of the results leads to a greater understanding of social media changes, trends, relationships between people, and more [Abdurakhmonova, N. 2019,2021,2022]. These results are fundamental to the knowledge and learning gained from the analysis of language change. Each step of this process requires challenge and reinforcement in data collection, analysis, and presentation of results.

When analyzing speech units in social networks using a corpus, it is necessary to pay attention to the following important aspects:

- topics of conversation in the language, such as new fashion, political events, technology or sports;

- in the process of analysis, the most used words and expressions in the network are people's opinions, important discussions in the network;

-how long will the exchange of ideas between users on the network change, what important negotiations will be conducted and what issues will be contested;

- the methods of language in social networks, writing, speaking, framing, memoirs and other features, changes in relations between people;

- the phonetics of the language, peculiarities, new words, unchanged words, changes in the style of speaking, specific analytical features of the language and directions of its development;

- discussions, opinions popular in social networks, words of people in social networks, reactions, changes and how they react to social processes in the world;

-in the analysis process, the data is analyzed using statistical analysis, related to many indications, such as how much speech units expand or decrease, which words or topics are used a lot, how the relationship between users' changes.

There are many factors to consider in the analysis process and can be customized for each project or request. The key here is to learn how to analyze language changes and to challenge the results in an advanced manner.

Practical significance

Corpus analysis of changes in social processes refers to the study of patterns, dynamics, and phenomena that develop within societies by analyzing large collections of texts or corpora [Abdurakhmonova, N. 2019,2021,2022]. This approach refers to the study of how language, behavior, attitudes, and interactions change over time under the influence of various social, cultural, economic, and technological factors.

There are some key aspects of changes in corpus analysis of social processes. Corpus analysis allows researchers to track linguistic changes in a given language or across languages over time. By examining linguistic features such as vocabulary, syntax, discourse structures, and stylistic conventions, researchers can identify changes in language use and how they reflect broader social trends [Abdurakhmonova, N. 2019,2021,2022].

Corpus analysis makes it possible to study social dynamics and laws of interaction within societies, groups and societies. Researchers can study how people communicate, share information, form relationships, express opinions, and negotiate in a variety of social contexts, including online platforms, formal institutions, and everyday interactions.

Also, through corpus analysis, scholars can investigate cultural evolution and innovation by tracking the spread, adoption, and adaptation of cultural artifacts, practices, and norms in societies. This includes the study of popular culture trends, media representations, consumer behavior, and cultural expressions over time.

Conclusion

In conclusion, the corpus analysis of changes in social processes creates a rich basis for studying the complex interaction between language, culture, and society. By systematically analyzing large data sets of textual materials, researchers can gain insight into the dynamics of human behavior, communication, and social organization in diverse temporal, spatial, and cultural contexts.

References:

1. Computational Linguistics: An Introduction (Studies in Natural Language Processing). Cambridge University Press.
2. Taweh Beysolow. Applied Natural Language Processing with Python.California, 2018.
3. Michele Zappavigna. Discourse of Twitter and Social Media: How We Use Language to Create Affiliation on the Web. A&C Black, 2012.
4. David Bamman, Noah A. Smith. Contextualized Sarcasm Detection on Twitter. International Conference on Web and Social Media. Computer Science, Linguistics, 2015.

5. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). Uzbek electronic corpus as a tool for linguistic analysis. In *Компьютерная обработка тюркских языков. TURKLANG 2022* (pp. 231-240).
6. Abduraxmonova, N. Z. Q., & Urazaliyeva, M. Y. (2022). O ‘zbek tili elektron korpusida (http://uzbekcorpus. uz/) og ‘zaki matnlar korpusini yaratishning nazariy va amaliy masalalari. *Academic research in educational sciences*, *3*(3), 644-650.
7. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, *11*(19), 9117.
8. Jacob Eisenstein. Introduction to Natural Language Processing Adaptive Computation and Machine Learning series. MIT Press, 2019.
9. Cynthia Van Hee, Els Lefever, Veronique Hoste. Exploring the Realization of Irony in Twitter Data.International Conference on Web and Social Media, 2016.
10. D. Nguyen, A. Seza Doğruöz, F. D. Jong. Computational Sociolinguistics: A Survey. Computational Linguistics, 2016.
11. L. Jenny, Davis, T. Graham. Emotional consequences and attention rewards: the social effects of ratings on Reddit. Information Communication and Society 24(4):1-18, 2021.
12. J. Nathan Matias. Going Dark: Social Factors in Collective Action Against Platform Operators in the Reddit Blackout. Microsoft Research, 2015.
13. S. Michael, E. Bakshy, B. Karrer. Quantifying the invisible audience in social networks. International Conference on Human Factors in Computing Systems, 2013.
14. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In *of the International Conference on Language Technologies for All (LT4All)*.
15. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).

MINNATDORCHILIK – MULOQOT MANBAI SIFATIDA

Oripova Parvina Valisherovna
Buxoro davlat universiteti
Filologiya fakulteti 2- bosqich talabasi.

Ilmiy rahbar: **Toirova Guli Ibragimovna**
BuxDU, O'zbek tilshunosligi va jurnalistika kafedrasi
Professori, filologiya fanlar doktori.
E-mail: tugulijon@mail.ru

Annotatsiya: Ushbu maqolada minnatdorchilik ko'rinishlari va ularning o'ziga xos xususiyatlari haqida ma'lumot keltirilgan. Qolaversa, muloqot va kommunikatsiya tushunchalariga izoh berilib,ular o'rtasidagi tafovutlar ochib berilgan. So'zlovchi va tinglovchi o'rtasidagi muloqotning ta'sir doirasi aks ettirilgan. Noverbal nutqning o'ziga xosliklari va bu nutqni yuzaga keltiruvchi vositalar va holatlarning namoyon bo'lishi mavjud. Kognitiv, emotiv, konativ muloqot turlari haqida fikr bildirilgan.

Kalit so'zlar: muloqot, kommunikatsiya, emotiv nutq, kognitiv nutq, konativ nutq, kommunikativ-axborotli nutq, interaktiv nutq, perseptiv nutq.

Аннотация: В данной статье представлена информация о выражениях благодарности и их специфических характеристиках, а также раскрыты понятия диалога и обшения и выявлены различия между ними. Отражена сфера влияния обшения и говоряшего и слушающего, выявлены особенности невербальной речи и проявления средств и ситуаций, создающих эту речь. Высказано мнение о типах когнитивной, эмоциональной и волевой коммуникации.

Ключевые слова: диалог, общение, эмоциональная речь, познавательная речь, конативная речь, коммуникативно-познавательная речь, интерактивная речь, перцептивная речь.

Annotation: This article provides information about expressions of gratitude and their specific characteristics. In addition, the concepts of dialogue and communication are explained and the differences between them are revealed. The sphere of influence of the communication between the speaker and the listener is reflected. There are features of non-verbal speech and the manifestation of means and thev situations that create this speech. An opinion was expressed about the types of cognitive, emotive, and conative communication.

Keywords: communication, emotive speech, cogonative speech, communicative-informative speech, interactive speech, perceptive speech.

Kirish

Insonning kim ekanligi, qanday xarakterga egaligi uning so'zlagan nutqi orqali namoyon bo'ladi. So'z shunday qudratli kuchki, inson so'zlayotganda tafakkuridagi so'zlar sandig'idan eng keraklisini tanlab olib,suhbatdoshiga ta'sir ko'rsatishni istaydi va buni o'z nutqi orqali amalga oshiradi. Shundagina uning asl xislati, xarakteri namoyon bo'ladi. Bu vaqtda tilda verbal va

noverbal muloqot vositalari muhim o'rin tutadi. Verbal aloqa – zarur shart sifatida tilni o'zlashtirishni belgilab beradigan faqat insonlarga xos bo'lgan jarayon. Muloqot (munosabat) birgalikda faoliyat ko'rsatuvchilar o'rtasida axborot ayirboshlashni o'z ichiga oladi. Bunda munosabatning kommunikativ (o'zaro aloqaga doir) jihati hisobga olinadi. Muloqot tushunchasini kommunikatsiyadan farqlash darkor. Kommunikatsiya tirik va o'lik tabiatdagi tizimlar o'rtasida axborot almashinuvini anglatadi. Hayvonlar o'rtasidagi signallar almashinuvi, insonning texnik vositalar bilan aloqa qilishi – bularning barchasi kommunikatsiya. Muloqot esa faqat insonlar o'rtasidagina amalga oshirilishi mumkin. Demak, muloqot jarayonida kommunikativ (axborot uzatish), interaktiv (o'zaro birgalikda harakat qilish), va perseptiv (o'zaro birgalikda) idrok etish amalga oshiriladi. Noverbal aloqa – so'zlardan foydalanmasdan xabarlarni yuborish va qabul qilish jarayoni. Noverbal (nutqsiz) kommunikatsiya vositalariga yuz ifodasi, mimika, ohang, pauza, poza (holat), ko'z yosh, kulgi va boshqalar kiradi. Imo – ishora, ma'noli qarash, rag'batlantiruvchi yoki istehzoli tabassum insonga ta'sir ko'rsatishda ko'p so'zli tushuntirish va e'tiroz bildirishga qaraganda ancha samarali muomala vositalari sanaladi. Bu vositalar verbal kommunikatsiya – so'zni to'ldiradi, kuchaytiradi va ba'zan uning o'rnini bosadi.

Asosiy qism

Minnatdorchilik beshta:

- Shaxslararo
- Kognitiv
- Kommunikativ-axborotli
- Emotiv
- Konativ tomonlarining birligida namoyon bo'ladi.

1. Muloqotning shaxslararo tomoni insonning bevosita atrof – muhiti: boshqa odamlar va o'z hayotida bog'liq bo'ladigan umumiyliklar bilan o'zaro ta'sirini aks ettiradi.
2. Muloqotning kognitiv tomoni suhbatdoshning kim, qanday odam ekanligi, undan nima kutish mumkinligi haqidagi, shuningdek, sherikning shaxsi bilan bog'liq bo'lgan boshqa ko'pgina savollarga javob berish imkonini beradi.
3. Muloqotning kommunikativ-axborotli tomoni tasavvurlar, g'oyalar, qiziqishlar, hissiyotlar, mayllar va shunga o'xshashlari turlicha bo'lgan odamlar o'rtasidagi almashinuvdan iborat.
4. Muloqotning emotiv tomoni sheriklarning shaxsiy aloqalaridagi his-tuyg'ular,kayfiyatning vazifa bajarishlari bilan bog'liq.
5. Muloqotning konativ (xulq-atvor) tomoni sheriklar qarashlaridagi ichki va tashqi qarama-qarshiliklarni muvofiqlashtirishga xizmat qiladi. Bunda quyidagilar muhim ahamiyatga ega: ohang, emotsional ifodalanganlik, bir xil iboraga turlicha ma'no berishga qodir. Suhbatdoshning mimikasi, gavda holati, nigohi ibora ma'nosini kuchaytirishi, to'ldirishi yoki rad etishi mumkin.Imo-ishoralar muloqot vositasi sifatida umumiy qabul qilingan,ya'ni belgilangan ma'noga ega bo'lishi yoki ekspressiv, ya'ni nutqning ifodaliligini oshirishi mumkin.
    - Qarasam, cholning ko'krak cho'ntagida boyagi yaltirab turgan soat zanjiri ko'rinmayapti. Dadam uning ko'ziga tikilib,sekin xo'rsindi. – Rahmat. Tirik bo'lsam , shu yaxshiligingizni qaytaraman.[50]

- Choynak ko'tarib kelsam, chol-kampir haliyam bir nimaning maslahatini qilishyapti. – Barakalla, qizim! - dedi bobom uzatgan piyolamni olib. – Iloyo baxting ochilsin! Ikkovlari bir-biriga zimdan qarab olishdi. Tashqarida eshak hangradi, molxonada sigir mo'radi… [136]
- - Rahmat,qizim! Yuzimni yorug' qilding, - dedi ovozi titrab. - Qo'sha qarigin, o'zingdan ko'paygin, bolam… [126]

Minnatdorchilik – bu yaxshi ish yoki boshqa har qanday yaxshilik uchun minnatdorchilik hissi. Bu biz boshqa odamga, odamlar guruhiga, jamiyatga, biror bir hodisaga, Xudoga, hayotga yoki koinotga nisabatan his qilishimiz mumkin bo'lgan ijobiy tuyg'u. Psixologlarning ta'kidlashicha, minnatdorchilik kimga qaratilganidan ko'ra uni boshdan kechirgan kishi uchun muhimroqdir. Bu tuyg'u hissiy kuchni oshiradi va ichki holatni uyg'unlashtiradi. Minnatdor odam odatda xotirjam va stressga chidamli bo'ladi, hamma narsada ijobiy tomonlarni sezishga intiladi va hamma narsaga nisabatan yanada oqilona munosabatda bo'ladi. Ta'kidlash joizki, dastlab shaxsiy rivojlanishning yuqori darajasida bo'lgan odamlar minnnatdorchilikni boshdan kechiradi. Bu tuyg'u o'z – o'zidan yoqimli, lekin u boshqa afzalliklarga ham ega. Xususan, u foydalidir, chunki: kayfiyatni yaxshilaydi. Minnatdorchilik bilan bir vaqtda kuchli salbiy his-tuyg'ularni boshdan kechirish mumkin emas. Biror kishi kimgadir chin dildan minnatdor bo'lsa, u deyarli barcha salbiy narsalarni tashlab yuboradi va buning o'rniga quvonch hissi paydo bo'ladi.

- Jayrona, hech kim o'zga yurtda dardga chalinmasin ekan. O'z yurtingda og'risang, yurtingni havosi ham dori bo'ladi. Ammo musofirlikda juda g'arib bo'lib qolarkansan. Yaxshiyam siz bor ekansiz. Ming rahmat. Bo'lmasam, kim biladi, nima bo'lardim. Hech qachon yaxshiligingizni unutmayman. O'lmasam, sizga bu yaxshiligingizni o'n chandon ortiq qilib qayataraman.[106]
- Buyuklar buyugi! Ko'z yoshlarimni, tavbamni qabul qil. Sen eshingingga kelganni quvmaysan. Qancha gunoh va xatolari bo'lsa, kechirasan. Zero, boshqa choram ham, eshigim ham yo'q. Agar qabul etmasang, qaysi eshikka boraman?... Men bir gunohkor qulingman, bugun dargohingga keldim. Menga shunday kunni nasib qilganing uchun o'zingga shukrlar bo'lsin!...[128]

Shuningdek, minnatdorchilik bizda bor narsani payqash va qadrlashni o'rgatadi, muammolardan chalg'itadi, o'zaro munosabatni yaxshilaydi, tushunish va kechirishni o'rgatadi. Minnatdorchilik ko'rinishlari: fe'lli so'z birikmalari (qisqartma shaklida): p.f. – (pour fe'liciter) tabriklashda; p.r. – (pour remercier) – minnatdorchilik bildirishda; p.p.c. – (pour prendre conge') - xayrlashuv belgisi sifatida; p.c. (pour condoleance) – hamdardlik bildirishda qo'llaniladi.

Xulosa

Minnatdorchilik o'ziga xos xususiyatga ega bo'lgan nutqiy birlik sanaladi. Minnatdorchilik ("Baraka berish") so'zidan – yaxshilik uchun , masalan, e'tibor yoki ko'rsatilgan xizmat uchun minnatdorchilik hissi, shuningdek, ushbu tuyg'uni ifodalashning turli usullari, jumladan rasmiy rag'batlantirish choralari (masalan, minnatdorchilik izhori). Minnatdorchilik- bu sodir bo'layotgan voqealarni hissiy va oqilona baholash asosida yuzaga keladigan murakkab tuyg'u bo'linb, uning ifodasi yuqori darajada marosimdir. Minnatdorchilikning manzili nafaqat aniq odamlar, balki odamlar jamoalari (masalan, ajdodlar) va shaxsiy bo'lmagan shaxslar bo'lishi mumkin: dunyo, Xudo. Nutrqiy muloqot jarayonida minnatdorchilik bir nechta pragmatic ma'no ifodalab keladi: suhbatni yakunlash, samimiy munosabat, shukronalik, rozilik.

Foydalanilgan adabiyotlar ro'yxati:

1. Бакалейников Г.Я. Речевой акт благоданости в современном русском языке // Р Бакалейников 1otsdamer Forschungen der Paed. Hochsch.Karl Liebknecht. R.A. Gesellschaftswiss. R. H. 106. Potsdam, 1990.- С. 85-101.
2. Бободжонова Б. Ш Категория вежливости в таджикском языке. Дисс. На соискание учёной степени канд. фил. наук / 2001.- 150с
3. Nordberg B. On closing in alarm calls // Sprak och stil. 1999 №8. S. 65-103.
4. G. Toirova, O. Jurayeva, Z.Abuzalova, M.Norova, F.Norova. Application of Innovative Technologies in Teaching Process. // International Journal of Psychosocial Rehabilitation, Vol.24, Special Issue 1, 2020. ISSN: 1475-7192.- P.386-390.
5. Toirova G., Yuldasheva M., Elibaeva I. Importance of Interface in Creating Corpus. // International Journal of Recent Technology and Engineering (IJRTE) ISSN:2277-3878, Volume-8 Issue-2S10, September 2019. – P.352-355.
6. O'. Hoshimov. Ikki eshik orasi.- Toshkent:Global Books,2021.-B.50

CHALLENGES IN CORPUS LINGUISTICS AND MACHINE TRANSLATION

Nazirova Elmira
DSc of the Tashkent university of information technologies.
E-mail: elmira_nazirova@mail.ru

Abdurakhmonova Nilufar
Doctor of Philology,
Professor of the National university of Uzbekistan.
E-mail: n.abdurakhmonova@nuu.uz

Usmonova Kamola
PhD of the Tashkent university of information technologies.
E-mail: kamolausmonova93@gmail.com

Abstract: This article addresses systematic cases of progress problems on corpora. The process of corpus generation creates linguistic and statistical issues, which eventually govern the entire process. During corpus generation, it needs careful attention to different factors such as corpus size, methods of data collection, organization of textual materials and others. These issues are significant not only for widely spoken languages like English and Turkish but also hold considerable vital Importance for less-resourced languages used in developing countries. We decided to explore these issues in detail in this article.

Keywords: corpus linguistics, language patterns, machine translation, morphological complexity lexical gaps, cultural nuances, linguistic diversity, data collection.

Аннотация: Данный статья рассматривает систематические случаи проблем прогресса в корпусной лингвистике. Процесс создания корпуса порождает лингвистические и статистические проблемы, которые в конечном итоге определяют весь процесс. Во время создания корпуса необходимо уделить внимание таким различным факторам, как размер корпуса, методы сбора данных, организация текстовых материалов и другие. Эти проблемы имеют важное значение не только для широко используемых языков, таких как английский и турецкий, но и имеют значительное значение для менее ресурсных языков, используемых в развивающихся странах. Мы решили подробно изучить эти вопросы в этой статье.

Ключевые слова: корпусная лингвистика, языковые шаблоны, машинный перевод, морфологическая сложность, лексические пробелы, культурные нюансы, лингвистическое разнообразие, сбор данных.

Introduction

In the linguistic field, the implementation of the corpus has increased exponentially for the last 10 years. As Cheng (2011) mentioned that text compilation which gathered for a specific purpose. Meanwhile, the definition of the corpus as a text collection, which is clear, among linguists there are arguments about corpus linguistics as a theory or methodology. According to Tognini-Bonelli (2001) it is necessary a methodological framework for studying language rather than corpus linguistics distinct field of research. There is serious attention involved in the process

of generating a corpus in digital form from written sources. It is necessary to take into consideration of various linguistic, linguistic and statistical considerations. As McEnery and Haedie (2011) and Crawford and Csomay (2015) stated the determination of target and careful execution by corpus designers to ensure the successful work of corpus. Likewise, as Weisser (2016) mentioned that maintains that corpus linguistics serves to enhance our understanding of language by analyzing real-life data, emphasizing the practical application of methodology. Therefore, conducting corpus linguistics transcends mere methodology. The advancement of Corpus mechanisms and systems makes easier for people around the world who want to do research in their areas. Over time these mechanisms have gone from being mainly used for linguistic studies and education to being used for research in other specific areas. More and more researchers are using corpus technology in their studies and making their own observations and finding out some challenges and this paper also looks at the progress in research findings in various studies and several challenges in machine translation systems from Uzbek into English within the realm of corpus linguistics.

Discussion

It is essential to understand what corpus linguistics is and what is not corpus linguistics if people can get enough information, then it is easy to enhance after deep observation. Corpus linguistics cannot provide negative evidence and all possible language at a one time as well as explain why. This means corpus tells us what is or is not have in the corpus. It cannot give information what is possible or incorrect in language. That is why many users believe erroneously that whether it cannot show all manners to convey the concept then they may consider the corpus is entirely flawed. Corpus linguistics helps spot patterns in language use, but it cannot always explain why those patterns exist. Figuring out the reasons behind language use often depends on the instincts of language speakers. Even though corpora are carefully put together and cover a wide range of texts, they cannot capture every possible way language is used [Abdurakhmonova, N. 2019,2021,2022]. Hence, while corpora are organized collections of language data, they have limits when it comes to representing the full diversity of language.

Challenges corpora linguistics from Uzbek into English

There are several issues occurring during the creation of machine translation. The issues are Morphological complexity, lexical Gaps, word order differences, and cultural nuances. The reason is there is a rich morphology with complex inflection and derivational patterns. While translating it can be challenging to do more accurately into English and make it simpler morphology. To be more precise, Uzbek has agglutination and extensive suffixation. To illustrate, “daftarlarimizdan” is Uzbek word, which consists of “daftar” (note-book) with suffix – “lar” (plural) and “imizdan” (from our). When we translate it, it is from our note-books that requires not only identifying the root but also correctly handling the plural and possessive markers [Abdurakhmonova, N. 2019,2021,2022].

Word order differences: Uzbek and English word order patterns is not similar. Uzbek word order follows S-O-V, while English has S-V-O order. It can be grammatical errors if not handling properly while translating from one language to another. To cite a relevant example is that “U o’qidi maqolani” is Uzbek sentence when we translate it, we have to rearrange it to She reads the article. In order to translate accurately the correct placement of subject, verb, and object is vital.

Lexical gaps: there are some words in Uzbek, which concepts do not have direct equivalents in English or vice versa. In this case, we need creative solutions sometimes. In Uzbek, the word "dustlik" means more than just friendship. It's about having close friends who support each other like family, showing loyalty and sticking together through thick and thin. While "friendship" is close, it does not fully capture the strong emotional bond and mutual care that "dustlik" represents in Uzbek culture. A useful example to mention here is that in Uzbek, the word "dustlik" means more than just friendship. It's about having close friends who support each other like family, showing loyalty and sticking together through thick and thin. While "friendship" is close, it does not fully capture the strong emotional bond and mutual care that "dustlik" represents in Uzbek culture.

Cultural Nuances: Languages often have words and phrases that carry unique cultural meanings, making direct translation difficult. Ensuring that translations capture both the intended meaning and cultural context is important but can be tough. In Uzbek, when someone says “Assalom alaykum, yaxshimisiz”(Hello. How are you), it is not just a greeting- it also shows politeness and hospitality. Translating it directly as "Hello, how are you?" might not capture all the cultural meanings behind the greeting [Abdurakhmonova, N. 2019,2021,2022].

Furthermore, researchers frequently encountered some issues in Data collection, annotation and tagging, domain adaptation, evaluation metrics and resource availability. In Data collection, it is clear that Uzbek and English cover various genres and domains. Thus, the issue can be collecting a corpus that contains various texts, new articles, and conversational speech that ensures machine translation models are trained on diverse linguistic data. In annotation and tagging, linguistic data is very important in order to train accurate machine translation models. Syntactic structures and entities in the corpus help machine systems understand the relationship within sentences. In domain adaptation, it is necessary to differ domain, such as legal or medical texts so demand training on domain-specific corpora. Thus, here machine translation tool should accurately translate those medical terminology and specific language patterns. In evaluation metrics and resource availability, it is very important to focus on measuring system performance in order to develop reliable evaluations that obtain the quality and accuracy. Moreover, if languages have limited resources, it also creates problems such as size and composition balancing of the corpus with the range of linguistic variations. Thus, it is necessary to take into consideration the coverage of diverse linguistic phenomena and sufficient data for robust analysis.

Conclusion

This article gives information on the systematic challenges encountered in corpus linguistics. Generation and application of corpora are emphasized, as well as it highlights the importance of considerations during corpus generation such as size, data collection methods and textual organization, which hold significant important not only for widely spoken languages but also for less-resourced languages, especially in developing countries. Moreover, certainly, the article also sheds light on the boundaries of corpus linguistics. It highlights its inability to provide negative evidence, cover all conceivable linguistic variations, or fully elucidate the underlying reasons for language patterns. Nevertheless, despite these limitations, corpus linguistics remains an indispensable tool for uncovering recurrent language patterns. However, the challenges extend further when considering machine translation from Uzbek to English within the corpus linguistics framework. These hurdles encompass the intricate morphological structure, lexical disparities,

divergent word order, and nuanced cultural subtleties present in Uzbek. These complexities demand innovative strategies and meticulous precision to achieve accurate translation into English.

References:

1. C. Chen, K.Chan, P.Wong, E.Chee, L.Wang, Q.Wang. A corpus-based online pronunciation learning system: The Pedagogical applications of a spoken corpus for improving Hong Kong/Mainland university students' English pronunciation. The Second Asia Pacific Corpus Linguistics Conference, 2014.
2. E.Tognini-Bonelli.Corpus linguistics at work. Amsterdam: J. Benjamins, 2001.
3. M. Weisser. Practical corpus linguistics: An introduction to corpus-based language analysis. John Wiley- Sons. 2016
4. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). Uzbek electronic corpus as a tool for linguistic analysis. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
5. Abduraxmonova, N. Z. Q., & Urazaliyeva, M. Y. (2022). O 'zbek tili elektron korpusida (http://uzbekcorpus. uz/) og 'zaki matnlar korpusini yaratishning nazariy va amaliy masalalari. Academic research in educational sciences, 3(3), 644-650.
6. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
7. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
8. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
9. N.Abdurakhmonova, U.Tuliyev, A.Gatiatullin. Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus. uz.International Conference on Information Science and Communications Technologies (ICISCT), 2021. pp. 1-4.
10. T.McEnery, A.Hardie. Corpus Linguistics: Method, Theory, and Practice. Cambridge: Cambridge University Press, 2011.
11. W.Crawford, E.Csomay. Doing Corpus Linguistics. London: Routledge, 2015.

FRAZEOLOGIK POLISEMIYA VA UNI KORPUS TIZIMIDA MODELLASHTIRISH MUAMMOLARI

Gulyamova Shahnoza
Turkiston yangi innovatsiyalar universiteti dotsenti, filologiya fanlari doktori.
E-mail: bunnycham@mail.ru

Nurboyeva Maftuna Vahobjon qizi
ToshDO‘TAU 2-bosqich tayanch doktoranti.
E-mail: maftunanurboyeva@gmail.com

Annotatsiya: Maqolada frazeologik polisemiyaning o‘zbek tilshunosligida o‘rganilishi, lug‘atlarda berilishi va polisemantik iboralarning korpus tilshunosligi tizimiga lingvistik model asosida kiritish hamda iboralarning ko‘p ma’noliligi tufayli yuzaga kelishi mumkin bo‘lgan noaniqliklarni bartaraf etish masalalari haqida so‘z boradi.

Kalit so‘zlar: frazeologiya, frazeologik polisemiya, frazeologik omonimiya, lingvistik model, variatsiya, morfo-semantik kollokatsiya, idioma, erkin birikma.

Аннотация: В статье речь идет об изучении фразеологической многозначности в узбекском языкознании, ее включении в словари и включении многозначных выражений в систему корпусной лингвистики на основе лингвистической модели, а также устранении неясностей, которые могут возникнуть из-за многозначности выражений.

Ключевые слова: фразеология, фразеологическая полисемия, фразеологическая омонимия, языковая модель, вариация, морфосемантическое словосочетание, идиома, словосочетание.

Abstract: The article deals with the study of phraseological polysemy in Uzbek linguistics, its inclusion in dictionaries and inclusion of polysemantic expressions into the system of corpus linguistics based on the linguistic model, and the elimination of ambiguities that may arise due to the polysemy of expressions.

Keywords: phraseology, phraseological polysemy, phraseological homonymy, linguistic model, variation, morpho-semantic collocation, idiom, free combination.

Har bir davr o‘z taraqqiy shakliga ega bo‘lar ekan, bu holat fanning barcha tarmoqlarida aks etmay qolmaydi va zotan, ilm-fanning har qanday turi zamon bilan hamnafas tarzda rivojlanishi ijobiy holat, albatta. Axborot texnologiyalarining jadal rivojlanishi fanning ko‘plab sohalarida, xususan, tilshunoslikda kompyuterlardan foydalanishni zaruratga aylantirdi. Bu holatning asosiy sababi, axborotni kompyuterlar orqali osonroq qayta ishlanishidir. Shu tufayli tilshunoslikda muayyan tilni kompyuter sistemasiga “tanitish” ehtiyoji paydo bo‘ldi va bu ehtiyoj 1900-yillarning ikkinchi yarmidan boshlab kompyuter lingvistikasi tarmog‘ining alohida shakllanishi va rivojlanishiga turtki bo‘ldi. Mazkur sohada tadqiqotlar doirasi kengayib, ahamiyati orta borgani sari tilshunoslikning barcha sathlari bo‘yicha qamrov yoya boshladi. O‘zbek tilshunosligi uchun birmuncha yangi hisoblangan bu sohaning ilk uchqunlari XIII asrda namoyon

bo‘lgani holda, elektr hisoblash mashinalari ixtirosi ommalashib, takomillashganidan so‘ng haqiqiy tadqiqot ko‘rinishiga ega bo‘ldi.

Bugunga qadar jahon tilshunosligida yetarlicha tadqiq bazasiga ega bo‘lgan korpus tilshunosligi o‘zbek lingvistikasi sohalariga ham jadal kirib kelmoqda. Davr zarurati sifatida dolzarblik kasb etayotgan bu soha muayyan tilning nufuzini va elektron qurilmalarda qo‘llanimini oshiribgina qolmay, uning yashovchanligini ham kafolatlaydi. Bugungacha mazkur sohada bir qancha tadqiqotlar olib borilib, o‘zbek tilida yaratilishi lozim bo‘lgan yaxlit korpus tizimi shakllantirilmoqda Bu esa fanimiz oldida turgan dolzarb vazifalardan biri – o‘zbek tili frazeologik birliklari avtomatik tahlilida omonimlik va ko‘p ma’nolilikni farqlashning lingvistik asoslarini yaratishni kun tartibiga qo‘yadi.

O‘zbek tilshunosligida kompyuter lingvistikasi, matnga leksikografik ishlov berish borasida M.Mahmudov, M.Ayimbetov, S.Karimov, G.Jumanazarova, A. Babanarov, D.O‘rinboyeva, A.Norov, S.Muhammedov, N. Abdurahmonovalarning tadqiqotlari mavjud. Sanalgan tadqiqotlar kompyuter lingvistikasi yo‘nalishidagi bir qator muammolarni tadqiq etganligi bilan ahamiyatli. Shuningdek, Sh.Gulyamovaning ham polisemik va omonimik leksemalarni korpus lingvistikasida semantik farqlovchi modellarini yaratish borasida qilgan ilmiy ishlari o‘zbek korpus sistemasida katta ahamiyatga ega.

Frazeologik polisemiyani korpus lingvistikasida farqlash masalasi nafaqat tilshunosligimizda, balki jahon tilshunosligida ham yechimini kutayotgan dolzarb masalalardan biri hisoblanadi. Chunki iboralar tilni avtomatlashtirish jarayonida ham, mashina tarjimonligida ham muammoli qism bo`lib kelgan. Iboralar o‘rasidagi omonimiya va polisemiya esa bu murakkablikni yanada oshiradi. Iboralarning polisemiyasi, umuman ularning leksik birliklar kabi so‘zlarning shakl va ma’no munosabatiga ko‘ra turlari mavjudligi frazeologiya alohida sath sifatida ajralib chiqqandan buyon tadqiqotlarda o‘z aksini topgan edi. Tadqiqotimizning obyekti bo‘lgan bu birliklarni aratib olishda ham shu ishlar natijasi bo‘lgan lug‘atlarga murojaat qilamiz.

Saralash jarayonida shunga guvoh bo‘lish mumkinki, lug‘atlarda polisemantik iboralarning berilishida broz noaniqlik borligini payqash mumkin. Jumladan, *boshini silamoq* iborasi izohiga diqqat qiladigan bo‘lsak, bu ibora Sh.Rahmatullayevning dastalbki frazeologik lug‘atida ibora sifatida uchramasligini, mualliflar jamoasi tomonidan yig‘ilgan yangi izohli lug‘atda polisemantik ibora sifatida quyidagicha berilganini ko‘rish mumkin: **Boshini silamoq** *kim?, kimning*? 1. kaft yurgizib, erkalatmoq, suymoq, yupatmoq. *Kampir uning boshini silab yostig‘ini to‘g‘irladi.* S.Ahmad. Ko‘klam chechaklari. 2. G‘amxo‘rlik qilmoq, siylamoq, hayotini yengillashtirish yo‘lida yordam bermoq. *U qurilishda ilg‘orlarni ishdan bo‘shatib, qoloqlarning boshini siladi.* A.Muxtor. Tug‘ilish.

Izohli lug‘atda ham yuqoridagi holatga yaqin mazmunni uchratish mumkin: **boshini silamoq.** - 1. erkalatmoq, yupatmoq. *Yo‘lchi Gulnorning boshini silab, voqealarni qisqacha aytib berdi.* Oybek. Qutlug‘ qon. 2. O‘z panohiga olib ardoqlamoq, boqmoq, tarbiya qilmoq. *Erga bersang oshingni, Erlar silar boshingni.* Maqol. Ammo fikrimizcha, iboraning polisemantik deya ko‘rsatilgan birinchi ma’nosi ibora emas, iboraga teng kelib qolgan so‘z birikmasidir, negaki uning tarkibi barqaror emas, ya’ni ibora tarkibiga boshqa gap bo‘laklarini kiritsa bo‘ladi, shuningdek, birinchi ma’noda aynan boshni kaft bilan silash jarayoni reallashyapti. Va unga qo‘shilayotgan, *boshni silash orqali erkalashni bildirish* semasi ibora tuzilmasi orqali emas, *silamoq* fe’lining mohiyatidagi erkalash semasi orqali yuklanyapti. Buni izohli lug‘atdagi *silamoq* fe’li izohidan ham dalillash mumkin: **Silamoq. 1**. Ustidan kaft yurgizib, siypamoq. *Inobat xudddi bo‘g‘ilib*

ketganday, ro‘molini yechib, tomog‘ini siladi. O.Yoqubov. Er boshiga ish tushsa. **2.** Ustidan qo‘lini asta yurgizib, siypab erkalatmoq. *G‘ulomjon otxonaga kirib, qorong‘ida ham samanining peshonasini silab, suyib qo‘ydi.* M.Ismoiliy. Farg‘ona tong otguncha. **3.** mayin, ohista tegmoq. *Mayin esgan shamol.. daraxtlar soyasida yugurib borayotgan Saodatning sochlarini, yuzlarini siladi*. N.Safarov. Saodatning qissasi. Ko‘rinib turibdiki, *boshini silamoq* iborasining polisemiya sifatida ko‘rsatilgan dastabki ma’nosi silamoq so‘zining ma’no imkoniyatidan kelib chiqqan holda *erkalatmoq* ma’nosini tashiyapti, shuningdek, dastlabki ma’noda iborani tashkil qiluvchi gap bo‘laklarining orasiga boshqa bo‘laklar qo‘yib bo‘ladi, ammo “o‘z panohiga olmoq, boqmoq, tarbiyalamoq, ardoqlamoq” ma’nosini berayotgan ikkinchi semada ibora birliklari orasiga boshqa gap bo‘laklari qo‘yib bo‘lmaydi. Buni iboraning 2-semasiga keltirilgan misoldan ham sinab bilib olish mumkin. Ammo birinchi semaga keltirilgan misolda ham, boshqa adabiyotlarda ham, namunalar dastabki semaning turg‘un emas, erkin birikma ekanini ko‘rsatmoqda, xususan, A.Hojiyevning “Fe’l” nomli tadqiqotida ham fe’l shaklidagi ibora va qo‘shma fe’lni farqlashga intilar ekan, qo‘shma fe’lda semantik va forma butunligi qat’iy bo‘lishini, iborada esa semantik butunlik qat’iy bo‘lib, forma butunligi ham qisman bo‘lishini ta’kidlab o‘tadi[Hojiyev A. 1973; 12-13], bu orqali qo‘shma fe’lni tashkil qilgan birliklarni shaxs-songa moslab bo‘lmaydi, grammatik jihatdan turlash yoki tuslashning imkoni yo‘q, ammo iboralarning tarkibidagi so‘zlarni jumla shaxsiga moslab o‘zgartirish mumkin, ammo uni tashkil qilgan birliklar orasiga boshqa gap bo‘laklari qo‘yib bo‘lmaydi, degan fikrni anglash mumkin. Namunalar orqali ham buni anglash mumkin: *Umrinisabibi uning boshini muloyim-muloyim siladi.* Cho‘lpon. Kecha va kunduz.

Shuningdek, *yo‘l bo‘lsin* iborasining ham lug‘atlardagi izohi biroz muhokamali. Lugatlarda mazkur ibora polisemantik ibora sifatida quyidagicha berilgan: **YO‘L BO‘LSIN 1** *kimga? qayerga ketayapsiz? qayerga bormoqchisiz? (so‘rash). «Yo‘l bo‘lsin?», – dedi chol. «Ko‘pdan quruq qolma deydilar. Anav vaqtda patir ushatildi degan edingiz, bugun kichik to‘y kelayotgan emish, shunga ketyapman».* Oydin. Shirin keldi. **2** qayoqda! (biror ishni) qila olarmidi?! *(ta’kid). Men bino qilganman bu uyni. Sizdaqa beshud erga yo‘l bo‘lsin uy qilish!* Oybek. Oltin vodiydan shabadalar. Bu izoh dastlabki izohli frazeologik lug‘atda ham, so‘nggi izohli lug‘atda ham xuddi shu shaklda berilgan. Ammo polisemiya, sinonimiya, antonimiyaning muhim jihati shundan iboratki, ularga mansub birliklar o‘zaro bir so‘z turkum doirasida bo‘lishi kerak. Ammo bu iboraning ma’nolarida turkum jihatdan mushtaraklik kuzatilmaydi, buning ustiga iboraning bir shakli matnda butun bir jumlani, so‘roq gapni ifodalab kelsa, ikkinchi holatda uning muayyan leksik o‘rindoshi yo‘q va uni grammatik ma’no tashuvchi ibora ham deyish mumkin. Chunki iboralarning semantik imkoniyati haqida so‘z ketganda ular ikki yoqlama lisoniy birlik bo‘lib, shakl va mazmunning dialektik birligidan iboratligi, frazeologik ma’no o‘ta murakkab tabiatli: ayrimlari leksema kabi denotativ xarakter kasb etsa, boshqalari grammatik ma’noga ega ekanligi aytiladi, jumladan, mustaqil frazeologizmlarga kiruvchi *kapalagi uchmoq, qovog‘idan qor yog‘moq, qo‘li ochiq, qo‘l ko‘tarmoq* kabi birikmalar denotatik tabiatga ega bo‘lsa, *shunga qaramay, turgan gap* kabi birliklar faqat grammatik vazifa bajarishi aytiladi[Raximova S. 2022; 42]. Bundan kelib chiqadiki, *yo‘l bo‘lsin* birikmasining ikkinchi smeasini ham grammatik ma’noda qo‘llanuvchi va dastabki ma’nosi bilan omonimlik hosil qiluvchi frazeologik birlik deyish mumkin. Bu korpus tizimiga olib kirish uchun tuzilgan modellarimizda ham aks etadi, modellar tuzilishini aniq belgilab olish uchun ham avvaliga, iboralarning morfo-semantik kollokatsiyasini ham ko‘zdan kechirish lozim bo‘ladi:

T/r	Ibora nomi	Semantikasi	Bog‘lanishi	Namunalar
I	Yo‘l bo‘lsin?	Qayoqqa ketilayotganligini so‘rash?	- boshqa birliklarga bog‘lanishi kuzatilmaydi. Yaxlit jumla bo‘lib keladi.	*Ha, mehmon, yo‘l bo‘lsin? - dedi Ikromjon unga yaqin kelib.* S.Ahmad Ufq.
II	Yo‘l bo‘lsin	Biror ishni eplay olishga ishonmaslik, loyiq ko‘rmaslik	Shaxs oti/kishilik olmoshi + j.k. ~ yo‘l bo‘lsin	*Atlas hozir otliqqa yo‘q-ku, biz piyodalarga yo‘l bo‘lsin? S.Ahmad Ufq.*

Shunga ko‘ra, iboraning lingvistik modelini quyidagicha shakllantirish mumkin:

[(W_{pm} =$P_{(v)phr}$)] - bu yerda iboraning dastlabki ma’nosi bitta gapga teng bo‘lganligi uchun biz uni sodda gapning eng kichik qolipiga tenglashtirgan holda shakllantirdik. Ya’ni biz tizimga shu kabi iboralarni alohida kiritib, jumla tarkibida uchraganida tizim alohida tanib olish funksiyasini yaratishimiz lozim bo‘ladi. Buning uchun shu kabi iboralarning matn tarkibida har ikki tarafidan tinish belgilar bilan ajratilishini asos qilgan holda algoritm tuzib chiqish ko‘zda tutilgan.

Ikkinchi sema uchun esa quyidagicha model tuziladi:

PN/PP + Dat_CS_aff + $P_{(v)phr}$ - Bunda PN(personal noun) - shaxs oti, PP(personal pronoun) - kishilik olmoshi, Dat_CS_aff - jo‘nalish kelishigi va $P_{(v)phr}$ - fe’l ibora. Shu o‘rinda savol tug‘ilishi tabiiy. Fe’l ibora frazeologizmlarning katta qismini tashkil qiladi, biroq ularni bir xil belgi ostida biriktirish ularni ajratib olishda noaniqliklar tug‘dirmaydimi? Bizning tuzayotgan modellarimiz ayni paytda ishchi holatida, sinov jarayonida bo‘lib, yakuniy bosqichga tatbiq etilganda qisman o‘zgarishlarga uchrashi mumkin, shuningdek, fe’l iboralar bitta belgi ostida umumlashtirilgani bilan ularning so‘z va qo‘shimchalarni biriktira olish imkoniyati alohida tavsiflanyapti va shu holatda ham teng kelib qolish ehtimoli juda kam. Teng kelganda ham bitta modelga mos tushuvchi iboralar o‘sha model ostiga o‘zi anglatadigan ma’nolari bilan birga biriktirilishi inobatga olinmoqda. Shu vaqtda korpus tizimi iborani tashkil qiluvchi so‘zlar orqali ham unga tegishli modelni va u asosda ochiqlanadigan ma’noni bexato aniqlashi ko‘zda tutilgan.

Adabiyotlar ro‘yxati :

1. Rahmatullayev Sh.O‘zbek frazeologiyasining ba’zi masalalari. – Toshkent: Fan, 1966.
2. Rahmatullayev Sh. Hozirgi zamon o‘zbek tilidagi obrazli fe’l frazeologik birliklarning asosiy grammatik xususiyatlari. Nomzodlik dissertatsiyasi. – Moskva,1952.
3. Rahmatullayev Sh. Nutqimiz ko‘rki. –Toshkent,1970.
4. Rahmatullayev Sh. O‘zbek tilining izohli frazeologik lug‘ati. –Toshkent: O‘qituvchi,1978.
5. Rahmatullayev Sh., Mahmudov N., Xolmonova Z., O‘razova I., Rixsiyeva K. O‘zbek tili frazeologik lug‘ati. –Toshkent: G‘afur G‘ulom, 2022.
6. Gulyamova Sh. O‘zbek tili semantik analizatorining lingvistik asoslari. Filol.fanlari doktori diss. – Farg‘ona, 2022.
7. Raximova S. O‘zbek tilida frazeologik polisemiya va uning leksikografik talqini. Filol. fan.bo‘y.fal-sa dok-ri diss. – Guliston,2022.

8. Sayfullayeva R.R., Mengliyev B.R., Boqiyeva G.H., Qurbonova M.M., Yunusova Z.Q., Abuzalova M.Q. Hozirgi o‘zbek adabiy tili. o‘quv qo‘llanma. – Toshkent, 2006. – 391 b.
9. Xojiyev. A. Fe’l. – T.: Fan, 1973.191.b
10. Cho‘lpon. Kecha va kunduz. – Toshkent: Sharq, 2000.
11. Oybek. Qutlug‘ qon. – Toshkent: G‘afur G‘ulom,1969.
12. Hoshimov O‘. Ikki eshik orasi. – Toshkent: Yangi asr avlodi, 2016.

TURKCHA - O‘ZBEKCHA PARALLEL KORPUSI UCHUN IBORALARNING SEMANTIK TAHLILI

Fattaxova Diloram Abduraxmanovna
Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti, p.f.n.
E-mail: fattaxovadilorom@gmail.com

Turayeva Umida Toxirjon qizi
Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti magistranti.
E-mail: umidasng@gmail.com

Annotatsiya: Mazkur maqolada iboralar va parallel korpusi uchun iboralarning semantik tahlili haqida fikr yuritiladi. Turkcha iboralarning qator ekvivalent tarjimalari quyida berilgan bo‘lib, unda o‘zbek tilidagi tarjimasini ko‘rish mumkin. Frazeologik iboralar o‘z ma’nodoshlari bo‘lmish so‘zlarga nisbatan, ma’noni kuchli darajada ifodalaydi hamda ularda obrazlilikni yorqin aks ettiradi. Frazeologik iboralar turmushdagi turli voqea-hodisalarga guvoh bo‘lish, kishilarning xilma-xil harakat-holatlariga baho berish, tajribalarini umumlashtirish asosida xalq chiqargan aniq-tiniq xulosalarning o‘ziga xos obrazli ifodalaridir. Maqolada frazeologizmlar borasida ilmiy tadqiqot olib borgan tadqiqotchilar haqida ham fikr yuritilgan.

Kalit so‘zlar: frazeologik birlik, parallel korpus, ekvivalent tarjima, semantik ma’no.

Abstract: This article discusses the semantic analysis of phrases and phrases for a parallel corpus. A number of equivalent translations of Turkish expressions are given below, where you can see the Uzbek translation. Phraseological expressions strongly express the meaning and vividly reflect imagery in comparison to the words that have their meaning. Phraseological expressions are unique figurative expressions of clear conclusions drawn by people based on witnessing various events in life, evaluating various actions of people, summarizing their experiences. During the article, the scientific researchers who conducted scientific research on phraseologisms were also discussed.

Keywords: phraseological unit, parallel corpus, equivalent translation, semantic meaning.

Аннотация: В данной статье рассматривается семантический анализ фраз для параллельного корпуса. Ниже приведен ряд эквивалентных переводов турецких выражений, а также дан перевод на узбекский язык. Фразеологические выражения ярко выражают смысл и ярко отражают образность по сравнению со словосочетаниями, имеющими прямое значение. Фразеологические выражения – это своеобразные образные выражения и выводы, сделанные людьми на основе наблюдения за различными событиями в жизни, оценки различных действий людей, обобщения их опыта. В статьи также говорится об ученых-исследователях, проводившие научные исследования о фразеологизмах.

Ключевые слова: фразеологизм, параллельный корпус, эквивалентный перевод, смысловое значение.

Kirish

Hozirda o‘zbek va turk tillaridagi so‘zlarni, ayniqsa, iboralarni har turli jihatdan o‘rganishga e’tibor kuchayib bormoqda. Aslida, iboralar biror bir tushuncha yoki gapning bir so‘z bilan ifodalanishini hamda og‘zaki va yozma nutqda tilning bor go‘zalligini namoyon qiluvchi omil hisoblanib, ma’noni aniq tarzda yetkazib berishda katta ahamiyatga egadir. Bundan kelib chiqqan holda, ikki tilda ham o‘ziga xos ajoyib iboralar mavjud, ikki tilning ham ildizi uzviy bog‘liq bo‘lib, ba’zi so‘zlar va iboralar ham shakliy, ham ma’no jihatidan juda o‘xshashdir. Ushbu maqolamizda iboralarning ayni shu jihatiga, ya’ni ikki tildagi bir-biriga o‘xshash bo‘lgan iboralarga ba’zi misollarni keltirib, ularning semantik tahliliga e’tibor qaratmoqchimiz.

Adabiyotlar tahlili va metodologiya

Ibora – ikki va undan ortiq so‘zning barqaror munosabatidan tashkil topgan, ma’nosi, odatda, bir so‘zga, ba’zan so‘z birikmasi, gapga teng keladigan, nutqqa tayyor holda olib kiriluvchi, ko‘chma ma’noli barqaror (turg‘un) birikma. Iboralar adabiy tilning badiiy va so‘zlashuv uslubida keng va ko‘proq qo‘llanadi, nutqni ta’sirchan, obrazli ifodalashga xizmat qiladi. Iboralar tilshunoslikda frazema, frazeologizm, frazeologik birikma deb ham yuritiladi. Iboralarni o‘rganuvchi bo‘lim esa frazelogiya deb ataladi. Turkcha va o‘zbekcha parallel korpusining semantik tahlili uchun kompyuterli tillar olami (NLP) foydalaniladi. Ushbu tillar matnlarni avtomatik ravishda tahlil qilish va ularning ma’nosini aniqlash imkonini beradi. Semantik tahlil, so‘zlarning ma’nosini, ma’nosi o‘rtasidagi bog‘lanishlarni tushuntirib beradi. Bu tahlil asosida, tarjimaviy dasturlar, so‘zlar orasidagi bog‘lanishlarni aniqlash, so‘z ma’nolari va matnlar bilan ishlash osonlashadi. Semantik tahlil asosan matnlarni, so‘zlarni va ularning ma’nolarini tahlil qilishni o‘z ichiga oladi. Bu tahlil jarayoni quyidagilardan iborat bo‘lishi mumkin:

1. Tavsif: Bu usulda iboralar ma’noga oid tavsiflar yordamida tahlil qilinadi. Odatda, iboralar haqida xususiyatlar, vazifalar, obyektlar va turli xil ma’nolarni ko‘rsatuvchi ma’lumotlar tushuniladi.

2. Taqsimlash: Bu usulda iboralar cheklanib, ularning o‘xshash yoki o‘xshash bo‘lmagan ma’no turiga qanchalik mos kelishadi, shuningdek, ularning boshqa o‘zbekcha iboralar bilan bog‘liq bo‘lishi aniqlanadi.

3. Sinonimlar va antonimlar: Bu usulda iboralar o‘zbek tilidagi sinonim va antonimlari bilan taqqoslanadi. Bunday tahlil orqali, bir ibora bo‘lgan ma’no boshqa iboralar orqali ifodalangan bo‘lishi, yoki ibora bilan o‘xshash ma’no yoki uning antonimi bo‘lishi aniqlanadi.

4. So‘zlararo munosabatlar: Bu usulda iboralar o‘zbek tilidagi boshqa so‘zlar bilan munosabati o‘rganiladi. Mazkur so‘zlar iboradagi ma’noni to‘liq aniqlashda yordam beradi.

5. Kontekstual tahlil: Bu usulda iboralar o‘zbek tilidagi iboralarni muvofiqligi bo‘yicha o‘rganiladi. Bunday tahlil jarayoni orqali iboralar o‘z ma’nolarini ham o‘rganishga imkon yaratiladi.

Bu tahlil usullari tufayli turkcha-o‘zbekcha parallel korpusi uchun iboralar semantik tahlili aniqroq amalga oshirilishi mumkin va tarjima texnologiyasining rivojlanishiga yordam beradi.

Frazeologizmlarni semantik xarakteriga ko'ra quyidagicha ajratish mumkin:

• Fe’lli iboralar
• Predlogli iboralar
• Gastronomik iboralar

• "Non" so‘ziga oid iboralar
• "Suv" so‘ziga oid iboralar
• Son komponentli iboralar
• Tana a’zolariga oid iboralar
• Hayvon nomlariga oid iboralar
• Rang - tus nomlariga oid iboralar
• Vaqt so‘ziga oid iboralar
• Oila mavzusiga oid iboralar
• Yer so‘ziga oid iboralar
• Oy so‘ziga oid iboralar
• So‘z va so‘zlashga oid iboralar
• Tinchlik so‘ziga oid iboralar kabi [B. Xudoyqulov. 2022].

Og‘zaki nutqda turkcha iboralardan foydalanish, ayniqsa, turk filologiyasi bo‘limlarida tahsil olayotgan chet ellik talabalar uchun til ta’limining eng qiyin bosqichini tashkil etadi. Buni osonlashtirish uchun, doimgidek, iboralarni shakli va ma’nosiga ko‘ra tasniflashdan tashqari, ularni mavzulariga ko‘ra ham tasniflash zarur. Masalan:

1.Odam tanasiga oid iboralar (qo‘l, miya, bosh, ko‘z, yurak kabi so‘zlardan tuzilgan iboralar);

2. insoniy munosabatlarga oid iboralar (sevgi, hurmat, nafrat, do‘stlik, adovat, or-nomus, g‘azab va boshqalar haqidagi iboralar);

3. hayvonot olamiga oid iboralar (hayvonlarning nomlarini, vujudini va harakatlarini ifodalovchi so‘zlar bilan tuzilgan iboralar);

4. Fe’lli iboralar (ish-harakat va jarayonni ifodalovchi so‘zlardan tuzilgan iboralar) va hokazo.

Ma’lumki, turkiy iboralarining ko‘pchiligi nutqqa oid iboralardan iboratdir. Bu iboralar, odatda, og‘iz, til, so‘z kabi asosiy so‘zlar ishtirokida hosil bo‘lib, so‘zlash, aytmoq, gapirish ma’nolarida qo‘llanadi; Og‘iz juftlamoq, og‘iz ochtirmaslik (turk. “Ağız açmak”, “ağız açtırmamak”.), tilga kelmoq (turk. “dile gelmek”), tili kalimaga kelmaslik (turk. "dili tutulmak".), gapga aralashmoq (turk. “söze karışmak”) va hokazo[H. Hamidov.2006:40]. Quyida ikki tildagi og‘iz, til, bosh, so‘z va fe’lli iboralarga oid muqobil ma’nodagilariga misollar keltiramiz:

Turkcha iboralar	O‘zbekcha iboralar
Ağzından dökülmek, ağzından laf kaçırmak	Og‘zidan chiqib ketmoq
Ağzı açık dinlemek, ağzı bir karış açık kalmak	Og‘zi lang ochilmoq
Ağzı gevşek	Og‘zi bo‘sh
Ağzından düşürmemek	Tilidan tushirmaslik
Ağza almamak	Og‘izga olmaslik
Ağzından (laf) kaçırmak	Og‘zidan gullash
Açtı ağzını, yumdu gözünü	Og‘ziga kelganini gapirdi
Dilini tutmak	Tilini tiymoq
Dili dönmemek	Tili aylanmaslik, tili kalimaga kelmaslik
Dili uzun, dilin kemiği yok	Tili uzun, tilning suyagi yo‘q
Sözünü kısa kesmek	Ovozini o‘chirmoq
Gönlünü açmak gibi	Ko‘nglini ochmoq
Açık gönüllü,	Ochiq ko‘ngil

Ağzından çıkanı kulağı duymamak	Og‘zidan chiqqanini qulog‘i eshitmaslik
Baş kaldırmak	Bosh ko‘tarmoq
Gözü Kör Olmak	Ko‘zi ko‘r, qulog‘i kar bo‘lmoq
Kulak Asmak	Quloq osmoq, quloq tashlab qo‘ymoq
El ele vermek:	Qo‘lni qo‘lga bermoq
Kulağını dikmek	Qulog‘ini ding qilmoq
Başı derde girmek	Boshi baloga kirmoq
Bir yastığa baş koymak	Bir yostiqqa bosh qo‘ymoq

Bunga o‘xshash misollar talaygina bo‘lib, bular orqali ikki tilning ma’naviy xazinasining naqadar boyligi, o‘xshashligi va uzviy bog‘liqligini namoyon qilishga harakat qilindi.

Iboralar taraqqiyoti va uning tarixiga nazar solsak. Tilshunoslikda frazeologiyaning tadqiqot obyektlarini turlicha tushunish va tavsiflash kuzatiladi. N.N.Amosova, A.M.Babkin, V.P.Jukov va boshqalar so‘z bilan o‘zaro bog‘lanishli munosabatda bo‘lgan so‘z birikmalarining semantik o‘zgarishini; A.V.Kunin, I.I. Chernisheva, A.D.Rayxshteyn va boshqalar ham so‘z ham gap bilan o‘zaro bog‘lanishli munosabatda bo‘lgan so‘z birikmalarining semantik o‘zgarishini; V.L.Arxangelskiy, N.M.Shanskiy va boshqalar yaxlit nominatsiyaga ega bo‘lgan semantik o‘zgarishga uchragan yoki uchramagan har qanday struktur-sintaktik tipli so‘z birikmalarini; M.M. Kopilenko, Z.D. Popova leksemalar birikuvini; V.N.Teliya o‘zgarmas leksik tarkibli har qanday barqaror so‘z birikmalarini frazeologiyaning obyekti sifatida hisoblashadi. Turkiy frazeologiyaning boshlovchilari turkiyshunos olimlar S.K.Kenesboyev va Sh.U.Rahmatullayevlar hisoblanishadi. Ularning o‘tgan asrning 40-yillarining ikkinchi yarmi va 50-yillarda yuzaga kelgan tadqiqotlari turkiy tillar frazeologiyasining shakllanishi va taraqqiyotida katta rol o‘ynaydi. Akademik S.K.Kenesboyevning fikricha, frazeologiya doirasiga keng ma’noda barcha turg‘un birikmalar (maqollar, matallar, iboraviy birikmalar, turg‘un ideomatik guruhlar) kiradi. Ularni birlashtiruvchi umumiy xususiyat - turg‘unlik va tilda tayyor holda mavjudlikdir. Shundan keyingi 30-40 yil mobaynida turkiy tillardagi frazemalarni intensiv o‘rganish bo‘yicha muhim yutuqlar qo‘lga kiritildi. Bu davrda S.N.Muratovning “Turkiy tillarda turg‘un so‘z birikmalari”, Sh.U.Rahmatullayevning “O‘zbek frazeologiyasining ayrim masalalari” (1966) kabi ishlari yaratildi Ozarbayjon tilshunoslari B.Cho‘ponzoda va F.Og‘azoda “Turk tili grammatikasi” asarida til bo‘limlari haqida fikr yuritib, “Semasiologiya”, “Stilistika” atamalari bilan birga o‘sha davrda tilshunoslikda keng qo‘llanilishda bo‘lgan “Idiomatizm” atamasini qo‘llashgan. Mualliflar mazkur asarda turk tilidagi iboralarga boshga solmoq, boshga tushmoq, ko‘z ko‘rmoq kabi birliklarni misol sifatida keltiradilar. Umuman, frazeologiya turkiyshunoslikda nisbatan keyin paydo bo‘lgan. Uni tizimli ravishda o‘rganish faqat o‘tgan asrning 40-50-yillariga kelib boshlandi [Y. Nurmuhammedov.2022].

Xulosa

Xulosa qilib aytganda, har bir til o‘z doirasidagi so‘zlarni o‘rnida qo‘llash, ma’no kengayish usullarini o‘rganish, o‘ziga xos bo‘lgan xususiyatlarini dunyoning tildagi ko‘rinishi orqali aks ettirishda muhim rol o‘ynaydi. Shuningdek, bugungi kunda o‘zbek tilida qo‘llanilayotgan frazeologik birlik atamasi bilan turk tilida qo‘llanilayotgan frazeologik birlik atamasi ham umumiylikni namoyon qilmoqda. Bunga ko‘ra ham, turk lahjalaridagi frazeologik taraqqiyot parallellik va turdosh lahjalarning aynan bir xil atamalar ishlatilmasa ham, bir doira ichida rivojlanayotgan uyg‘unligini ko‘rsatayotganini ko‘rish mumkin. Turk tilining bu ikki

lahjasidagi ibora va frazeologiyaga oid masalalar yuzasidan turli tadqiqotlar olib borilgan bo‘lsa-da, ma’lumki, hali o‘rganilishi kerak bo‘lgan masalalar mavjud. Masalan, barcha turkiy shevalardan frazeologlar bir joyga to‘planib, turkiy iboralarni morfologiya va semantika nuqtai nazaridan tekshirib, ta’rif ma’nosida aniq chegaralar bilan umumiy tavsif yaratishlari mumkin [B. Yilmaz.2021].

Adabiyotlar ro‘yxati:

1. N. Abduraxmonova. O‘zbek tili elektron korpusining kompyuter modellari. (Monografiya). GlobeEdit, 2021.
2. B. Xudoyqulov. Frazeologizmlarning asosiy xususiyatlari.Uchinchi renessans: ilm-fan va ta'lim taraqqiyoti istiqbollari konferensiya materiallari, 2021. 312 b.
3. B. Xudoyqulov. Frazeologik birliklarning lingvomadaniy xususiyatlari. Oriental renaissance: Innovative, educational, natural and social sciences, 1(11): 2021. 948 b.
4. G‘. Salomov. Tarjima nazariyasiga kirish. - T.: O'qituvchi, - 1978. - 291b
5. H. Hamidov. Türkçede İnsan Vücuduyla İlgili Deyimler, Türkçemin Sesi, Dergi, Bahar, 2003, s.4-6
6. H. Hamidov. Türkçede İnsan İlişkileriyle İlgili Deyimler.Türk Dili, Dil ve Edebiyat Dergisi, sayı:610, Ekim 2002, s. 831-835.
7. B. Xudoyqulov, M. Jumayeva. Frazeologik iboralarning tematik va semantik tahlili. Academic research in Educational Sciences: Volume 3. 2/2022.
8. H. Hamidov. Türkçe deyimlerin anlam özellikleri: Konuşma ile ilgili deyimler. Türk Dünyası Dil ve Edebiyat Dergisi 22. Sayı. s. 40-42.
9. Kanmaz, Aslıhan. Fransızcanın Yabancı Dil Olarak Öğretiminde Red Sözedimi ve Kültürlerarası Nezaket”, Turkish Studies – İnternational Periodikal for The Languages, Literature and History of Turkish or Turkic Volume 8/10 Fall 2013, p. 343-353, Ankara-Turkey.
10. Türkçe deyimler yoki Turk tilidagi iboralar https://forum.ziyouz.com/index.php?topic=4657.0
11. https://tdk.gov.tr
12. https://turkoloji.cu.edu.tr/pdf/ozbekce_bildiri.pdf

MUALLIFLIK KORPUSINING AHAMIYATI

Muradova Nilufar Baxodir qizi
ToshDO‘TAU Kompyuter lingvistikasi
mutaxassisligi 1-kurs magistranti.
E-mail: nilufarmurodova818@gmail.com

Annotatsiya: Bu maqolada mualliflik korpusining nima ekanligi, undan qanday maqsadlarda foydalanishi, uning ilmiy va amaliy muhimligi haqida batafsil ma’lumot berilgan. Mualliflik korpusi mualliflarning asarlarini jamlash orqali yaratilgan ma’lumotlar to‘plamidir. Ushbu ma’lumotlar to‘plami muallifning tildan foydalanish, yozuvchi uslubi, so‘z tanlash mahorati kabi xususiyatlarini o‘rganish va tahlil qilish uchun foydalaniladi. Aniq misollar orqali, mualliflik korpusining tilshunoslik sohasi va boshqa sohalardagi ahamiyati yoritildi.

Kalit so‘zlar: korpus, mualliflik korpusi, lingvistik hodisalar, badiiy o‘ziga xoslik, ta’limiy maqsad.

Annotation: This article provides detailed information about what the author's corpus is, for what purposes it is used, and its scientific and practical importance. An authorship corpus is a collection of data created by compiling the works of authors. This data set is used to study and analyze the characteristics of the author, such as language use, writing style, and word choice skills. Through specific examples, the importance of the author corpus in the field of linguistics and other fields was highlighted.

Keywords: corpus, author's corpus, linguistic phenomena, artistic originality, educational purpose.

Аннотация: В данной статье представлена подробная информация о том, что представляет собой авторский корпус, для каких целей он используется, а также о его научном и практическом значении. Авторский корпус — это совокупность данных, созданная путем компиляции произведений авторов. Этот набор данных используется для изучения и анализа характеристик автора, таких как использование языка, стиль письма и навыки выбора слов. На конкретных примерах была подчеркнута важность авторского корпуса в области языкознания и других областях.

Ключевые слова: корпус, авторский корпус, языковые явления, художественное своеобразие, образовательная цель.

Korpus lingvistikasi jahon kompyuter lingvistikasining juda tez rivojlanib ketayotgan sohasi bo‘lib, bu borada ancha yutuqlarga erishilgan. Dunyo oliy ta’lim muassasalarida korpus lingvistikasi fan sifatida ham o‘qitiladi. Bu sohaning predmeti korpus yaratish nazariyasi va amaliyoti bo‘lsa, fan sifatida korpusning o‘ziga xosligi, dasturlash asoslari kabi jihatlari o‘qitiladi. Korpus lingvistikasi kompyuter lingvistikasining tarkibiy qismi, til korpusini yaratish, kompyuter texnologiyasi yordamida ulardan foydalanishning umumiy nazariyasi va amaliyoti bilan shug‘ullanadi [Захаров В. 2005:27].

Mualliflik korpusi - mualliflarning asarlarini jamlash orqali yaratilgan ma’lumotlar to‘plami. Ushbu ma’lumotlar to‘plami muallifning tildan foydalanish, yozuvchi uslubi, so‘z

tanlash mahorati kabi xususiyatlarini o‘rganish va tahlil qilish uchun foydalaniladi. Ular odatda tilshunoslik tadqiqotlari, yozma matn tahlili, tilni qayta ishlash tadqiqotlarda qo‘llaniladi. Ushbu korpuslar odatda har xil turdagi matnlarni o‘z ichiga oladi, masalan, romanlar, maqolalar, she’rlar va boshqalar. Mualliflik korpusi (author corpus) – alohida shaxs ijodiga, mualliflik asarlariga bag‘ishlanuvchi korpus turi. Unda muayyan ijodkorga tegishli asarlargina joy oladi. Mualliflik korpusi bir qancha sohalar uchun ahamiyatli. Jumladan, tilshunoslik, adabiyot va ta’lim sohasida keng foydalanish mumkin.

Mualliflik korpusi bir qator sohalar uchun foydali bo‘lishi mumkin:

1. *Lingvistik ahamiyati:* Tilshunoslar muallifning tildan foydalanishi bo‘yicha tadqiqot o‘tkazishi mumkin. Masalan, ular ma’lum bir davrdagi til o‘zgarishlarini yoki muallifning ijodining o‘ziga xos til xususiyatlarini haqida aniq xulosaga kelishi mumkin.
2. *Badiiy ahamiyati:* Adabiyotshunoslar muallifning ijod uslubini, mavzusini yoki muallif davrining adabiy xususiyatlarini tahlil qilish uchun mualliflik korpusidan foydalanishlari mumkin.
3. *Ta’limiy ahamiyati:* Mualliflik korpusi o‘quvchilarga ma’lum yozuvchi asarlarini o‘rganish va ularning uslublarini tushunishga yordam beradi. Shu bilan birga yozuvchi asarlarining bir saytda yig‘ilgan tayyor elektron kutubxona vazifasini ham bajaradi.
4. *Milliy korpuslar uchun ma’lumotlar bazasi:* Sun’iy intellekt va kompyuterlashtirilgan tilni qayta ishlash sohalarida mashina o‘rganish algoritmlari ushbu korpuslardan katta hajmdagi matn ma’lumotlaridan foydalangan holda til modellarini o‘rgatish uchun foydalanishi mumkin. Ma’lum til korpuslari va milliy korpuslar uchun ham tayyor ma’lumotlar bazasi sifatida foydalanish imkonini beradi. Bu avtomatik matn yaratish va tarjima qilish kabi ilovalarda muhim ahamiyatga ega.
5. *Madaniy xulosalar:* Mualliflik korpusidan ma’lum bir davr yoki madaniyat yozuvlarini o‘rganish orqali o‘sha davr yoki madaniyatning til va adabiy xususiyatlarini tushunish uchun foydalanish mumkin.

Lingvistik ahamiyati

Mualliflik korpusida muallif tilini to‘liq, keng, obyektiv ko‘rsatish imkoniyati bor, shuning uchun bunday korpuslar boshqa axborot bankidan afzalligo bilan ajralib turadi. Bunday korpuslarning afzalligi shundaki, ular yordamida nafaqat biron bir so‘z yoki jumlaning, balki butun boshli asar tilini bilish imkoniyati mavjud [Rahimov A., 2011: 156]. Mualliflik korpuslari tilshunoslikda quyidagi maqsadlarda foydalaniladi.

Tilning o‘zgarishi va rivojlanishi: Til korpusi tilning ma’lum bir davr ichida qanday o‘zgarishi va rivojlanishini kuzatish uchun foydalanish mumkin. Yozuvchilarning turli davrlarda qo‘llagan til xususiyatlarini o‘rganib, til o‘zgarishi haqida xulosa chiqarish mumkin.

So‘zdan foydalanish va chastotasi: Yozuvchi ijodiga xos bo‘lgan so‘zlarini tahlil qilish tilshunoslarga ma’lum so‘zlarning foydalanish chastotasi va semantik o‘zgarishlarini kuzatish imkonini beradi. Bu til taraqqiyoti uchun muhim, chunki ba’zi so‘zlar ma’lum davrda ko‘p istemolda bo‘ladi va ma’lum davrda eskiradi va arxaik yoki tarixiy so‘zga aylanib boradi.

Stilistik tahlillar: Mualliflik korpusi mualliflarning uslublarini va ma’no ko‘chish usulini tekshirish uchun ham ahamiyatlidir. Metafora, kinoya, o‘xshatish kabi stilistik xususiyatlarni tahlil qilish orqali muallifning niyatini, u yetkazmoqchi bo‘lgan fikrni tushunish uchun muhim omil bo‘lib xizmat qiladi.

Dialekt va tilshunoslik: Mualliflik korpusi turli geografik mintaqalar yoki ijtimoiy guruhlarga mansub yozuvchilarning tildan foydalanishini taqqoslash orqali dialekt va til xilma-xilligini o‘rganish uchun foydalanish mumkin.Bu fikrlar lingvistik nuqtai nazardan mualliflik korpusining bir nechta misolidir. Mualliflik korpusi tilning ko‘p jihatlarini o‘rganish va tushunish uchun qimmatli manba bo‘lib xizmat qiladi.

Zahiriddin Muhammad Bobur mualliflik korpusidan ijodkor so‘z birliklarining lingvistik xususiyatlarini, aynan o‘sha davrda istemolda bo‘lgan so‘zlar lug‘atini, gap qurilishi va so‘z tarkibi kabi lingvistik hodisalarni kuzatish va tahlil qilish imkoniyatini beradi.

Badiiy ahamiyati

Mualliflik korpusi adabiyot sohasida bir qator muhim maqsadlarga ega:

Muallif ijodini tahlil qilish va asarlarni taqqoslash: Mualliflik korpusi muayyan mualliflarning asarlarini tahlil qilish va turli asarlar orasidagi o‘xshashlik yoki farqlarni aniqlash uchun foydalanish mumkin. Bu muallifning uslubini, mavzularini, xarakter rivojlanishini va hatto vaqt o‘tishi bilan muallif ijodining evolyutsiyasini o‘rganish uchun muhimdir.

Tarixiy va madaniy aloqalarni o‘rganish: Mualliflik korpusidan ma'lum bir davrning adabiy asarlarini birlashtirib, o'sha davrning tarixiy va madaniy holatini o‘rganish uchun foydalanish mumkin. Bu esa muayyan davrdagi adabiy oqimlar, ijtimoiy-siyosiy voqealar oqibati, madaniy o‘zgarishlarni o‘rganish uchun qimmatli manba bo‘lib xizmat qiladi.

Adabiy tanqid va tahlil: Adabiyotshunoslar mualliflik korpusidan foydalangan holda ma’lum bir muallifning asarlarini chuqur o‘rganishlari mumkin. Bu tahlillar asarlarning mavzui, uslubi, ramziyligi, hikoya qilish texnikasi va boshqa adabiy xususiyatlarini baholash imkonini beradi. Mualliflik korpusi adabiyot sohasida chuqur izlanishlar olib borish va adabiy asarlarni grammatika, uslub va mazmun jihatidan tekshirish imkonini beradi. Shuning uchun mualliflik korpusi adabiyotshunoslikda muhim manba sanaladi. Badiiy matnlarning statistic tahlili natijasida matnda tez-tez qo‘llaniluvchi til birliklari (otlar, sifatlar, kalit so‘zlar, fe’llar, grammatik shakllar, jumla qurilishi, bir so‘z bilan aytganda, yozuvchining uslubi- o‘ziga xos yozish uslubini ko‘rsatuvchi vositalar) lingvostatistik tahlil yordamida aniqlanadi. Turli matnlardan olingan dalillarning qiyosiy tahlili bizga o‘sha matnning mazmun-mundarijasini, matn tuzilgan davrni, dalillarni argumentlash xarakterini va hatto muallifni aniqlashga imkon beradi [Rahimov A., 2011: 154]. Adabiyotshunoslar muallifning ijod uslubini, mavzusini yoki muallif davrining adabiy xususiyatlarini tahlil qilish uchun mualliflik korpusidan foydalanishlari mumkin.

Ta’lımıy ahamıyatı

Mualliflik korpusining nisbatan ta’lim sohasidagi ahamiyati ancha yuqori. Negaki maktab o‘quvchilari uchun mavjud korpus tayyor elektron kutubxona vazifasini bajaradi. Shuningdek, ta’lim jarayonida muhim manba sanaladi. O‘quvchi ma’lum bir yozuvchi asarlari bilan tanishni maqsad qilgan bo‘lsa mualliflik korpusidan foydalanish uning vaqtini tejaydi va muallifning ijodiy mersini bilan yaxlit tanishish va foydalanish imkoniyatini beradi. Maktab darsligida Zahiriddin Muhammad Bobur hayoti va ijodi bo‘yicha 2023-2024-o‘quv yili uchun jami 5 soat ajratilgan (1-jadval). Ko‘rstilgan dars soatlarida ham o‘qituvchi, ham o‘quvchi uchun manba bo‘lib xizmat qiladi.

	Dars mavzusi	Sinf	Soat
1	Boburning hayoti va ijodi. Ruboiylari.	5	1
2	Bobur ruboiylarini o‘qish, tahlil qilish	5	1
3	Bobur lirikasi. G‘azallari	11	2
4	Bobur she’riyatida badiiy san’atlar	11	1

1-jadval

Bu maqolada bir nechta sohalarda mualliflik korpusining ahamiyatini o‘rganildi, xalos. Bundan tashqari mualliflik korpusi turli maqsadlarda, jumladan, til va yozuvning turli sohalarida tadqiqot va tahlil qilish uchun qimmatlidir. Ta’kidlash joizki, mualliflik korpuslarining ahamiyati-keng qamrovli bo‘lib, undan bir qancha soha vakillari hamda tarjima dasturlarini ishlab chiqishda foydalanish mumkin. Xulosa qilib aytganda, maqola adabiyotshunoslik va fanlararo tadqiqot ishlarida mualliflik korpusining ahamiyati muhimligini ta’kidlaydi.

Foydalanilgan adabiyotlar ro‘yxati:

1. Ahmedov S., Qosimov B., Qochqorov R., Rizayev Sh. Adabiyot (Darslik). – Toshkent: Sharq, 2020.
2. Hamroyeva Sh. O‘zbek tili mualliflik korpusini tuzishning lingvistik asoslari: Monografiya. Globe edits, 2020.
3. Rahimov A., Kompyuter lingvistikasi asoslari. –Toshkent: Akademnashr, 2011.
4. To‘xliyev B., Karimov B., Usmonova K. Adabiyot (darslik), Toshkent: Davlat ilmiy nashriyoti, 2018.
5. Tursunov U., Muxtorov A., Rahmatullayev Sh. Hozirgi o‘zbek adabiy tili. –Toshkent: O‘zbekiston, 1992.
6. Xolmanova Z. Kompyuter lingvistikasi. – Toshkent, 2020.

KORPUSDAN QIDIRUV TIZIMI SIFATIDA FOYDALANISH

O‘roqova Dilnoza Odil qizi
Mirzo Ulug‘bek nomidagi O‘zbekiston
Milliy universiteti 2-bosqich magistranti.
E-mail: adilovnadilnoza09@gmail.com

Annotatsiya: Mazkur maqola korpus tushunchasi, uning tildagi o‘rni, korpusning qidiruv turlari, ishlash prinspi va ularni qidirish, o‘rganishga bo‘lgan turlicha yondashuvlarni o‘z ichiga oladi.

Kalit so‘zlar: korpus, konteks, lug’at, intertekstuallik, yashirin modellar, tillarni taqqoslash, metama’lumot, matn razmetkasi va lingvistik annotatsiya

Abstract: This article includes the concept of the corpus, its place in the language, types of corpus searches, the principle of operation and different approaches to their search and study.

Keywords: corpus, context, vocabulary, intertextuality, implicit models, language comparison, metadata, text markup and linguistic annotation

Аннотация: В данной статье представлены понятие корпуса, его место в языке, виды корпусных поисков, принцип работы и различные подходы к их поиску и изучению.

Ключевые слова: корпус, контекст, словарный запас, интертекстуальность, неявные модели, языковое сравнение, метаданные, текстовая разметка и лингвистическая аннотация.

Dunyo tilshunosligida tilning imkoniyatlarini kengroq o‘rganish, til grammatikasining muammoli tomonlarini kontekstda aniqlash, tilda grammatik qoliplarni belgilash, ko‘p tarmoqli elektron lug‘atlar yaratish ishini yengillashtirish, tilni o‘rganishda zamonaviy axborot texnologiyalardan foydalanish samaradorligini oshirish, tilda avtomatik tarjima, qidiruv va komyuter tahlilini yo‘lga qo‘yish, elektron darsliklar va lug’atlar tayyorlash kabi masalalarini hal qilish uchun tillarda korpuslar yaratishning nazariy va amaliy asoslarini ishlab chiqish, tilning maxsus sohalar bo‘yicha korpusini qurish zaruratining mavjudligi maqolamizning dolzarbligini belgilaydi. Korpus matnlaming katta to‘plami bo‘lishi bilan birga, bu lingvistik tahlilga asoslanadigan yozma yoki og‘zaki materiallar asosi hamdir. Korpusda:

1) aniq bir yozuvchining yoki yozuvchilarning matnlari;
2) aniq o‘n yil yoki asrga oid matnlar;
3) muayyan mavzudagi zamonaviy matnlar;
4) til yoki jamiyatda yetarlicha mavjud bo‘lgan zamonaviy matnlardan foydalaniladi.

Yuqorida belgilangan matn turlaridan korpus shakllantirilganda quyidagilami qidirish mumkin:

1) kontekstdagi so‘zlaming barcha shakllari;
2) lug‘atdagi o‘zgarish va izchillik;
3) eng ko‘p tanlangan so‘z birikib kela oladigan so‘zlar;
4) ikkita matnlar o‘rtasidagi eng muhim farqlar;

5) muayyan yozuvchining so‘z va iboralami qo‘llashdagi o‘ziga xoslik;
6) intertekstuallik: so‘zning ma’nosi uning qo‘llanilish yig‘indisi sifatida;
7) so‘z birliklaridan foydalanadigan yashirin modellar;
8) tillarni taqqoslash.

Korpusning qidiruv tizimi va uning ishlash prinsipi 3 turdagi asosiy qismni o‘z ichiga oladi, bu korpusdagi ma’lumotlar bazasini tekshirishga yordam beradi. Bular: metama’lumot, matn razmetkasi va lingvistik annotatsiya (izohlash). Metama’lumotlar - matn kim tomonidan yozilgani, qachon nashr qilingani, qaysi tilda yozilgani to‘g‘risida ba’zi ma’lumotlami beradi. Metama’lumot korpus matnida kodlanishi yoki alohida hujjat sifatida ma’lumotlar bazasida saqlanishi mumkin. Matn razmetkasi yoki matnni belgilash matnni formatlashni ifodalash uchun ishlatiladi masalan, jumlaning boshlanishi va tugash nuqtasini belgilashda muhim hisoblanadi. Metama’lumot matndagi ma’ruzachilarni aniqlab beradi va har birining yoshi, jinsi haqida foydali ma’lumot beradi. Matn razmetkasi keyinchalik har bir ma’ruzachi gapirishni boshlaganini va tugatganligini ko‘rsatish uchun ishlatiladi. Korpusga kiritilgan metama’lumot va matn razmetkasi birgalikda bir qator tadqiqot savollariga javob beradi. Matn korpusi ichida lingvistik ma’lumotlami kodlashimiz mumkin, keyinchalik bu tahlilni muntazam va aniq tarzda olib boriladi, bu holatda korpus analitik yoki lingvistik nuqtayi nazar bilan izohlanadi. Annotatsiya kodlash qoidalaridan matn razmetkasi sifatida foydalanadi, masalan, XML (extensible markup language) qirrali-briket teglari ishlatilishi mumkin, bu yerda otli fraza boshlanadigan nuqtada (<np>) shakli va tugash nuqtasida (</np>) shakli ishlatiladi. Masalan, *np O‘quvchi </np> <np> doskaga </np> chiqdi* (<np> *The children* </np> *sat in* <np> *the room* </np>)[Abdurakhmonova, N. 2019,2021,2022].

Korpus lingvistikasida korpuslar foydalanuvchiga tez va ishonchli qidirishga imkon beradigan dasturlami qo‘llashda o‘zgarmasdir. Bu dasturlarning ba’zilari, chunonchi, konkordanserlar foydalanuvchiga matndan so‘zlarni qidirishga imkon beradi. Shunga o‘xshash dasturlarning ko‘pchiligi matnlardagi ma’lumotlar chastotasini yaratishga imkon beradi. Masalan, korpusda paydo bo‘ladigan barcha so‘zlarning ro‘yxati va har bir so‘zning maxsus o‘sha korpusda necha marta uchrashini ko‘rsatadigan so‘zlarning chastotali ro‘yxatini yaratadi. Konkordanser va chastota faktlari tahlilning ikki muhim tomoni hisoblanadi, ya’ni sifat va miqdoriy jihati sifatida korpus lingvistikasining muhim qismlaridir.

Korpus qidiruv tizimidagi eng muhim saytlardan biri *uzbekcorpus.uz*.O‘zbek tili korpusi uchun lemme va token asosiy qidiruv birligi sanaladi. Lekin korpusning qidiruv menejeri ham token, ham so‘z birikmasi va gap qidiruviga mo‘ljallangan. Lingvistik ma’lumotlar bazasida o‘zbek tilining morfologik bazasi bo‘lib, barcha qidiruv algaritmi uchun asos vazifasini bajaradi. U orqali korpusda mavjud so‘zlarning statistikasini chiqarish imkoniyati tug‘iladi. Shuningdek, korpusning qidiruv tizimi orqali ixtiyoriy so‘zning n-gram bo‘yicha qo‘llanilishi natijasida so‘zlarning konkordanslarni hosil qilish mumkin. Lemma, so‘zning lug‘atdagi shakli bo‘lib, aksariyat tillar uchun yasalish va grammatik jihatdan shakllanish materiali bo‘la oladi. O‘zbek tilining turkiy tillarga xos xususiyatlardan biri morfem birliklarini kombinatsion shakllanishi bo‘yicha natijaga erishildi[Abdurakhmonova, N. 2019,2021,2022].

O‘zbek tili elektron korpusning funksiyalari quyidagidan iborat:

1) Bosh sahifa
2) Izlash (lemma, token, konkordans bo‘yicha)
3) Morfoanalizator

4) Subkorpuslar (ta’limiy, parallel korpus)
5) Foydalanuvchilar yo‘riqnomasi
6) Elektron lug‘atlar
7) Tezaurus

Korpusning qidiruv tizimi uch asosiy birlik bo‘yicha funksional vazifa bajaradi



1) Matnlarni uslub bo‘yicha qidirish
2) Matnlarni xronologik ya’ni davr tartibida qidirish
3) Konkordanslar uchun n-gram modeli bo‘yicha o‘ng va chap tarafdan qidirish

Ushbu electron korpus statistic ma’lumotlarni quyidagi jihatlarga ko‘ra chiqarish imkoniyatiga ega:

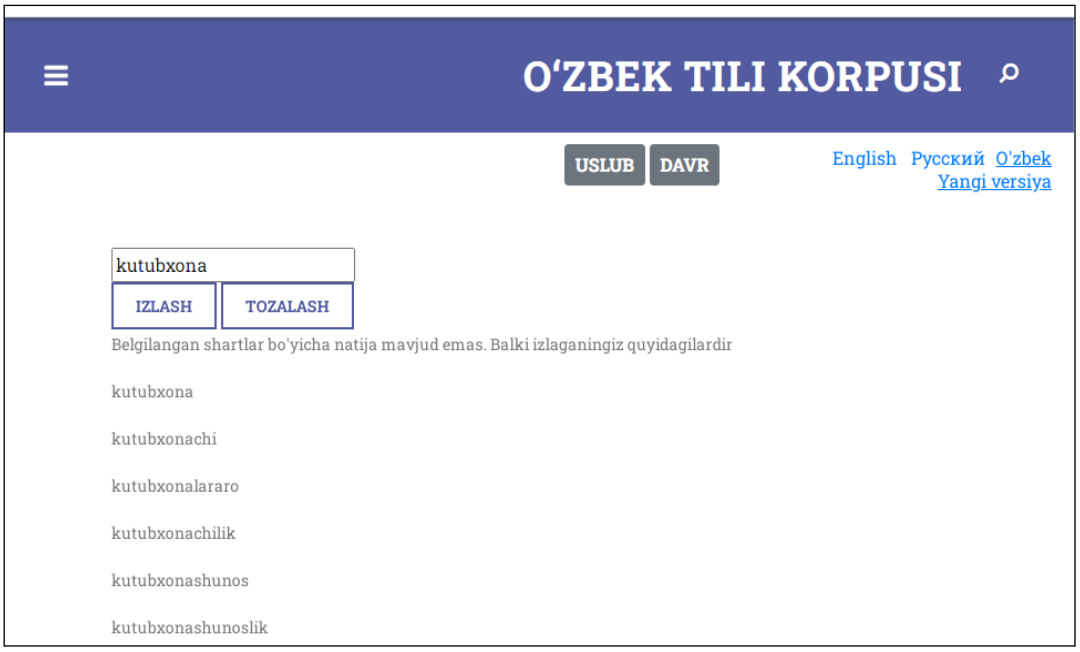
1) lemma bo‘yicha qidirish – masalan, kitob so‘zi qidirilganda, ushbu so‘z o‘zak holatda matn tarkibida aniqlanadi



Lemma bo‘yicha qidirishda so‘zlarni lug‘at shaklida 0 grammatik ko‘rsatkichda yoki aynan grammatik qo‘shimchalar bilan qidiruvda so‘rovga berilgan shakliga qarab natija uzatiladi[Abdurakhmonova, N. 2019,2021,2022].

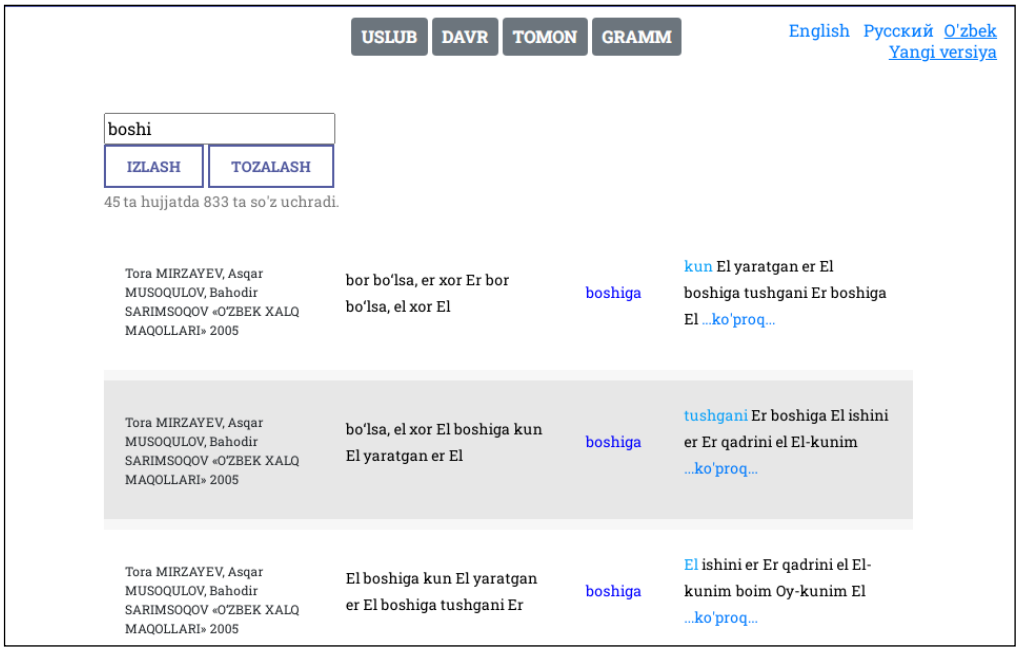
Lemma korpusdan ma’lumot qidirishning asosiy birligi sanaladi. Morfologik tahlil natijasida mustaqil so‘z turkumining statistik natijasi olinadi. Morfologik shakllangan so‘zlar kombinatsion jihatdan turli shakllarga ega bo‘ladi.

2) token bo‘yicha qidirish – [asos + grammatik kategoriya] modeli bo‘yicha so‘rov yaratish. Mazkur holatda yasovchi qo‘shimchalar yoki shu so‘z qatnashgan birlik morfologik jihatdan tahlil qilinadi



(asosan, Grammatik belgilari) bo‘yicha toppish imkoniyati mavjud.

3) konkordans bo‘yicha qidirish – n-gram left [W_1+W_2+W_3+Lt]/ n-gram right [Lt + W_1 + W_2 + W_3 + W_n]. Bunda W – so‘z, W_1 qidirilayotgan so‘zga birish masofadagi n-so‘z, Lt – [lemma + Grammatik kategoriya]



Konkordans qidirishda qidiruv tizimidagi so‘zning o‘zidan oldingi yoki keying keluvchi birliklar bilan birikish holati n-gram bo‘yicha ko‘rsatiladi.

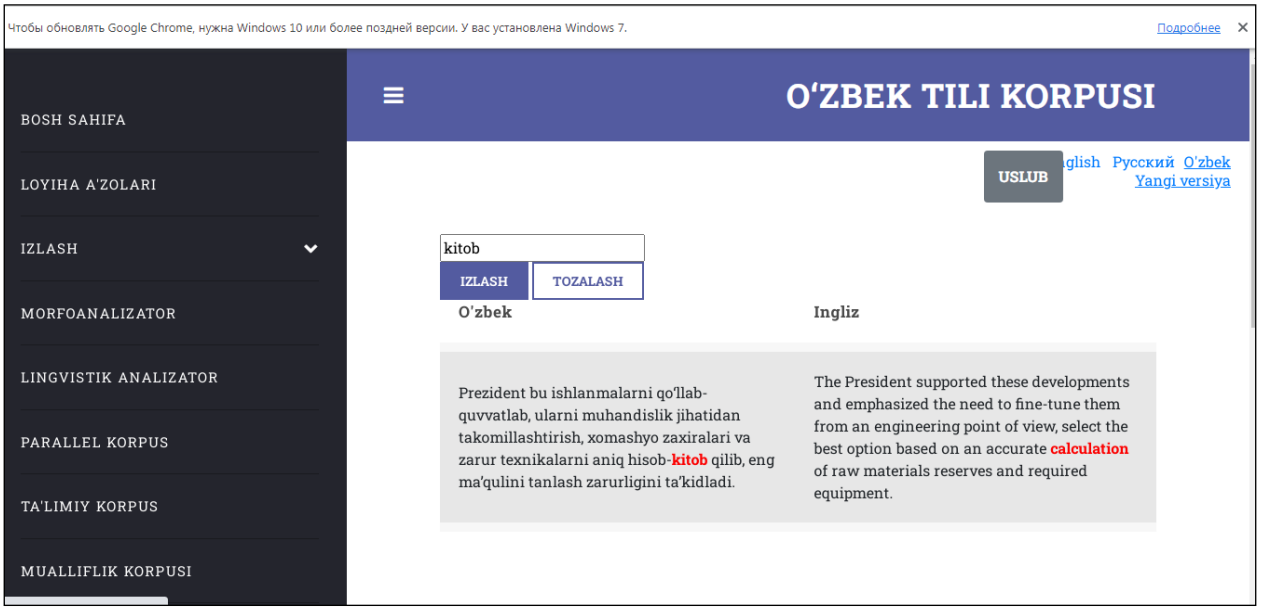


Matndagi ixtiyoriy so‘zning lemmasi, uning turkumi va FST metodidan olingan morfologik annotatsiyasi oynada o‘z ifodasini topadi

4) so‘z birikmasi bo‘yicha qidirish – [W + W + W_n+ Grammatik kategoriya].Ushbu qidiruvda turg‘un birikmalarni ajratishda qo‘l keladi. Masalan, o‘rmon qiroli, dala malikasi kabi perifrazalarda gab bo‘lagi vazifasida grammatik qo‘shimchalar bilan kelish imkoniyati hisobga olinadi.

Ushbu tizim ham lotin, ham kirill grafikasi bo‘yicha qidirish imkoniyatiga ega.

O‘zbek tili elektron korpusining ichki korpusi parallel korpus bo‘lib, unda o‘zbekcha-inglizcha va inglizcha-o‘zbekcha parallel matnlarning tarjimon xotirada segmentlangan birliklar bo‘yicha toppish imkoniyati mavjud.



Korpus qidiruv tizimi sifatida so‘zning bexato ekanligini tasdiqlash yoki uning xatoligini unga yaqin so‘zlar bazasidan qidirish korpus texnologiysi uchun muhim sanaladi[Abdurakhmonova, N. 2019,2021,2022].

Korpus lingvistikasining rivojlanish evolyutsiyasi shuni ko‘rsatadiki, ushbu fan kompyuter lingvistikasi bilan uzviy bog‘liq. Chunki u yoki bunda yaratilgan ilmiy yutuqlar bir-birini

RAQAMLI TEXNOLOGIYALAR VA SUN'IY INTELLEKTNI RIVOJLANTIRISH ILMIY-TADQIQOT INSTITUTI

to‘ldirishga, imkoniyatlardan foydalanishning optimal usullaridan foydalanishga omil bo‘lib xizmat qiladi. Korpuslar uchun berilgan ta’riflar ichida Webster lug‘atida tilga oid ta’rifida “bilimlar yoki dalillar to‘plami; tilni tavsiflovchi tahlil uchun foydalanilgan so‘zlar to‘plami” izohi berilgan, boshqa manbalarda korpus (ko‘pligi - corpora) yozma matnlar yoki yozib olingan og‘zaki nutqning transkripsiyasidan tuzilgan lingvistik ma’lumotlar to‘plami bo‘lib, asosiy maqsadi tilda mavjud farazlarni tasdiqlash deya, ta’riflanadi. Korpus haqidagi ta’riflar va farazlar biroz farq qilsa-da, ammo, barcha fikrlar birlashib, korpus til birliklarining xususiyatlarini aniqlash maqsadida qidiruv dasturiga bo‘ysundirilgan matnlar majmui, tabiiy tildagi elektron shaklda saqlanadigan yozma yoki og‘zaki, kompyuterlashtirilgan qidiruv tizimiga dasturiy ta’minot asosida joylashtirilgan matnlar jamlanmasi ta’rifini shakllantiradi.

Foydalanilgan adabiyotlar ro‘yxati:

1. Мирзиёев Ш. Миллий тараққиёт йўлимизни қатъият билан давом эттириб, янги босқичга кўтарамиз. – Тошкент: Ўзбекистон, 2017. – Б.168.
2. Abdurahmonova N. O‘zbek tili elektron korpusining kompyuter modellari. Monografiya. – Toshkent, 2021.
3. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). Uzbek electronic corpus as a tool for linguistic analysis. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
4. Abduraxmonova, N. Z. Q., & Urazaliyeva, M. Y. (2022). O ‘zbek tili elektron korpusida (http://uzbekcorpus. uz/) og ‘zaki matnlar korpusini yaratishning nazariy va amaliy masalalari. Academic research in educational sciences, 3(3), 644-650.
5. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
6. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
7. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
8. https://www.merriam-webster.com/dictionary/corpus
9. Crystal D. An Encyclopedic Dictionary of Language and Languages. - Oxford, 1992.

THE INSTRUMENTAL OBJECTIVE ROLE OF CORPUS IN RESEARCHES OF VARIOUS FIELDS

Vasliddinova Kamola
Phd student of Uzbekistan State World Languages University.

Annotation: In this article the instrumental objective role of corpus in the scientific researches is given and shows how the corpus helps to improve efficiency in researches and using of the corpus in the language learning process in education system and its fairy effective aspects in other fields are highlighted.

Keywords: Corpus, computational linguistics, corpus theory, corpus and discource, corpus-based methods, research, assessments, chat bot, server.

Аннотация: В данной статье представлена инструментальная объективная роль корпуса в научных исследованиях и показано, как корпус помогает повысить эффективность исследований и использования корпуса в процессе изучения языка в системе образования, а также выделены его эффективные аспекты в других областях.

Ключевые слова: Корпус, компьютерная лингвистика, корпусная теория, корпус и дискурс, корпусные методы, исследования, оценки, чат-бот, сервер.

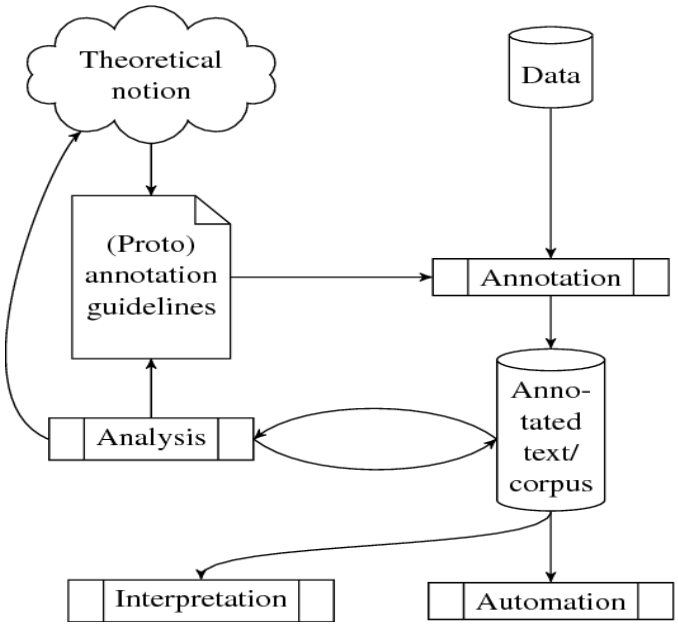
Linguists have relied on various methodologies for many years to unravel the mysteries of human communication from deciphering ancient scripts to analyzing contemporary discourse. Among these methodologies, corpus linguistics stands out as a powerful tool that has revolutionized the study of language in the digital age.

Corpus linguistics involves the systematic analysis of large collections of text, known as corpora, to investigate patterns of language use. These corpora can encompass a wide range of texts, including written works, transcripts of spoken language, social media posts, and more. By examining these vast repositories of linguistic data, researchers can uncover insights into vocabulary usage, grammatical structures, semantic relationships, and even cultural phenomena.

The interest to a corpus in the field of linguistics has been growing rapidly over the ten-year periods. Corpus in the simple term is defined as the compilation of texts that has been gathered for a specific reason [W.Cheng., 2011: 2].

The meaning of the term "corpus" varies slightly throughout academic areas. It usually refers to a grouping of texts; in literary studies; in theology, any collection of data (narrative texts or single sentences) elicited for linguistic research [Sebba, Fligelstone.,1994: 769].

It is clearly that early corpus-based research was carried out in English and then became the main focus of research by linguists around the world. Although corpus-based methods emerged in the 1960s, they have emerged as a new field in linguistics since the 1980s and corpora have developed new theories and directions in various linguistic fields.



1-rasm. Framework of corpus-linguistic techniques in the process of theorizing.

The first researches carried out in the field of corpus were in the areas of educational processes, in the field of linguistics, cross-cultural studies and corpus & discourse. On learning processes Barker (2010), Bieber (2012), Boulton & Tyne (2015), Chen (2014), Conrad (2009), Fuster Márquez & Clavel, Arroitia (2010), K Hyland (2012), K Hyland & Wong (2013), Lee & Webster (2012), Ma (2012), Walsh (2010) carried out early basic scientific research. In the field of linguistics, the scientific works of Gries (2009), Kaszubski (2003), Moisl (2015), Cheng (2011), Lukin (2017) are highly recognized. Wiegand & Mahlberg (2019) investigated the importance of corpora in cross-cultural research. Baker (2006) and Partington & Marchi (2015) elaborated on the differences and similarities between corpus and discourse [A.Mohammed Saleh Al-Hamzi, A.Gougui, Y.S.Amalia, T.Suhardijanto., 2020: 176-181].

If we think over that corpus helps to evaluate scientific researches, can any researches use to carry out their searches with the help of corpus? Before answering this question, we should analyze how to do it and with which method a researcher go to a destination. Elena Tognini-Bonelli (2001) has made a distinction between “corpus-based” research and “corpus-driven” research [J. Jitpaiboon1, A.Rungswang., 2022: 151].

In corpus- based research a researcher looks for in corpus with a specific intended approach. That is to say, by grammar or statistical rules a researcher tries to analyze how is the rule working? how it should be in fact? how many language users follow the rules or how many of them use wrong in context or speech.

In contrast to the method described above, “corpus-driven” approach serves to see what types of language patterns surface from the corpus. Different language variation examples come out during the analysis like lexical bundles other language varieties and collocations. Corpus - driven method especially evaluates the efficiency of language teaching in classrooms. Because they serve as the authentic materials in teaching.

A number of advantages of using corpora in language teaching and learning have been identified by many corpus linguists and thereby its implementation into the language classroom has been highly recommended. The advocates of using corpora have argued that corpora can

provide a powerful tool with which learners can explore and discover patterns of authentic language, providing such information as collocations, colligation, and semantic prosody that are hardly obtainable [Lee, Shinwoong., 2011: 159-178].

According to famous corpus researchers Aijmer, Kaltenbock, Mehlmauer-Larcher, it has been also contended that corpus-based language teaching has potentials to motivate learners and promote learner autonomy that are highly valued in pedagogy. I absolutely agree the mentioned ideas that at present we may take real life materials to enhance classroom environment with the help of corpus. In the following we may see some agreements by a few scolars who are doing on corpus.

Due to those potentials of corpora in language teaching and learning, a number of researchers (Aston, 1997; Braun, 2007; Conrad, 2004; Hunston, 2002; Tribble, 2001) offered them as an inventive teaching tool and a useful resource, and language specialists have begun to view their use as fairly trendy.

Many projects also review relevant factors related to the use of corpora in higher education. In a certain study titled “Corpus Linguistics and its Applications in Higher Education” by Fuster Márquez & Clavel Arroitia (2010), they are set out to depict implied essentials of corpus linguistics and its progress in relevance to theoretical linguistics and its implementations in modern teaching pursuits.

Although the first researches have been done in linguistics and methodology, later corpus has become one of the main objects to investigate in different fields such as politics, economics and especially medicine in recent years.

Firstly, we should mention that the latest researches in medicine especially statistical ones are being done on corpus. For example, based on 214,340 written patient comments (14,403,694 words) about National Health Service (NHS) cancer care in England from 2015 to 2018, the analysis was conducted with the help of corpus-based method. Patients are categorized based on the length of their therapy, and their qualitative assessments of their experiences are contrasted based on the keywords that best describe their language, the themes of their positive and negative comments, and the feedback ratings they provided [G. Brookes, P. Baker., 2021:1].

Another copus-based quantitative analysis is related to expressions of epistemicity. That is to say, a category covering the expression of commitment to the information transmitted and comprising epistemic modality and evidentiality, in a corpus of 400 newspaper articles from The Guardian concerning the COVID-19 pandemic.

Breakdown of comments and words, according to duration of treatment.

Duration of treatment	Number of comments	Words	Average words per comment
Under 1 year	128,804	8,866,673	68.84
1-5 years	59,577	3,960,335	66.50
Over 5 years	16,390	1,049,834	64.05

2-rasm. Breakdown of comments and words according to duration of treatment in the Covid-19 pandemic [G. Brookes, P. Baker., 2021:1].

200 articles were written in April 2020; the other 200 were written between January and April 2022, after massive vaccination and an extraordinary increase in medical knowledge. The analysis distinguishes between a number of subtypes of epistemic expressions and three kinds of

authorial voice. The results show that the April 2020 articles contain more epistemic expressions, of both weak commitment (might, perhaps, apparently …) and strong commitment (know, clearly, surely), which suggests a greater need to distinguish the known from the unknown in this period, due to the pervasive state of uncertainty. The analysis has social implications, since it gives readers an opportunity to appreciate the careful assessments of epistemicity found in the corpus and therefore to consider the convenience of obtaining information from quality media. These social implications, together with the methodology of the analysis, contribute to the potential of the paper for pedagogical applications [C.Jones, D.Oakey, K.O'Halloran., 2023:1].

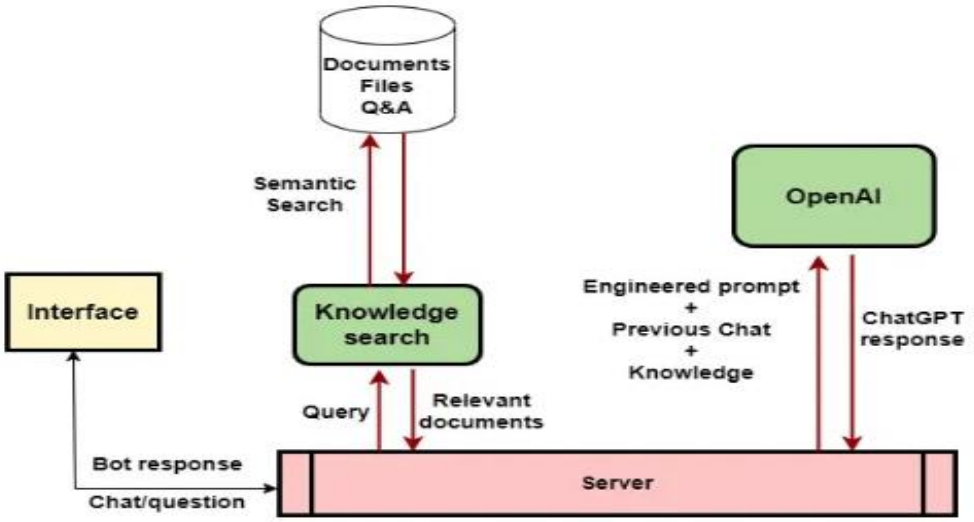
Number of epistemic expressions in the two subcorpora.

Expression type	Subcorpus A (April 2020) 191,716 words		Subcorpus B (January-April 2022) 159,550 words	
	Total	Ratio/ptw	Total	Ratio/ptw
Epistemic-modal auxiliaries	524	2.73	334	2.09
Epistemic-modal adverbs	131	0.68	59	0.37
Evidential adverbs	70	0.37	50	0.31
Epistemic-modal adjectives	147	0.77	133	0.83
Evidential adjectives	63	0.33	65	0.41
Epistemic-modal lexical verbs	340	1.77	204	1.28
Evidential verbs of appearance	165	0.86	96	0.60
TOTAL	1440	7.51	941	5.90

Chi-square with six degrees of freedom = 22.0551; $p = 0.0011$.

3-rasm. Number of epistemic expressions in the two subcorpora. [C.Jones, D.Oakey, K.O'Halloran., 2023:1].

We know that these days chat bots have become popular in this digital era. A chatbot is a machine conversation system which interacts with human users via natural conversational language. Software to machine-learn conversational patterns from a transcribed dialogue corpus has been used to generate a range of chatbots speaking various languages and sublanguages including varieties of English, as well as French, Arabic and Afrikaans.



4-rasm. A functional model of chat bot.

One of the most powerful language models ChatGPT was created by OpenAI in 2022 and this model has the ability to understand and interpret human language, making it a valuable tool for a wide range of applications. The model is trained on a large corpus of text data, which enables it to generate human-like responses to user queries.

In conclusion we can see that corpus has been becoming a central object to conduct researches not only in linguistics, but also in various fileds like economy, medicine, politics and others. Corpus focuses and characterizes authenticity, for that reason it is so fairy tool or method to carry out, investigate or analyze.

References:

1. A.Mohammed Saleh Al-Hamzi, A.Gougui, Y.S.Amalia, T.Suhardijanto. Corpus Linguistics and Corpus-Based Research and its Implication in Applied Linguistics: A Systematic Review Parole: Journal of Linguistics and Education, 10 (2), 2020. pp. 176-181.
2. B.A.Shawar, E.S.Atweel. Using corpora in machine-learning chatbot systems. International journal of corpus linguistics, vol.10, 4, 2025. pp. 489-516. https://doi.org/10.1075/ijcl.10.4.06sha
3. C.Jones, D.Oakey, K.O'Halloran. I will say the picture of the background is not related to the words: using corpus linguistics and focus groups to reveal how speakers of English as an additional language perceive the effectiveness of the phraseology and imagery in UK public health tweets during COVID-19. Applied Corpus Linguistics 3, 2023. P.1
4. G. Brookes, P. Baker. Patient feedback and duration of treatment: A corpus-based analysis of written comments on cancer care in England. Applied Corpus Linguistics 1, 2021. P.1. https://doi.org/10.1016/j.acorp.2021.100010
5. J. Jitpaiboon1, A.Rungswang. Trends of Corpus Linguistics Used in English for Specific Purposes Research: A Case of Asian ESP Journal.2022.20(2), P.151
6. Lee, Shinwoong. Challenges of Using Corpora in Language Teaching and Learning. Linguistic Research.2011. 28(1), P.159-178
7. Sebba, Fligelstone conference ,1994. P.769
8. W.Cheng. Exploring corpus linguistics, London: 2011. – P.2

O‘ZBEKCHA SO‘ZLARNING GRAMMATIK TAHLIL QILISH ALGORITMLARI

Qarshiyev Abduvali
f.-m.f.n., professor. Muhammad al Xorazmiy nomidagi
Toshkent axborot texnologiyalari universiteti Samarqand filiali.

Karimov S.A.
Filologiya fanlari doktori, professor
Samarqand davlat universiteti.

Tursunov M.S.
Muhammad al Xorazmiy nomidagi
Toshkent axborot texnologiyalari universiteti
Samarqand filiali, ilmiy izlanuvchi.

Annotatsiya. Ushbu maqolada o‘zbekcha so‘zni grammatik tahlil qilish algoritmlari, usullarni tavsiflash va kodlashga umumiy yondashuv muhokama qilinadi. Ishda so‘zni grammatik tahlil qilishda dastavval so‘z shaklidan so‘zning asosini va qo‘shimchalar qismini ajratib olish. Undan so‘ng qo‘shimchlarni bo‘laklarga ajratish hamda so‘z asosida mos turkumni aniqlash hamda qo‘shimchalar tahlilini bajarishdan iborat. Ishlab chiqilgan algoritm asosida nafaqat so‘zlarni balki matnlarni grammatik tahlil qilish uchun foydalansa bo‘ladi. Bu esa korpus tilshunosligida ham muhim ahamiyat kasb etadi.

Kalit so‘zlar: so‘z shakli, so‘z asosi, model, algoritm, ma’lumotlar bazasi, lug‘at, grammatika, morfologik tahlil.

Аннотация. В данной статье рассматриваются алгоритмы грамматического анализа узбекских слов, описание методов и общий подход к кодированию. В работе при грамматическом анализе слова прежде всего необходимо отделить от словоформы основу слова и суффиксальную часть. После этого он состоит из разделения наречий на части и определения на основе слова соответствующей категории и выполнения анализа наречий. На основе разработанного алгоритма его можно использовать для грамматического анализа не только слов, но и текстов. Это также важно для корпусной лингвистики.

Ключевые слова: словоформа, словесная база, модель, алгоритм, база данных, словарь, грамматика, морфологический анализ.

Annotation. This article discusses algorithms for grammatical analysis of Uzbek words, description of methods and a general approach to coding. At work, in the grammatical analysis of the word, first of all, it is necessary to separate the base of the word and the suffix part from the word form. After that, it consists of dividing the adverbs into parts and determining the appropriate category based on the word and performing the analysis of the adverbs. Based on the developed algorithm, it can be used for grammatical analysis of not only words, but also texts. This is also important in corpus linguistics.

Keywords: word form, word base, model, algorithm, database, dictionary, grammar, morphological analysis.

O‘tgan asrning 90-yillarida morfologik tahlil *generativ model, paradigmatik model, ikki tarkibli morfologik model* [Mahlov C., Piotrowski M., 2009, 15] kabi metodlarga asoslanilgan. O‘ttiz yildan oshiqroq vaqt mobaynida kompyuter morfologiyasi uchun turli texnologiyalarni yaratishga doir tadqiqotlar olib borilmoqda. Tillar uchun morfologik tahlil qilishning o‘nlab algoritmlari bo‘lishi tabiiy hol. Algoritmning bir yo‘nalishi so‘z oxiridagi harfiy birikmalarni qamrab oluvchi ma’lumotlar bazasidan foydalanadi. Bunda ehtimollik darajasiga ko‘ra, lingvistik resursdan olingan natijalar statistik tahlil qilinadi. Avtomatik usulda morfologik tahlil qilish jarayoni quyidagi turlarga ko‘ra tasniflanadi [A.Po‘latov, S.Muhamedova, 2008]:

1. Mantiqiy ko‘paytirish metodi orqali morfologik analiz.
2. Lug‘atsiz jadval orqali morfologik analiz.
3. So‘zlar asosiga asoslangan morfologik tahlil (so‘zlar asosiga tayanilgan morfologik tahlil Yevropa tillari uchun moslashtirilgan. Unda qo‘shimcha yordamchi jadvallardan foydalaniladi).
4. So‘zshakllarga (morfotaktik qoidalarga ko‘ra) ko‘ra morfologik analiz.

Yuqoridagilardan so‘zshakllariga ko‘ra morfologik tahlil qilish usuli keng qo‘llanadi. Ushbu algoritmga muvofiq so‘zlar morfemalarga ajratilganda qidirilayotgan so‘z lug‘atdan topilmaslik ehtimoli kuzatiladi. Bu holatda o‘ngdan chapga qarab tahlil qilish jarayonida ko‘makchi morfemalarning turi va modeliga qarab uning qaysi so‘z turkumiga oidligi aniqlanadi. U holda o‘sha so‘zning qaysi so‘z turkumiga tegishli ekanligi belgilanadi [Tursunov M.S., Qarshiyev A.B., 2019, 290-291]. So‘z shakllarni tahlil qilish jarayonida quyidagi bosqichlar amalga oshiriladi:

1. So‘zshakl morfem segmentlarga ko‘ra zanjirsimon birikadi.
2. So‘zshakl morfem segmentlar zanjiri sifatida tartiblanmagan yangi elementlar to‘plami - morfemalarga almashtiriladi. So‘z shakllar morfemalar to‘plami sifatida ma’lumotlar yoziladi.
3. Ushbu ma’lumot so‘z shakllari haqidagi yakuniy natija bilan qayta to‘ldiriladi.
4. Mantiqiy ko‘paytirish metodi orqali morfologik tahlil qilish flektiv tillar uchun foydalaniladi

O‘zbek tilidagi so‘z shakllarini grammatik asosini aniqlash muhim ahamiyatga ega.

O‘zbek tili so‘z shakllarini grammatik tahlil qilish morfotaktik qoidalarga asoslangan holda amalga oshiriladi.

O‘zbek tilidagi so‘z shakllarini morfologik asosini aniqlashda so‘zning strukturasi muhim ahamiyat kasb etadi. Shu munosabat bilan quyidagi matematik ifodalar orqali tasvirlangan lingvistik qoidalar asosidagi affiksasiya usuli bilan yasalgan so‘z shakllarining morfemik modellari taklif etildi [Baqayev I.I., 2021, 11]:

$$w_{sy} = K + \left\{\sum_{l=1}^{m} sx_l\right\} + \left\{\sum_{i=1}^{n} af_i\right\}.$$

Bu yerda K- sodda tub so‘z (o‘zak), sx - so‘z yasovchi affiks (suffiks), af - affikslar, $\sum - 0,1,\ldots n$ - ketma-ket konkatenasiya operasiyasi, l – suffikslar miqdori indeksi, i – affikslar miqdori indeksi w_{sy} - sodda yasama so‘z. Fikrimizni dalillash uchun ot so‘z turkumidagi «gul-don-lar-imiz-ni» so‘z shaklini tahlil qilamiz. Bu misolda K- «gul», sx - «-don», af - «-lar «, «-imiz», -ni «, «gul-don-lar-imiz-ni» - wsy so‘z shakli [4].

Bunda so‘z shaklidan so‘zning ajratib olinadi, hamda keyingi qadamda qo‘shimchalardan tashkil topga bo‘lagi tahlil qilinadi. So‘z shaklini grammatik tahlil qilish jarayonida albatta

ma’lumotlar bazasiga murojaat qilinadi. Biz ilmiy ishimiz uchun ma’lumotlar bazasida ikkita jadvaldan foydalanganmiz: 1) so‘zlar asosi lug‘ati, 2) qo‘shimchalar lug‘ati (1-rasm). So‘zlar asosi lug‘ati 3ta maydondan iborat: 1) tartib raqam (So‘z asosining IDsi), 2) So‘z asosi, 3) So‘z asosining turkumi. Qo‘shimchalar lug‘ati ham 3ta maydondan iborat: 1) tartib raqam (qo‘shimchaning IDsi), 2) Qo‘shimchaning nomi, 3) Qo‘shimchaning grammatik ma’noga ega izohi.

ID	Asos	Turkum
1	kitob	ot
2	qizil	sifat
…	…	…

a)

ID	Qo’shimcha	Izoh
1	lar	ko’plik
2	im	o’zlik
…	…	…

b)

1-rasm. a) So‘zlar asosi lug‘ati, b) Qo‘shimchalar lug‘ati.

So‘zlarni ma’lumotlar bazasidan o‘qib olish uchun ularning simvollar ketma- ketligiga amal qilinadi. Belgilash o‘sha so‘z yaxlit birlik sifatida aniqlanmagunicha bu holat takrorlanaveradi. Morfologik tahlil uchun komponent tahlil metodidan ham foydalanish mumkin. Unga ko‘ra so‘zning uzunligi aniqlanadi. Shu tarkibdagi simvollar soniga qarab ma’lumotlar bazasidan qidiriladi. So‘zning asosini topish va uni grammatik tahlilini qililshda ishlab chiqilgan algoritm quyidagi ketma ketlikdan iborat:

1-qadam. Dastur ishini boshlash.

2-qadam. <*input*> tegiga so‘z shakli kiritiladi.

3-qadam. Input tegiga kiritilgan so‘z shakli *“suz”* nomli o‘zgaruvchiga ta’minlanadi.

4-qadam. *“suz”* o‘zgaruvchining qiymati, *“suz1”* nomli o‘zgaruvchiga ta’minlanadi, ya’ni so‘zning nusxasi yaratiladi.

5-qadam. *“l”* o‘zgaruvchisiga *“suz1”* o‘zgaruvchisi qiymatining uzunligi, ya’ni so‘z shaklining belgilar soni ta’minlanadi.

6-qadam. *“suz1”* o‘zgaruvchisiga ta’minlangan so‘z shaklidan so‘z asosini aniqlash algoritmi boshlanadi. Bunda berilgan so‘z shakli ma’lumotlar bazasidan qidiriladi. Agar mavjud bo‘lsa 11-qadamga o‘tiladi, aks holda so‘zning oxiridan boshlab bittadan belgi kesib tashlash tartibida lug‘atdan qayta qidirish shaklida siklimiz davom etadi. Bunda sikl aylanishi toki so‘z asosi topilgunga qadar yoki *“suz1”* o‘zgaruvchining qiymati bo‘sh bo‘lib qolgunga qadar davom etadi. *“suz1”* o‘zgaruvchisining qiymati bo‘sh bo‘lib qolsa, u holda so‘z shaklidan uning asosini aniqlashni imkoni bo‘lmagan, va “so‘z asosini aniqlab bo‘lmadi” degan xabarni berib, 12-qadamga o‘tiladi. *“suz1”* o‘zgaruvchisidagi so‘z shaklidan uning asosi topilsa siklimiz rost qiymatni olib, davom etish to‘xtatiladi va 7-qadamga o‘tiladi.

7-qadam. *“suz1”* o‘zgaruvchisidagi so‘z shaklidan uning asosi o‘zgaruvchining o‘ziga ta’minlangan bo‘ladi. “suz” o‘zgaruvchisi qiymatidan *“suz1”* o‘zgaruvchisi qiymati kesib olinib, *“qush1”* o‘zgaruvshisiga ta’minlanadi, ya’ni *“suz1”* o‘zgaruvchisida so‘z shaklining asosi, *“qush1”* o‘zgaruvchisida so‘z shaklining qo‘shimchalar qismi yaratiladi.

8-qadam. *“qush1”* o‘zgaruvchisi qiymati sifati so‘z shaklini qo‘shimchalar qismini olgan va undan qo‘shimchalarni aniqlash algoritmi boshlanadi. Bunda qo‘shimchalar lug‘atidan qo‘shimchani izlaydi. Agar topilsa 11-qadamga o‘tiladi, aks holda qo‘shimchalarni Grammatik tahlil qilish uchun bo‘laklash boshlanadi

9-qadam. *“qush1”* o‘zgaruvshisi qiymatining birinchi belgisidan boshlab “qo‘shimchalar lug‘atidan qidiriladi. Agar lug‘atdan topilmasa keyingi belgini avvalgi belgiga qo‘shgan holatda

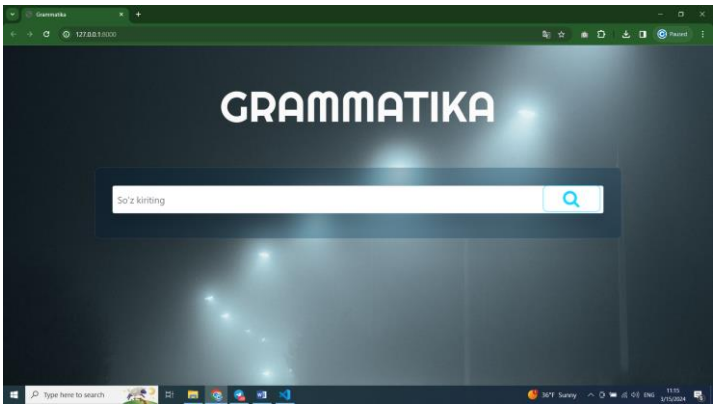
lug‘atdan qidirish davom etadi. Bu amal toki *“qush1”* ning qiymatidagi belgilarning oxirigacha davom etadi. Agar topilmasa 12-qadamga o‘tiladi, aks holda dastur ishini davom ettiradi.

10-qadam. *“qush1”* qo‘shimchalar ketma-ketligidan iborat satrda birinchi belgidan boshlab tekshirish (9-qadamda) davomida birinchi qo‘shimcha aniqlanganda ushbu qo‘shimcha va uning izohini massivga yozadi va *“qush1”* o‘zgaruvchisining qiymatidan aniqlanga qo‘shimcha kesib tashlanadi. Masalan: -larining, -lar+-i-ning. Bunda –lar qo‘shimchasi aniqlanganda “qush1”ning qiymati –ining bo‘ladi va 9-qadamdagi algoritmga qaytadi.

11-qadam. Bunda morfolig tahlil qilingan ma’lumotlar so‘z sahklining asosi, turkumi, qo‘shimchalar ro‘yxati va ularning izohi foydalanuvchiga taqdim etiladi.

12-qadam. Dastur ishini tugatadi.

Ushbu algoritm asosida morfologik razmetkalash dasturi ishlab chiqilgan.



2-rasm. Dasturning asosiy interfeysi

So’z shaklini Grammatik razmetkalash uchun so’z shakli kiritiladi. Misol sifatida “kitoblarga” so’zni tahlil qilaylik. Bunda 3-rasmda ko’rsatilgandek so’z shaklini Grammatik tahlil qilib beradi.

Grammatik tahlil xulosalarini quyidagicha ko’rish mumkin:

So’z shakli – “*kitoblarga”;*

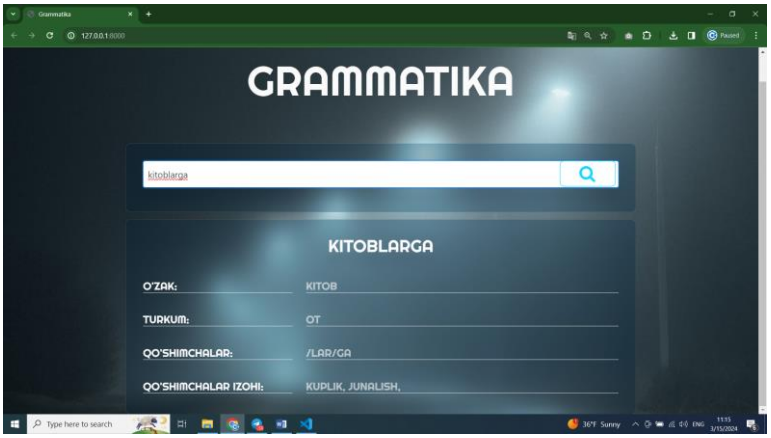
So’z asosi – *“kitob”;*

So’z turkumi – *“ot so’z turkumi”;*

Qo’shimchalar ro’yxati – *“ -lar, -ga”;*

Qo’shimchalarga berilgan izoh – *“ko’plik, jo’nalish kelishigi”;*

Bundan o’rganayotgan kalit so’zimiz ot so’z turkumiga taalluqligi, unda ko’plik va jo’nalish kelishiklaridagi qo’shimchalar to’g’risidagi lingvistik ma’lumotlar olingan.



3-rasm. Grammatik ma’lumotlar taqdimoti

Adabiyotlar ro‘yxati:

1. Mahlov C., Piotrowski M., JSLIM – Computational Morphology in the Framework of the SLIM, Theory Language / State of Art in Computational Morphology, Zurich, 2009, 15-p.
2. A.Po‘latov, S.Muhamedova. Kompyuter lingvistikasi. -T.: 2008.
3. M.Tursunov, A.Qarshiyev. O‘zbekcha matnlarni morfologik va leksik tahlil qilish algoritmlari va dasturlari / International scientific conference “Actual problems of applied mathematics and information technologies. –Toshkent:2019, 290-29.
4. Baqayev I.I., O‘zbek tili so‘z shakllarini morfologik tahlil qilish modellari va algoritmlari / avtoreferat, Toshkent: TATU, 2021, b. 11

KORPUS LINGVISTIKASIDA INTERTEKSTUALLIKNING ROLI

Xomidova Mahfuzaxon Farhodjon qizi
PhD, O‘zMU dotsenti v.b.

Annotatsiya: Bugun amaliy tilshunoslik sohalari qatorida korpus lingvistikasi ham o‘z o‘rniga ega va ayni damda juda zarur yo‘nalishlardan biridir. Korpusda jamlangan materiallar asosida tadqiqotchilar, til o‘rganuvchilar juda katta imkoniyatga ega bo‘ladilar. Ushbu maqolada korpus lingvistikasi va korpusning zarurati hamda til korpuslarida intertekstuallikning o‘rni haqida atroflicha fikr yuritilgan. Xususan, ingliz tili uchun yaratilgan korpusdagi intertekstual biurliklar, ularning manbasi, statistikasi kabilar mavzuga bog‘liq holda berilgan. Shu bilan birga muayyan til korpuslarida intertekstlarning berilishi va uning foydali tomonlari tahlilga tortilgan.

Kalit so‘zlar: koprus lingvistikasi, milliy korpus, matn, iqtibos, Shekspir, intertekstuallik.

Annotation. Today, corpus linguistics has its place among the fields of applied linguistics and is one of the most necessary areas at the moment. Based on the materials collected in the corpus, researchers and language learners will have a great opportunity. In this article, corpus linguistics and the need for corpora and the role of intertextuality in language corpora are discussed. In particular, the intertextual units in the corpus created for the English language, their source, statistics, etc., are given depending on the topic. At the same time, the provision of intertexts in certain language corpora and its useful aspects are analyzed.

Keywords: Corpus linguistics, national corpus, text, quotation, Shakespeare, intertextuality.

Аннотация: Сегодня корпусная лингвистика занимает свое место среди областей прикладной лингвистики и является одним из наиболее востребованных направлений на данный момент. На основе материалов, собранных в корпусе, у исследователей и изучающих язык появится прекрасная возможность. В этой статье обсуждаются корпусная лингвистика, необходимость корпусов и роль интертекстуальности в языковых корпусах. В частности, в зависимости от темы приводятся интертекстуальные единицы в корпусе, созданном для английского языка, их источник, статистика и т.д. При этом анализируется обеспеченность интертекстами в отдельных языковых корпусах и ее полезные стороны.

Ключевые слова: Корпусная лингвистика, национальный корпус, текст, цитата, Шекспир, интертекстуальность.

Til korpusi qidiruv dasturiga bo‘ysundirilgan, ma’lum bir tilga oid matnlar majmuyi hisoblanadi. Matnlar korpusi bu - bir yoki bir qancha tildagi matnlarning elektron shakli hamdir.

Korpuslar 10 milliondan 100 milliongacha, hatto milliardlab so‘z qo‘llanishlarini o‘z ichiga oladi. Misol uchun Britaniya milliy korpusi, Vengriya milliy korpuslarida 100 milliondan ortiq so‘z mavjud [6]. Tilning mukammal korpuslarida til paydo bo‘lgandan buyon hozirgi kungacha yaratilgan barcha yozma, og‘zaki, (audio, video) matnlar joylashtiriladi. Til korpusining ahamiyati shundaki, unda barcha nutq uslublariga, tilning turli davrlariga, shevalarga xos matnlar o‘z aksini topadi, shuningdek, tilga xos eng mayda detallarni ham qidiruvchiga taqdim eta oladi.

Korpuslarning asosiy vazifasi tilllarning grammatika, leksiklogiya, umuman barcha yo‘nalishlari bilan shug‘ullanuvchi mutaxassislar uchun bebaho manba rolini bajarishida ko‘rinadi. Shuningdek, korpus yordamida til sathlarida ma’lum bir davrda, bir asrda yuz bergan o‘zgarishlarni kuzatish mumkin bo‘ladi. Korpusning yana bir ahamiyatli jihati shuki, ona tili yoki chet tilini o‘rganuvchilar uchun tilning barcha sathlari bo‘yicha o‘quv manbasi bo‘la olishidadir. Qisqacha aytganda, korpuslar muayyan tilning barcha boyliklarini bir joyga to‘playdigan asosiy vositadir. Muayyan bir tilning milliy korpusi yaratildi degani bu – shu tilning elektron manbalari mavjud deganidir. Hozir dunyoda ingliz, ispan, rus, vengr, italyan, xitoy kabi tillarning milliy korpuslari yaratilgan. Shuningdek, bir qator turkiy tillarning ham milliy korpuslari yaratilmoqda.

Korpusning asosiy, parallel korpuslar, shevaga xos matnlar korpusi, poetik korpuslar, og‘zaki nutq korpusi kabi turlari mavjudligi adabiyotlarda qayd etiladi.

Korpus tarkibini shakllantiruvchi asosiy obyekt bu – turli matnlar hisoblanadi. Misol uchun shevaga xos matnlar korpusida ma’lum bir shevaga oid so‘zlarning qo‘llanilishi matnlar yordamida beriladi. Yoki og‘zaki korpuslarda yozib olingan og‘zaki matnli misollar yordamida korpus shakllantiriladi. Chunki ma’lum bir so‘zning ma’nosi, aynan amaldagi ma’nosi kontekstda, ya’ni matnda namoyon bo‘ladi. Korpus matnlardan iborat ekan, umuman, matnlar bilan ish ko‘rar ekan, unda matnlararo matn, ya’ni intertekstual matnlarning ham o‘z o‘rni bor.

Intertekstuallik matn tahlilida asosiy tamoyillardan biri bo‘lib, bugun u lingvistika va adabiyotshunoslikning asosiy tadqiq obyekti hisoblanadi. Ilmiy muomalaga bu bu atama fransuz olimasi Y. Kristeva tomonidan 1967-yilda olib kirilgan. Uning “Baxtin, so‘z, dialog va roman”[Kristeva, 1967: 438.] nomli ilmiy risolasida ilk bor bu atama qo‘llangan va intertekstuallik hodisasiga ta’rif berilgan. Olima intertekstuallik asosan ilmiy va badiiy matnda uchrashini ta’kidlab, ko‘proq badiiy matndagi intertekstuallik tamoyiliga urg‘u beradi. Y.Kristevaning intertekstuallikka bergan ta’rifi quyidagicha:

“Har qanday matn sitatalar mozaikasidan tuzilgan va har qanday matn boshqa bir matn bilan o‘zaro aloqada bo‘ladi[Kristeva, 1967: 440]”. Garchi olima aytganidek har bir matn sitatalar mozaikasidan iborat bo‘lmasa ham, har qanday matnda, albatta, intertekstuallik elementlariga duch kelish mumkin. Umuman olganda, muayyan matn tarkibida boshqa matn parchasining kelishi intertekstuallikdir. Intertekstuallik faqat badiiy uslubdagina emas, tilning barcha shakllarida, ya’ni shevaga xos matnlarda, yoshlar nutqida, qariyalar nutqida, og‘zaki va yozma so‘zlashuv uslubida, ilmiy, rasmiy, siyosiy, publististik matnlarda ham faol qo‘llanuvchi hodisa hisoblanadi. Muayyan matnlar korpus uchun ajratib olinganda, albatta, uning tarkibidagi o‘zga matn parchalari ham korpusdan joy oladi.

Aksariyat tilshunoslar korpusdan “misollar banki” sifatida foydalanishadi. Chunki korpusda til materiallarining katta qismi to‘plangan, shuningdek, maxsus qidiruv tizimi orqali ishning tezlashuviga yordam beradi[1, 95]. Mana shu jarayonda asl matnlar bilan, ular tarkibida kelgan intertekstlar ham qidiruv davomida chiqib keladi. Quyida ingliz korpusida uchragan intertekstlarning asosiy xususiyatlari haqida fikr yuritiladi.

Amerika ingliz tilining milliy korpusida eng ko‘p tarqalgan intertekstual matnlar bu taniqli shoir, dramaturg va mashhur ingliz adibi Uilyam Shekspir asarlaridan olingan iqtiboslardir [6]. Xususan, uning mashur “Romeo va Juliyetta” tragediyasidagi *“Vidolashish - bu shirin qayg‘u” (Parting is such sweet sorrow) [5] jumlasi Amerika korpusidagi turli janrlardagi matnlarda 13 marta uchraydi. Shuni ta’kidlash o‘rinliki,* Shekspir asarlaridan iqtiboslarni boshqa kontekstlarda ishlatish ularning semantikasida o‘zgarish bilan birga keladi. Allaqachon ma’lum ma’nolarga ega

bo‘lgan havolalarni batafsilroq o‘rganib chiqib, uning qaysi kontekstda qanday ma’nolarda kelganini ajratish mumkin. Bu jumla beshta nutq uslubining deyarli barchasida: og‘zaki, ilmiy, publististik, badiiy uslublarda uchragan hamda kontekst bo‘yicha ma’lum ma’no nozikliklariga ega bo‘lgan.

Shekspirning yana bir mashur asari “Gamlet” dan olingan – “tirik qolish yo o‘lish: shudir masala” jumlasi, ya’ni hayot haqidagi abadiy savol korpusda 14 marta uchragan. Bu mashhur jumla ko‘proq hayot va o‘lim haqidagi falsafiy asarlarda uchragan. Xususan, “Poeziya” ilmiy-adabiy jurnalining bir maqolasida bu jumladan hayot, o‘lim, o‘z joniga qasd qilish kabi psixologik jarayonlarni tasvirlashda intertekst sifatida foydalanilgan.

Amerika korpusida, shuningdek, Shekspirning o‘zbek kitobxonlari uchun ham yaxshi tanish bo‘lgan “Hayot bu – teatr” degan jumlasi ham intertekst sifatida turli janrdagi matnlarda uchraganligini ko‘rish mumkin. Umuman olganda, aksariyat intertekst matnlarning Shekspir ijodidan ekanligini, ularning asosiy matn ko‘rinishidan tashqari, intertekst sifatida ham korpus tarkibidan joy olishini til egalari uchun uning asarlari va iboralari mashurligi, shuningdek, Shekspir asarlari tilining ekspressiv va fatik funksiyalari mukammalligi bilan ham izohlash mumkin.

Elektron matn korpusiga asoslanib intertekstuallik hodisasi o‘rganish muayyan bir iqtibos turli vaziyatlarda va qanday hollarda ulardan foydalanishga ehtiyoj tug‘ilishini hamda matn tarkibida qanday vazifa bajarayotganini qiyoslash imkonini beradi. Shuningdek, ma’lum bir matndan olingan iqtibos, ya’ni intertekstni turli janrdagi matnlarda uchrashini qiyoslash, ularning o‘ziga xosliklarini aniqlash imkoniyati paydo bo‘ladi. Chunki intertekstuallik hodisasidan tilimizning turli uslublarida shunchalik keng va ko‘p foydalanamizki, bugun ularning ayrimlari tayyor til materiali sifatida lisonimizdan o‘rin egallagan. Intertekstuallik hodisasi o‘rganayotgan tadqiqotchi minglab qog‘oz kitoblardan shu hodisaga misolni qidirib vaqt yo‘qotmaydi. Elektron qidiruv yordamida o‘ziga kerakli manbani soniyalarda topa olish imkoniga ega bo‘ladi. Bu esa korpusning amaliy ahamiyati qanchalik muhimligini yana bir bor ko‘rsatuvchi omildir. Dunyo tilshunosligidagi ayrim sohalar qatori matn tadqiqining ham kuchayganligi aksariyat tillarda milliy korpuslarning mavjudligi uchundir. O‘zbek tilining ham milliy korpusi bu – tilshunoslar, adabiyotshunoslar, til o‘ganuvchilar uchun juda katta imkoniyat deganidir. Bu borada amalga oshirilayotgan tadqiqotlar, natijalar ko‘p sohalar rivojiga, birinchi o‘rinda esa o‘zbek tilining butun boyligi elektron ko‘rinishda avlodlarga yetib borishiga xizmat qiladi. Shuningdek, muayyan millat dunyoqarashi, tili va madaniyati haqida xabar beruvchi matniy hodisa – intertekstuallikni ham kengroq va chuqurroq tahlil qilish, o‘rganish imkoni beradi.

Foydalanilgan adabiyotlar ro‘yxati:

1. Захаров, В. П. Корпусная лингвистика. – СПб.: Изд-во СПбГУ, 2012. – 188 с.
2. Kristeva, J. Bakhtine, le mot, le dialogue et le roman Texte. / J. Kristeva // Critique, 1967. T. 23, № 239, pp. 438 465.
3. Что такое Корпус? Национальный корпус русского языка [Электронный ресурс]. — Режим доступа: http://www.ruscorpora.ru/corpora-intro.
4. M. Xomidova. Intertextuality and literary translation. Scientific Bulletin of Namangan State University, 1(6), 2019. 309-311.
5. The Complete Works of William Shakespeare [Electronic resource]. URL: http://shakespeare.mit.edu/

6. The Corpus of Contemporary American English [Electronic resource]. URL: http://corpus.byu.edu/coc
7. M.Xomidova. The problem of intertextuality in modern linguistics //Central Asian Problems of Modern Science and Education. – 2019. – T. 3. – №. 4. – C. 259-263.
8. M.Xomidova. Intertekst tushunchasi, tadqiqi va turlari. Educational Research in Universal Sciences, 2(4 SPECIAL), 2023.902–908.
9. M.Xomidova. Erkin Vohidov ijodida intertekstuallik. Research and education, 1(9), 2022. 525–533.

CHILDES KORPUSLAR TIZIMIDA O‘ZBEK TILI UCHUN MA’LUMOTLAR BAZASI YARATISH

O‘rinboyeva Ro‘za Toliboyevna
Mirzo Ulug‘bek nomidagi
O‘zbekiston Milliy universiteti magistranti.
E-mail: orinboyevaroza@gmail.com

Allaberdiyeva Durdona Gurbanmurat qizi
Mirzo Ulug‘bek nomidagi
O‘zbekiston Milliy universiteti magistranti.
E-mail: durdonallaberdiyeva@gmail.com

Annotatsiya: Bolalar nutqini tahlil qilish tilshunoslikdagi murakkab vazifalardan biri hisoblanib, jahon korpusshunosligida bu masalaga oid tajribalar CHILDES tizimida jamlangan. Ushbu maqolada o‘zbek tilidagi bolalar nutqini o‘rganishga qaratilgan tadqiqotlar asosida CHILDES tizimi uchun ma’lumotlar bazasi yaratish masalasi yoritib berilgan.
Kalit so‘zlar: transkript, audio, video, klinik nutq, Praat, Phon.

Abstract: Analysis of children’s speech is considered one of the most complex tasks in linguistics, and experiences related to this issue in world corpus studies are gathered in the CHILDES system. This article describes the issue of creating a database for the CHILDES system based on research aimed at studying children’s speech in the Uzbek language.
Keywords: transcript, audio, video, clinical talk, Praat, Phon.

Аннотация: Анализ детской речи считается одной из самых сложных задач в лингвистике, и опыт, связанный с этим вопросом в мировых корпусных исследованиях, собран в системе CHILDES. В данной статье описан вопрос создания базы данных для системы CHILDES на основе исследований, направленных на изучение речи детей на узбекском языке.
Ключевые слова: стенограмма, аудио, видео, клинический разговор, Праат, Фон.

Kirish

Dunyodagi jamiki tillar o‘zining fonetik, fonologik, sintaktik, semantik, morfologik kabi xususiyatlariga ega bo‘lib, tilshunoslikda ular ma’lum qonun-qoidalarga solingan. Nutqning kishilik jamiyatidagi barcha ko‘rinishi muayyan tadqiqot ishlari uchun obyekt vazifasini o‘tamoqda. Ular orasida bolalar nutqini o‘rganish, ularning transkripsiyasini yaratish ham qiziqarli, ham mashaqqatli vazifalardan biridir.

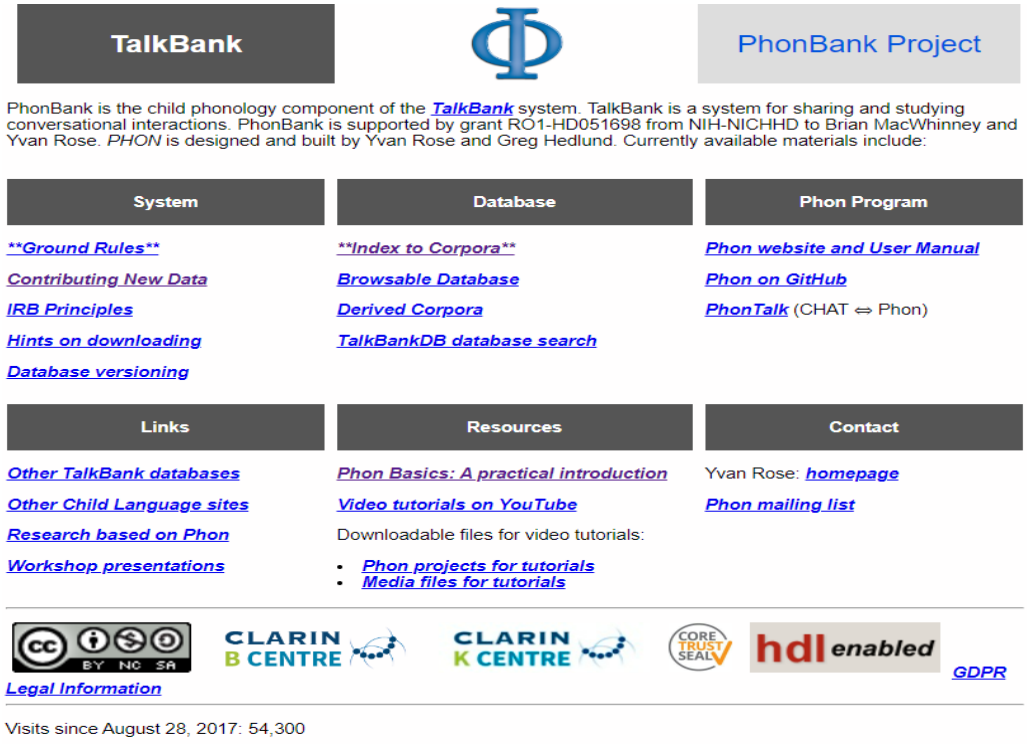
CHILDES – bola tilidagi ma’lumotlar almashinuvi tizimi bo‘lib, bolalar nutqini o‘rganishga qaratilgan korpuslar jamlanmasi hisoblanadi. Olimlar bolalarda uchraydigan afaziya, ikkinchi tilni o‘zlashtirish hamda nutqni tahlil qilishda yuzaga keladigan muammolarga yechim topish maqsadida turli xil tajribalarni amalga oshirishgan. Ularning nutqi, har bir so‘zi audio, video shaklida yozib olingan. Muayyan darajada transkripsiyada ifodalangan. Ammo CHILDES

tarkibiga o‘zbek tilidagi bolalar nutqini o‘rganishga qaratilgan tadqiqotlar o‘rin olmagan. Ushbu maqolada bu masalaning dolzarbligi va uning amaliy ahamiyati yoritib berilgan.

Adabiyotlar tahlili va metodlar

1984-yilda Karnegi Mellon universitetining psixologiya bo‘limida doktor Brayan MakVinni va doktor Ketrin Snou tomonidan tashkil etilgan CHILDES dasturi bolalar tilini o‘zlashtirish transkriptlari va yozuvlarini tadqiqotchilarga bepul, ommaviy ma’lumotlar to‘plami sifatida taqdim etishni maqsad qildi [MacWhinney., Snow K., 1985]. Mualliflar bu boradagi yuzaga keladigan muammolarni bartaraf etishga kirishdi. Dasturning eng dastlabki transkriptlari 1960-yillarda taqdim etilgan. 1990-yillarning boshida CHILDES dan foydalanish qulayligi sababli bolalar tilini o‘zlashtirish bo‘yicha tadqiqotlar soni sezilarli darajada oshgan [Alessandro S., Stephan C. M., 2019]. 2015-yilga kelib CHILDES tarkibidagi 230ta korpus 26 tildagi transkriptlar, audio va video materiallarni o‘zida jamladi. Bu butun dunyo bo‘ylab bir vaqtning o‘zida bir necha tillar kesimida bolalar nutqini o‘rganish imkonini beradi.

CHILDESning bosh menyusi TalkBankdan iborat. Bunda o‘rganilgan dunyo tillari 24 kategoriyaga ajratilgan. Ularning har biri korpusning yaratilish tarixi, muallifi haqida qisqacha ma’lumot, o‘rganilgan bolalarning yoshi, soni, media ma’lumotning turi, izohi va qo‘shimcha ma’lumotlar berilgan. CHILDESning asosiy ma’lumotlar bazasi PhonBankda joylashgan:



Bu sahifadagi “System” qismida korpus haqidagi asosiy qoidalar, foydalanish bo‘yicha tavsiyalar, versiyalari haqida ma’lumotlar berilgan. “Database” qismida mavjud korpuslarga hissa qo‘shgan olimlar va ularning ishlari ko‘rsatilgan. “Phon Program” qismida mualliflar bilan bog‘lanish uchun havolalar, minnatdorchilik xatlari hamda ma’lumotlarning har xil turdagi qurilmalar uchun moslangan shakllari o‘rin olgan. “Links” asosiy qismlardan biri bo‘lib, unda TalkBankning, bola tilini o‘rganayotgan tadqiqot institutlari hamda olimlarning internet saytlariga havolalar keltirilgan. “Resources”da turli qo‘llanmalar, You Tubedagi video darsliklar va media

fayllar ko‘rsatilgan. So‘nggi qismda CHILDES va TalkBank muammolari bilan bog‘liq yettita Google guruhlarining elektron pochtalari berilgan.

Tadqiqotning amaliy ahamiyati

CHILDES korpuslar tizimida o‘zbek tili uchun bolalar nutqiga asoslangan ma’lumotlar bazasini yaratish orqali korpus lingvistikasining taraqqiyotiga hissa qo‘shish mumkin. Bu masalani hal etishda bir necha dasturiy ta’minotlardan foydalaniladi.

Dastlab bolalar nutqi bilan ishlashda Phondan foydalanamiz. Phon – bu matn va fonologik ma’lumotlar korpusini yaratishni qo‘llab-quvvatlaydigan dasturiy ta’minot. U bolalar tilini rivojlantirishga qaratilgan bo‘lsa-da, Phondan bolalar yoki kattalar tili ma’lumotlariga asoslangan korpus tadqiqotlarining deyarli barcha turlarida (masalan, matn, fonologik, akustik, klinik) foydalanish mumkin. Shuningdek, Phon fonologik birliklar (masalan, fonologik xususiyatlar, stress, ohanglar) bo‘yicha tadqiqotlar uchun maxsus yordamni taklif qiladi va klinik nutqni tahlil qilish uchun ayrim maxsus funksiyalarni o‘z ichiga oladi [Rose Y., Carol S.G., 2016].

Korpusshunoslikda nutqni tahlil qilish uchun, asosan, Praatdan foydalaniladi. Praat – bu nutq va boshqa audio signallarni tahlil, sintez va manipulyatsiya qilish uchun ishlatiladigan bepul dasturiy ta’minot. U Amsterdam universitetida Pol Boersma va Devid Venink tomonidan ishlab chiqilgan va tilshunoslik, fonetika hamda nutq patologiyasi kabi sohalarda tadqiqotchilar tomonidan keng qo‘llaniladi. Praatning asosiy xususiyatlari quyidagilar:

1. Ovoz yozish va tahrirlash: Praat foydalanuvchilarga audio fayllarni yozib olish va tahrirlash imkonini beradi, jumladan, tovush to‘lqinlari va spektrogrammalarni ko‘rish imkoniyati.
2. Nutqni tahlil qilish: Praat nutq signallarini tahlil qilish uchun keng ko‘lamli vositalarni taqdim etadi: tovush balandligi, formantlar, intensivlik va boshqa akustik parametrlarni o‘lchash.
3. Matnni nutqqa sintez qilish: Praat matn kiritishdan nutqni sintez qilish uchun ishlatilishi mumkin, bu foydalanuvchilarga tadqiqot maqsadlari uchun maxsus nutq stimullarini yaratishga imkon beradi.
4. Skript tili: Praat o‘rnatilgan skript tiliga ega bo‘lib, u foydalanuvchilarga vazifalarni avtomatlashtirish va maxsus tahlillarni yaratish imkonini beradi.

Tadqiqot jarayonida to‘plangan audio va video materiallarning transkriptlari yaratiladi. Bolalardagi til o‘rganish qobiliyati yosh darajasiga ko‘ra tahlil qilinadi, ulardagi muammo va kamchiliklar o‘rganilib, jahon tajribalari bilan taqqoslanadi. O‘zbek tilidagi bolalar o‘rganishi uchun qiyin bo‘lgan xususiyatlar aniqlanadi. Ayni shu jihatlar korpusshunoslikning til taraqqiyotidagi ahamiyatini belgilab beradi [Abdurakhmonova, N. 2019, 2021, 2022].

Xulosa

CHILDES tizimida o‘zbek tili uchun korpus yaratish orqali bolalar nutqini o‘rganish borasidagi tadqiqotlarning amaliy natijasini taqdim etish mumkin. Bu kelgusi tadqiqotlar uchun asos bo‘la oladi. Bolalar nutqida uchraydigan nuqsonlarning oldini olishga, turli yoshdagi bolalarda ikkinchi tilni o‘zlashtirishga doir muammalarning bartaraf etilishiga yordam beradi.

Foydalanilgan adabiyotlar:

1. Rose Y., Carol S.G. Using PhonBank and Phon in Studies of Phonological Development and Disorders, 2016. Using PhonBank and Phon in Studies of Phonological Development and Disorders - PMC (nih.gov)

2. MacWhinney B. Morphosyntactic Analysis of the CHILDES and TalkBank Corpora. poster2 (psu.edu)
3. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). Uzbek electronic corpus as a tool for linguistic analysis. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
4. Abduraxmonova, N. Z. Q., & Urazaliyeva, M. Y. (2022). O ‘zbek tili elektron korpusida (http://uzbekcorpus. uz/) og ‘zaki matnlar korpusini yaratishning nazariy va amaliy masalalari. Academic research in educational sciences, 3(3), 644-650.
5. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
6. Sanchez A., Stephan M.C., Mika B., MacDonald K.E., Daniel Y., Michael C.F. Childes-db: A Flexible and reproducible interface to the child language data exchange system, 2019.childes-db: A Flexible and reproducible interface to the child language data exchange system
7. MacWhinney B., Snow C. The child language data exchange system//Journal of Child Language: 1985.
8. CHILDES Corpora (talkbank.org)
9. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
10. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).

KOMPYUTER LINGVISTIKASIDA AUDIOKORPUS MASALALARI

Urazaliyeva Mavluda
Mustaqil izlanuvchi (PhD)
O‘zbekiston Milliy universiteti.

Annotatsiya: Maqolada zamonaviy kompyuter texnologiyalar yordamida yaratilgan korpus va uning imkoniyatlarini takomillashtirishga doir bir qancha fikrlar tahlilga tortilgan. O‘zbek tili elektron korpusining audiomatnli korpuslarini yaratishda xorij tajribasi o‘rganilib, og‘zaki nutq aktlarini korpus bazasiga kiritishning amaliy jihatlariga e’tibor qaratilgan.

Kalit so‘zlar: kompyuter lingvistikasi, elektron korpus, audiokorpus, multimediyali korpus, o‘zbek tili.

Аннотация: В статье анализируется корпус, созданный с помощью современных компьютерных технологий, и несколько идей по улучшению его возможностей. Изучен зарубежный опыт создания аудиотекстовых корпусов электронного корпуса узбекского языка, а также уделено внимание практическим аспектам включения устных речевых актов в базу корпуса.

Ключевые слова: компьютерная лингвистика, электронный корпус, аудиокорпус, мультимедийный корпус, узбекский язык.

Abstract: The article analyzes a corpus created using modern computer technologies and several ideas for improving its capabilities. Foreign experience in creating audio-text corpora of the electronic corpus of the Uzbek language has been studied, and attention has also been paid to the practical aspects of including oral speech acts in the corpus database.

Keywords: computational linguistics, electronic corpus, audio corpus, multimedia corpus, Uzbek language.

Tilshunoslik va boshqa sohalarda elektron korpusdan foydalanish kompyuter lingvistikasi rivojlanish bosqichi davriga kelib, aynan XX asrning ikkinchi yarmida yanada keng tus oldi.

Tilning lisoniy muammolarini o‘rganish va ularga yechim topishda kompyuter texnologiyalardan foydalanish ma’lumotlarning elektron bazasini yaratishda katta yordam beradi. Buning uchun korpus yaratish va unga tilda mavjud so‘zlar elektron bazasini yaratish muhim.

Dunyo tilshunosligida korpus yaratishning lingvistik, matematik va dasturiy jihatlari olimlar tomonidan qilingan bir qancha ishlarda o‘z ifodasini topgan [Abduraxmonova N., 2021; 7-8.]. Chunonchi, rus va ingliz tillari bo‘yicha korpus lingvistikasi turli sohalar kesimida V.Zaxarov, A.Sedov, A.Baranov, R.Potapova, V.Rikov, U.Frensis, N.Leontyeva, V.Martin, S.Kubler, A.Laurans, E.Etwell, S.Hunston, L.Boizou, McKenneri, J.Grafmiller, J.Grieve, N.Grum, S.Hansson, K.McAulif, M.Malberg, P.Milin, A.Murakami, R.Peych, A.Shembri, P.Tompson, B.Vinter, G.Lich kabi xorijiy olimlar tomonidan ham turkologiyada korpus lingvistikasi bo‘yicha ilmiy tadqiqotlar olib borilgan.

O‘zbek korpus lingvistikasi bo‘yicha Sh.Hamrayeva, N.Abduraxmonova, Sh.Gulyamova, G.Toirova kabi olimlarning ishlarini keltirish o‘rinlidir.

Korpus ko‘plab yo‘nalishlarda kompyuter lingvistikasi, tarjimashunoslik, pedagogika kabi sohalarning tadqiqot obyekti vazifasini bajargani bois mazkur sohada olib borilayotgan ishlarning oxirgi o‘n yillikda sezilarli darajada oshdi.

Ilk bor Factored va MLCommons tomonidan MSWC – Ko‘p tilli og‘zaki so‘zlar korpusining birinchi versiyasi yaratildi [https://mlcommons.org/en/multilingual-spoken-words/]. Bu korpus 50 xil tildagi katta hajmdagi ovozli ma’lumotlarni o‘z ichiga oladi. Bu tillarda 5 milliarddan ortiq kishilar so‘zlashadi va ko‘pgina tillar uchun bu ovozli interfeys ta’lim olish uchun mo‘ljallangan ilk cheklanmagan bepul ma‘lumotlar bazasidir. MSWCda kalit so‘zlarni aniqlash, og‘zaki termin orqali qidirish va turli sohadagi odamlarga foyda keltiruvchi boshqa dasturlardagi akademik darajadagi tadqiqotlarni va tijorat ishlarida foydalanishga mo‘ljallangan. Bunda har qanday tildagi kalit so‘zlar uchun ovozli interfeys yaratish maqsad qilib qo‘yilgan.

Bunday dasturlar katta ma’lumotlar bazasining kompyuter modellarini o‘rganishni talab qiladi. Aslida korpus bunday dasturiy ta’minot uchun kalit so‘zlar turli kontekstlardagi minglab so‘zlarni to‘plash va tekshirish uchun resurs bo‘lib xizmat qiladi.

Ovozli dasturlar allaqachon kundalik hayotga kirib kelgan. Masalan, foydalanuvchi atrofidagi holatlarni aniqlash ko‘plab aqlli ilovalar (masalan, Apple Siri, Amazon Alexa yoki Google ovozli yordamchisi) zimmasiga yuklatilgan. Chiroqni o‘chirish yoki murakkabroq interfeysni ishga tushirish kabi harakatlarni boshqarishda buyruq ohangidagi so‘zlarni to‘xtovsiz eshitish uchun kalit so‘zlarni aniqlash tizimi yaratilgan. Bunday ovozli dasturlar ba’zi odamlar uchun axborot asrida qulaylik hisoblansa, ko‘zi ojiz kishilar uchun muhim ta’lim olish vositasi hamdir.

Bunday dasturlar katta ma’lumotlar bazasining kompyuter modellarini o‘rganishni talab qiladi. Aslida korpus bunday dasturiy ta’minot uchun kalit so‘zlar turli kontekstlardagi minglab so‘zlarni to‘plash va tekshirish uchun resurs bo‘lib xizmat qiladi. Hozirda mavjud ovozli platformalardan biri bo‘lgan MLCommons MSWC 50 ta tildagi nutqni aniqlash uchun katta hajmdagi ma’lumotlar bazasini yaratishda tabiiy tilning audiomatnli korpusidan foydalanmoqda va u doimiy ravishda yangilanib boradi [https://mlcommons.org/en/multilingual-spoken-words/].

Rus milliy korpusidagi multimediyali korpusning hajmi 5 763 881 ta so‘zni tashkil qiladi [https://ruscorpora.ru/]. Mazkur subkorpus doimiy ravishda yangilanib boradi. Har bir berilgan video 8-30 soniyalarda aks etgan. Har bir tovush ohangi, unlilar talaffuzi alohida-alohida keltiriladi. Har bir uslubdan olingan matn va audiolardagi ovoz egasining yoshi, jinsi, millati ko‘rsatiladi. Bu esa dialektologiya uchun juda zarur va juda muhim manba bo‘lib xizmat qiladi.

N.Abdurahmonovaning “O‘zbekcha matnlarni ovozlashtirish dasturining lingvistik ta’minotini ishlab chiqishda ayrim masalalar tadqiqi” nomli maqolasida so‘z turkumlari, tinish belgilari, arab va rim raqamlarini yozish va o‘qishda uchrovchi bir qator kamchiliklar sifatida keltiriladi. Bunda bazaga ma’lumot kiritishda matnning qaysi bandida *chiziqcha*, qaysi birida *tire* ekanligi va *-inchi* qo‘shimchalariga ham e’tiborli bo‘lish kerak. Yaratiladigan dastur esa buni tushunib olishi lozim. Tinish belgilari yozilgan paytda qo‘yiladigan belgilar ovozli matnda o‘qilmaydi. O‘zbekcha matnlarni ovozlashtirish dasturining har qanday o‘zbek tilidagi matnlarni hech qiyinchiliksiz o‘qib berishda uning lingvistik ta’minotining qay darajada muakammal ishlab chiqilgani katta ahamiyatga egadir. Shuningdek, o‘zbek tiliga boshqa tillardan, asosan, rus tili va u orqali boshqa tillardan o‘zlashgan ruscha internatsional so‘zlarni tadqiq etish va bunday so‘zlarni dastur lingvistik ta’minotiga kiritish masalalarini o‘rganish vazifasi ham oldimizda ko‘ndalang turibdi. O‘zlashma so‘zlarning talaffuzi o‘zbek tili so‘zlari talaffuzidan farq qilgani bois ham

ularning audio formatdagi va yozma shaklini lingvistik ta’minotga kiritish dasturning bunday so‘zlarni xatosiz o‘qishiga imkon yaratadi [Abduraxmonova N., Turklang, 2018].

Bundan tashqari o‘zbek tilida unli harflarning qisqa cho‘ziqligi ham og‘zaki nutqda ta’sir etmay qolmaydi: *ilm* [il:m]⇔*bilim [*bılım] *o‘lim* [olim] ⇔ *o‘lka* [ölka] kabilar.

Iste’moldan chiqish xavfi ostiga kelib qolgan tillar uchun ularning elektron bazasi va korpusini yaratish, shu tilga taalluqli bo‘lgan ilmiy va badiiy adabiyotlar yillar davomida asrashga, ular ustida bir qancha ilmiy ishlar qilishga imkon beradi.

Ma’lumotlar bazasida berilgan so‘zlar orfografik jihatdan to‘g‘ri yozilishi ham hisobga olinishi kerak. Natoijada morfoanalizator to‘g‘ri va sifatli ishlashiga zamin yaratilgan.

Korpusning kundalik hayotdagi ahamiyati yetarlicha bo‘lib, uni avtomatik tarjimada sintaktik tahlil uchun yordam, audio korpusda matnn tahlil qilish uchun baza, qolaversa, multimediyali darslarda elektron doskalarda til o‘rgatishda keng qo‘llash mumkin.

Umuman olganda, audio korpuslar ta’limga ayniqsa, maktab yoshidagi bolalar nutqini kuzatib borishda yuqori samaradorlikka erishishga yordam beradi. Sababi til ijtimoiy hodisa sifatida doimiy ravishda o‘zgarib turadi, qaysidir so‘zlar neologizm sifatida kirib kelsa, ba’zilari esa tarixiy so‘zlarga aylanadi. Bu jarayonni esa multimediali korpus orqali bevosita kuzatib borish mumkin. Ko‘rinib turibdiki, korpus nafaqat soha kishilarning, balki tilni rivojlantirishda umummilliy masala hisoblanadi.

Foydalanilgan adabiyotlar:

1. Aripov, M., Sharipbay, A., Abdurakhmonova, N., Razakhova B.: Ontology of grammar rules as example of noun of Uzbek and Kazakh languages. In: Abstract of the VI International Conference “Modern Problems of Applied Mathematics and Information Technology - Al-Khorezmiy 2018”, pp. 37–38, Tashkent, Uzbekistan (2018).
2. Kubedinova L. Khusainov A., Suleymanov D., Gilmullin R., Abdurakhmonova N. First Results of the TurkLang-7 Project: Creating Russian-Turkic Parallel Corpora and MT Systems. Proceedings of the Computational Models in Language and Speech Workshop (CMLS 2020) co-located with 16th International Conference on Computational and Cognitive Linguistics (TEL 2020) .2020/11: 90-101.
3. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. InProceedings of the International Conference on Language Technologies for All (LT4All) 2019.
4. N. Abdurakhmonova, U. Tuliyev and A. Gatiatullin, "Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus.uz," 2021 International Conference on Information Science and Communications Technologies (ICISCT), 2021, pp. 1-4, https://doi.org/10.1109/ICISCT52966.2021.9670043
5. https://ruscorpora.ru/new/
6. http://www.turklang.net
7. https://mlcommons.org/en/multilingual-spoken-words/

O‘ZBEK TILINING TA’LIMIY KORPUSI UCHUN BOLALAR ADABIYOTINI YARATISHNING DIDAKTIK ASOSLARI YUZASIDAN AYRIM MULOHAZALAR

Ergasheva Gulhayo
O‘zbekiston Milliy universiteti
4 - bosqich talabasi.

Ilmiy rahbar: **Inogamova Nargiza**
O‘zMU Kompyuter lingvistikasi
va amaliy tilshunoslik kafedrasi katta o‘qituvchisi.

Annotatsiya: Mazkur maqolada ta’limiy korpusda bolalar adabiyotini kiritishning ahamiyati, o‘zbek tili ta’limiy korpusi uchun bolalar adabiyotini tasniflash masalalari ko‘rib chiqilgan. Ertaklarni yosh bo‘yicha tahlil qilish va uni lingvistik jihatlari o‘rganilgan, ta’limiy korpus va uning lingvodidaktikadagi ahamiyati.

Kalit so‘zlar: korpus lingvistikasi, ta’limiy korpus, lingvodidaktika, metama’lumot, bolalar adabiyoti, janr.

Аннотация: В данной статье рассматривается важность включения детской литературы в учебный корпус вопросы классификации детской литературы для учебного корпуса узбекского языка. Изучен анализ сказок по возрастам и его лингвистические аспекты учебный корпус и его значение в языковой дидактике.

Ключевые слова: корпусная лингвистика, учебный корпус лингводидактика, метаданные, детская литература, жанр.

Annotation: This article considers the importance of including children’s literature in the educational corpus, the issues of classifying children’s literature for the educational corpus of the Uzbek language. The analysis of fairy tales by age and its linguistic aspects have been studied, the educational corpus and its importance in language didactics.

Keywords: corpus linguistics, educational corpus, linguodidactics, metadata, children’s literature, genre.

Insoniyat taraqqiyotidagi har bir davr til va tilshunoslik zimmasiga muayyan ijtimoiy vazifalarni yuklaydi. Til va kompyuter integratsiyasi natijasida tabiiy tilni qayta ishlash, milliy, ta’limiy va sheva korpuslarini yaratish, terminologik ma’lumot bazalarini loyihalashtirish, mashina tarjimasining lingvistik ta’minotini shakllantirish masalalari o‘zbek tilshunosligining kun tartibidagi asosiy vazifalari qatoridan joy oldi. Prezidentimiz Sh.M.Mirziyoyev ta’kidlaganidek: “Ona tilimizni jozibali va qiziqarli tarzda taqdim etish - o‘zbek tilining kompyuter dasturlarini, onlayn darsliklar, elektron lug‘atlarni yaratishni taqozo etadi” [Mirziyoyev Sh.M., 2019]. XXI asrning global muammolaridan biri tabiiy tillarning milliy xususiyatini saqlab qolishdir. Shuning uchun ham bugungi kunda mutaxassislar tomonidan tillarning elektron korpuslari yaratilmoqda. Korpus ma’lum bir tildagi mavjud katta-yu kichik jihatlarni jamlovchi elektron xazinadir. Til pedagogikasida korpusning afzalliklari xorijda keng tadqiq etilgan va ko‘pgina tillar o‘z ta’limiy korpuslariga ega. “Учебный корпус русского языка” [http://ruscorpora.ru], CORYL – Corpus of Young Learner Language, Commented Leaner Corpus Academic Writing, Tisus corpus, The

Advanced Finnish Learners' Corpus, The Dresden corpus, Open Cambridge Learner Corpus kabilar shular jumlasidandir. Bugungi kunda o'zbek tilshunosligida ta'limiy korpusni yaratishga doir ko'plab tadqiqotlar amalga oshirilmoqda, shuningdek, Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universitetida AM-FZ-201908172 raqamli "O'zbek tilining ta'limiy korpusini yaratish" nomli amaliy loyiha doirasida O'zbek tilining Milliy korpusi va ta'limiy korpusi yaratilmoqda.

Ta'limiy korpus-materiallari ma'lum bir tilni o'rganishga yo'naltirilgan, lingvodidaktik xususiyatga ega til korpusi. O'zbek tilining ta'limiy korpusi - o'zbek tili imkoniyatlarini o'rgatishga yo'naltirilgan, lingvodidaktik xususiyat kasb etgan elektron materiallarni qamrab olgan, maxsus sayt ko'rinishida amal qiladigan o'zbek tilidagi korpus. Korpusdan unumli foydalanuvchi soha –lingvodidaktika; u ona tili, xorijiy tilni o'rganishda birdek ahamiyatli. Ta'limiy korpus elektron darslik, o'quv qo'llanma, badiiy asarlar, lug'atlar va boshqalarni o'z ichiga oladi. Shular qatoriga bolalar adabiyotining alohida ahamiyati bor. Bolalar adabiyoti so'z san'ati va tarbiya vositasidir. Yoshlarni imon-e'tiqodli inson sifatida hamda vatanga muhabbat ruhida tarbiyalashda mustaqil davlatimizning qudratli qurolidir. Bolalar adabiyotini katta qismini ertaklar tashkil qiladi. Ertaklar qadim- qadimdan xalq og'zaki ijodi sifatida qadrlanib va og'izdan-og'izga ko'chib hozirgi kunga qadar yetib kelgan. Bolalarni dunyoni anglashida, tasavvurlarini kengayishida, ezgulik va yovuzlik haqida bilishlarida ertaklar muhim ahamiyat kasb etadi. Ertaklar zamonlar osha turli xalqlarda millatlarda turlicha aks etilsada, ularning zamirida asosan ezgulik, mehr oqibatlik, halollik, poklik kabi hislatlarning namoyon bo'lishida xizmat qiladi. Ertak tinglab ulg'aygan bolalarning fikrlash qobilyati individual salohiyati boshqa bolalarga nisbatan rivojlangan bo'ladi. Xo'sh, bolalarning ertak eshitishi yoki o'qishi ularga qanday foyda beradi?

- nutqi rivojlanadi, so'z boyligi oshadi;
- tasavvuri kengayadi;
- yangi axborotlarni qabul qiladi;
- hayotiy qiyin vaziyatlardan chiqish yo'llarini o'rgatadi;
- yaxshi va yomonni ajratishni o'rganadi;
- ijtimoiy-hissiy rivojlantiradi, u o'zgalarni tushunish, hamdardlik va kechira olishni o'rganadi;
- ertak misolida bola kimdandir ko'mak so'rashni, shu bilan birga kimgadir yordam berish zarurligini tushunib yetadi;
- asabiylashishga moyil bolalar ertak tinglash jarayonida tinchlanadi, ularda xotirjamlik hissi paydo bo'ladi.

Ertaklar orqali bolada yuqorida keltirilgan xususiyatlar shakllanadi. Bundan tashqari, bola o'zi mustaqil fikrlaydi, uning dunyoqarashi kengayadi. Bu esa bolalarning boshqa fanlarni ham puxta egallashlariga yordam beradi.

Bolalar kitobxonligini o'quvchilarning pedagogik-psixologik xususiyatlariga ko'ra quyidagicha guruhlarga ajratish mumkin:

1. Maktabgacha ta'lim yoshidagi bolalar kitobxonligi (2 yoshdan 7 yoshgacha)
2. Maktab yoshidagi kichik bolalar kitobxonligi (7 yoshdan 11-12 yoshgacha)
3. O'rta va katta yoshdagi bolalar kitobxonligi (13-14 yoshdan 15-17 yoshgacha)

Maktab yoshidagi kichik bolalar maktabgacha yoshdagi bolalardan bir oz farq qiladi. Bu farq bolalarning yoshi va bilim saviyasi bilan bog'liq. Maktabgacha ta'lim yoshidagi bolalar asosan tarbiyachilari, ota-onalari yordamida badiiy asar bilan tanishsalar, boshlang'ich sinf

o‘quvchilari bu ishni mustaqil bajaradilar. Bolalar adabiyoti yosh jihatdan ajratilganda maktabgacha ta’lim va boshlang‘ich sinf bolalariga ko‘proq o‘qitilgan. Masalan, 3- sinf o‘qish kitobi, (2016) 4- sinf o‘qish kitobi (2017) yildagi nashrlarida o‘zbek adiblaridan A.Obidjon, A.Abdurazzoq, F.Jalilova, X.To‘xtaboyev va boshqa ijodkorlarning 15 dan ortiq ertaklari kiritilgan. Bu darsliklarni 2021-yildagi nashr bilan solishtirsak ular bir-biridan tubdan farq qiladi. Bugungi boshlang‘ich sinf darsliklarida ertaklar deyarli kiritilmagan. Bu juda achinarli holat, albatta. Fikrimizcha, o‘quv darsliklariga ertaklarni kiritish lozim, chunki o‘zbek xalq ertaklarida milliy madaniyatimiz, qadryatlarimiz aks etadi. Bu esa yosh avlod tarbiyasida muhim rol o‘ynaydi va ularning dunyoni teran anglashlariga zamin yaratadi.

Ertak — xalq og‘zaki poetik ijodining asosiy janrlaridan biri; to‘qima va uydirmaga asoslangan sehrli sarguzasht va maishiy xarakterdagi epik badiiy asar. Asosan, nasr shaklida yaratilgan. Mahmud Koshg‘ariyning "Devonu lug‘otit turk" asarida *etuk* shaklida uchraydi va biror voqeani og‘zaki tarzda hikoya qilish ma’nosini bildiradi. Ertak Surxondaryo, Samarqand, Farg‘ona o‘zbeklari orasida matal, Buxoro atrofidagi tuman va qishloqlarda ushuk, Xorazmda varsaqi, Toshkent shahri va uning atrofida cho‘pchak deb ataladi. Ertak hayot haqiqatining hayoliy va hayotiy uydirmalar asosida tasvirlanganligi, tilsim va sehr vositalariga asoslanishi, voqea va harakatlarning ajoyib-u g‘aroyib holatlarda kechishi, qahramonlarning g‘ayritabiiy jasorati bilan folklorning boshqa janrlaridan farq qiladi. Ertaklarda uydirma muhim mezon bo‘lib, syujet voqealarining asosini tashkil etadi, syujet chizig‘idagi dinamik harakatning konflikt yechimini ta’minlaydi. Uydirmalarning turli xil namunalari ta’limiy estetik funksiyani bajaradi, janr komponenti sifatida o‘ziga xos badiiy tasvir vositasi bo‘lib xizmat qiladi. Uydirmalar voqea va hodisalarni hayotda bo‘lishi mumkin bo‘lmagan yoki mavjud bo‘lgan hodisalar tarzida tasvirlaydi. Xayoliy va hayotiy uydirmalarning ishtiroki, syujet chizig‘ida tutgan o‘rni va vazifasiga ko‘ra, ertaklarni 2 guruhga — xayoliy uydirmalar asos bo‘lgan ertaklar, hayotiy uydirmalar asos bo‘lgan ertaklarga bo‘lish mumkin. Xayoliy uydirmalar asosidagi ertaklar syujeti mo‘jizali, sehrli; hayotiy uydirmalarga asoslangan ertaklar syujeti esa hayotiy tarzda bo‘lib, unda real voqea-hodisalar tasvirlanadi. Ertakda, asosan, uch maqsad hikoya qilinadi. Birinchisida ideal qahramonning jasorati, yovuz kuchlarga qarshi chiqib, xalq manfaatini himoya qilishi, ikkinchisida, asosan, bosh qahramonning o‘zga yurt malikalariga yoki parizodlariga oshiq bo‘lib uylanishi, kasalga dori topishi, ajdar va devlar olib qochgan kishilarni ozod qilishi, uchinchisida esa adolatsizlikka, zulmga qarshi chiqishi kabi maqsadlar bayon qilinadi. Birinchi maqsad pahlavonning faol harakati va g‘ayritabiiy kuch-qudrati bilan amalga oshsa, ikkinchi maqsad passiv qahramonga yordam beruvchi tilsim vositalari yordamida, uchinchi maqsad esa bosh qahramonning aql idroki, tadbirkorligi bilan amalga oshadi.

Ertak janri obrazlar talqini, g‘oyaviy mazmuni va konflikti, syujet va kompozitsiyasi, uydirmalarning o‘rni va vazifasi, tili va uslubiga ko‘ra, shartli ravishda hayvonlar haqidagi ertaklar, sehrli ertaklar, maishiy ertaklar, hajviy ertaklarga bo‘linadi. Hozirda o‘zbek tilining ta’limiy korpusi uchun bolalar adabiyotini yaratish borasida bazaviy ma’lumotlar jamlanmoqda. Har bir ertak metama'lumot (qo'shimcha ma’lumot) lar bilan ta’minlanadi. Metama’lumotda ertaklar quyidagicha tasniflanadi.

t/r	Nomi	Janr	Tur	Mavzu	Yoshi	Muallifi	Manba	Nashri
1	“Bo‘rining tabib bo‘lgani”	Ertak	Adabiy ertak	Hayvon lar haqida	11	Anvar Obid jon	O‘qish kitobi 4- sinf	Toshkent.: “Yangiyo‘l

								poligraf servis”, 2017.
2	“Danak ichidagi daraxt”	Ertak	Ertak	Sehrli fantastik ertak	11	Aziz Abdurazzoq	O‘qish kitobi 4- sinf	Toshkent.: “Yangiyo‘l poligraf servis”, 2017.
3	“Sharq hikoyasi”	Ertak	Ertak	Hayotiy maishiy	11	Abdulla Oripov	O‘qish kitobi 4- sinf	Toshkent.: “Yangiyo‘l poligraf servis”, 2017.
4	“Laqma it”	Ertak	Adabiy ertak	Hayvonlar haqida	11	Shukur Sa'dullo	O‘qish kitobi 4- sinf	Toshkent.: “Yangiyo‘l poligraf servis”, 2017.
5	“Ko‘za Cha bilan tulki”	Ertak	Ertak	Hayvon lar haqida	11		O‘qish kitobi 4- sinf	Toshkent.: “Yangiyo‘l poligraf servis”, 2017.

Biz taklif qilayotgan korpus o‘zbek tili ta’limiy korpusi interfeysida joylashadi. Uning konkardansi uchun umumta’lim maktablarining boshlang‘ich sinf o‘quvchilarining o‘qish kitobidagi ertaklar matnidan foydalaniladi.

Ta’limiy korpusning bolalar adabiyoti bo‘limidan kimlar foydalana oladi? Ta’limiy korpusning bolalar adabiyoti bo‘limidan o‘rta maktablarning boshlang‘ich ta’lim o‘qituvchilari, maktabgacha ta’lim tarbiyachilari, uslubchilar, difektologiya yo‘nalishi talabalari va o‘qituvchilari foydalanishlari mumkin. O‘zbek tilining ta’limiy korpusi uchun bolalar adabiyotini yaratish soha mutaxassislari uchun katta ma’lumotlar bazasi bo‘lib xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Jumaboyev M. “Bolalar adabiyoti va folklor”. Toshkent. 2006-y.
2. O‘zbekiston Milliy ensiklopediyasi. — O‘z.R. 12 jildli, universal. 1-jild. 2000-2005-yy.
3. Jamilova B., “O‘zbek bolalar adabiyoti”. O‘quv qo‘llanma. Toshkent. “Innovatsiya – ziyo”, 2020-y.
4. Matjonov S., Shojalilov A. O‘qish kitobi. 4-sinf. 2017-y
5. http://ruscorpora.ru

II SHO‘BA
TIL TA'LIMIDA KOMPYUTER TEXNOLOGIYALARI

БАЗА ЗНАНИЙ ПОРТАЛА «ТЮРКСКАЯ МОРФЕМА » КАК ЛИНГВИСТИЧЕСКИЙ РЕСУРС ДЛЯ СОЗДАНИЯ ИНСТРУМЕНТОВ ОБУЧЕНИЯ ТЮРКСКИМ ЯЗЫКАМ

Гатиатуллин Айрат Рафизович
Институт прикладной семиотики Академии Наук
Республики Татарстан.

Абдурахмонова Нилуфар
заведующий кафедрой Компьютерной и прикладной лингвистики,
доктор наук, профессор Национального университета Узбекистана.

Аннотация: В статье представлены элементы графов знаний портала «Тюркская морфема», которые представляют собой лингвистический ресурс для разработки учебных курсов по изучению тюркских языков. Сам интернет-портал «Тюркская морфема» – это web-сайт (modmorph.turklang.net), который включает набор различных сервисов на базе лингвистических ресурсов по тюркским языкам, и ориентирован на работу с тюркскими языками в разных аспектах: морфонологическом, морфологическом, синтаксическом, семантическом. Создание образовательной среды требует предметно-ориентированные графы знаний, для получения которых не подходят методы создания общих и открытых графов. В данной работе описываются лингвистические графы знаний, которые, с одной стороны, отображают потенциальные возможности тюркских языков, а с другой стороны, примеры реального использования в текстах на тюркских языках. Особенность этих графов знаний в том, что, с одной стороны, они содержат лингвистические единицы разного языкового уровня, а с другой стороны – концепты, соответствующие значениям этих лингвистических единиц, которые встроены в тезаурус концептов. Структура такого графа знаний позволяет формировать контент учебного курса, строить индивидуальную образовательную траекторию, а также формировать задания и средства автоматизированной их проверки в рамках контроля знаний при обучении тюркским языкам. Это дает возможность разрабатывать впоследствии, на основе этих графов, программы обучения с учетом структурно-функциональных особенностей тюркских языков, а также способствует реализации индивидуальных целей обучающихся.

1.Введение

В настоящее время наблюдается бурное развитие систем автоматической обработки естественных языков (распознавания речи, автоматического перевода, анализа тональности,), которые основаны на технологиях машинного обучения. Данные технологии требуют большого количества ресурсов для эффективной работы. Однако языков с достаточной степенью изученности и готовыми большими наборами данных сравнительно мало. В список высокоресурсных языков входят такие языки, как английский, китайский (мандарин), испанский, французский, немецкий, японский, и несколько других языков.

Среди тюркских языков в данный список можно включить только турецкий язык, хотя оценка специалистов по степени оснащенности ресурсами турецкого языка неоднозначная. Одни авторы, называют его малоресурсным языком, а обзор сделанный в работе [1], показывает наличие большого количества лингвистических ресурсов для турецкого языка.

Сам термин малоресурсные языки был введен еще в 2003 год Краувером [2].

Согласно его определению малоресурсные языки – это естественные языки, обладающие следующими свойствами:

1. недостаток своей системы письменности или устойчивой орфографии;
2. нехватка квалифицированных лингвистов и переводчиков для данного языка;
3. ограниченное распространение в сети Интернет;
4. нехватка электронных ресурсов для обработки языка и речи, в том числе одноязычных корпусов, двуязычных электронных словарей, орфографических и фонетических транскрипций речи, словарей произношения и т. д.

Низкий уровень развития речевых технологий для малоресурсных языков объясняется рядом причин.

1. Задача описания и формализации структур малоресурсного языка достаточно трудоемка с научной точки зрения и сопряжена с высокими временными, людскими и финансовыми затратами для подготовки необходимых ресурсов (речевых корпусов, программного обеспечения).
2. Существующие на сегодня способы моделирования и создания программ для обработки речи направлены на решение узкого круга задач и не учитывают особенности работы с малоресурсными языками.

Подготовка многоязычных ресурсов требует разного уровня сопоставления или выравнивания лингвистических ресурсов разного уровня. В настоящее время можно выделить три уровня выравнивания лингвистических ресурсов на разных языках: документ, предложение и слово. Ясно, что выстраивание соответствий между этими языковыми единицами приводит к увеличению объема информации (в указанном порядке) за счет добавления новых структурных элементов. Полученные структуры с выстроенными соответствиями представляют собой графы знаний в узлах которых может храниться информация разного уровня в зависимости от уровней выравнивания. При выравнивании документов в узлах графа знаий могут храниться целые документы, при выравнивании на уровне предложений узлами графа могут быть предложения. При выравнивании на уровне более мелких лингвистических единиц соответствия не всегда выстраиваются между отдельными словами, в некоторых случаях отдельному слову может соответствовать аналитическая форма или аналитическая конструкция.

Разные типы соответствий используются для решения разного типа задач. Так выравнивание на уровне документа используется для поиска информации. Выравнивание на уровне предложения используется в задачах машинного перевода. Выравнивать можно, как неструктурированные, так и структурированные предложения представленные в виде синтаксических деревьев или семантических фреймов[Abdurakhmonova, N. 2019,2021,2022].

Выравнивание на уровне слов или лексем лежит в основе большого количества лингвистических задач.

Базы знаний, полученные в виде графов знаний, с выровненными лингвистическими единицами имеют достаточно широкий спектр применения, к примеру, они являются необходимым компонентом при разработке обучающих систем иностранным языкам. Для обучения тюркским языкам необходимы ресурсы, содержащие тюркские лингвистические базы знаний, что позволяет использовать такие базы знаний для создания многоязычных обучающих систем при обучении одному из тюркских языков. Например, это может быть обучение одному из тюркских языков лиц, владеющих другими тюркскими языками. Это достаточно актуальная задача поскольку существующие тюркские языки зачастую достаточно сильно отличаются друг от друга, и носители одного тюркского языка не всегда понимают носителя другого тюркского языка.

В данной работе рассматривается лингвистическая база данных в виде графа знаний портала «Тюркская морфема» (modmorph.turklang.net) [3] с точки зрения потенциала его использования в задачах образования.

2. Графы знаний

В последнее десятилетие, одним из эффективных способов представления лингвистической информации в ресурсах являются графы знаний. Имеется целый ряд работ с описанием лингвистических графов знаний. Рассмотрим один из примеров, который на наш взгляд наиболее близок к требованиям, предъявляемым к графам знаний для представления тюркского лингвистического ресурса. Таким примером является лингвистический граф, описанный в работе [4] (модель данного лингвистического графа знаний представлена на рисунке 1). По утверждению авторов, данный граф позволяет моделировать:

1) Отношения между понятиями и словами;
2) Информацию о встречаемости слов;
3) Диахроническую информацию как понятий, так и слов.

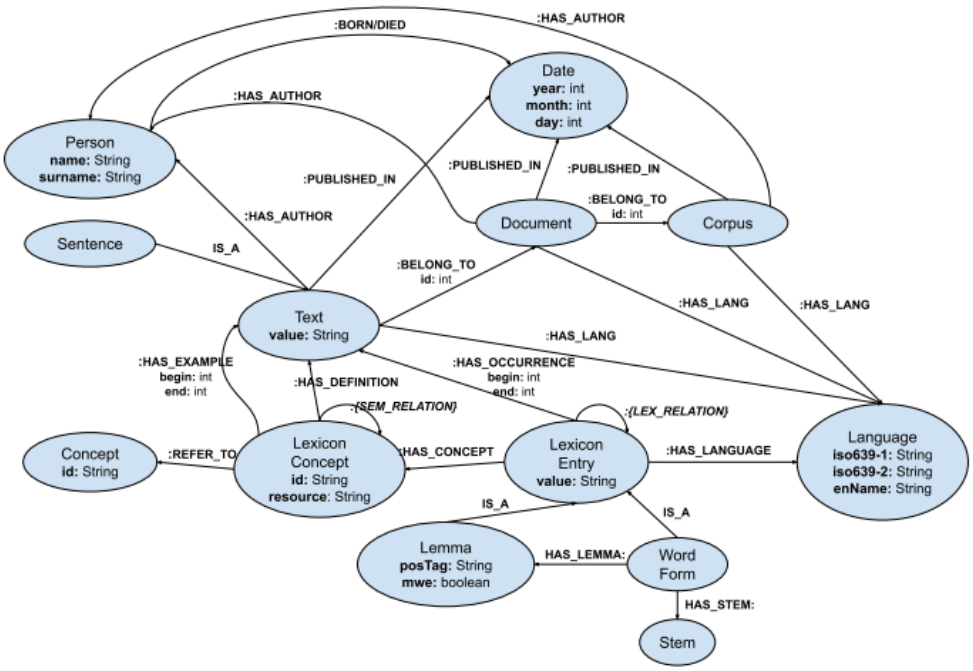


Рисунок 1. Модель лингвистического графа знаний

Описанный в данной работе лингвистический граф включает такие вершины, как Концепты (Concept), Лексические концепты (Lexicon Concept), Лексемы (Lexicon Entry). Лексические концепты взаимосвязаны между собой семантическими отношениями типа гипонимии и гиперонимии. Лексемы взаимосвязаны с леммой (Lemma) и основой словоформы (Stem). Особенность данного графа знаний в том, что он не описывает ситуационную семантику, поэтому его семантическое описание является достаточно ограниченным и отражает только таксономические отношения, аналогичные тем, что представлены в известном ресурсе WordNet. Для описания ситуационной семантики подходят графы знаний фреймового типа, например такой достаточно известный ресурс, как FrameNet, отсюда возникла идея добавления ситуационных фреймов к структуре данного лингвистического графа при реализации ресурса портала «Тюркская морфема».

Также в данном графе знаний не представлена грамматическая (морфологическая) структура словоформ, а для тюркских языков это является необходимым свойством, так как данные языки обладают богатой морфологией агглютинативного типа. В отличие от языков флективного типа, к которым относится русский язык, в тюркских языках существует четкое деление на структурные компоненты слова, которые называются морфемами. Такое деление позволяет представить морфологическую структуру словоформы в виде подграфа, вершинами которого являются морфемы, что также реализовано в графе знаний портала «Тюркская морфема».

3. Архитектура лингвистической базы знаний портала

Лингвистическая база данных портала «Тюркская морфема» представляет собой единый граф знаний, который подразделяется на несколько подграфов. Разделение на подграфы сделано в связи со структурными особенностями каждого из этих подграфов, а также с тем, что каждый из подграфов содержит наборы вершин одного типа. Вершины одного подграфа связаны между собой отношениями одного типа, а с вершинами из других подграфов отношениями иного типа. Схема разделения на подграфы представлена на рисунке 2. Такое разделение связано и с задачами, для решения которых используются каждый из подграфов единого графа знаний портала. Далее рассмотрим подграфы знаний портала, объединяемые в единый граф знаний.

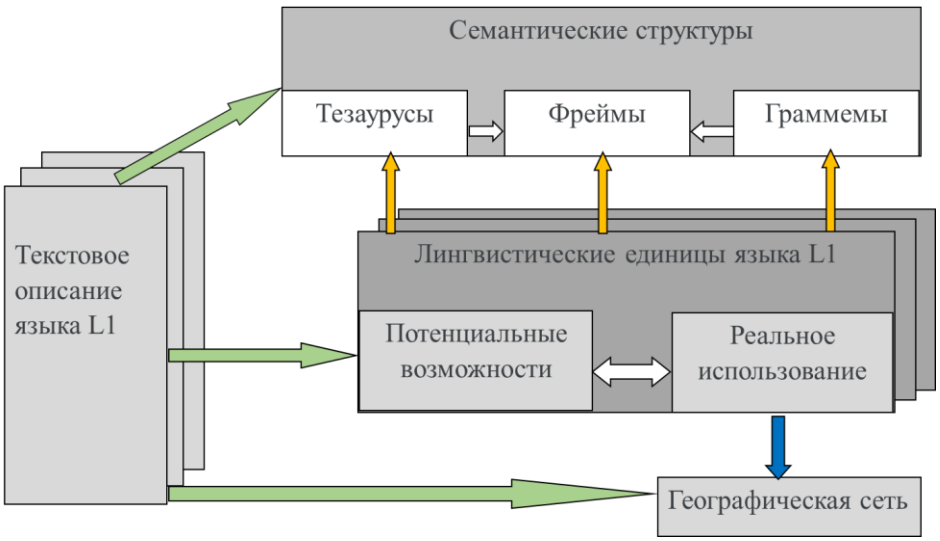


Рисунок 2. Архитектура подграфов графа знаний портала

Элементы всех подграфов портала «Тюркская морфема» объединяются в единый граф знаний, схема которого представлена на рисунке 3. В центре данного графа изображен элемент «Морфема» (morpheme), который является основной лингвистической единицей графа знаний портала.

Элементы «Корпус» (corpus), «Документ» (document), «Предложение» (sentence), относятся к текстовому описанию языка (см. рисунок 2) и указывают на реальное использование лингвистических единиц в текстах на некотором тюркском языке.

Элементы «Морфема» (morpheme), «Корень» (stem), «Аффикс» (affix), «Частица» (particle), «Послелог» (postposition), «Многословное выражение» (MWE) являются лингвистическими единицами языка, а их связи с текстовым описанием языка отображают реальное использование [Abdurakhmonova, N. 2019,2021,2022].

Элементы «Грамматическая категория» (grammatical category), «Граммема» (grammeme), «Семантема» (semanteme) относятся к семантическим структурам грамматики. Элементы «Лексема» (lexeme), «Концепт» (concept), «Онтология» (ontology) относятся к семантическим структурам тезауруса. Элементы «Ситуация» (situation), «Роль» (role) относятся к семантическим структурам фреймов. Связь семантических структур с лингвистическими единицами, а также морфотактическая связь лингвистических единиц между собой выражает потенциальные возможности графа знаний по генерации новых текстовых описаний (разработке анализаторов и синтезаторов текстов на различных уровнях: морфологическом, синтаксическом, семантическом).

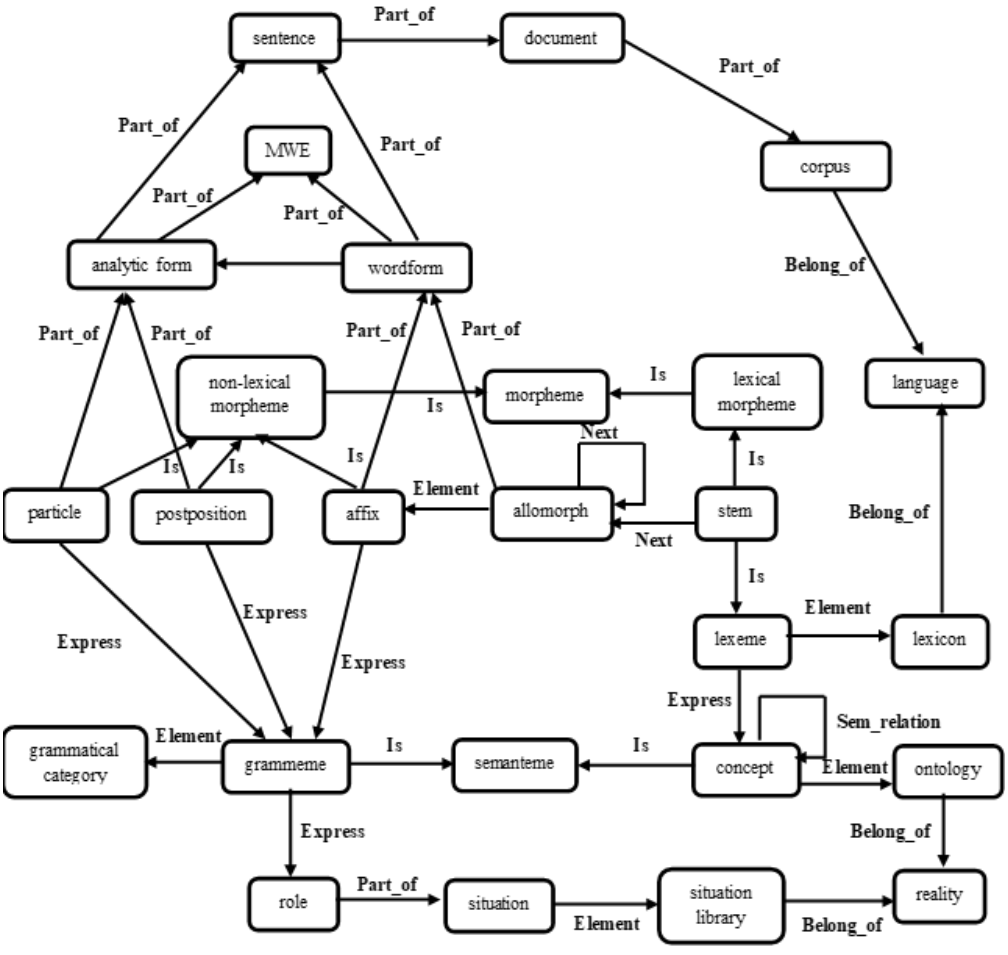


Рисунок 3. Схема графа знаний портала «Тюркская морфема»

4. Использование графов знаний портала в процессе обучения.

Одним из основных блоков графа знаний портала является подграф для описания ситуационных фреймов. Ситуационные фреймы представляют собой описание типовых ситуаций в языке, которые представляют собой более 1000 различных фреймов. Для каждой ткой ситуации дается текстовое описание ситуации (Рис.4) и наборы ролей типового фрейма (Рис.5).

Motion : Двигаться

Название (Русский)	Двигаться
Описание (Русский)	Некоторая сущность (Тема) начинается в одном месте (Источник) и заканчи месте (Цель), покрыв некоторое пространство между ними (Путь). В качест указать Область или Направление, в котором движется Тема, или Расстоян
Название (Английский)	Motion
Описание (Английский)	Some entity (Theme) starts out in one place (Source) and ends up in some other space between the two (Path). Alternatively, the Area or Directionin which the The movement may be mentioned.

Рисунок 4. Описание типовой ситуации

Роли в фреймах

Роль	Описание (Русский)	Описание (Английский)
Source : Источник	Источник-это место, которое Тема занимает изначально до ее изменения местоположения.	The Source is the location the Theme occupies initially before its change of location.
Theme : Тема	Тема-это объект, который меняет местоположение. Обратите внимание, что это не обязательно самодвижущийся двигатель.	The Theme is the entity that changes location. Note that it is not necessarily a self-mover.
Direction : Направление	Это ФЕ используется для выражений, которые указывают на движение по линии от центра божества к опорной точке (которая	This FE is used for expressions that indicate motion along a line from the deitic center towards a reference point (which may be implicit) that is neither

Рисунок 5. Набор ролей типового фрейма

За основу базы знаний портала “Тюркская морфема” взяты набор типовых ситуационных фреймов FrameNet и наборы типовых ролей этого ресурса. Все семантические роли выражаются грамматическими элементами: аффиксальными или аналитическими средствами (Рис.6). Для каждого языка будет свой набор грамматических форм, выражающих определенные семантические роли.

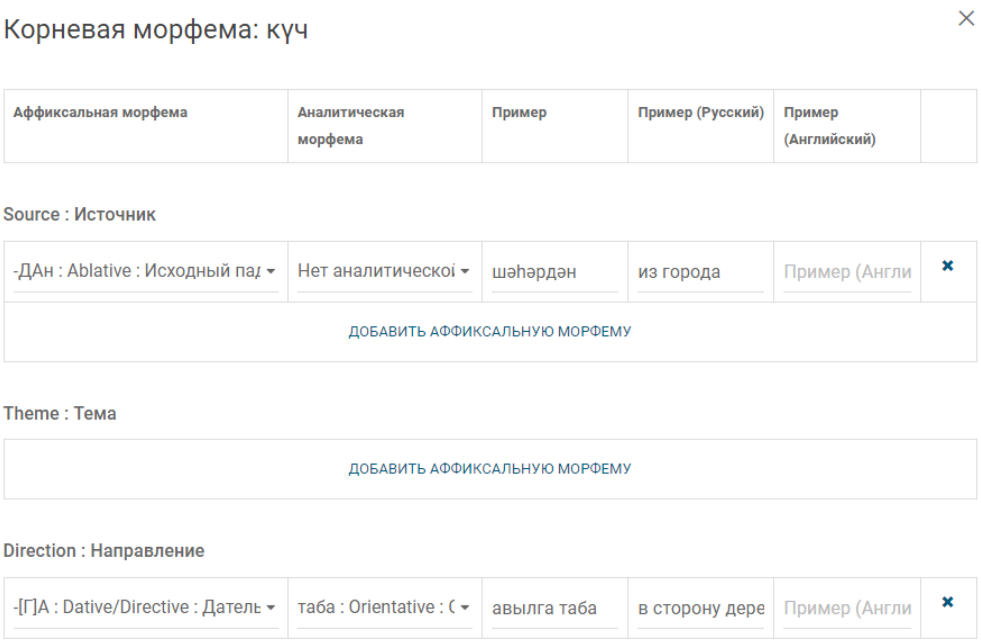


Рисунок 6. Грамматическое выражение типовых ролей фрейма

Для каждого из тюркских языков будет свой набор глагольного (ситуационного) управления.

Рассмотрим прямое использование графа знаний как информационно-справочной системы для составления заданий при обучении некоторому тюркскому языку студентов, уже знающих один тюркский язык на примере рассмотрения разницы в ситуационных фреймах татарских и турецких глаголов с использованием морфогенератора предложений.

1. В зависимости от ролевой схемы глагола выбирается вариант перевода
vurmak – убить / ударить

o insanı vuruyor *'он убивает человека'* → PN(o) N(insan)+**ACC(-yI) V(vur)**+PRES(-Iyor) → PN(ул) N(кеше)+**ACC(-нЫ) V(үтер)**+PRES(-Й) → ул кешене үтерә
o insana vuruyor *'он ударяет человека'* → PN(o) N(insan)+**DIR(-yA) V(vur)**+PRES(-Iyor) → PN(ул) N(кеше)+**DIR(-ГА) V(сук)**+PRES(-Й) → ул кешегә суга

2. Разные ролевые схемы в разных языках.

o bunu Ayşeye sordu *'он спросил это у Айшы'* → PN(o) N(bu)+ACC(-yI) N(Ayşe)+**DIR(-yA)** V(sor)+PST_DEF(-du) → PN(ул) N(бу)+ACC(-нЫ) N(Әйшә)+**ABL(-ДАн)** V(сора)+PST_DEF(-ДЫ) → ул моны Әйшәдән сорады
o işe başlıyor *'он начинает работу'* → PN(o) N(iş)+**DIR(-yA)** V(başla)+PRES(-Iyor) → PN(ул) N(эш)+**ACC(-нЫ)** V(башла)+PRES(-Й) → ул эшне башлый

Заключение

В данной статье представлен очередной способ использования фреймового блока портала “Тюркская морфема”.

Список литературы:

1. Çöltekin Ç, Doğruöz AS, Çetinoğlu Ö. Resources for Turkish natural language processing: A critical survey. Lang Resour Eval. 2023;57(1). –P. 449-488. http://doi.org/10.1007/s10579-022-09605-4 Epub 2022 Aug 26. PMID: 36060268; PMCID: PMC9417072.
2. Krauwer S. The basic language resource kit (BLARK) as the first milestone for the language resources roadmap. Proceedings of International workshop on speech and computer (SPECOM-2003). 2003. -P. 8–15.
3. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 1790-1793). IEEE.
4. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
5. Gatiatullin, A., Suleymanov, D., Prokopyev, N., & Abdurakhmonova, N. (2022, November). “Turkic Morpheme”: From the Portal to the Linguistic Platform. In World Conference Intelligent System for Industrial Automation (pp. 181-189). Cham: Springer Nature Switzerland.
6. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
7. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
8. Gatiatullin A., Kubedinova L., Prokopyev N, Suleymanov D. Linguistic knowledge graphs of the “Turkic morpheme” portal // 2023 8th International Conference on Computer Science and Engineering (UBMK), 2023. –P. 408-413
9. Basile P., Cassotti P., Ferilli S., McGillivray B. New Time-sensitive Model of Linguistic Knowledge for Graph Databases // Proceedings of the 1st Workshop on Artificial Intelligence for Cultural Heritage co-located with the 21st International Conference of the Italian Association for Artificial Intelligence (AixIA 2022), CEUR Workshop Proceedings. 2022. –P. 3286.

O‘QUVCHILARDA MEDIA SAVODXONLIK KO‘NIKMASINI SHAKLLANTIRISH (ona tili ta’limi misolida)

Mavlonova Klaraxon
Pedagogika fanlari doktori, dotsent
Toshkent davlat o‘zbek tili va adabiyoti universiteti.

Annotatsiya: XXI asr ko‘nikmalari – bu zamonaviy dunyoda, ayniqsa, yosh avlodning katta hayotga qadam qo‘yishi, muvaffaqiyat qozonishida hal qiluvchi ahamiyatga ega bo‘lgan bilimlar, hayotiy va kasbiy ko‘nikmalar, odatlar jamlanmasidir. Bu ko‘nikmalar hayotiy ko‘nikmalar deb ham yuritiladi. Umumiy o‘rta ta’limda o‘quvchilar tarbiyasiga alohida e’tibor qaratish, XXI asr ko‘nikmalarini shakllantirish orqali ularni hayotga tayyorlash, maktablarda yuz berayotgan huquqbuzarliklarning oldini olish, turli vaziyatlarda yuzaga keladigan muammolarini mustaqil hal etish, zo‘ravonlik va unga qarshi kurashishga o‘rgatish, yetakchilik, mediasavodxonlik masalalarining motivatsion yechimlari, global fuqarolik ko‘nikmalarini yoshlarga singdirish bugungi kunning dolzarb masalalaridan biridir. Mazkur maqolada ona tili (o‘zbek tili) ta’limida hayotiy ko‘nikmalar tarkibiga kiruvchi mediasavodxonlikni shakllantirish haqida so‘z boradi.

Kalit so‘zlar: XXI asr ko‘nikmalari, ona tili ta’limi, mediasavodxonlik, sun’iy intellekt, ijtimoiy tarmoqlar, audiomatn, imloviy savodxonlik.

Annotation: 21st century skills are a set of knowledge, life and professional skills, habits that are of decisive importance in the modern world, especially for the young generation to step into a big life and succeed. These skills are also called life skills. Paying special attention to the education of students in general secondary education, preparing them for life by forming the skills of the 21st century, preventing violations that occur in schools, independently solving problems that arise in various situations, teaching them to fight against violence, motivational solutions to the issues of leadership, media literacy, inculcating global citizenship skills in young people is one of the urgent issues of today. This article talks about the formation of media literacy, which is a part of life skills in mother tongue (Uzbek language) education.

Keywords: 21st century skills, mother tongue education, media literacy, artificial intelligence, social networks, audio text, spelling literacy.

Umumiy o‘rta ta’lim o‘quv dasturi metodologiyasining asosiy vazifalari sirasida “o‘zining “men”ligi, intellektual salohiyati, ma’naviyati va dunyoqarashi ustida mustaqil ishlash hamda XXI asr ko‘nikmalari (madaniyatli muloqot, insoniy fazilat, kitobxonlik savodxonlik, tanqidiy-kreativ fikrlash, ijodkorlik, hamkorlikda ishlash, muammolarga yechim topish kabi)ni o‘quvchilarda shakllantirish” belgilab berildi [Umumiy o‘rta ta’limning ona tili fanidan Milliy o‘quv dasturi loyihasi, 2020.]

O‘quvchilarni yoshligidan axborotlarni saralashga o‘rgatish ham bugunning talabi sanalmoqda. Chunki axborotning oluvchisi, tarqatuvchisi va yaratuvchisi ham aynan, siz-u biz tarbiyalayotgan maktab o‘quvchisi bo‘lishi mumkin. To‘g‘ri, bugunga qadar, axborot va xabarlarni rasmiy ommaviy axborot vositalari tomonidan amalga oshirilar edi, ammo bugunga kelib, bu chegaralanmagan.

Media va axborot savodxonligining asosiy elementlari quyidagilarni qamrab oladi:

- axborot savodxonligi (ma'lumotga bo'lgan ehtiyojlarni aniqlash va ifodalash; ma'lumotni topish va undan foydalanish; ma'lumotga baho berish; ma'lumotni tartiblash; ma'lumotdan axloqiy jihatdan to'g'ri foydalanish; ma'lumot almashish)
- OAV savodxonligi (OAV hamda internetning demokratik jamiyatdagi o'rni hamda vazifalarini anglash; OAV o'z vazifalarini bajara oladigan shart-sharoitlarni tushunish
- OAV vazifalari nuqtayi nazaridan ularning faoliyati hamda kontentini tanqidiy baholash; o'zini namoyon qilish, demokratik huquqlardan foydalanish uchun OAVlardan foydalanish; shaxsiy kontentni yaratish uchun zarur AKT ko'nikmalariga ega bo'lish)

Raqamli dunyo savodxonligi (raqamli vositalardan foydalanish; raqamli dunyoda o'zlikni anglash; raqamli dunyo huquq va majburiyatlarini tan olish; raqamli vositalar orqali muloqotni yaxshilash; raqamli vositalardan foydalanishda salomatlikni asrash; raqamli dunyo shaxsning xavfsizligi hamda daxlsizligini qo'llash) [Media and Information Literate Citizens., 2021:15].

Bugungi kunda, mediata'lim, mediasavodxonlik, mediatanqid ko'nikmalarini shakllantirishdan maqsad axborot asrida tarqatilayotgan axborotning mohiyatini anglash, foydalanishga o'rgata olishdir.

Axborot ishonchli, xolis, tushunarli va foydali bo'lishi kerak. Har qanday reklama, ommaviy axborot vositalaridagi xabarlar, ijtimoiy tarmoqlardagi izohlar mediamatn sanaladi. Shuning uchun boshqa fanlar tarkibida ham yoshlarimizga bu ko'nikmalarni, ya'ni axborotning to'g'ri va soxtaligini, ya'ni feyk (fake-soxta) yangiliklarini ajratish, soxta faktlarni aniqlash va tarqatmaslik masalalarini anglatish, ma'lumotlardan to'g'ri foydalanish ko'nikmalarini rivojlantirish lozim. Bu yoshdagi bolalar qiziquvchanligi sababli, turli mavzulariga oid ko'p sonli materiallar bilan tanishish mumkin. Media axborot savodxonligining (MAS) umumiy konsepsiyasidan tortib, axborot manbalari bilan ishlashning o'ziga xosligini tushuntirib borish lozim.

Mediasavodxonlik ko'nikmalarini egallashda ma'lumotlarni taqqoslash; tanqidiy fikrlash; axborotni olish va saralash; tahlil qilish; baholashni o'rganish; sharhlash; asoslash; xulosalarni ifodalash kabi tushunchalar qamrab olingan.

Bu ko'nikmalarini egallash davomida, nafaqat yangi bilimlarni egallaydi, balki o'z qiziqishlari doirasida bunyodkorlik, ijodkorlik qobiliyatini o'zlashtiradi, raqamli texnologiyalarning sir-asrorini anglaydi. Tanqidiy fikrlashni rivojlantirgan o'quvchilar o'zlashtirgan bilimlari doirasida aniq va asosli ma'lumotlarni qidirib topib, o'z qarashlari asosida yangiliklar (ixtirolar) yaratishga qodir bo'ladi. Shuning uchun ham o'quvchi yoshlarda bu ko'nikmalarni rivojlantirib borish muhim sanaladi [Mavlonova K., 2021: 15].

Quyida 7-sinf "Ona tili" [Mavlonova K., 2021: 5–12] darsligiga kiritilgan "Ijtimoiy tarmoqlar" nomli nutqiy mavzu doirasida berilgan o'quv topshiriqlari mazmuni va ularni amalga oshirish yo'llari haqida 2 soatlik (1-va 3-soat) dars misolida metodik tavsiyalar beriladi:

Nutqiy mavzuning maqsadi ijtimoiy tarmoqlar va ulardan foydalanishga doir ko'rsatmalar, axborot xavfsizligi; diagramma ko'rinishida berilgan ma'lumotlarni o'qib tushunish, ularni tahlil qilish, bir shaklda berilgan ma'lumotlarni ikkinchi shaklda ifodalash, ijtimoiy tarmoqlar, ulardan foydalanish, matnda mavjud bo'lgan so'z va so'z birikmalarining ma'nolari ustida ishlash, tinglab tushunish, mavzu yuzasidan shaxsiy qarashlarini dalillar asosida ifodalay olish, fikrni tartib bilan ifodalashda qo'llanadigan til birliklari, ularni qo'llash, yozma nutq bilan bog'liq imlo va tinish belgilarining qo'llanishiga oid lingvistik bilimlardan iborat.

1-soat

DARS JARAYONI

Kirish

O‘qituvchi yangi nutqiy mavzu nomini e’lon qiladi. O‘quvchilarni mavzuga olib kirish uchun quyidagi savollar bilan murojaat qiladi:

1. Ijtimoiy tarmoqlar deganda nimani tushunasiz?

2. Bugungi kunni ijtimoiy tarmoqlarsiz tasavvur etish mumkinmi? Uning inson hayotidagi o‘rni va ahamiyati haqida gapiring.

Asosiy qism

Og‘zaki nutq



Rasmdagi belgilarga diqqat bilan qarang va savollarga javob bering.

O‘quvchilar darslikda berilgan rasmni diqqat bilan kuzatadilar. Undagi belgilar nimani anglatishi haqida o‘ylab ko‘radilar. Berilgan savollar asosida og‘zaki suhbat olib boriladi.

Ushbu suhbatni tashkillashtirishdan maqsad o‘quvchilarning yangi nutqiy mavzuga olib kirish, mavzu doirasidagi tasavvurlarini aniqlashtirish, mavzu bilan bog‘liq atamalar ustida ishlash.

1. Qaysi belgilar sizga tanish? Ular haqida qanday ma’lumotga egasiz?

Rasmda berilgan belgilar haqida so‘raladi, ular anglatgan ma’no to‘g‘risida ma’lumot beriladi. O‘quvchilar ijtimoiy tarmoqlar bilan bog‘liq yana qanday belgilarni bilishlari haqida gapirib berishlari mumkin. Ushbu belgilarning ahamiyati va vazifasi to‘g‘risida suhbat olib boriladi.

2. Ijtimoiy tarmoqlarning hayotimizga ta’siri haqida fikr bildiring.

O‘qituvchi o‘quvchilarni mavzu doirasida fikrlashga undaydi. Suhbat davomida ijtimoiy tarmoqlarning foydali va zararli jihatlari haqida o‘quvchilarning fikrlarini tinglaydi. Undan foydalanish odobi haqida tavsiyalar ishlab chiqish kerakligi haqida taklif beradi.

3. Ijtimoiy tarmoqlardagi yozishmalarda muloqot odobi deganda nimani tushunasiz?

O‘qituvchi avval o‘quvchilarning bu boradagi fikrini so‘raydi. So‘ng muloqot odobiga rioya qilingan va ushbu qoidalar buzilgan bir nechta vaziyatni taqdim etadi. Imkon qadar ushbu vaziyatlar o‘quvchilarning kundalik turmushi bilan bog‘liq holda berilishiga erishish lozim. So‘z qo‘llash, so‘zlarni to‘g‘ri yozish, imlo va uslub qoidalari singari qoidalarning qanchalik ahamiyati borligi haqida xulosalar chiqarish imkonini berishi kerak.

4. Siz ham biror ijtimoiy tarmoqdan foydalanasizmi? Javobingizni izohlang.

O‘qituvchi o‘quvchilardan ijtimoiy tarmoqdan nima maqsadda foydalanayotganlari haqida so‘raydi. Suhbat jarayonida o‘quvchilarning imkon qadar muloqotda erkinlikni his qilishiga erishish lozim.

Imlo va talaffuz

Ijtimoiy tarmoqlar nomini yozing va talaffuz qiling.

Nutqiy mavzu doirasidagi atamalar, ular talaffuzi va imlosi bilan bog‘liq bilimlarni rivojlantirish ustida ishlash.

O‘quvchilar o‘zlari biladigan ijtimoiy tarmoq nomini yozadilar. O‘qituvchi ba’zi o‘quvchilar bilan sinf taxtasi yordamida ushbu topshiriqni bajarishi mumkin. O‘quvchilarning e’tibori ijtimoiy tarmoqlarning nomlarining yozilishi va talaffuzi o‘rtasidagi farqqa qaratiladi. Bu boradagi qoidalar yodga olinadi.

O‘qib tushunish va yozma nutq

Berilgan diagramma bilan tanishing va unga sarlavha qo‘ying

Matn ko‘rinishida yozing.

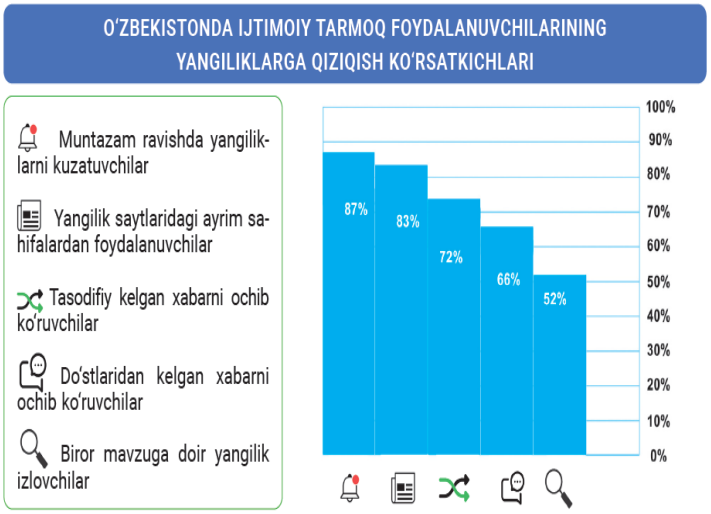
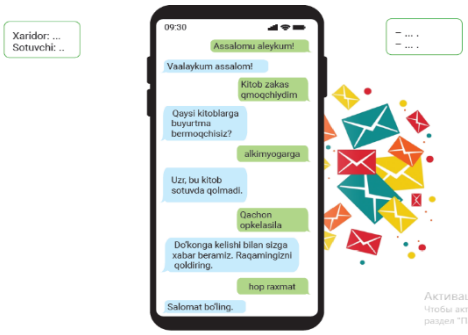


Diagramma ko‘rinishida berilgan ma’lumotlarni o‘qib tushunish, ularni tahlil qilish, bir shaklda berilgan ma’lumotlarni ikkinchi shaklda ifodalash bilan bog‘liq yozma nutq ko‘nikmalarini mustahkamlash.

O‘quvchilarning e’tibori diagrammada berilgan ma’lumotlarga qaratiladi

Diagramma necha qismdan iborat ekani, undagi ko‘rsatkichlar nimani ifodalayotgani, ko‘rsatkichlar qanday ifodalangani, ularni o‘zaro taqqoslash orqali xulosalar chiqarishga ko‘maklashuvchi og‘zaki suhbat olib boriladi. Ushbu suhbatni yozma matn ko‘rinishida ifodalash kerakligi aytiladi. Matnda qo‘llanishi mumkin bo‘lgan (ikki nuqta, nuqtali bergul, tire) yozma nutq qoidalari yodga olinadi.



Imlo qoidalari

Yozishmadagi xatolarni to‘g‘rilang, suhbatni quyidagi qoliplardan birini tanlab, yozing.

Ijtimoiy tarmoqlar, ulardan foydalanish bilan bog‘liq nutqiy ko‘nikmalarni shakllantirish.

O‘quvchilar yozishmada berilgan xatoliklarni aniqlaydilar, ularni qanday to‘g‘rilash kerakligini aytadilar va yozishmani qayta yozadilar. Shundan so‘ng berilgan qoliplardan biridan foydalangan holda mustaqil ravishda o‘zlari yozishlari mumkin.

Imlo qoidalari

Uyga vazifa. Matndagi gaplar chegarasini aniqlab, tinish belgilarini qo‘ying. Bo‘sh o‘rinlarga zarur til birliklarini qo‘yib, matnni ko‘chiring.

Gap va uning chegarasini aniqlash, tinish belgilari va til birliklarining qo‘llanishi bilan bog‘liq lingvistik ko‘nikmalarni mustahkamlash.

O‘quvchilar matnni o‘qib chiqadilar, gaplarning chegarasini aniqlaydilar. So‘ng qo‘yish uchun berilgan til birliklaridan mosini tanlaydilar, ularni o‘zaro taqqoslaydi, o‘z tanlovini izoh

shga tayyorlanadi.

3-soat

Kirish

Darsning tashkiliy qismi o‘tkazilgandan so‘ng quyidagi savol asosida kirish suhbati o‘tkaziladi.

"YouTube" haqida nimalarni bilasiz?

Ushbu platformaning afzalligi, undan foydalanish bilan bog'liq ayrim qonun-qoidalar, foydalanish maqsadi haqidagi fikrlar tinglanadi.

Asosiy qism

Tinglab tushunish

"Sun'iy intellekt" audiomatnini tinglang va savollarga javob bering.

Audiomatnni tinglab tushunish, tinglangan monologik nutq va suhbat haqidagi savollarga javob bera olish, unga munosabat bildira olish bilan bog'liq nutqiy ko'nikmalar rivojlantiriladi.

SUN'IY INTELLEKT NIMA?

"You Tube"ga kirsangiz, tarmoq sizning qiziqishlaringizga qarab turli videolarni taklif qila boshlaydi. Masalan, siz robot va kompyuterlarga qiziqsangiz, "You Tube" ham shu mavzularga oid videoroliklarni ko'rsatadi. Qanday qilib, "You Tube"ning bir xodimi faqat sizning qiziqishlaringiz bilan band bo'lmaydi-ku? Ha, aslida bu aqlli kompyuter. Uning sun'iy intellekti bor. U sizni kuzatib turadi! Internetda nimalarni izlaganingiz, nimalarni ko'rganingizga qarab qiziqishlaringizni aniqlab oladi. Siz internetdan qancha ko'p foydalansangiz, aqlli kompyuter ham sizni yaxshiroq tanib oladi.

Ko'pchilik sun'iy intellekt deganda insonlarga o'xshab fikrlovchi narsani tasavvur qiladi. Ammo sun'iy intellekt biznikiday haqiqiy ongga ega emas. U oddiy dastur, xolos. Faqat u shunday yaratilganki, o'z vazifalarini bajarish yo'llarini o'zi mustaqil o'rganadi va xatolaridan xulosa chiqara oladi. Shu tariqa, sun'iy intellekt borgan sari yaxshilanib, bilimlarini ko'paytirib boradi. Buning uchun unga insonning ko'magi deyarli kerak emas.

Hozirda sun'iy intellekt deyarli har bir qurilmada bor. Telefoningizda, kir yuvish mashinasida va muzlatkichda. Sun'iy intellektli avtomobillar haydovchisiz ham yurib ketmoqda. Xullas, sun'iy intellekt hayotimizning ajralmas qismiga aylanmoqda. Balki, siz katta bo'lguncha sun'iy intellekt ham ancha ulg'ayib bir necha kasblarning yo'q bo'lib ketishiga sababchi bo'lar. Masalan, suxandonlar, hisobchilar va haydovchilarning keragi bo'lmay qolar. Ammo sun'iy intellekt yaratuvchisi, sun'iy intellekt ta'mirlovchisi kabi yangi kasblar paydo bo'lishi ham tabiiy.

"Jajji akademik" jurnalidan

1. Aqlli kompyuter sizni qanday tanib oladi?

2. Sun'iy intellekt odamlardek haqiqiy ongga egami? Javobingizni izohlang.

O'qituvchi o'quvchilarning matnda berilgan ma'lumotlardan qanday xulosa chiqara olganliklarini aniqlashi kerak. Bunda o'quvchilar javoblarga berilgan izohlarning asosli ekaniga e'tibor qaratadi.

3. Ayting-chi, hozirgi kunda sun'iy intellekt qaysi qurilmalarda bor?

4. Sun'iy intellekt qaysi kasblarning yo'qolib ketishiga sabab bo'lishi mumkin ekan? Bu fikrga qo'shilasizmi? Nima uchun?

Muhokama jarayoni tashkillashtiriladi. O'quvchilarning fikrlari tinglanadi. Shu orqali sun'iy intellektning yaratilishi va uning insonlarga berayotgan yordamiga nisbatan o'quvchilarning munosabati aniqlanadi.

Tinglab tushunish

Berilgan gaplarni o'qing. To'g'ri fikrlarni aniqlab, yozing.

Tinglab tushunish, berilgan ma'lumotlar matnda qanday mazmunda ifodalanganiga e'tibor qaratish, matn mazmunini anglash bilan bog'liq ko'nikmalar mustahkamlanadi.

O‘quvchilar berilgan ma’lumotlarni bir necha marotaba o‘qib chiqadilar. Ma’lumotlar matn mazmuni bilan taqqoslanadi, to‘g‘ri va noto‘g‘ri ma’lumotlar aniqlanadi. Noto‘g‘ri ma’lumotlar yuzasidan izoh berish so‘raladi.

1. “YouTube”ning bir xodimi doimiy ravishda sizning qiziqishlaringiz bilan band bo‘ladi. – Noto‘g‘ri

2. “YouTube” sizni qiziqtirgan mavzularga oid videolavhalarni ko‘rsatadi. – To‘g‘ri

3. Ayrim kasblar yo‘qolib ketishi aniq. – Noto‘g‘ri (Chunki ushbu fikr taxminiy mazmunga ega, keltirilgan jumlada esa “aniq” so‘zi orqali fikr qat’iy tasdiqlangan mazmunga ega).

4. Sun’iy intellekt internetda izlagan, ko‘rgan narsalaringizga asosan qiziqishlaringizni aniqlab oladi. – To‘g‘ri

5. Avtomobillar haydovchisiz, sun’iy intellekt orqali boshqarilishi mumkin. – To‘g‘ri

Imlo qoidalari

Quyidagi so‘zlarni talaffuz qiling. Imlo qoidalari asosida yozilgan so‘zlarni aniqlang va ko‘chiring.

O‘quvchilarning imlo qoidalari va talaffuz bilan bog‘liq bilimlari mustahkamlanadi.

Berilgan so‘zlarning qaysilar to‘g‘ri yozilgani aniqlangach, xato yozilgan so‘z tahlil qilinadi. Nima uchun aynan ushbu so‘zlar tanlab olinganiga e’tibor qaratiladi.

Video, kompyuter, intellekt, avtomobil, suxandon, haydovchi, ta’mirlovchi, hisobchi, muzlatkich.

Yozma nutq

“Sun’iy intellekt” audiomatnidan olgan xulosangiz asosida quyidagi jumlalarni davom ettirib yozing. Fikringizni misollar bilan asoslang.

Xulosa chiqarish, xulosani matnda berilgan hodisaga bog‘lay olishga qaratilgan tinglab tushunishga oid ko‘nikmalar shakllantiriladi.

Berilgan jumlalar bilan tanishtirilgach, o‘quvchilar mavzu bilan bog‘liq xulosaviy fikrlarini yozadilar.

Yozma nutq

Uyga vazifa. Sun’iy intellektning hayotimizga ta’sirini ikki ustunga ajratib yozing.

Xulosa chiqarish, xulosani matnda berilgan hodisaga bog‘lay olishga qaratilgan tinglab tushunishga oid ko‘nikmalar shakllantiriladi.

Matn asosida olingan xulosalar hamda shaxsiy qarashlar asosida sun’iy intellektning ijobiy va salbiy jihatlari yoziladi. O‘quvchilar keyingi darsda o‘z fikrlarini izohlashga tayyorlanishi kerakligi haqida topshiriq beriladi [Mavlonova K., 2022: 9–16].

Xulosa sifatida shuni aytish mumkinki, medisavodxonlikning maqsadi har bir medianing ustuvorliklari va kamchiliklarini tushungan holda ular tomonidan tarqatilayotgan axborotni saralay bilish va zarurini qabul qilish ko‘nikmalarini shakllantirish bo‘lsa, asosiy vazifasi insonlar tomonidan iste’mol qilinadigan har qanday axborotning manipulyativ kuchini anglagan holda undan chegaralanishdir [https://uzhurriyat.uz]. Ona tili ta’limi jarayonida bunday ko‘nikmani shakllantirish va rivojlantirish imkoniyatlaridan to‘la foydalana olish zamon talabidir. Ona tili ta’limi bugungi kunda kommunikativ yondashuv asosida olib borilmoqda, ya’ni tilning muloqot vazifasi birlamchidir. Mazkur yondashuvning mohiyati real vaziyatda nutqni tushunish, o‘z fikrini to‘g‘ri ifodalashga o‘rgatishdir. Shuning uchun har bir mavzu uchun alohida vaziyat beriladi va shu vaziyatda nutq vositasida masalani hal qilish talab qilinadi. Hayotiy ko‘nikmalar ma’lum bir nutqiy mavzular doirasida amaliy yo‘naltirilgan holda shakllantiriladi.

Foydalanilgan adabiyotlar:

1. Azimova I., Mavlonova K., Quronov S., Tursun Sh., Hakimova N., Siddiqov M. Ona tili. Umumiy o‘rta ta’lim maktablarining 7-sinfi uchun darslik. Yangi nashr. – Toshkent: Respublika ta’lim markazi, 2022. – B.224
2. Mavlonova K., Tursunaliyeva B., Musurmonkulova O. va b. Maktab o‘quvchilarida hayotiy ko‘nikmalarni rivojlantirish (filologiya fanlari). O‘qituvchilar uchun o‘quv qo‘llanma. – T.: Respublika ta’lim markazi (YUNESKO), 2021. – B.154
3. Mavlonova K. Ona tili darslarida matn tuzilishini o‘rgatish metodikasi. Monografiya. – Toshkent: Anorbooks, 2023. – B.256
4. Mavlonova K. va b. Ona tili. 7-sinf o‘qituvchilari uchun metodik qo‘llanma. – T.: Respublika ta’lim markazi, 2022. – B.192
5. Media and Information Literate Citizens. Think Critically, Click Wisely! 2021. United Nations Educational, Scientific and Cultural Organization, 7, place de Fontenoy, 75352 Paris 07 SP, France, P.9
6. Umumiy o‘rta ta’limning ona tili fanidan Milliy o‘quv dasturi loyihasi (1–11-sinf). – Toshkent, 2020. http://uzviylik.tdi.uz/
7. https://uzhurriyat.uz/2016/04/05/mediasavodxonlik-va-u-nima-uchun

COMPUTER TECHNOLOGIES IN LANGUAGE EDUCATION

Jumaboev Asadbek Axmed o'g'li
Student of Translation faculty of
Uzbekistan state world languages university.

Scientific adviser: **Saitkhanova Aziza**
Uzbekistan State World Languages University
Translation Faculty Department of
"Practical translation of English Language" ESL teacher.

Abstract: In the article, it is one of the urgent issues to apply information technologies to the educational process, which is rapidly entering, using computers and information technologies to create new opportunities in the field of education, educational activities and the development of students' creative thinking, about the child's knowledge through computer tools, to realize moral, creative, communication and aesthetic possibilities and abilities, and for this the teacher himself should be well aware of the field of information technologies, the information of structured computer training programs should be mastered by the student it is explained that if psychological and pedagogical rules are taken into account, it allows for individualization and differentiation of teaching, stimulation of children's interests and independent activity in learning.

Keywords: computer, internet project, interactive, communicative competence, resource, optimization.

The growth of social processes taking place in society at the present stage determines a new approach to the formation of the personality of a modern specialist. There is a growing demand for a specialist who is able to take responsibility, think critically and creatively, and successfully and fruitfully cooperate with representatives of other cultures. The ability to adapt to changing life situations, work with information, independently acquire the necessary knowledge, and apply it in practice when solving various tasks occupies an increasingly important place in the life of a modern person. Such tasks must be implemented with the involvement of modern pedagogical strategies and learning technologies. The ability of educational programs to meet the needs of modern society is of great importance while maintaining the accumulated positive experience. Information resources and services on the Internet contribute to the introduction of a fundamentally new approach to the educational process, providing an opportunity to realise the opportunities inherent in new pedagogical technologies. Teachers' access to modern Internet resources makes it possible to develop their knowledge based on a wide range of educational, search, and research work, which, in turn, is also associated with the application and expansion of language knowledge. In fact, it is a socio-cultural education aimed at the self-knowledge of a person through the knowledge of other people, the study of cultures, cultural differences, and cultural unity, and a and a tolerant attitude towards cultural differences existing in the world. By reporting the search results, the student can express his opinion and discuss the issue under study. Ideas can be discussed with other students in the group. Participants should be able to substantiate their ideas, draw conclusions, and put forward hypotheses. During the discussion, students are asked to answer one general question for discussion. The most common forms of

telecommunications (i.e., communication via Internet technologies) include e-mail, forums, chat technologies, and video conferences. Originally created for communication, they are now used for educational purposes. It is a powerful tool for stimulating education and an ideology of global thinking. The simplest and most accessible form of telecommunications—e-mail—helps to involve listeners in communication practice in a foreign language and contributes to the improvement of written speech. In addition, this type of work is recommended for students who find it difficult to build verbal statements, as it allows them to consolidate and rewrite the text. While the most important task of education is to teach the ability to search for and select the necessary information and creatively apply it to solve complex problems, the advantages of introducing new information technologies into the educational process are obvious. Internet projects contribute to the development of creative and research activities among students, the formation of new socio-cultural knowledge, and, as a result, the formation of socio-cultural competencies.

The global computer network Internet is an incredibly valuable resource for the implementation of various models of language learning, that is, one of the best sources of information, which is often presented in an interactive form and is practically inaccessible. Speaking about language teaching in neophilological higher educational institutions, it should be borne in mind that it belongs to general education subjects and has a limited duration of its study. However, modern socio-political processes in society require an increase in the linguistic culture of the individual and dictate the need for practical language acquisition.

In recent years in the field of foreign language teaching the question of the importance and expediency of using Internet resources in foreign language education has been increasingly raised, which implies not only analyzing the use of new technical means, but also researching the positive and negative sides of the introduction of innovative forms and methods of teaching.

Modern methods of teaching foreign languages are connected with the ongoing technological progress, as well as with the technological renewal of the learning process. Recent advances in high technology and the spread of the global Internet provide foreign language teachers, methodologists, and students themselves with tremendous opportunities for further improvement of the teaching process.

That is why it is so important to improve the methodology of using computer information technologies in teaching English. Modern information technology is becoming part of the learning process. Computer technologies and the English language class are an actual direction that requires modern approaches and innovative solutions.

Modern pedagogical technologies such as collaborative learning, project methods, the introduction of modern information technologies and Internet resources can help bring to life a person-centered approach to learning, provide individualization and differentiation of training, taking into account the capabilities of children and their level of learning.

The possibilities of Internet resources are endless. The universal Internet creates conditions for obtaining any important information for students and teachers from anywhere in the world: country study material, announcements from the life of young people, notes from printed editions and magazines, etc. At English lessons with the help of the Internet you can solve the whole range of didactic tasks: to form reading skills, improve the knowledge of written language, enrich the vocabulary of students, create motivation in teenagers to learn English.

The idea of introducing Internet technologies in the course of theoretical and practical classes in a foreign language, according to E. Y. Sokolova, has been widely spread among teachers, methodologists around the world. The didactic aspects of computerization of education have been developed by well-known scientists and educators E. G. Azimov, V. P. Bespalko, B. S. Gershunsky, I. O. Loginov, E. I. Mashbitz, R. P. Mil-rud, E. S. Polat, N. F. Talizina, I. V. Robert, A. V. Khutorskoy and others.

All this forces us to take a fresh look at this problem. Many university staff are introducing a new paradigm of higher education aimed at training specialists in accordance with modern requirements. New goals and objectives of higher education are the realization of the creative potential of the student's personality through the development of a culture of personality in which the word is understood in a broad sense, the development of an understanding of responsibility for the culture of speech, interethnic communication; formation of the need for self-education, awareness of the need for continuous replenishment of knowledge through obtaining information and applying them in practice; it also consists in organizing training on an interactive basis. The global computer network Internet is a very valuable resource for the implementation of various models of language learning. At its core, the Internet is one of the best sources of information, which is often presented in an interactive format and is not available in a traditional textbook or textbook. However, speaking about online learning, we must understand that the Internet is a means of implementing didactic tasks, and the main purpose of learning as a non-linguistic specialty is the formation of communicative competence, that is, abilities. The peculiarity of language learning is that the purpose of learning is not to know the subject itself, i.e. language (language competence), and the development of certain skills and abilities of various kinds. When teaching a language, it is necessary to organize independent actions of students (and each student) in the form of speech activity in which they study. If students are taught to read, then each student should be given the opportunity to study and practice reading. When learning to speak, everyone should be given the opportunity to speak out, to express their thoughts in a foreign language. When teaching listening, each student must listen to speech in the language being studied. The famous methodologist I.V. It is important to keep in mind Rakhmanov's opinion: auditory motor skills are the basis for learning any type of speech activity, therefore, when forming skills, verbal exercise is necessary in any type of speech activity. Students have different levels of language proficiency. In this regard, teachers need to know where to find Internet resources that meet such a variety of tasks, and properly organize work in the classroom in order to increase its effectiveness without changing the program itself. There are a huge number of resources on the World Wide Web that can help expand learning opportunities within the existing program. The teacher's use of original material obtained from Internet resources in lessons, a diverse information search based on real data, allows you to form reading skills and abilities. In addition, the teacher determines the complexity of the articles in accordance with the curriculum and selects topics and content depending on the purpose of the work with reading. Modern articles about the latest achievements of students in their specialty, as a rule, arouse the interest of future engineers, broaden their horizons, stimulate further work in the classroom, and allow them to feel the connection of today with the problems of science and technology. Reading such material motivates students, activates the state of their preparation for their future profession. A special place in language learning is occupied by students' independent search for material to obtain information, its processing and use in preparing an abstract. Such an organization of individual training gives the student the

opportunity not only to choose a topic of interest to him, but also to determine the complexity of the resource that allows him to fully realize his potential. Undoubtedly, using the Internet to prepare for an essay will help in solving a wide variety of tasks of training a future engineer, starting from getting acquainted with the latest technologies of the world and ending with searching and reading various information and ending with web pages and websites in a multimedia laboratory equipped with a computer with a direct Internet connection not only for the teacher, but also for For every student, using a language is a big step in optimizing learning. It is so versatile that it can include individual work, work in pairs, threes and large groups, as well as various levels of student training. The possibilities of using Internet technologies are endless. However, it should not be forgotten that the Internet is only an auxiliary technical means of learning, and in order to achieve optimal results, it is necessary to correctly combine its use in the lesson process. It is necessary to competently direct the student to use such valuable and enormous potential, to motivate correctly and in a timely manner and the desire to use the resources obtained during language learning.

In today's rapidly evolving society, the demand for modern professionals who possess critical thinking skills, adaptability, and cross-cultural competence is on the rise. To meet these demands, educational programs must integrate modern pedagogical strategies and learning technologies while retaining valuable traditional methods.

The internet has emerged as a powerful tool in education, offering a wealth of resources and opportunities for interactive learning experiences, particularly in language education. However, effective guidance from teachers is essential to navigate this vast online landscape.

While internet technologies hold great promise for enhancing language education, it's crucial to remember that they are tools to support the fundamental goals of learning. By striking a balance between leveraging internet resources and maintaining a focus on core learning objectives, we can create dynamic and effective learning environments that prepare students for success in a globalized world.

References:

1. Akramova, N. M., Nigmatullina, A. S., & Galiakberova, A. R. (2019). Fostering the process of learner autonomy in foreign languages classrooms. Problemy sovremennoi nauki i obrazovaniya, (11-2), -53-55 p.
2. Bogomolov A. N. Virtual'naya sreda obucheniya russkomu yazyku kak inostrannomu: lingvokul'turologicheskii aspekt. M.: MAKS Press, 2008, -319 s.
3. Crystal, D. (2002). Language and the Internet. IEEE Transactions on Professional Communication, 45(2), -142-144 p. https://doi.org/10.1109/TPC.2002.1003702
4. Rinatovna, G. A., Shamsunovna, N. A., & Muzaffarovna, A. N. (2019). Using information and communication technologies to develop writing competence of students at the lessons of the English language. Vestnik nauki i obrazovaniya, (20-3 (74)), -8-10p.
5. Erenchinova, E. B. (2014). Ispol'zovanie seti Internet pri obuchenii inostrannomu yazyku. In Innovatsionnye pedagogicheskie tekhnologii: materialy Mezhdunar. nauch. konf. (g. Kazan', oktyabr' 2014g.). Kazan', -325-327 s.

O‘ZBEK TILI MORFOLOGIK LUG‘ATINING LINGVISTIK TA’MINOTI

Fattaxova Diloram Abduraxmanovna
p.f.n., Mirzo Ulug‘bek nomidagi
O‘zbekiston Milliy universiteti.

Jumanazarova Dildora
Mirzo Ulug‘bek nomidagi
O‘zbekiston Milliy universiteti magistranti.

Annotatsiya: Lingvistik korpus orqali tillarni tahlil qilish va o‘rganish, qonuniyatlarini tadqiq qilish bugungi tilshunoslik sohasidagi yutuqlardan biri hisoblanadi. Ushbu maqolada o‘zbek tili morfologik lug‘atining dasturiy ta’minoti strukturasi va vazifalari keltirilgan. Ta’lim jarayonida ushbu dasturiy ta’minotning ahamiyati foydalanuvchilar uchun qulay va samarali tomonlari yoritib berilgan.

Kalit so‘zlar: kompyuter lingvistikasi, dasturiy ta’minot, lingvistik ta’minot, morfologik lug‘at (lingvistik lug‘at), model, elektron lug‘at, morfologik analiz.

Abstract: Analyzing language through the linguistic corpus and researching the ways and laws of its learning is one of the achievements in today’s field of linguistics. The aim of this article is to increase the number of these achievements by one more by bringing innovation to computer linguistics. Also, the article presents the structure and functions of the software of the morphological dictionary of the Uzbek language, the importance of the software, user-friendly and effective aspects of computer linguistics will be highlighted in the educational process.

Keywords: computer linguistics, linguistic support, morphological dictionary (linguistic dictionary), model, electronic dictionary, morphological analysis.

Аннотация: Анализ языков посредством языкового корпуса и исследование путей и закономерностей его изучения является сегодня одним из достижений в области языкознания. В статье представлены структура и функции программного обеспечения морфологического словаря узбекского языка. Также подчеркивается важность и эффективность программного обеспечения в образовательном процессе.

Ключевые слова: компьютерная лингвистика, программное обеспечение, лингвистическая поддержка, морфологический словарь (лингвистический словарь), модель, электронный словарь, морфологический анализ.

Tilshunoslikning kompyuter texnologiyalari bilan bog‘langan yangi bir sohasi -bu kompyuter lingvistikasidir. Kompyuter lingvistikasi fani XX asrning o‘rtalarida matematik lingvistika fani asoslaridan kelib chiqqan bo‘lib, tabiiy tilning kompyuter imkoniyatlari asosida reallashgan ko‘rinishidir. Bugungi kunda mamlakatimizda kompyuter lingvistikasi rivojlanishi, u bilan bog‘liq muammolarni bartaraf qilish borasida jadal ishlar olib borilmoqda. O‘rni kelganda A.K.Po‘latov [1], S.Muxammedova, A.Rahimov[2], S.Rizayev, L.Abduhamidova, N.Abdurahmonova [3] singari olim va olimalarning ishlarini alohida ta’kidlab o‘tish lozim. Kompyuter lingvistikasi masalalarida umumiylik shundaki, ularning har biri amaliy jihatdan matn

muammosiga, matnni analiz hamda sintiz qilishga borib taqaladi. Bu muammolarni kompyuterda xal qilishda maxsus ishlab chiqilgan dasturiy ta'minot muhimdir. N. Abdurahmonovaning "Kompyuter lingvistikasi" kitobida [4] yuqorida keltirib o'tilgan termin quyidagicha izohlangan: "Lingvistik ma'lumotlar bazasi (ta'minoti) deyilganda, tilga oid barcha ma'lumotlar tushuniladi. Til tuzilishi hamda matnni analiz va sintez qilishning aniq vazifalariga asosan lingvistik ta'minotida quyidagilarning mavjud bo'lishi talab qilinadi: matn − leksik asos, leksema shakl − so'z birikmasi, sintaktik gap strukturalari. Tarjimadan avvalgi tahlil jarayonida uch asosiy iyerarxik bosqich amalga oshiriladi: matn-jumla-leksema shakl. Bunda leksema shakli keng ma'noda matndan olingan istalgan segment birlik sifatida qaraladi. Gap va so'z birikmalarining barcha strukturaviy hamda boshqa xususiyatlari leksema shakl tushunchasida ifodalanadi. Bundan ko'rinib turibdiki, lingvistik ta'minot kompyuyter lingivtikasi va tilni doimiy bog'liqligini ta'minlab turadi "[Abdurahmonova., 2021; 31]. Dasturiy ta'minot 3 guruhga bo'linadi: 1-Sistema dasturlari (unga turli yordamchi vazifalarni bajaruvchi dasturlar kiradi: Task Manager (Windows OSda mavjut), 2-Amaliy (unga foydalanuvchiga aniq bir foydalanish sohasida ma'lumotlarga ishlov berish va qayta ishlashni amalga oshiruvchi dasturlar, masalan: Microsoft Office, Adobe CC), 3-Uskunaviy dasturlar (bular dasturlash uchun ishlatiladigan dasturlar). Shu o'rinda aytib o'tish lozimki, mashina tarjimasiga asoslangan dasturlarni birinchi tomonda uning lingvistik baza tomoni tursa, ikkinchi tomonda uning dasturiy ta'minoti jihati e'tiborga olinadi.

Ma'lum bir til yoki ikki va undan ortiq til so'zlarini yig'ib, sistemalashtirilgan holda lug'at tarzida chiqarish leksikografiya hisoblanadi. Lug'atlar mohiyati jihatdan quyidagicha turlarga ega: qomusiy lug'atlar (ensiklopedik) hamda lingvistik lug'atlar (morfologik). Eng keng tarqalgan lug'atlar lingvistik lug'atlar hisoblanadif. Muayyan til doirasida lingvistik lug'atlar imlo, orfoepik, izohli, frazeologik, etimologik va hokazolarga bo'linadi. Morfologik lug'atlarning obyekti so'zdir. Ular so'zlarning to'g'ri talaffuzi, yozilishi, semantikasi, etimologiyasi, grammatik xususiyatlari talqini bo'yicha ma'lumotlarni olib yuradilar. Lingvistik lug'atlarda ma'lum bir tilning deyarli barcha so'zlari mavjud holda bo'ladi, ya'ni lingvistik lug'atlar so'zligida barcha mustaqil va yordamchi so'zlar qayd etiladi. Tabiiy tilni avtomatik qayta ishlash muammosi yarim asrdan oshdiki, dolzarb bo'lib qolmoqda. Lingvistik analizatorlar esa matnga avtomatik ishlov berish vositalari sifatida muhim[Iskandarova, O. 2019,2021]. Morfologik analizator – lug'atdagi alohida so'z/so'zshakllarni taqqoslovchi va so'zlarning grammatik xususiyatlarini topuvchi algoritmlar majmui. Morfologik analizator morfologik lug'atlar yordamida tahlil qilingan jumladagi so'zlarning ketma-ketligi, morfologik xususiyatini aniqlaydi. Sintaktik analizator esa shajara daraxti tuzishni amalga oshiradi. Morfologik analizning vazifasi matndagi har bir so'zning qaysi turkumga mansubligini avtomatik tarzda muayyanlashtirish; so'zlarning qaysi leksik-grammatik sinfga tegishliligini aniqlashdan iborat. Morfologik analiz va sintez uchun lingvistik protsessorning tarkibiy qismlaridan biri – morfologik analizatordan foydalaniladi. Morfologik va semantik analizatorlar orasidagi oraliq bog'lanish sintaktik analizatordir. Sintaktik analizator, bir tomondan, o'rganilgan so'z shakllarining morfologik strukturalarini sintaktik strukturaga o'tkazish uchun mas'ul bo'lsa, ikkinchi tomondan, matnni yanada aniqroq semantik ifodalash usulidir. Sintaktik analizator – leksemalarning ketma-ketligini grammatik xususiyatlari bilan taqqoslovchi dastur. Har qanday parserning eng muhim vazifasi morfologik tahlil davomida olingan leksemalarning morfologik xususiyatlari, lug'at ma'lumotlaridan foydalanib, tabiiy tilda gapning sintaktik strukturasini tuzishdir. Xuddi morfologik analizator singari parser ham analiz va sintezni o'z ichiga oladi [6]. Tabiiy til matnlarini bir vaqtning o'zida morfologik va sintaktik tahlil

qilish uchun parserlar matnni so‘z va jumlalarga ajratish, son shakli va boshqa grammatik xususiyatlarni aniqlash, so‘zning kontekstini hisobga olgan holda so‘zni dastlabki shakliga keltirish, jumlada so‘zlarning sintaktik munosabatlarini aniqlash, ega-kesimni topish kabi vazifalarni bajaradi. Bundan tashqari, morfologik analizatorda matndagi imlo xatolarini tuzatish rejimi ham mavjud[Iskandarova, O. 2019,2021]. Masalan, ot so‘z turkumidan iborat imlo lug‘atidagi so‘zlarni misol sifatida keltirsak va morfologik analiz qilsak. *Avomshunos – ot so‘z turkumi*, yasalishiga ko‘ra *sodda yasama ot*, yasalish usuli *affiksatsiya ko‘rinishida*, *yakka ot, aniq ot, sanaladigan ot*; *abzalsoz – ot so‘z turkumi*, yasalishiga ko‘ra *sodda yasama ot*, yasalish usuli *affiksatsiya ko‘rinishida, yakka ot, aniq ot, sanaladigan ot*; *agroiqtisodiyot – ot so‘z turkumi*, yasalishiga ko‘ra *qo‘shma ot*, yasalish usuli *kompozitsiya ko‘rinishida*, *aniq ot, sanalmaydigan ot*; *azm-iroda – ot so‘z turkumi*, yasalishiga ko‘ra *juft ot*, yasalishiga ko‘ra *sodda ot, jamlovchi ot, mavhum ot, sanalmaydigan ot* va hokazolar. Bu kabi leksemalar qatorini minglab davom ettirish va morfologik jihatdan tahlil qilish mumkin. O‘zbek tilidagi bu kabi leksemalarning har birini morfologik analiz qilib uning elektron lug‘at versiyasini yaratish orqali kelajakdagi bo‘lajak ishimiz uchun lingvistik baza yaratish mumkin. Istalgan tildagi matnlarni avtomatik tahrir va tahlil qilishda yoki elektron lug‘at tizimi orqali qidiruv berilganda kompyuter xotirasida muayyan lingvistik ta’minot bo‘lishi talab qilinadi. Ta’minot asosini esa lingvistik qoidalar va me’yorlar tashkil etadi. Shuningdek, lingvistik ta’minotdan muayyan tilning lingvistik va filologik lug‘at boyligi ham o‘rin oladi. Shu bois lingvistik protsessorni yaratishda qayta ishlanayotgan tabiiy til haqida yetarli darajada ma’lumotga ega bo‘lish talab qilinadi. Bu, o‘z navbatida, lingvistik protsessorning mukammal bo‘lishini ta’minlaydi. O‘zbek tilidagi matnlarni avtomatik tahrir va tahlil qilish dasturining lingvistik protsessorini yaratishda tilshunosga o‘zbek adabiy tilning lingvistik lug‘atlari va grammatik qoidalar jamlammasi zarur bo‘ldi [Абжалилова, Рашидова., 2021: 5:22-25].

Tilimizdagi barcha so‘zlar o‘zbek tilining lug‘at boyligini tashkil qiladi. Bu leksemalarni matnni qayta ishlash, matnning avtomatik tahlili va tahriri, mashina tarjimasini amalga oshiruvchi dasturlarning lingvistik ta’minotida alohida maydonni tashkil etishini namoyon qiladi[Iskandarova, O. 2019,2021]. Shu bois avtomatik tarjima qilish dastur, mobil ilova va platformalarning lingvistik tahminotini yaratishda ham morfologik analiz qilingan leksemalar bazaga leksik birlik shaklida kiritiladi va o‘zga tillarda ularning tarjimasi beriladi. Leksemalarni kompyuterli modellashtirish quyidagi imkoniyatlarni beradi:

— analiz qilingan leksemalarni avtomatik tahlil qilishni osonlashtiradi;

— matn tarkibida ularning tarjimasini osonlashtiradi;

— lingvistik protsessorda semantik tarzda so‘z yoki jumla identifikatori tarzida tushuniladi;

— lingvokulturologik ahamiyati saqlanib qoladi.

Foydalanilgan adabiyotlar:

1. Пулатов А. Глобаллашув даврида лингвистикага оид муаммолар// Тилшуносликнинг долзарб масалалари. Т., 2004.
2. Рахимов А. Компьютер лингвистикаси асослари. Т.: Akademnashr, 2011.
3. Abdurahmonova N. Kompyuter lingvistikasi darslik, “Nodirabegim”, – Toshkent: 2021. Abduraxmonova N. Kompyuter lingvistikasi (darslik), Toshkent: Nodirabegim, - 200 b.

4. Iskandarova, O. R. (2021). OLIY TA'LIMDA RUSIYZABON GURUH TALABALARIGA "O 'ZBEK TILI-DAVLAT TILI","O 'ZBEK TILINING IMLO QOIDALARI" MAVZUSINI O'RGATISH BO 'YICHA AYRIM TAVSIYALAR. Oriental renaissance: Innovative, educational, natural and social sciences, 1(9), 1039-1047.
5. Iskandarova, O. R. (2021). MAQSUD SHAYXZODA SHE'RIYATIDA OZARBAYJON VA O 'ZBEKISTON MADHI. Oriental renaissance: Innovative, educational, natural and social sciences, 1(9), 683-688.
6. Iskandarova, O. (2019). THE ARTISTIC BURDEN OF THE STORY "THE SIX AND SEVENTH" IN THE NOVEL. Theoretical & Applied Science, (11), 121-124.
7. Iskandarova, A. (2019). THE ARTISTIC BURDEN OF THE STORY "THE SIX AND SEVENTH" IN THE NOVEL. Theoretical and Applied Science. https://doi.org/https://dx.doi.org/10.15863/TAS
8. Abdurahmonova N. Mashina tarjimasining lingvistik ta'minoti, Akademnashr, Toshkent, 2018 -31b.
9. https://cyberpedia.su/3x7ca2.htm

BERT MODELINING QO‘LLANILISHI VA VAZIFALARI

Berdiyev Jahongir Botir o‘g‘li
Alisher Navoiy nomidagi
Toshkent davlat o‘zbek tili va adabiyoti universiteti,
kompyuter lingvistikasi yo‘nalishi magistranti.

Annotatsiya: Til modellari mashinaga tabiiy tillarni yaxshiroq tushunishga yordam beradigan vosita hisoblanadi. Hozirgi kungacha bir qancha til modellari yaratilgan va ularning barchasi ma’lum bir vazifani bajarish uchun mo‘ljallangan. Shunday modellardan biri BERT modelidir. Bu model boshqa modellardan o‘zining afzalligi, yutuqlari, imkoniyatlari bilan ajralib turadi. Ushbu til modeli matnlar ustida ko‘plab amallar bajarish va juda aniq darajadagi natijalarni taqdim eta olishi bilan mashhurdir. Misol tariqasida matnlardagi va mediadagi hissiyotlarni aniqlash, matnlarni va tasvirlarni tasniflash, mashina tarjimasi, tillarni va nutqni aniqlash, spamni aniqlash, savollarga javob berish, matn yaratish kabi ko‘plab amaliyotlarni bajarishda muhim vosita ekanligini ta’kidlab o‘tish joiz.

Mazkur maqolada BERT modelining yaratilish tarixi, bu modelning yaratilishiga bo‘lgan ehtiyoj, uning imkoniyatlari haqida fikr yuritiladi. Bundan tashqari, BERT modeli qanday ishlaydi, BERT qanday o‘qitiladi, BERT modeli o‘zi nima kabi savollarga ham javob berilgan. Aytish mumkinki, BERT modelining qo‘llanilishi tabiiy tilga ishlov berishda samarali natijalarni beradi.

Kalit so‘zlar: BERT, tasniflash, mashina tarjimasi, mashinali o‘qitish

Abstract: Language models are tools that assist machines in comprehending natural languages. Several language models have been created to date, each designed to perform a specific task. One such model is the BERT model, which is distinguished from others by its advantages, achievements, and capabilities. This language model is renowned for its ability to perform various operations on texts and provide highly accurate results. As an example, it is important to note that BERT is a valuable tool for various practices, including emotion detection in texts and media, text and image classification, machine translation, language and speech recognition, spam detection, question answering, and text creation. This article discusses the history of BERT’s creation, the need for its development, and its potential applications. Additionally, this text answers questions about the functionality, training, and purpose of the BERT model. It can be concluded that implementing the BERT model yields effective results in natural language processing.

Keywords: BERT, classification, machine translation, machine learning

Аннотация: Языковые модели — это инструменты, которые помогают машинам лучше понимать естественные языки. На сегодняшний день создано несколько языковых моделей, и все они предназначены для выполнения конкретной задачи. Одной из таких моделей является модель BERT. Данная модель отличается от других моделей своими преимуществами, достижениями и возможностями. Эта языковая модель известна своей способностью выполнять множество операций над текстами и предоставлять очень точные результаты. В качестве примера следует отметить, что это важный инструмент во многих

практиках, таких как обнаружение эмоций в текстах и медиа, классификация текстов и изображений, машинный перевод, распознавание языка и речи, обнаружение спама, ответы на вопросы, создание текста.

В данной статье рассматривается история создания модели BERT, необходимость создания этой модели и ее возможности. Кроме того, даются ответы на такие вопросы, как, как работает модель BERT, как обучается BERT и что такое модель BERT. Можно сказать, что применение модели BERT дает эффективные результаты при обработке естественного языка.

Ключевые слова: BERT, классификация, машинный перевод, машинное обучение

Inson mavjudki, u doim rivojlanib, yangilikka intilib yashaydi. Inson qiyinchilik ko‘rgani sari o‘zi uchun yengillik yaratishga, yaxshi yashash uchun o‘ziga qulay imkoniyatlar yaratishga intilaveradi. Bu esa inson oldiga ko‘plab vazifalar, maqsadlar qo‘yadi. Har bir inson o‘z imkoniyati va qobiliyatidan kelib chiqib qaysidir sohani rivojlantirishga intiladi. Kimdir tabiatshunoslik bilan shug‘ullansa, yana kimdir sanoatni rivojlantirishga hissa qo‘shadi. Xuddi shunday sohalardan biri kompyuter texnologiyalari sohasidir. Kundan kunga ushbu sohaga bo‘lgan ehtiyoj tobora ortib bormoqda. Bu esa ko‘plab sohalarda bo‘lgani kabi tilshunoslik sohasiga ham o‘z ta’sirini o‘tkazmoqda. Buning natijasi o‘laroq tabiiy tilni qayta ishlashda ham ko‘plab izlanishlar, tadqiqotlar yuzaga keldi. Jahon tajribasida bu bo‘yicha samarali ishlar amalga oshirilgan bo‘lsada, yurtimizda bu sohada yetarlicha tadqiqotlar amalga oshirilmagan. Shunday ekan, tabiiy tilni qayta ishash uchun eng mashhur til modellaridan biri BERT modeli orqali tabiiy tilga ishlov berish eng samarali natijalarni beradi.

Jahonda til modellarini yaratish va ulardan foydalanish bo‘yicha ko‘plab tadqiqotlar olib borilgan. Jumladan, BERT modelining yaratilishi tabiiy tillarni qayta ishlashda yangi bosqich bo‘ldi. Ushbu model Google kompaniyasining tadqiqotchilari, xususan, uning Brain jamoasi (Y.Devling, M.V. Chang, K.Li, K.Toutanova) tomonidan 2017-yilda ishlab chiqilgan va 2018-yilda ochiq manbaali transformator sifatida e’lon qilinadi.[1] Shuningdek, jahonda BERT modelidan foydalanib matnlar bilan bog‘liq turli xil ishlar amalga oshirildi. Jumladan, D.Nozza va boshqalar[2], M.Baz va boshqalar[3], N.Baratalipour va boshqalar[4], Y.SH. Chuang[5], P.Jain va boshqalar[6],M.Ashraf va boshqalar[7], D.Kiela va boshqalar[8], A.S.Talaat[9], Z.Guo va M.L.Nguyen[10],

[1] Devling J., Chang M.W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding / J.Devling, M.W.Chang, K.Lee, K.Toutanova - Proceedings of NAACL-HLT 2019, pp. 4171–4186
[2] https://towardsdatascience.com/sentiment-analysis-and-emotion-recognition-in-italian-using-bert-92f5c8fe8a2
[3] Baz M., Sami A. , Prottasha N., Kowsher Md, Murad S., Bairagi A., Masud M. (2022). Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning. Sensors. 22. 10.3390/s22114157.
[4] Baratalipour N., Suen Y., Ormandjieva O. (2020). Abusive Language Detection Using BERT Pre-trained Embedding. 10.1007/978-3-030-59830-3_60.
[5] https://arxiv.org/abs/2307.15331
[6] Jain P., Taneja K., Taneja H. (2024). Advertisement Image Classification Using Deep Learning with BERT: A Novel Approach Exploiting Textual Features. 10.1007/978-981-99-6547-2_34.
[7] Ashraf M., Zhou X., Vivone G., Chen L., Chen R., Majdard R. (2024). Spatial-Spectral BERT for Hyperspectral Image Classification. Remote Sensing. 16. 539. 10.3390/rs16030539.
[8] https://arxiv.org/abs/1909.02950
[9] https://rdcu.be/dzKUk
[10] Guo Z., Nguyen M.L. / Proceedings of the1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10 th International Joint Conferenceon Natural Language Processing: Student Research Workshop, pages 101–107 December 4 - 7, 2020.

T.Patzak[11], I.V.Selivanova[12], R.Buoy va boshqalar[13] kabi tadqiqotchilar BERT modeli yordamida hissiyotlar tahlili, tasvirlar va matn tasniflari, tarjima jarayonlari, kiritilgan matnning qaysi tilga tegishli ekanligini aniqlovchi va shu kabi matnlar ustida turli xil amallar bajaruvchi bir qancha dasturlar yaratgan.

O‘zbek tadqiqotchilari tomonidan ham BERT modeli asosida olib borilgan bir qancha ishlarni ko‘rishimiz mumkin. Masalan, matnlarni tasniflash uchun E.Kuriyozov va boshqalar[14] tomonidan yaratilgan BERTbek modelini ko‘rishimiz mumkin.

Tabiiy tillarni qayta ishlash sohasidagi ilk qadamlar XX asrning 50-yillaridan boshlangan va bu hozirgi kunigacha bir qancha bosqichlarni bosib o‘tgan. Tabiiy tilni qayta ishlash uchun harakatlarning boshlanganiga yarim asrdan oshdi va bu davr mobaydida tabiiy tilni qayta ishlash uchun ko‘plab til modelllari yaratildi. Bu til modellarining barchasi o‘z vaqtida ma’lum bir yutuqlarga erishgan bo‘lsa-da, shunga yarasha kamchiliklardan ham holi emas edi. BERT modelining yaratilishi esa tabiiy tilni qayta ishlash borasidagi ko‘plab muammolarni hal qildi hamda tabiiy tilni qayta ishlashda boshqa modellarga nisbatan ancha yaxshi natijalarni bera boshladi.

BERT tabiiy tilni qayta ishlovchi til modeli bo‘lib, u 2017-yilda Google tadqiqodchilari jamoasi tomonidan ishlab chiqilgan va 2018-yilda ochiq manbaa sifatida ommaga e’lon qilingan. Uning kengaytmasi Bidirectional Encoder Representations from Transformers. Tarjimasi transformatorlardan ikki tomonlama kodlovchi tasvirlar deyiladi. Ikki tomonlama kodlovchi deganda model kiritilgan ma’lumotlarni chapdan o‘ngga va o‘ngdan chapga harakatlangan holda o‘qiydi. Katta hajmdagi oldindan o‘qitilgan matnlar GPT-2, BERT, RoBERTA, ALBERT kabi keng tarqalgan ko‘plab til modellari yordamida tabiiy tillarni tushunish vazifalarida qo‘llaniladi. Ular orasida BERT eng samarali hisoblanadi[Guo Z. va boshqalar, 2020: 2]. Misol tariqasida BERT modelini GPT bilan solishtirishimiz mumkin. GPT BERTdan farqli o‘laroq ma’lunotlarni bir tomonlama, ya’ni chapdan o‘ngga harakatlanib o‘qiydi, ya’ni so‘zning faqat chap tomonini hisobga oladi. BERT modeli esa kontekstdagi butun so‘zlar to‘plamidan foydalanadi(1-rasmga qarang). Shuning uchun BERTni yo‘nalishsiz harakatlanuvchi model ham deyishadi. Buni quyidagi rasmda ko‘rishimiz mumkin:

BERT o‘zgaruvchan neyron tarmoq modellari asosida yaratilgan katta miqdordagi axborotni qayta ishlaydigan modellardan biridir. “Bu model orqali matnni tasniflash, his-tuyg‘ularni tahlil qilish, spamni aniqlash, savollarga javob berish kabi vazifalarni yaratish uchun BERT modeli arxitekturasiga faqatgina ba’zi o‘zgarishlarni kiritish kifoya qiladi[Berdiyev J., 2024: 176].” BERT zamonaviy neyron tarmoq modeli bo‘lib, ma’lumotni tekshirish va belgilashni bir qadamda bajaradi. Bu esa ma’lumotlar ustida ko‘plab amallar bajarishga imkon beradi. Jumladan, BERT modeli hissiyotlarni tahlil qilish, savollarga javob berish, tasvirlarni va matnlarni tasnif qilish, mashina tarjimasi va shu kabi ko‘plab amallarni bajara oladi. Misol uchun, biz BERT

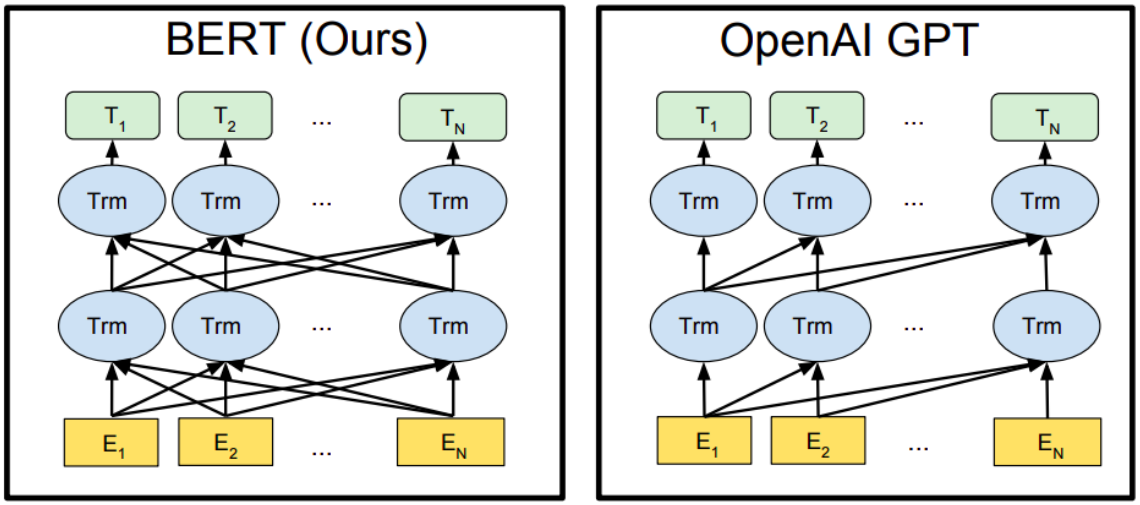
[11] Patzak T. AutomatischeKlassifikationvon Alltagsszenarienmiteinem BERT-basiertenSiamese-Netzwerk. T.Patzak - Humboldt-Universitet zu Berlin, 2021.

[12] Селиванова И.В. Методы тематической классификации научных текстов на основе теоретико-информационного подхода

[13] https://arxiv.org/abs/2112.06748

[14] https://www.researchgate.net/publication/368877295_Text_classification_dataset_and_analysis_for_Uzbek_language

modelini birinchi navbatda Vikipediya (maskali LM vazifasi) kabi korpusda o‘rgatamiz va keyin bir nechta qo‘shimcha qatlamlarni qo‘shish orqali sharhlarni salbiy, ijobiy yoki neytral deb tasniflash kabi vazifasini bajarish uchun modelni o‘z ma’lumotlarimizga moslashtiramiz. Amalda, biz faqat tasniflash vazifasi uchun [CLS] tokenidan foydalanamiz[1]. Bundan tashqari BERT modeli orqali mashina tarjimasi yaratish ham juda yaxshi samara beradi va tarjima aniqligini oshiradi. Yuqorida ta’kidlaganimizdek, BERT modeli oldindan o‘qitilgach, boshqa vazifalar singari mashina tarjimasi uchun ham BERT modelini o‘rgatish uchun bir nechta tillardagi katta hajmli matnlarda oldindan tayyorlanadi. Oldindan tayyorgarlik jarayonida BERT modeli jumlalar ichida yetishmayotgan so‘zlarni bashorat qilishni o‘rganadi, turli tillarning grammatikasi, sintaksisi va semantikasini samarali o‘rganadi. Bundan tashqari, nozik sozlash usullari oldindan tayyorlangan BERT modelini aniq tarjima vazifalariga moslashtiradi va uning ishlashini aniq va ravon tarjimalar uchun optimallashtiradi.



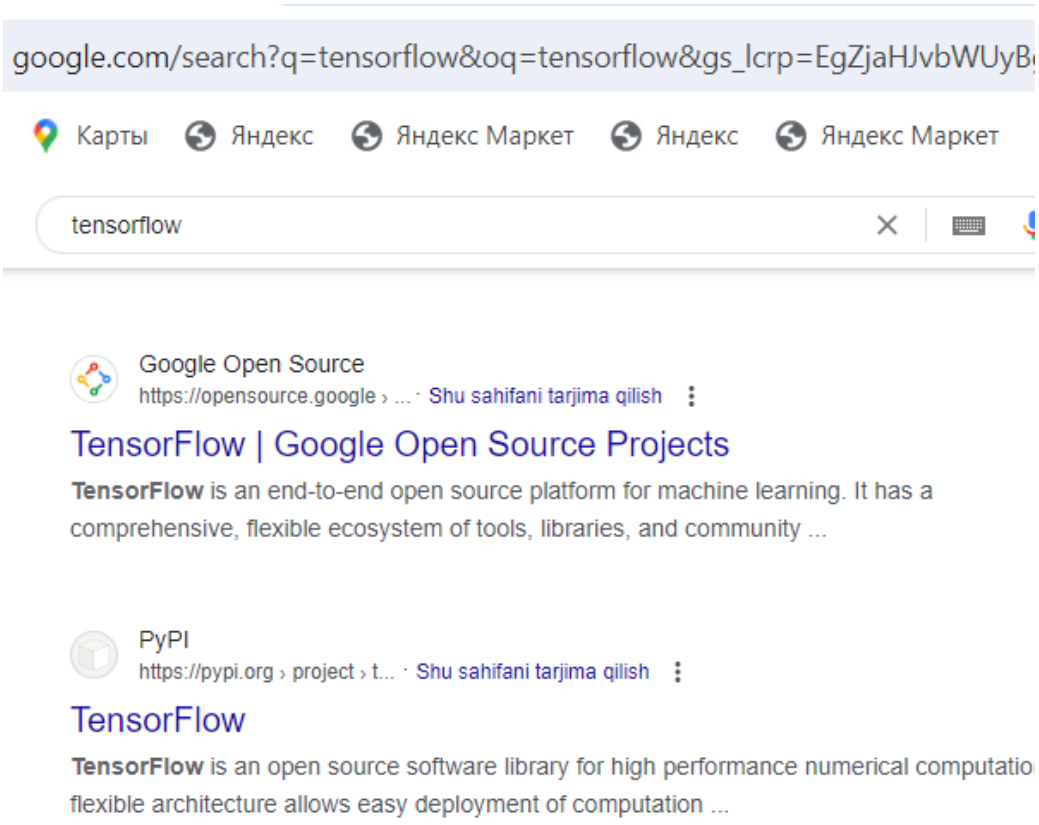
1-rasm. BERT va GPT modellarnining ishlash sxemasi

BERT modelining asosiy xususiyati matn ustida amallarni bajarishda konteksdagi so‘zlarning butun to‘plamidan foydalanishidir. Boshqa til modellari esa matnning chap yoki o‘ng tomonidagi so‘zlardan foydalangan holda natijani taqdim etadi. BERT modelining boshqa til modellaridan yana bitta farqi shuki, bu model bir vaqtning o‘zida ikkita vazifa bo‘yicha o‘qitiladi, ya’ni keying jumlani bashorat qilish va yetishmayotgan tokenni yaratish. Buni D.Samuel va boshqalar shunday tushuntiradi: “Asl BERT modeli ikkita o‘z-o‘zini boshqaradigan o‘quv maqsadi - niqoblangan tilni modellashtirish (MLM) va keyingi jumlani bashorat qilish (NSP) orqali o‘qitiladi”[Samuel D. Va boshqalar, 2023: 1958]. Buni Googledan foydalangan paytimizda qidiruv uchun jumlani yozish jarayonida tizim biz uchun taxminiy tavsiya berishida ko‘rishimiz mumkin.



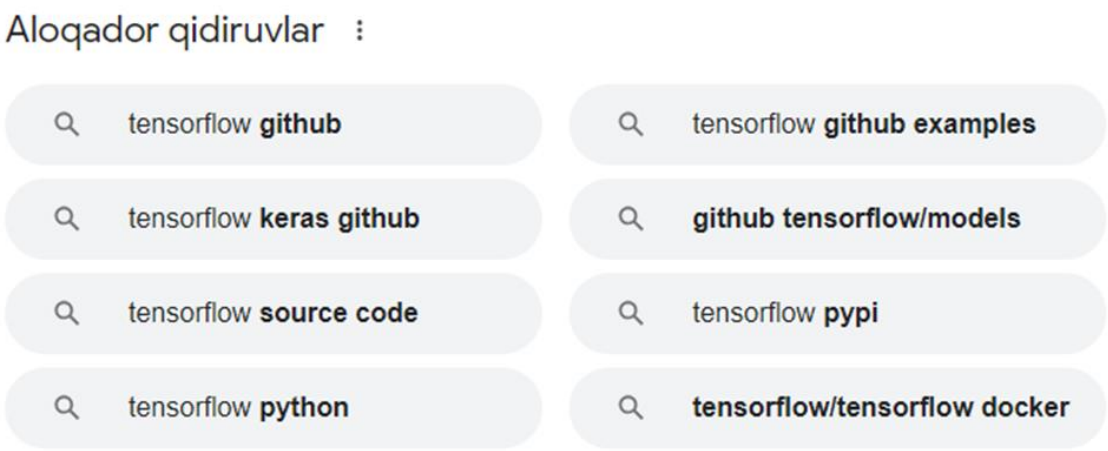
2-rasm. Googlening qidiruvda yetishmayotgan jumlaning taxminiy tavsiyasi

Bu yerda Googlega *ten* degan bo‘g‘inni kiritishimiz bilan bizga so‘zning taxminiy davomini va biz qidiruvda ko‘p foydalanadigan so‘zlarga tegishli taxminiy natijalarni tavsiya qilib beryapti(2-rasmga qarang). Agar biz *tensorflow* so‘zini qidiruvga bersak, quyidagicha natija chiqarib beradi:



3-rasm. Qidiruv natijasida foydalanuvchiga berilgan tavsiyalar

Ko‘rinib turganidek, Google *tensorflow* so‘ziga tegishli bir qancha natijalarni chiqarib berdi(3-rasmga qarang), bundan tashqari, aloqador qidiruvlar qismida qidirilayotgan so‘zga aloqador, foydalanuvchiga kerak bo‘lishi mumkin bo‘lgan yoki qiziq bo‘lgan bir qancha tavsiyalarni ham berdi(4-rasmga qarang).



4-rasm. Foydalanuvchiga foydasi tegishi mumkin bo‘lgan tavsiyalar

BERT modeli matnlar bilan bog‘liq bir qancha vazifalarni bajaradi. Jumladan, matnlarni tasniflash, hissiyotlarni tahlil qilish, tarjima jarayoni, tilni aniqlash (matnning tili qaysi tilga tegishli ekanligi) kabi amaliyotlarni misol keltirishimiz mumkin. Yuqorida ta’kidlaganimizdak, BERT modeli qanchalik yuqori aniqlikda ishlamasin, baribir u ham kamchiliklardan mustasno emas. Misol uchun ushbu model yordamida yaratilgan tillarni aniqlash va tarjima qiluvchi dasturlarga e’tibor beraylik. Tillarni aniqlovchi dastur deyarli xatosiz ishlaydi, chunki ushbu dastur bir qancha tillar bilan tanishtiriladi, ammo ularning ma’nosi ahamiyatsiz qolaveradi. Tarjima qiluvchi dastur esa bir tildagi matnning tarjimasini boshqa tilga tarjima qilishda so‘zlarning ma’nosiga diqqat qiladi. Mashina fikrlay olmaganligi uchun hatto model qanchalik katta korpusda o‘qitilmasin, ba’zan xato ishlay boshlaydi. Bu, ayniqsa, badiiy tarjimada yaqqol ko‘rinadi. Chunki badiiy asarlarda so‘zlarning ma’nosi butunlay boshqa ma’noda kelishi kuzatiladi, hatto ikki asarda kelgan aynan bir so‘z bir-biriga o‘xshamagan boshqa-boshqa ma’nolarda kelishi kuzatiladi. Bu esa mashinaning yaxshi natija bermasligiga olib kelishi mumkin.

BERT modeli yorliqsiz (oldindan tayyorlangan) katta ma’lumotlar to‘plamlarida oldindan o‘qitiladi va keyinchalik murakkabroq vazifalarga (nozik sozlash) moslashtirilishi mumkin. Ya’ni bu jarayonlarning barchasi, avvalo, matnlarni qayta ishlashdan so‘ng bajariladi, matn birinchi navbatda raqamli token identifikatorlariga aylantiriladi va tensorlarga ajratiladi[Patzak T., 2021: 3]. Matnlarni BERT modeliga kiritishdan oldin ularni raqamli token identifikatorlariga aylantirish va bir nechta Tensorlarga ajratish kerak. Buning uchun tanlangan matn TensorFlow yoki PyTorch kutubxonalaridan biriga yuklanadi. Yuklangan matn esa ushbu kutubxonalarda qayta ishlanadi, ya’ni kiritilgan matn tokenizatsiya qilinadi, so‘ngra zichlangan vektor ko‘rinishga aylantiriladi. TensorFlow matn ma’lumotlari uchun bir qator oldindan ishlov berish vositalarini, jumladan, normallashtirish (kichik harflar, tinish belgilarini olib tashlash), bir xil uzunlikdagi to‘ldirish ketma-ketligi va lug‘atdan tashqari (OOV) tokenlar bilan ishlashni taklif etadi. TensorFlow matn tasniflash modellari va ketma-ketlik modellarini (masalan, takroriy neyron tarmoqlar (RNN), uzoq

qisqa muddatli xotira tarmoqlari (LSTM) va transformatorlar) yaratish uchun Keras kabi yuqori darajadagi APIlarni taqdim etadi. TensorFlow transformator modellarini, shu jumladan BERT, GPT va T5-ni, shuningdek, maxsus ma'lumotlar to'plamlarida uzatishni o'rganish yoki nozik sozlash uchun oldindan o'qitilgan nazorat nuqtalarini ta'minlaydi. Ayniqsa, BERT kabi transformator arxitekturalari tomonidan ommalashgan e'tibor mexanizmlari modellarga kirish ketma-ketligining tegishli qismlariga e'tibor qaratishga imkon beradi va turli NLP vazifalarida ishlashni yaxshilaydi.

Xulosa sifatida shuni aytish mumkinki, BERT modeli yaratilgunga qadar bir qancha til modellari yaratilgan va ularning kamchiliklarini tuzatish maqsadida ushbu modelga ehtiyoj sezilgan va yaratilgan. BERT modelining tabiiy tilni tushunish va qayta ishlashdagi o'rni muhim ahamiyat kasb etadi. Uning kontekstni tushunishi, oldindan tayyorlangan taqdimotlar va o'rganish qobiliyati turli vazifalar va ilovalar bo'yicha NLP modellarining ishlashini yaxshilashga xizmat qiladi.

BERT modelining turli xil vazifalar bo'yicha o'qitilishi tabiiy tilni qayta ishlashda inson mehnatini sezilarli darajada kamaytiradi va osonlashtiradi. Bundan tashqari, ba'zan inson ilg'amay qoladigan xato va kamchliklarni tuzatishi mumkin. Ammo BERT modelining o'zi ham kamchiliklardan holi emasligini unutmaslik lozim. Bu modelning qay darajada aniq ishlashi uning bajaradigan vazifasiga ham bog'liq bo'ladi. Ayniqsa, uning juda aniq ishlashi va ba'zi kamchiliklar bilan ishlashini tillarni aniqlash vazifasi va mashina tarjimasida yaqqol ko'rish mumkin. Ammo shunday bo'lsa-da, tabiiy tillarni qayta ishlash bo'yicha tadqiqotlar va ishlanmalar davom etar ekan, BERT modeli inson tilini tushunishda yana-da ko'proq aniqlik bilan ishlash imkoniyatiga ega bo'la boradi.

Foydalanilgan adabiyotlar:

1. Guo Z., Nguyen M.L. / Proceedings of the1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10 th International Joint Conferenceon Natural Language Processing: Student Research Workshop, December 4 - 7, 2020.- B.101–107
2. Berdiyev J.B. BERT va TensorFlowning matnlarni tasniflashdagi ahamiyati / J.B.Berdiyev. - Educational Research in Universal Sciences (ERUS) journal, 2024. – B.176.
3. David Samuel, Andrey Kutuzov, Lilja Øvrelid, and Erik Velldal.2023. Trained on 100 million words and still in shape: BERT meets British National Corpus. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 1954–1974, Dubrovnik, Croatia. Association for Computational Linguistics.
4. Patzak T. Automatische Klassifikation von Alltagsszenarien mit einem BERT-basierten Siamese-Netzwerk / T.Patzak. - Humboldt-Universität zu Berlin, Institut für Informatik, 2021.
5. https://dzone.com/articles/bert-transformers-how-do-they-work

РОЛЬ И ФУНКЦИЯ КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ В ЯЗЫКОВОМ ОБРАЗОВАНИИ

Ахмедова Мукаддас
Кандидат педагогических наук, доцент.

Фаттахова Дилорам
кандидат педагогических наук
Национальный университет Узбекистана.

Аннотация: В статье рассматриваются вопросы использования современных информационно-коммуникационных технологий в преподавании языковых дисциплин в вузе. Информационная компетентность, ставшая одним из основных показателей качественного обучения, является одним из главных условий современного образования, и ее формирование напрямую зависит от активной деятельности обучающихся в открытой информационно-образовательной среде. Современные ИКТ позволяют создавать уникальные дидактические условия для разработки и упорядочения индивидуальных образовательных траекторий в обучении языковым дисциплинам в вузе. В работе проанализирована научно-педагогическая литература о применении средств ИКТ для языкового обучения. Анализ показал, что основными технологическими особенностями ИКТ-инструментов, которые открыли широкие возможности применения информационно-коммуникационных технологий для формирования и развития иноязычной коммуникативной компетенции являются интерактивность, мультимедийность, публичность и гипертекстовая структура.

Ключевые слова: информатизация образования, информационные технологии, языковая и речевая компетенции, информационная компетентность, компьютерная грамотность.

Annotatsiya: Maqolada universitetda til fanlarini o'qitishda zamonaviy axborot-kommunikatsiya texnologiyalaridan foydalanish masalalari ko'rib chiqiladi. Sifatli ta'limning asosiy ko'rsatkichlaridan biriga aylangan axborot kompetentsiyasi zamonaviy ta'limning asosiy shartlaridan biri bo'lib, uning shakllanishi to'g'ridan-to'g'ri talabalarning ochiq axborot va ta'lim muhitidagi faol faoliyatiga bog'liq. Zamonaviy AKT sizga universitetda til fanlarini o'qitishda individual ta'lim traektoriyalarini ishlab chiqish va tartibga solish uchun noyob didaktik sharoitlarni yaratishga imkon beradi. Ishda tilni o'rganish uchun AKT vositalaridan foydalanish bo'yicha ilmiy va pedagogik adabiyotlar tahlil qilindi. Tahlil shuni ko'rsatdiki, xorijiy kommunikativ kompetentsiyani shakllantirish va rivojlantirish uchun axborot-kommunikatsiya texnologiyalaridan foydalanishning keng imkoniyatlarini ochgan AKT vositalarining asosiy texnologik xususiyatlari interaktivlik, multimediya, oshkoralik va gipermatnli tuzilishdir.

Kalit so'zlar: ta'limni axborotlashtirish, axborot texnologiyalari, til va nutq kompetentsiyalari, axborot kompetentsiyasi, kompyuter savodxonligi

Abstract: The article discusses the use of modern information and communication technologies in teaching language disciplines at the university. Information competence, which has become one of the main indicators of quality education, is one of the main conditions of

modern education, and its formation directly depends on the active activity of students in an open information and educational environment.Modern IT makes it possible to create unique didactic conditions for the development and ordering of individual educational trajectories in teaching language disciplines at the university. The paper analyzes the scientific and pedagogical literature on the use of IT tools for language learning. The analysis showed that the main technological features of ICT tools that have opened up wide opportunities for the use of information and communication technologies for the formation and development of foreign language communicative competence are interactivity, multimedia, publicity and hypertext structure.

Keywords: informatization of education, information technologies, language and speech competencies, information competence, computer literacy.

Информатизация образования является многоуровневым, многоаспектным процессом, предполагающим не только использование инновационных форм учебных материалов и средств ИКТ в процессе обучения, но и выдвижение особых критериев информационно цифровой компетентности педагогов. Информационная компетентность, ставшая одним из основных показателей качественного обучения, является одним из главных условий современного образования, и ее формирование напрямую зависит от активной деятельности обучающихся в открытой информационно-образовательной среде.

В современном словаре дается следующее определение термина: «ИКТ включают различные программно-аппаратные средства и устройства, функционирующие на базе компьютерной техники, а также современные средства и системы информационного обмена, обеспечивающие сбор, накопление, хранение, продуцирование и передачу информации» [1, 27]. Определение ИКТ имеет множество формулировок, но в контексте анализируемой проблемы базовым является их понимание как совокупности сетевых технологий, методик и программно-технических средств, созданных и используемых с целью сбора, анализа, сохранения, распространения использования информации.

Прогрессивность обучения с применением ИКТ доказана тем, что, во-первых, их включение в образовательный процесс значительно ускоряет преемственность социального опыта и знаний от поколения к поколению; во-вторых, ИКТ способствуют более мобильной и успешной адаптации обучающихся к изменениям, происходящим в социальной сфере; в третьих, использование ИКТ в системе образования позволяет обновить и «осовременить» обучение в соответствии с социальным заказом общества. Кроме того, избранная предметная область осваивается студентами на различных уровнях, вырабатываются навыки и умения решения типовых практических задач, скорости принятия решений в нестандартных проблемных ситуациях, развиваются способности к определенным видам деятельности. [2, 156]

Процесс информатизации высшего образования в республике осуществляется с помощью электронных образовательных ресурсов, информационных систем. При анализе методической литературы [3,5] определены следующие преимущества применения средств ИКТ при обучении языкам: повышение качества и эффективности учебного процесса; повышение активности познавательной деятельности у обучаемых; реализация межпредметных связей; увеличение объёма и оптимизация поиска необходимой информации; создание единой информационно образовательной среды для студентов, учитывая их уровень и индивидуальную траекторию обучения.

Использование современных информационных технологий в процессе обучения позволяет интенсифицировать и повышать качество преподавания языковых дисциплин, увеличивать количество и объём предъявляемых студентам обучающих и контролирующих программ за счёт сокращения времени на их освоение и выполнение, развивать интерес к учебной деятельности и разнообразить её. Следует отметить, что использование в учебном процессе информационно-коммуникационных технологий (ИКТ) является эффективным способом повышения мотивации студентов к изучению учебных дисциплин. Использование компьютера, как средства отображения наглядности и поддержки процесса обучения, обладает целым рядом уникальных возможностей, облегчая студентам восприятие и понимание лексико-грамматического материала. Компьютер позволяет расширить возможности предъявления и закрепления учебного материала, вовлечь студентов в ход учебного процесса, усиливая мотивацию процесса получения новых знаний. Компьютер качественно изменяет подход к контролю за аудиторной и самостоятельной работой студентов, обеспечивая при этом гибкость управления процессом обучения и делая его объективным и доступным.[3]

Наглядность представляемого материала повышает степень его усвоения, так как оказываются задействованными зрительный и слуховой каналы восприятия студентами внешней информации. Данная задача успешно решается с использованием многократно апробированных и получивших широкое распространение мультимедиатехнологий.

Мультимедиа — это совместное представление текстовой, графической, числовой и звуковой и видеоинформации. Графическое представление учебного материала позволяет акцентировать внимание студентов на ключевых моментах изучаемой темы, даёт возможность преподавателю повысить их мотивацию и усилить интерес к изучению и запоминанию нового материала, т.е. реализовать основополагающий дидактический принцип обучения - принцип наглядности.

Использование современных технологий обучения в сочетании с мультимедийными материалами, адаптированными к специфике обучения студентов, а также образовательными ресурсами, имеющимися в свободном доступе, позволяет повысить мотивацию, уровень и качество подготовки студентов при изучении практического курса языковых предметов. Особую роль в формировании мотивации играют новизна и форма представления получаемой студентом информации, а также их включение в поисковую, творческую деятельность

Следует отметить одну из тенденций современного этапа информатизации образования — это стремление к интеграции аппаратных и программных средств ИКТ.

Попытаемся систематизировать, где и как целесообразно использовать информационные технологии в обучении, учитывая, что современные компьютеры позволяют интегрировать в рамках одной программы тексты, графику, звук, анимацию, видеоклипы, высококачественные фотоизображения, достаточно большие объемы полноэкранного видео, качество которого не уступает телевизионному:

1) при изложении нового материала — визуализация знаний (демонстрационно - энциклопедические программы; программа презентаций PowerPoint)

2) закрепление изложенного материала (тренинг — разнообразные обучающие программы, практические работы)

3) система контроля и проверки (тестирование с оцениванием, контролирующие программы);

4) самостоятельная работа учащихся (обучающие программы, энциклопедии, развивающие программы);

5) при возможности отказа от классно-урочной системы: проведение интегрированных уроков по методу проектов, результатом которых будет создание Web-страниц, проведение телеконференций, использование современных Интернет-технологий;

6) тренировка конкретных способностей учащегося (внимание, память, мышление и т.д.).[5, 106]

Подводя итог сказанному, можно сделать вывод о том, что использование современной вычислительной техники и информационных технологий позволяет достаточно успешно решать следующие задачи:

1. обеспечить индивидуальную траекторию обучения студентов
2. обеспечить визуализацию учебной информации;
3. обеспечить в реальном режиме времени самоконтроль за степенью и качеством усвоения учебного материала;
4. усилить мотивацию обучения за счёт новизны работы с современной вычислительной техникой и обучающими программами.

Информационные технологии позволяют существенно изменить организацию учебного процесса, использовать возможности вычислительной техники и электронных обучающих и контролирующих материалов с целью построения индивидуальной траектории обучения студентов

Использование возможностей существующего программного обеспечения, в частности Microsoft PowerPoint, позволяет создавать красочно оформленные и информационно насыщенные презентации, которые обладают широкими возможностями при обучении студентов фонетике, лексике языка и по осмыслению грамматических правил. Представление лексико-грамматических правил в виде текстового материала, а также таблиц и схем способствуют повышению эффективности усвоения теоретического материала, активизации мыслительной деятельности, развитию логического и абстрактного мышления студентов. Следует также отметить, что при использовании презентаций существенно уменьшаются потери времени, например, отпадает необходимость записи учебного материала на доске. При использовании тренировочных тестов реализуется возможность многократного повторения одного и того же теста, возможность использования учебников и справочных материалов без ограничения времени, отведённого на его выполнение. При выполнении контрольно-тренировочного теста студент может видеть в динамике результаты усвоения учебного материала.[4]

Проанализировав опыт применения ИКТ при проведении учебных занятий, можно сделать вывод о том, что ИКТ ускоряют процесс обучения, способствуют повышению интереса студентов к изучаемой дисциплине, улучшают качество усвоения материала, позволяют индивидуализировать процесс обучения и обеспечивают возможность объективности оценки. Увеличивается объём и улучшается качество знаний студентов, повышается мотивация к обучению, процесс обучения становится более наглядным и понятным.

Использованная литература:

1. Новый словарь методических терминов и понятий (теория и практика обучения языкам). [Электронный ресурс]. – URL: methodological_terms. academic.ru
2. Тарамова, Э.А. Проблемы и перспективы использования ИКТ в высшей школе // Актуальные задачи педагогики: материалы VI Междунар. науч. конф. (г. Чита, январь 2015 г.). – Чита: Издательство Молодой ученый, 2015. -С. 156.
3. Сластенин, В.А. и др. Педагогика: учеб. пособ. для студ. высш. пед. учеб. завед. /. - М.: Академия, 2005. –С.245.
4. Полат Е.С. Новые педагогические и информационные технологии в системе образования. – М., 2000.
5. Буримская Д. В., Харламенко И. В. Комплексный подход при обучении студентов неязыковых факультетов иностранному языку на базе информационных систем // Вестник Московского государственного областного университета. Серия: Педагогика. 2019. № 2. -С. 96–111.
6. Владимирова, Л.П. Современные информационно-коммуникативные и педагогические технологии в образовании/ - Режим доступа: www.relarn/ru/conf.
7. Ахмедова М.Х., Хосилова Ф.Р. Основные направления оптимизации обучения русскому языкув вузе на базе ИКТ.Ta’lim, fan va innovatsiya. ЎзР ОАК журнали 2020й, 3-сон. –С.80-85.
8. Титова С. В., Харламенко И. В. Метод совместного написания эссе и их взаимного оценивания при обучении письменно-речевым умениям // Вестник Московского университета. Серия 19. Лингвистика и межкультурная коммуникация. 2017. № 3. - С. 26–40.
9. Казакова, Т.А. Использование информационно-коммуникационных технологий в процессе преподавания иностранного языка в вузе // Лингвистика и межкультурная коммуникация. – 2015. – Выпуск №3 (17). – С. 78.

ПОВЫШЕНИЕ ОБРАЗОВАТЕЛЬНЫХ РЕЗУЛЬТАТОВ С ПОМОЩЬЮ МУЛЬТИМОДАЛЬНЫХ ТЕКСТОВ: НА ПРИМЕРЕ УЗБЕКИСТАНА

Odilova Kamola Taxirovna
Кандидат наук, научный сотрудник
Узбекский Государственный Университет мировых языков.

Аннотация: В этом исследовании изучается влияние мультимодальных текстов на результаты обучения в Ташкенте, Узбекистан. В связи с растущей важностью мультимодальной грамотности в 21 веке [4] это исследование изучает, насколько эффективно разработанные мультимодальные тексты могут улучшить обучение за счет управления когнитивной нагрузкой [9]. Исследование также углубляется в потенциальные преимущества использования мультимодальных текстов при изучении языка, уделяя особое внимание контексту Ташкента, Узбекистан. Результаты подчеркивают необходимость дальнейших исследований в этой области, перекликаясь с призывом [9] к постоянному исследованию мультимодального дискурса. Результаты этого исследования могут стать основой для разработки учебных программ и педагогических стратегий в Ташкенте и за его пределами, способствуя более широкому обсуждению мультимодальной грамотности и образования.

Ключевые слова: мультимодальные тексты, мультимодальная грамотность, теория когнитивной нагрузки, результаты обучения, мультимодальный дискурс, цифровой, режим, веб-сайт.

Annotatsiya: mazkur tadqiqot O‘zbekiston, hususan, Toshkent shahridagi ta’lim natijalariga multimodal matnlarning ta’sirini o‘rganadi. XXI asrda multimodal savodxonlikning ortib borayotgan ahamiyati bilan [4], ushbu tadqiqot kognitiv yukni boshqarish orqali o'rganishni qanchalik samarali tarzda ishlab chiqilgan multimodal matnlarni yaxshilashi mumkinligini o'rganadi [9]. Tadqiqot shuningdek, til o‘rganishda multimodal matnlardan foydalanishning mumkin bo‘lgan afzalliklarini o‘rganadi, bunda alohida e’tibor Toshkent, O‘zbekiston kontekstiga qaratilgan. Natijalar [9] multimodal diskursni doimiy ravishda o'rganishga chaqiruviga mos ravishda ushbu sohada qo'shimcha tadqiqotlar o'tkazish zarurligini ta'kidlaydi. Ushbu tadqiqotning natijalari Toshkent va undan tashqarida o'quv dasturlari va pedagogik strategiyalarni ishlab chiqishda ma'lumot berishi mumkin, bu esa multimodal savodxonlik va ta'lim bo'yicha kengroq muhokamaga hissa qo'shishi mumkin.

Kalit so‘zlar: multimodal matnlar, multimodal savodxonlik, kognitiv yuk nazariyasi, ta'lim natijalari, multimodal nutq, raqamli, rejim, veb-sayt.

Abstract: This study explores the impact of multimodal texts on educational outcomes in Tashkent, Uzbekistan. With the increasing importance of multimodal literacy in the 21st century [4], this research investigates how effectively designed multimodal texts can enhance learning by managing cognitive load [9]. The study also delves into the potential benefits of using multimodal texts in language learning, with a specific focus on the context of Tashkent, Uzbekistan [8]. The findings underscore the need for further research in this area, echoing [9] call for ongoing exploration of multimodal discourse. The implications of this study could inform curriculum

design and pedagogical strategies in Tashkent and beyond, contributing to the broader discourse on multimodal literacy and education.

Keywords: multimodal texts, multimodal literacy, cognitive load theory, educational outcomes, multimodal discourse, digital, mode, website.

Введение.

Мультимодальные тексты становятся важной частью нашего повседневного общения, особенно с появлением цифровых медиа. Эти тексты, объединяющие две или более семиотических систем или способов общения, предлагают богатый и сложный способ передачи смысла [7]. Эта статья призвана предоставить исчерпывающее руководство по идентификации мультимодальных текстов. То, как мы общаемся, резко изменилось в эпоху цифровых технологий. Традиционная текстовая форма, которая в основном полагалась на письменные или печатные слова, расширилась и включила в себя различные стили, что привело к появлению мультимодальных текстов. Эти тексты не ограничиваются использованием языка, но включают несколько других способов передачи значения, таких как изображения, звуки, жесты и пространственное расположение. Эта сложная форма общения называется мультимодальностью [7]. Концепция мультимодальности не нова. Его корни уходят в область социальной семиотики, области, которая изучает, как мы создаем смысл из различных способов общения. Однако с появлением цифровых носителей широкое распространение получило использование мультимодальных текстов. Сегодня мы сталкиваемся с мультимодальным текстом в различных формах, от сообщений в социальных сетях и веб-сайтов до цифровой рекламы и онлайн-учебных материалов. Понимание и идентификация мультимодальных текстов имеет решающее значение в современном мире. Это позволяет нам лучше интерпретировать представленную информацию и взаимодействовать с ней. Кроме того, он дает нам навыки эффективного создания собственных мультимодальных текстов, квалификация, которая все больше ценится как в академической, так и в профессиональной среде.

Эта статья призвана предоставить исчерпывающее руководство по идентификации мультимодальных текстов. Он исследует характеристики мультимодальных текстов, обсуждает различные используемые стили и объясняет, как эти способы взаимодействуют для создания смысла. К концу этого руководства вы сможете надежно идентифицировать и анализировать мультимодальные тексты.

Основная часть. Мультимодальность-это концепция, которая широко изучается и обсуждается в области социальной семиотики. Это означает использование нескольких режимов или семиотических ресурсов для создания и интерпретации значения [5]. Эти режимы могут включать, среди прочего, письменный текст, устные слова, визуальные эффекты, аудиоэлементы, жесты и пространственные порядки. Концепция мультимодальности основана на представлении о том, что значение формируется не только языком. Вместо этого он признает, что мы используем разные способы общения и понимания окружающего мира. Например, в новостной статье письменный текст предоставляет фактическую информацию, в то время как добавленные изображения могут вызывать эмоции или предоставлять дополнительный контекст [7]. Рост цифровых медиа еще больше повысил важность понимания мультимодальности. Цифровые платформы по своей природе поддерживают интеграцию различных режимов, что делает

мультимодальную связь нормой, а не исключением. Например, сообщение в социальной сети может включать текст, изображения, смайлики и хэштеги для передачи сообщения, каждый из которых уникальным образом способствует общему смыслу [3]. Понимание мультимодальности также важно для образования. Это позволяет учителям распознавать и использовать различные способы понимания учащимися информации. Например, включение мультимодальных текстов в обучение может соответствовать различным стилям обучения и повышать вовлеченность и понимание учащихся [5]. Понимание мультимодальности очень важно в современную цифровую эпоху. Это не только улучшает нашу способность интерпретировать мультимодальные тексты и работать с ними, но также дает нам навыки для эффективного создания собственных мультимодальных коммуникаций. Прежде чем определять мультимодальные тексты, важно понять, что означает мультимодальность. Мультимодальность относится к использованию нескольких режимов или семиотических ресурсов для создания значения [5]. Эти режимы могут включать, среди прочего, письменный язык, устный язык, визуальные образы, аудио, жесты и пространственное расположение.

Результаты. Мультимодальное распознавание текстов.

1. Наличие нескольких режимов. Первым этапом определения мультимодального текста является поиск наличия нескольких режимов. Например, текст, использующий только письменный язык, не является мультимодальным. Однако, если текст сочетает в себе письменный язык с изображениями, аудио, видео или другими режимами, он считается мультимодальным [6].

Мультимодальные тексты, сочетающие письменную речь с другими способами, такими как изображения, аудио или видео, становятся все более узнаваемыми благодаря своим возможностям для улучшения общения и обучения. Это особенно верно в области образования, где мультимодальные тексты обеспечивают более увлекательный и интерактивный опыт обучения [2]. Использование нескольких режимов в тексте может служить разным целям. Например, изображения можно использовать для иллюстрации или выделения идей в письменном тексте, а аудио и видео могут предоставить дополнительный контекст или объяснение. Это может сделать контент более доступным и понятным, особенно для читателей.

Использованная литература:

1. Беземер, Дж. и кресс, г. (2008). Письмо в мультимодальных текстах: социальный семиотический учет дизайна для обучения. Письменное общение, 25(2), 166-195.
2. Коуп, Б. и Каланцис, М. (2009). "Множественная грамотность": Новая грамотность, новое обучение. Педагогика: международный журнал, 4(3), -С.164-195.
3. Данези, М. (2016). Семиотика эмодзи: рост визуального языка в эпоху интернета. Лондон: Блумсбери.
4. Джуитт, К. (2008). Мультимодальность и грамотность в школьных классах. Обзор исследований в области образования, 32(1), 241-267.
5. Джуитт, К., Беземер, Дж., & О'Халлоран, К. (2016). Введение в мультимодальность. Лондон: Рутледж.
6. Кресс, Г. (2010). Мультимодальность: социально-семиотический подход к современной коммуникации. Лондон: Рутледж.

7. Кресс, Г. и Ван Левен, т. (2001). Мультимодальная речь: современные методы и средства общения. Арнольда.
8. Майер, Р. Е. (2009). Мультимедийное образование (2-е изд.). Издательство Кембриджского университета.
9. Майер, Р. Е. и Морено, Р. (2003). Вот девять способов снизить когнитивную нагрузку в мультимедийном обучении. Педагогический психолог, 38(1), -С.43-52.
10. Паас, Ф., Renkl, А. и Свеллер, Дж. (2003). Теория когнитивной нагрузки и дизайн обучения: последние разработки. Педагогический психолог, 38(1), -С.1-4.
11. Паивио, А. (1986). Ментальное представление: подход двойного кодирования. Издательство Оксфордского университета.
12. Серафини, Ф. (2012). Расширение четырех исходных моделей: визуального и мультимодального чтения текстов. Педагогика: международный журнал, 7(2), -С.150-164.
13. Свеллер, Дж. (1988). Когнитивная нагрузка при решении проблем: влияние на обучение. Когнитивная наука, 12(2), -С. 257-285.
14. Уолш, М. (2010). Мультимодальная грамотность: что это значит для практики в классе? Австралийский журнал языка и грамотности, 33(3), -С.211-239

XORIJIY TILLARNI O‘RGANISHDA KOMPYUTER TEXNOLOGIYALARINING O‘RNI

Asadova Shahinabonu
O’zbekiston davlat jahon tillari universiteti talabasi.

Ilmiy rahbar: **Abdurahmonova Hayitgul**
O’zbekiston davlat jahon tillari universiteti stajyor-o’qituvchisi.

Annotatsiya: Ushbu maqolada xorijiy tillarni o’rganishda kompyuter texnologiyalarining o’rni xususida fikr yuritilgan. Unda til ta’limida kompyuter texnologiyalaridan foydalanishning kommunikativ ahamiyati, samaradorligi, didaktik o’rni va shu asnoda xorijiy tillarni o’zlashtirish tendensiyasining imkoniyatlari xususida misollar asosida ma’lumotlar berilgan.
Kalit so‘zlar: texnologiya, kompetensiya, tendensiya, kognitif, funksiya, motivatsiya, innovatsiya, o’quv motivlari.

Аннотация: В данной статье рассматривается роль компьютерных технологий в изучении иностранных языков. На основе примеров представлена информация о коммуникативной значимости, эффективности, дидактической роли использования компьютерных технологий в языковом образовании, а также о возможностях овладения иностранным языком в этом контексте.
Ключевые слова: технология, компетентность, тенденция, познавательная, функция, мотивация, инновация, мотивы обучения.

Abstract: This article discusses the role of computer technology in learning foreign languages. It provides information based on examples about the communicative importance, efficiency, didactic role of using computer technologies in language education, and the possibilities of foreign language acquisition in this context.
Keywords: technology, competence, tendency, cognitive, function, motivation, innovation, learning motives.

Kompyuter texnologiyalari dasturlashtirilgan ta'lim g'oyalarini rivojlantirmoqda, zamonaviy kompyuterlar va telekommunikatsiyalarning o'ziga xos imkoniyatlari bilan bog'liq bo'lgan mutlaqo yangi, hali o'rganilmagan texnologik ta'lim imkoniyatlarini ochib beradi. Kompyuterni o'qitish texnologiyalari - bu axborotni o'quvchiga tayyorlash va uzatish jarayonlari bo'lib, uning vositasi kompyuterdir.

O‘quv jarayonida kompyuter texnologiyalaridan, ayniqsa, rasmli o‘quv qo‘llanmalaridan foydalanish samaradorligi, ularning didaktik imkoniyatlarini hayotga tatbiq etish, shubhasiz, eng muhim talab hisoblanadi. Og‘zaki ma’lumotni to'ldiribgina qolmay, balki axborot tashuvchisi vazifasini bajaradigan vizualizatsiya shakllaridan foydalanish o'quvchilarning aqliy faolligini oshirishga yordam berishi kerak. Jadvallar, grafiklar, diagrammalar, audiovizual vositalar va boshqalar. bosma va elektron o'quv materiallarining tarkibiy qismlari bo'lib, intellektual va kognitiv faolligini rivojlantirishda muhim rol o'ynaydi.

Kompyuter texnologiyalari an'anaviy o'qitishga xos bo'lgan o'qitish faoliyati turlaridan voz kechish imkoniyatini ochib, unga intellektual mehnat shakllaridan foydalanish imkoniyatini beradi. Yangi texnologiyalardan foydalanish nafaqat fanni yaxshiroq o'rganish, balki egallagan ko'nikmalarni ham o'zlashtirish imkonini beradi. Bunda:

1. Chet tilini o‘qitishning eng muhim muammolaridan biri og‘zaki nutqni o‘rgatish bo‘lib, u tilning kommunikativ funksiyasini ochib berishga sharoit yaratadi va o‘quv jarayonini real o‘rganish sharoitlariga yaqinlashtirish imkonini beradi, bu esa o‘rganish motivatsiyasini oshiradi.

2. Chet tilini o'rganish, eng avvalo, talabaning mustaqil ishi bo'lgani uchun, muammo mustaqil ishlash ko'nikma va malakalarini rivojlantirishdadir. Mustaqil ishning samaradorligi o‘qituvchi va talabaning samarali o‘zaro hamkorligining to‘rtta komponentiga, qo‘llaniladigan pedagogik texnologiyalarga, ishlab chiqilgan o‘quv materiallari va ularni qo‘llash usullarining samaradorligiga bog‘liq.

3. Ta’lim jarayonini maqbul tashkil etish uchun, birinchi navbatda, talabalarning o‘quv motivlarini chuqur bilish, ikkinchidan, ularni to‘g‘ri aniqlash va ularni oqilona boshqarish qobiliyati muhim ahamiyatga ega.

Motivatsion jihat barcha psixologik jarayonlarni - fikrlash, idrok etish, chet tili materialini tushunish va o‘zlashtirishni faollashtirish uchun ham hal qiluvchi ahamiyatga ega. Buning uchun motivatsiya darajasini oshirish, bilish va intellektual faollikni rivojlantirishga ko‘maklashish, pirovardida o‘quv jarayonining samaradorligini oshirishga intilish kerak.

Chet tillarini o‘qitish motivatsiyasini boshqarish o‘qitish usullarining asosiy muammolaridan biridir. Chet tili fan sifatida bir qator o‘ziga xos xususiyatlarga ega, ulardan biri chet tilida muloqot qilish qobiliyatini o‘rgatish orqali chet tillarini o‘zlashtirishdir. Afsuski, hozirgi vaqtda maktab o‘quvchilarida chet tilida muloqot qilish uchun tabiiy ehtiyoj yo‘qligi sababli chet tillarini o‘rgatish asosan sun'iy tarbiyaviy xarakterga ega.

Shu munosabat bilan o‘qituvchi oldida til o‘rganish jarayonida tabiiy sharoitlarga imkon qadar yaqin bo‘lgan chet tilidagi nutqiy muloqot muhitini yaratish vazifasi turibdi. Chet tilidagi nutqiy muloqot jarayonini rag‘batlantiruvchi eng muhim omil chet tilini o‘zlashtirish uchun motivatsiya deb hisoblanishi kerak.

Chet tilini kognitiv funksiyada o‘zlashtirish muvaffaqiyati lingvistik tuyg‘uni shakllantirishga yordam beradi. Til o‘rganishda kognitiv motivlarni qondirish u bilan doimiy ishlash uchun barqaror motivatsiyani shakllantiradi.

4. Nazariy jihatdan aytganda, zamonaviy tadqiqotlar shuni ko‘rsatdiki, o‘quvchilarga chet tili madaniyatini singdirmasdan turib, chet tilini zamonaviy o‘qitish mumkin emas. Aksariyat metodistlar shaxsiyatni har tomonlama rivojlantirish va ma’naviy qadriyatlarini rivojlantirishga yordam beradigan aniq kommunikativ yo‘nalishga ega bo‘lgan chet tilini o‘qitish nazariyasi va amaliyotining hozirgi holatiga ustuvor ahamiyat beradi.

Chet tilini o‘rganishning shu va boshqa muammolari o‘quvchilarning chet tili darslariga bo‘lgan qiziqishini kamaytirish tendensiyasini keltirib chiqaradi.

Axborot texnologiyalari asrida kundalik muloqot va kasbiy faoliyatda chet tilidan amaliy bilimlarni o‘rgatish bo‘yicha yuqori talablar, ayniqsa o‘quv materialining hajmi ortib bormoqda va uni uzatish va o‘qitishning an’anaviy usullari ko‘pincha samarasiz bo‘lmoqda.

Chet tillarini o‘qitishning asosiy maqsadi talabalarning kommunikativ kompetensiyasini shakllantirish va rivojlantirish, chet tilini amaliy o‘zlashtirishga o‘rgatishdir. Chet tili o‘qituvchisining vazifasi har bir talaba uchun tilni amaliy o‘zlashtirishi uchun zarur shart-

sharoitlarni yaratish, chet tili darsida talabaning kognitiv faolligini rag‘batlantiradigan tegishli o‘qitish usullarini tanlashdir.

Chet tillarini o‘rgatishning ustuvor yo‘nalishlari - aloqa, haqiqiylik, interaktivlik, madaniy kontekstda til o‘rganish va madaniyatlararo kompetensiyani rivojlantirish hisoblanadi. Chet tillarini o‘rgatishning yakuniy maqsadi chet tili muhitida bepul navigatsiya bo‘lib, bunga har qanday shakldagi audiovizual ma’lumotlarni birlashtiradigan, foydalanuvchi va foydalanuvchi o‘rtasida interaktiv muloqotni amalga oshiradigan ta’lim axborot texnologiyalari sifatida belgilangan yangi multimedia innovatsiyalari katta yordam beradi.

Chet tilini o‘rganish jarayonida axborot-kommunikatsiya texnologiyalari tobora ommalashib bormoqda. Til ta’limida ulardan faol foydalanish talabalarning motivatsiyasi va bilim faolligini oshirish, dunyoqarashini kengaytirish uchun dastak bo‘lib xizmat qiladi. Bundan tashqari, chet tilini o‘qitish jarayonida kompyuter texnologiyalaridan foydalangan holda o‘qituvchi o‘qitilayotgan fanga qiziqishini oshiradi. Shuningdek, axborot-kommunikatsiya texnologiyalari o‘quv materialini taqdim etish vositasi sifatida ham, bilimlarni nazorat qilish vositasi sifatida ham harakat qilishi mumkin.

Zamonaviy kompyuter texnologiyalari an’anaviy til o‘qitish usullariga nisbatan bir qator afzalliklarga ega. Chet tilini o‘rgatishda zamonaviy texnologiyalardan foydalanish o‘qituvchilar uchun katta imkoniyatlar ochmoqda. Masalan, ular nutq faoliyatining har xil turlarini mashq qilish va ularni turli xil kombinatsiyalarda birlashtirish imkonini beradi. Bundan tashqari, kompyuter texnologiyalari lingvistik ko‘nikmalarni rivojlantirishga, kommunikativ vaziyatlarni tartibga solishga, til va nutq harakatlarini avtomatlashtirishga yordam beradi, shuningdek, individual yondashuvni amalga oshirishni va talabaning mustaqil ishini faollashtirishni kafolatlaydi.

Shunday qilib, kompyuter texnologiyalari bilan ishlashda o‘qituvchining roli o‘zgaradi, uning asosiy vazifasi talabalarning shaxsiyatini rivojlantirish va ijodiy izlanishlarini qo‘llab-quvvatlash va boshqarishdir. Talabalar bilan munosabatlar hamkorlik va birgalikda ijodkorlik tamoyillari asosida quriladi.

Kompyuter va kompyuter texnologiyalari yordamida amalga oshiriladigan bir nechta asosiy uslubiy funksiyalar mavjud:

1) Til ta’limida kompyuterning asosiy afzalligi katta hajmdagi ma’lumotlarni saqlash va qayta ishlash qobiliyatidir, buning natijasida kompyuter chet tili darslarida faol qo‘llaniladi;

2) o‘qitish funksiyasi - chet tilini o'rganish jarayonida talabalarning universal harakatlarini mashq qilish, o'qitish maqsadida shaxsiy kompyuterdan foydalanish. Ba'zi hollarda kompyuter o'qituvchiga nisbatan o'quv materialini o'qitish va mustahkamlash paytida bir qator afzalliklarga ega bo’lib, kompyuter bilan ishlashda cheksiz vaqti borligi va kompyuter ishni obyektiv baholash va cheksiz ishlashi bilan izohlash mumkin;

3) o'quv faoliyati natijalarini joriy va yakuniy nazorat qilish uchun kompyuter texnologiyalaridan foydalangan holda talabalar bilimini kuzatish va tuzatish funksiyasidan foydalaniladi, individual va differensial yondashuvlarni amalga oshirish imkoniyatiga ijobiy ta'sir ko‘rsatadi. Fikr-mulohazalarni amalga oshirish, natijalarni maksimal obyektiv baholash, bu o‘qituvchiga nazorat qilish uchun sarflangan vaqtni minimallashtirish va test natijalarini qayta ishlashda ko'p mehnat talab qiladigan ishlardan ozod qilish imkonini beradi.

4) kommunikativ, bu funksiya global Internetda muloqot qilish qobiliyatini anglatadi;

5) tashkiliy-rag‘batlantiruvchi funksiya quyidagicha amalga oshiriladi: talabaning javobi kompyuter tomonidan avtomatik ravishda qabul qilinadi, so‘ngra javob tahlil qilinadi va javobning to‘g‘ri ekanligi aniqlanadi, natija esda qoladi va tashkilotga xabar qilinadi.

Bundan kelib chiqadiki, kompyuter chet tilini o‘rgatish jarayonida har bir o‘rganuvchini pedagogik jarayonning faol ishtirokchisiga aylantirish imkonini beradi. Masalan, tinglashni o‘rgatishda har bir talaba chet tilidagi haqiqiy nutqni quloq orqali idrok etish imkoniyatiga ega. Nutqni ‘yoki fonetik ko'nikmalarni mashq qilishda mikrofon bilan ishlashni o'z ichiga olgan o'quv dasturlari qo'llaniladi. Talaba chet tilidagi so'z yoki iboralarni tinglagandan so'ng, undan ushbu iboralarni o'zi talaffuz qilishi so'raladi, shundan so'ng u ushbu iboralarni talaffuzi va ularning intonatsiyasi ona tilida so'zlashuvchining talaffuziga qanchalik yaqin ekanligini bilib oladi. Ma’ruzachi o‘z nutqini so‘zlovchining nutqi bilan taqqoslab, talaba talaffuzidagi noaniqliklarni ko‘rishi va keyinchalik talaffuz ko‘nikmalarini oshirish uchun ularni tuzatish ustida ishlashi mumkin. Bunday ish turining kamchiligi shundaki, bunday ish o’qituvchidan ko’p vaqt oladi, lekin darsning bir necha daqiqalari shu ishga ajratilsa, o’z samarasini beradi. Grammatik ko‘nikmalarni o‘rgatishda har bir talaba grammatik mashqlarni bajarish, xatolarini aniqlash va to‘g'ri javoblarga erishish imkoniyatiga ega. Bundan tashqari, kompyuter yangi lesik yoki grammatik material bilan tanishishda, dialogik nutqni, o‘qish va yozishni o‘rgatishda, shuningdek testlarni bajarishda darsda yordamchi bo‘lishi mumkin. Agar sinfda faqat o‘qituvchining kompyuteri bo‘lsa, undan lug‘at bilan tanishish va mustahkamlashda ko‘rgazmali kompyuter sifatida foydalanish mumkin.

Chet tilini o‘qitishning yordamchi vositasi sifatida kompyuter texnologiyalari o'zining afzalliklariga ega. Masalan, har bir o‘quvchi, hattoki eng passiv o‘quvchida ham o‘z mahoratini boshqa o‘quvchilar bilan musobaqalashtirib ko‘rsatish imkoniyati mavjud bo‘lib, unda yuqori natijalarga erishish istagi va rag‘bati uyg‘onadi. Kompyuter texnologiyalari yordamida o‘qituvchi o‘z vaqtida barcha o‘quvchilarning imkoniyatlari va qiziqishlarini hisobga olgan holda o‘quv dasturini to‘g‘irlashi mumkin. O‘z navbatida, multimedia elementlaridan o‘quvchilar o‘z uy vazifalarini bajarishda foydalanishlari mumkin, ular o‘qituvchiga elektron shaklda yuborishlari mumkin. Darsdan tashqari vaqtlarda esa yoki uyda kompyuter va tegishli dasturlar mavjud bo‘lsa, talaba o‘z imkoniyati va xohishidan kelib chiqib, kerakli darajada mashq qilishi mumkin.

So‘nggi o‘n yilliklarda til o‘rgatish nazariyasi va amaliyoti internet resurslari bilan sezilarli darajada boyidi. Shuni ta’kidlash kerakki, global tarmoq resurslari lingvodidaktika uchun katta rol o‘ynaydi, chunki ular o‘rganilayotgan tillarda ona tilida so‘zlashuvchilar bilan real vaqt rejimida muloqot qilish imkonini beradi, haqiqiy materiallar matn, audio, video formatlarida va ko‘p sonli ta’lim resurslaridan foydalanishni ta’minlaydi. Multimedianing rivojlanishi va Internetning tarqalishi bilan biz kompyuterga asoslangan til o‘rganish muhitini yaratishda bir qancha yondashuvlarni ko‘rishimiz mumkin. Ushbu yondashuvlar aniq ta’lim maqsadlari uchun sinchkovlik bilan tanlangan va o‘quv materiallaridan kompleks foydalanishni o‘z ichiga oladi.

Pedagogik jarayonga o‘qitishning innovatsion shakllarini joriy etish va ulardan muvaffaqiyatli foydalanishda o‘qituvchiga o‘quvchilarning o‘quv faoliyatini tashkil etuvchi, yo‘naltiruvchi va rag‘batlantiruvchi faol yordamchi roli yuklanadi. Bundan kelib chiqadiki, zamonaviy kompyuter texnologiyalari talabalarning chet tilining kommunikativ kompetensiyasini shakllantirishda muhim ro‘l o‘ynaydi, chunki chet tilini samarali o‘rganish haqiqiy audio, video va bosma materiallar bilan ishlashga asoslanadi.

Kompyuterlar maktabda o‘rganish va bilim olish uchun juda samarali yordamdir, ular o‘quvchilarning o‘rganganlarini va bilganlarini aks ettirish uchun kognitiv vositalar sifatida

foydalaniladi. Axborotni tarqatishda kompyuter texnologiyalari kuchidan foydalanish o‘rniga, kompyuterdan bilimning barcha sohalarida o‘quvchilarga o‘rganayotgan tushunchalar haqida o‘ylangan va tanqidiy fikr yuritishga yordam beradigan vosita sifatida foydalanish kerak. Amaliy dasturlardan bilimlarni taqdim etishda rasmiyatchilik sifatida foydalanish orqali kompyuterdan o‘qitish vositasi sifatida foydalanish, mavjud bo‘lgan barcha o‘quv kompyuter dasturlarini ishlatishdan ko‘ra, materialni tezroq va to‘liqroq o‘zlashtirishga yordam beradi. Biroq, amaliyot shuni ko‘rsatadiki, kompyuter hali maktabda to‘liq o‘quv quroliga aylanmagan. Xususan, axborot texnologiyalaridan foydalanishning ijobiy ta'siriga erishish uchun ma’lum shartlarga rioya qilish kerak:

Foydalanilgan adabiyotlar:

1. Information and communication technologies in distance learning: Special training course / trans. from English Michael G. Moore, W. McIntosh, L. Black and others. M.: Publishing House "Education-Service", 2006.
2. Dmitrieva E.I. Didactic capabilities of computer telecommunication networks for teaching foreign languages // Foreign languages at school. 2007. -22 –26 b.
3. Нелунов А.И. Личностно-ориентированное обучение иностранному языку на основе информационных технологий // Информатика и образование. 2008.-31- 33с.
4. Polipova T.A. Introduction of computer technologies in foreign language teaching // Foreign languages at school. 2009. -11-16 p.
5. Rahimov A.E. Chet tillarini o'qitishda kompyuter texnologiyalaridan foydalanishning afzalliklari // Maktabda chet tillari. 2012. -56–59 b.
6. Problems of application of multimedia technologies in higher education // High technology in pedagogical process: abstracts Interuniversity Scientific Conference of university teachers, researchers and specialists. Frolova N.H. Nizhni Novgorod, VSPI, 2000. -96–98b.

ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ ВЕБ-КВЕСТОВ ДЛЯ ОРГАНИЗАЦИИ САМОСТОЯТЕЛЬНОЙ РАБОТЫ НА РАЗЛИЧНЫХ УРОВНЯХ

Жупарбек Д.
Евразийский Национальный университет
имени Л.Н. Гумилёва, Астана, Казахстан.

Аннотация: В современном образовательном контексте, акцент смещается с традиционных методов преподавания в сторону более интерактивных и студентоцентричных подходов. Веб-квесты представляют собой инновационный инструмент, который может эффективно внедряться для организации самостоятельной работы на различных уровнях образования. Целью нашего исследования является изучение возможностей использования веб-квестов в образовательном процессе для стимулирования самостоятельной работы студентов. Основные направления включают в себя анализ эффективности веб-квестов на разных уровнях образования и исследование их влияния на активность и мотивацию студентов на занятиях иностранного языка. Практическая значимость исследования заключается в разработке рекомендаций для преподавателей и образовательных учреждений по успешной интеграции веб-квестов в учебные программы по изучению иностранного языка.

Ключевые слова: веб-квесты, самостоятельная работа, образовательные технологии, мотивация, интерактивные методы, инновации, учебные программы.

Abstract: In today's educational context, the focus is shifting from traditional teaching methods to more interactive and student-centered approaches. Web quests are an innovative tool that can be effectively implemented to organize independent work at different levels of education. The aim of our study is to explore the possibilities of using web quests in the educational process to stimulate students' independent work. The main directions include analyzing the effectiveness of web quests at different levels of education and researching their influence on students' activity, motivation in foreign language classes. The practical significance of the study lies in the development of recommendations for teachers and educational institutions on the successful integration of web quests into foreign language teaching programs. To achieve this goal, we use a combined research method that includes analyzing existing pedagogical practices, conducting experimental projects with the use of web quests, as well as collecting and analyzing data on students' reactions to this training.

Keywords: web quests, independent work, educational technologies, motivation, interactive methods, innovations, training programs.

Как цифровые учебные среды, веб-квесты предлагают возможность структурировать исследования в цифровом пространстве. Это облегчает нагрузку на учащихся благодаря организованной среде, которая поэтапно знакомит их с комплексным заданием и предлагает материал в ориентированном на студента виде. Это тип веб-исследования, сочетающий в себе проблемно-ориентированное обучение, проектный подход и игровые элементы, а также активно использующий информационные и коммуникационные технологии [1]. Кроме того, в веб-квесты можно интегрировать специальные предложения,

которые поддерживают учащихся при самостоятельной работе, не перегружая учебную среду в присущей ей структуре. Так веб-квест обычно используется для обучения навыкам чтения и письма. На начальном этапе можно сформировать регулятивные универсальные учебные действия, например, такие как: целеполагание, планирование, коррекция [2]. Более того, использование веб-квеста заставляло студентов полагаться на себя и сотрудничать друг с другом, а также осознавать, что они должны делать во время обучения. Кроме того, использование парной и групповой работы оказалось жизненно важным для успешного выполнения собственных заданий. Наконец, нетрадиционная роль учителя во время программы сделала процесс обучения поддерживающим и интересным. Учитель или исследователь был проводником, помощником, наставником и контролером, а использование студентами журнала рефлексии для самооценки сделало их зависимыми и уверенными в себе.

Использование веб-квеста - один из наиболее эффективных способов реализации интерактивного обучения иностранному языку с помощью интернет-технологий, и он, безусловно, более выгоден, чем традиционное обучение по учебнику [3]. Однако применительно к обучению иностранным языкам главное заключается в том, что веб-квест служит целям лингвистического обучения, способствующего развитию четырех навыков - чтения, аудирования, письма и говорения. Важно, чтобы учащиеся практиковали новую языковую структуру, особенно если у них нет фоновых знаний о ней. При этом приоритет отдается коммуникативно-интеркультурным и конструктивистским методам, в которых коммуникация, аутентичность и цифровые медиа занимают центральное место. Студенты неязыковых направлений подготовки нуждаются в стимулировании мотивации к изучению иностранного языка, поскольку, чаще всего, демонстрируют низкий уровень заинтересованности непрофильной дисциплиной [4].

Для достижения цели были использованы следующие методы исследования:

- разделение студентов на группы для оценки эффективности веб-квестов.
- применение веб-квестов в учебном процессе для каждой группы.
- сравнительный анализ:
- опросы среди студентов для оценки педагогической эффективности веб-квестов.

Веб-квесты – это ориентированная на изучение деятельность, в которой часть или вся информация, с которой взаимодействуют учащиеся, поступает из интернет-платформ. Технология разработки веб-квестов - процесс длительный и непростой [6]. Студенты работают самостоятельно и проблемно-ориентированно с информацией из интернета на основе поставленной задачи. Цель - познакомить студентов с работой с источниками из интернета, используя структурированную учебную среду. По тематике веб-квесты могут быть разнообразными [7]. Характерная структура веб-квеста представлена в таблице ниже, стоит отметить, что с течением времени в названия разделов, последовательность и организацию содержания вносились коррективы. Однако основная идея структурированной учебной среды для онлайн-исследований всегда оставалась неизменной.

Таблица 1 – Разделы веб-квеста

Вступление	Вводит в курс дела и предоставляет исходную информацию
Задание	Представляет собой интересную и выполнимую задачу

Материал	Обращается к информационным ресурсам в интернете и требуется для выполнения задания
Порядок действий	Описывает и разделяет процесс обучения на четкие шаги во время работы над заданием
Помощь	Предлагает конкретную помощь в обработке собранной информации
Заключение	Оформляет заключение веб-квеста в подходящей форме

Самостоятельная работа над веб-квестом обеспечивается подготовкой структурированной учебной среды, предварительным отбором подходящих источников и вспомогательных материалов. Таким образом, предотвращаются такие проблемы, как информационная перегрузка, фальшивые новости и т.п. Это позволяет студентам сосредоточиться на работе с онлайн-источниками. Несомненным преимуществом включения веб-квеста в процесс обучения является то, что он не требует специальных технических знаний и может быть выполнен как группой, так и индивидуально [9]. Некоторые веб-квесты предлагают возможность загрузить лингвистические ресурсы и предложения по оказанию помощи и сделать их доступными в шаблоне документа. Они могут включать рабочие листы с фрагментами, которые основаны на речевых актах, ожидаемых в учебном продукте. Также возможны дополнительные средства структурирования контента, которые учащиеся могут загружать по мере необходимости.

При составлении материала для организации самостоятельной работы следует обращать внимание на тематическую релевантность заданию, надежность, аутентичность и соответствие возрасту. Если эти аспекты учитываются в процессе отбора, это также снижает нагрузку на студентов, которые могут сосредоточиться на поиске информации в предоставленных материалах. Этот поиск информации происходит в контексте проблемы или вопроса, над которым учащиеся могут работать с помощью информации.

Рассматривая процесс обучения и самостоятельной работы, который учащиеся проходят в рамках веб-квеста, можно заметить, что в нем представлены требования из всех областей компетенции образовательных стандартов:

1. Для выполнения задания учащиеся получают несколько, часто разных источников информации, чтобы найти нужную информацию (чтение, работа с текстами и СМИ).

2. Затем отфильтрованная информация обычно обобщается в письменном виде (письмо), чтобы в итоге представить ее другим (говорение и аудирование).

Материал, предоставляемый студентам для этой цели в веб-квесте, обычно представляет собой тексты, в смысле концепции расширенного текста, которые интегрированы в цифровую форму. Каждой группе дается область для работы. На выбор предлагаются следующие области:

- культура города;
- история;
- размещение и питание;
- наука и образование.

Каждая группа должна решить следующие вопросы в своей области:

Группа 1 - культура: выбрать наиболее интересные культурные и архитектурные достопримечательности Праги, кратко расскажите о каждой из них и укажите их особенности (например, художественные коллекции, выставки).

Группа 2 - история: создать исторический обзор развития города и его текущее состояние (краткий обзор от истоков до XIX века, подробное представление Рима в XX веке до наших дней).

Группа 3 - размещение и питание: поиск и бронирование различных вариантов размещения (студенческие хостелы, отели, частные пансионы), проверка гастрономических предложений и рекомендация лучших из них (рестораны, предложения фаст-фуда, трактиры, пабы).

Группа 4 - наука и образование: поиск и представление различных образовательных учреждений (колледж, университет, курсы), которые интересны.

Студенты кроме интернет-ресурсов, также могут использовать другие источники информации, такие как журналы, статьи, книги, фотографии и т.д. Для каждой группы было подготовлено несколько ссылок, чтобы облегчить начало работы:

- https://www.nomadicmatt.com/travel-blogs/the-cost-of-traveling-western-europe/
- https://hotelchantelle.com/what-is-european-plan-in-hotel/
- https://www.educations.com/search/institutes/europe

В конце работы каждая группа должна представить свои результаты. Результаты не должны длиться более 20 минут. Это способствовало развитию у учащихся медиа-навыков (визуализация и презентация) и творческих способностей. Критерии оценки должны быть известны, понятны и однозначны всем участникам до начала работы. Так, например, по результатам обеих оценок выставляется групповая оценка. Оценивается следующее:

- язык: ясный, понятный, беглый, грамматически правильный (35%);
- содержание: разнообразная информация (30%);
- представление: презентация, рисунки, раздаточный материал (35%).

Экспериментальную базу составили студенты первого курса второго семестра факультета международных отношений Евразийского национального университета имени Л.Н. Гумилева. Всего в педагогическом исследовании приняли участие 38 студента (2 групп), обучающихся на первом курсе международного факультета. Студенты были разделены на две группы: контрольную (20 человек) и экспериментальную (18 человек). Выборка соответствует гендерному критерию.

Метод тестирования по английскому языку на базе бесплатной платформы English House (https://e24.kz/free-online-english-level-test). Тест состоит из 30 вопросов: вопросы по грамматике, фонетике, лексике (базовой и профессиональной). Исходный уровень языковой компетенции студентов анализировался в соответствии с критериями, представленными в таблице ниже. Ввод и обработка данных осуществлялись в Microsoft Excel. Все данные представлены в относительных величинах (% от числа опрошенных).

Таблица 2 – Критерии уровня языковой компетенции студентов

Критерии уровня языковой компетенции студентов по курсу, преподаваемому в вузе	
Фонетика	Оценка четкости и точности артикуляции звуков, умение преодолевать основные трудности в произношении для достижения более нейтрального звучания.
Грамматика	Оценка умения использовать основные грамматические структуры, такие как времена, модальные глаголы, артикли и предлоги.
Лексика	Оценка общего уровня словарного запаса студента в контексте общения на учебном или профессиональном уровне, умение использовать специализированный лексический запас в соответствии с областью знаний или профессиональной областью.

Что касается оценки, то студенты должны были заполнить самооценку, а также оценить своих членов группы. Как уже говорилось, чтобы сделать эту задачу проще и привлекательнее, были созданы две Google-формы. Для оценки себя и членов своей группы студенты должны были заполнить четыре обязательных поля: собранная информация, организация / выполнение задания, действие / автономность и сотрудничество.

На контрольном этапе на основе данных, полученных в результате наблюдения и оценки учебной деятельности студентов, определялись количественные пропорции сформированности умений поиска и анализа информации у учащихся разного уровня.

Таблица 3 – Уровень сформированности у студентов навыков поиска и анализа информации при изучении английского языка как иностранного с использованием веб-квестов на контрольном этапе эксперимента

Методы	Уровень владения навыками поиска и анализа информации		
	Высокий (%)	Средний (%)	Низкий (%)
Наблюдение	31,5	48,1	20,4
Оценка результатов деятельности	27,1	44,3	28,6
Всего	29,3	46,2	24,5

Диагностика, проведенная на заключительном этапе исследования, позволила определить уровень сформированности навыков получения и анализа информации у изучающих английский язык с помощью веб-квест.

Таблица 4 – Уровень сформированности у студентов навыков поиска и анализа информации при изучении английского языка как иностранного с помощью веб-квестов на суммирующем этапе эксперимента

Методы	Уровень владения навыками поиска и анализа информации		
	Высокий (%)	Средний (%)	Низкий (%)
Наблюдение	58,1	30,4	11,5
Оценка результатов деятельности	50,6	36	13,4
Всего	54,35	33,2	12,45

Анализ результатов показал, что после проведения обучающего эксперимента увеличилось количество студентов с более высокими навыками поиска и анализа информации. Визуальное сравнение результатов, проанализированных на обобщающем и контрольном этапах, представлено на рисунке.

Сравнительный анализ результатов констатирующего и контрольного этапов показал, что количество студентов с высоким уровнем навыков поиска и анализа информации увеличилось на 25,5% после внедрения веб-опосредованного подхода. С другой стороны, количество студентов с низким уровнем развития навыков поиска и анализа информации уменьшилось на 12,4%. Таким образом, в результате экспериментального исследования можно сделать вывод, что метод веб-квестов помогает

улучшить информационно-поисковые и аналитические навыки студентов при изучении английского языка.

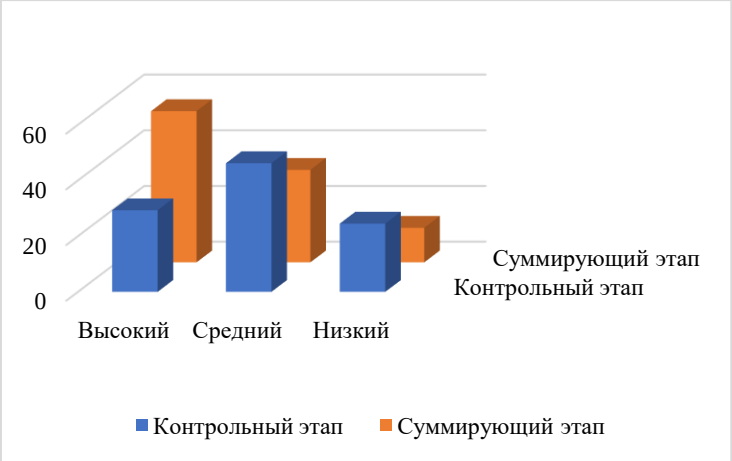


Рисунок 1 – Динамика уровня сформированности у студентов умений поиска и анализа информации.

Кроме того, визуальный анализ графиков на рисунке 2 подтверждает, что студенты контрольной и экспериментальной групп достигли одинакового уровня коммуникативной компетенции по критерию «фонологическая компетенция» на этапе подведения итогов. На рисунке 3 представлены результаты уровня коммуникативной компетенции студентов по критериям «грамматика» и «лексика» в контрольной и экспериментальной группах. На этапе подведения итогов баллы по критериям «грамматическая компетенция» и «лексическая компетенция» оказались ниже критических значений на уровне значимости 0,05 (0,67 < 7,78; 0,13 < 7,82). Таким образом, на уровне значимости 0,05 группы характеризовались согласованностью на этапе подведения итогов. Визуальный анализ графиков также свидетельствует о том, что уровень владения языком у студентов контрольной и экспериментальной групп схож.

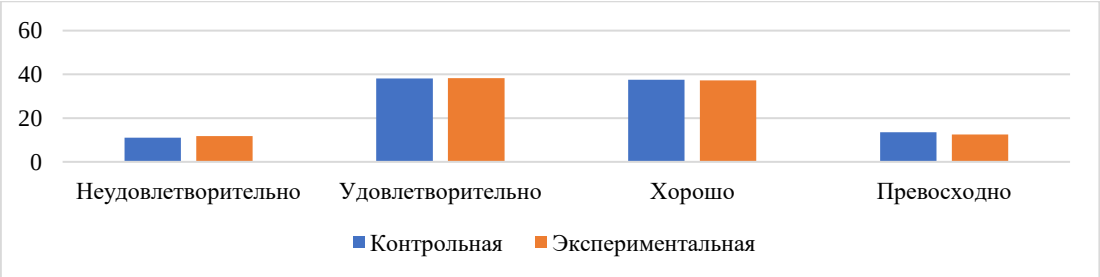


Рисунок 2 – Диаграмма уровня языковой компетенции студентов по критерию фонетика на суммарном этапе

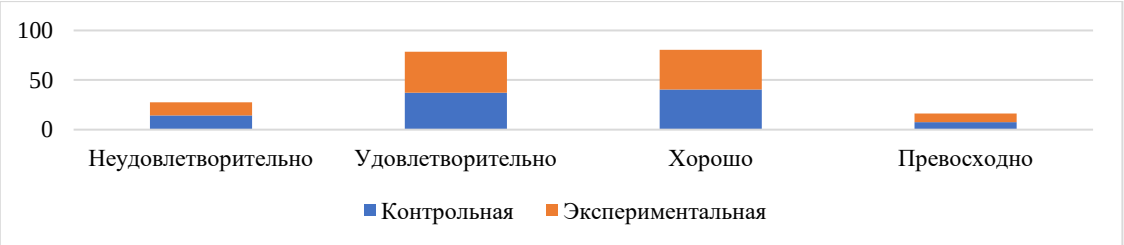


Рисунок 3 – Диаграмма уровня языковой компетенции студентов по критерию грамматика на суммарном этапе.

Эти положительные результаты и значительные различия между контрольной и экспериментальной группой были обусловлены эффективностью веб-квестов.

Кроме того, отвечая на вопросы о проекте применения веб-квестов, 95,2% студентов считают, что веб-квест отличается от других групповых заданий, которые они выполняли раньше, а 78,1 % - что он лучше других групповых заданий. Некоторые студенты отметили, что им понравилось то, что веб-квест был разработан таким образом, что каждая группа имела свой собственный раздел ресурсов. Другие студенты отметили, что у них было больше свободы при выполнении веб-квеста и что это было более творчески, и признались, что, по их мнению, это давало темы, отличные от тех, что изображались на занятиях иностранного языка. Некоторые признались, что, поскольку у них были необходимые ресурсы, им было легче выполнять задания и организовывать.

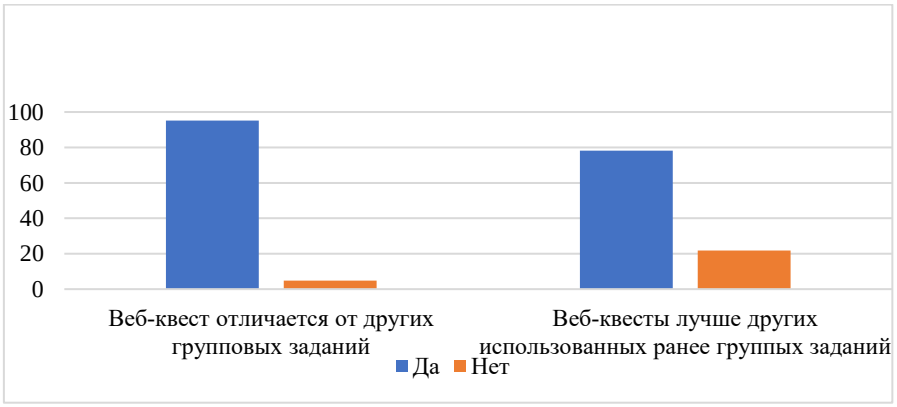


Рисунок 4 – Результаты опросов студентов после применения веб-квестов

Таким образом, при разработке веб-квестов желательно иметь четкие инструкции, которым можно было бы следовать без дополнительных разъяснений со стороны преподавателя. Студенты могут работать самостоятельно и сотрудничать друг с другом. В дополнение к положительному эффекту веб-квест, продемонстрированному количественными статистическими результатами, положительное отношение и качественные данные, полученные в результате собственных наблюдений и заметок исследователя, также подтверждают эффективность. Например, некоторые участники выразили свое положительное отношение к веб-квесту. Вот некоторые из их комментариев:

1. «Веб-квест - удобный инструмент. Его легко использовать».
2. «Я бы хотел, чтобы веб-квест преподавался и по другим предметам».

3. «Веб-квест помог мне взаимодействовать с одногруппниками».
4. «Это помогло мне расширить свои знания».
5. «Я научился проводить исследования в интернете».

Таким образом, на основе практического эксперимента было установлено, что применение веб-квестов оказывает положительное воздействие на процесс обучения английскому языку у студентов первого курса.

Результаты настоящего исследования подтверждают предположение о том, что использование методики веб-квестов является эффективным средством обучения иностранному языку на всех этапах учебного процесса. Следовательно, применение метода онлайн-квестов оказалось эффективным при изучении как базовой, так и специализированной лексики. Дальнейшие перспективы исследования направлены на усовершенствование выявленных методов и средств обучения английскому языку для студентов первого и второго курсов, а также на разработку индивидуальных методик, учитывающих особенности обучения студентов разных специальностей. Полученные результаты могут быть полезны для преподавателей, работающих с учащимися в высших учебных заведениях.

Список использованной литературы:

1. Рассадина М.Н. Совершенствование организации образовательного процесса посредством применения веб-квест-технологии в самостоятельной работе курсантов // Пенитенциарная наука, 2017, 4. –С.40.
2. Зонова С.А., Новосёлова Ю.А. Образовательный веб-квест как средство формирования универсальных учебных действий на уроках различного типа // E-Scio, 2020, 7. –С.46.
3. Восковская А.С., Карпова Т.А. Использование веб-квестов для организации самостоятельной работы студентов в процессе обучения иностранному языку в неязыковом ВУЗе // МНКО, 2019, 6.-С.79.
4. Ильина В.Ю. Веб-квест как одна из технологий организации самостоятельной работы по иностранному языку в неязыковом вузе (опыт внедрения на компьютерных направлениях подготовки) // Проблемы современного педагогического образования, 2018. –С.59-1.
5. Абрамова И.В. Дидактические возможности и условия применения технологии веб-квест как дидактической формы в профессиональной подготовке студентов // Juvenis scientia, 2019. –С. 2.
6. Нурмиева Р.Р., Першина Н.О. Использование веб-квест технологий в обучении иностранным языкам // Казанский вестник молодых учёных, 2018, 5. –С.8.
7. Ефанова О.А., Лаврищев А.И. Использование веб-квест-технологии для организации самостоятельной работы студентов при обучении иностранным языкам в ВУЗе // Наука и школа, 2022, 5.
8. Харунжева Е.В., Козлова Ю.Д. Потенциал использования образовательных веб-квестов в процессе обучения // Педагогическое искусство, 2021, 1(2).
9. Левченко Н.В., Нечаева О.А. Веб-квест как средство повышения учебной мотивации студентов ВУЗа // Проблемы современного педагогического образования, 2021.-С 70-4.

10. Королева Е.В., Конышева Ю.Р. Веб-квесты как средство развития коммуникативных навыков на уроках английского языка // Проблемы современного педагогического образования, 2021. -С. 72-3.
11. Шевченко Г.И., Иванцов Е.А. Образовательные веб-квесты в организации и управлении самостоятельной работой студентов // МНКО, 2015, 6. –С. 55.

TIL TA'LIMIDA KOMPYUTER LINGVISTIKASI IMKONIYATLARIDAN FOYDALANISH: MUAMMO VA YECHIMLAR

Hasanova Malika G'ofir qizi
O'zbekiston Milliy universiteti talabasi.

Ilmiy rahbar: **Abduraxmonova Nilufar**
Mirzo Ulug'bek nomidagi O'zbekiston milliy universiteti
Professori, filologiya fanlari doktori.
E-mail: n.abduraxmonova@nuu.uz

Annotatsiya: maqolada kompyuter lingvistikasi imkoniyatlaridan foydalangan holda til ta'limi bilan bog'liq muammolarni hal qilishga doir taklif va mulohazalar bayon qilingan.
Kalit so'zlar: til ta'limi, kompyuter texnologiyalari, kompyuter lingvistikasi, algoritmlash, sun'iy intellekt, mashina tarjimasi, korpus tilshunosligi, leksik minimum, matn, matnning o'qishlilik darajasini aniqlash

Abstract: The article presents suggestions and considerations for solving problems related to language education using the capabilities of computational linguistics.
Keywords: language education, computer technology, computational linguistics, algorithmization, artificial intelligence, machine translation, corpus linguistics, lexical minimum, text, definition of the level of readability of the text

Аннотация: В статье изложены предложения и соображения по решению задач, связанных с языковым образованием, с использованием возможностей компьютерной лингвистики
Ключевые слова: языковое образование, компьютерные технологии, компьютерная лингвистика, алгоритмизация, искусственный интеллект, машинный перевод, корпусная лингвистика, лексический минимум, текст, определение уровня читаемости текста

Har qanday til rivojlanishi, zamon talablariga moslashishi uchun ilmiy tadqiqotga ehtiyoj sezadi. Jumladan, til ta'limi masalasida ham tilshunoslikning eng so'nggi yutuqlaridan foydalanish, sohaga doir muammolarni yechishga yo'naltirish dolzarb vazifa hisoblanadi.

Keyingi yillarda mamlakatimizda til siyosati yangi bosqichga ko'tarildi. O'zbek tilini rivojlantirish bo'yicha ketma-ket islohotlar olib borildi. Xususan, O'zbekiston Respublikasi Prezidenti Shavkat Mirziyoyevning 2019-yil 21- dekabrdagi "O'zbek tilining davlat tili sifatini, nufuzi va mavqeini tubdan oshirish chora-tadbirlari to'g'risida"gi PF-5850-sonli Farmoni, O'zbekiston Vazirlar Mahkamasining 2019-yil 12-dekabrdagi "Davlat tilini rivojlantirish departamenti to'g'risidagi nizomni tasdiqlash haqida"gi 984-sonli Qarori, O'zbekiston Respublikasi Oliy va o'rta maxsus ta'lim vazirligining 2019-yil 2-dekabrdagi 87-06- 52 sonli xatida "2020-2030 yillarda o'zbek tilini rivojlantirish hamda til siyosatini takomillashtirish Konseptsiyasi"[O'zR Prezidentining 20.10.2020 y. PF-6084son Farmoniga 1-ilova] va "Yo'l xaritasi" loyihasida davlat tilini rivojlantirishning ustuvor yo'nalishlari, vazifalari sohaga oid dasturlar va kompleks chora-tadbirlarni ishlab chiqish o'z ifodasini topgan. Bu hujjatlar davlat

tilini o‘rgatish va uni bilish darajasini yanada takomillashtirish uchun oliy ta’lim muassasalarida 2020-2021 o‘quv yilidan boshlab "O‘zbek tilini sohada qo‘llanilishi" asosiy fan sifatida o‘quv rejasiga kiritildi.

Texnologiyalar asrida taraqqiyot sur’atining tezlashgani, intellektual salohiyat, texnikaning yuqori darajada yuksalishi kishilik jamiyati oldiga qator vazifalarni qo‘ymoqda. Mintaqamizdagi globallashuv jarayoni barcha sohalarda tezkorlikni, jadal taraqqiyotni talab etmoqda. Kompyuter tizimi qulayliklar, imkoniyatlar majmuyiga va tarkibiy qismiga aylandi. Xususan kompyuter lingvistikasining kirib kelishi til bilan bevosita bog‘liqdir. Kompyuter lingvistikasi kompyuterdan unumli foydalanish hamda lingvistika va tilshunoslikka doir masalalar (axborot uslubini o ‘zlashtirish, tilning funksional doirasi haqida bilimga ega bo‘lish, tillarni o‘qitish, bilimlarni baholash, matnlarni tahrir va tahlil qilish, bir tildan ikkinchi tilga tarjima qilish)ni kompyuter vositasida hal qilish yo‘llarini belgilash, kompyuter savodxonligi darajasini oshirish, mantiqiy to‘g‘ri, izchil fikrlashga o‘rgatish, nazariy bilimlarni shakllantirish va muayyan yo‘nalishlar bo‘yicha amalda qo‘llashga oid ko‘nikmalami hosil qilishni nazarda tutadi [Xolmanova, 2; 8].

Kompyuter lingvistikasi informatika va lingvistikaning shunchaki birlashuvidan shakllangan soha emas, balki o‘zida kengroq doiradagi masalalarni qamrab oladi. Ya’ni tilshunoslikka bevosita bog‘lanadi. Kompyuter lingvistikasi tilning kompyuter nuqtayi nazaridagi ilmiy talqinidir. Ya’ni til ta’limida kompyuter texnologiyalarining o‘rni. Olimlar turli tipdagi til hodisalarining kompyuter (formal) modellarini tuzish ustida izlanmoqdalar. Bunday modellar bilimlarga, qoidalarga yoki ma’lumotlarga, faktlarga asoslanadi. Kompyuter lingvistikasi bo‘yicha mutaxassislar masalaga ilmiy jihatdan yondashishlari mumkin, ya’ni ma’lumotni formal jihatdan, kompyuter tizimi nuqtayi nazaridan ifodalashlari, lingvistik yoki psixolingvistik fakt orqali izohlashlari mumkin bo‘ladi. Hozirda kompyuter lingvistikasining vazifasi amaliy ahamiyatga ega bo‘lgan nutqni farqlash, nutqni sintezlash, ovozli avtomatik qidiruv mashinalari, matn muharrirlari, tillarni o‘rganish uchun materiallar tizimlarini yaratish va mashina tarjimasini sifatini yaxshilash bilan belgilanadi. Biroq bu jarayonlarda turli xil muammolar uchramoqda, bu muammolarga mashina tarjimasining lingvistik muammolari, o‘zbek tilidagi birikmalarning boshqa tillarga tarjima qilishdagi muammolar, avtomatik tarjima muammolari, o‘zbek tilidagi yozuvlarni o‘qishdagi muammolar, o‘zbek tilidagi lug‘atlar muammolari va boshqa shunga o‘xshash muammolar hisoblanadi.

Mashina tarjimasi yoki avtomatik tarjima bu muayyan so‘z yoki matnlarning kompyuter yordamida bir tildan boshqa tilga qilingan tarjimasi. Mashina tarjimasi kompyuter lingvistikasining asosiy yo‘nalishlaridan biri hisoblanadi, avtomatik tarjima tizimi prinsip va metodlari haqidagi bilimlarni hosil qilish, lingvistik ta'minot yaratish, mavjud tizimlar va lug‘at yaratish tarixi, til haqidagi bilimlar, ko‘nikmalarni shakllantirish, mashina tarjimasi tizimining lingvistik protsessorini shakllantirishda lingvistik tahlilni algoritmlash tizimini amalga oshiradi. Mashina tarjimasi oldiga quyidagi vazifalar qo‘yiladi:

- avtomatik tarjima tizimini yaratish va undan foydalanishning asosiy prinsiplarini o‘rgatish;
- tilning nazariy asoslari (formalizatsiya) metodlari va vositalari bilan tanishtirish;
- avtomatik tarjima tizimi lingvistik protsessorni yaratishda so‘z va uning ma'nosi,
- gap strukturasi, semantikasi va matn tahlilida lingvistik algoritm bosqichlarini

o‘rgatish.

Biroq aynan shu vazifalarda o‘zbek tilida avtomatik tarjimada kamchiliklar mavjud. Bundan tashqari 2003-yilda A.Po‘latov va boshqalarning “Dunyoviy o‘zbek tili” kitobi bosmasiga

e'tibor qiladigan bo'lsak, unda o'zbek tilidagi 1 ta ishlamoq fe'lining taxminan 100000 ta shakli (leksik grammatik paradigmalari) rus va ingliz; tillariga o'girilgan. Ingliz tilida esa fe'l taxminan 100 ta shakli borligi hammaga ma'lum. Mana qiyos: 100 000 ta va 100 ta. Kompyuter uslubida ushbu 100 000 ta shakldan faqat eng zaruriy eng maqbul (optimal) varianti tanlab olinadi, qolganlari esa imkoniyat tarzida saqlanadi. [A. Po'latov., S.Muhamedova., 2019: 60]. Bundan kelib chiqadiki hali tilimizda shunga o'xshash so'z va muammolar mavjud.

Shu sababli, o'zbek tilining milliy korpusini yaratish hayotimizda muhim ahamiyatga ega. Zero, kompyuter lingvistikasi bo'yicha mutaxassislar – kompyuter dasturchilari va filologlar (lingvistlar) hamkorligida yaratiladi. Kompyuter lingvistikasining ijobiy jihati ham aynan shunda ko'rinadi: u filolog va kompyuter mutaxassislarini birlashtiradi. Shuningdek, Nilufar Abduraxmonova, Azamat Raximov, Shahlo Hamroyeva, Zulxumor Xolmonova va boshqa mutaxassislar kompyuter lingvistikasi faniga o'z hissalarini qo'shib, yangi izlanish va tadqiqotlarni olib borishmoqda.

Xulosa qilib ko'rishimiz mumkinki kompyuter lingvistikasida o'zbek tilini qayta ishlashdagi muammolarni quyidagilar tashkil etadi:

Turg'unlik va morfologik murakkablik: O'zbek tili juda turg'un til hisoblanadi, unda so'zlar turli morfologik o'zgarishlarga ega. Bu morfologik tahlil va so'zlarni tanib olishni qiyinlashtiradi.

Imlo va transkripsiya masalalari: O'zbek tilida yozilish standartlari bir necha marta o'zgargan, bu esa matnlarni qayta ishlash va yozish turlari o'rtasida tafovutlarga olib keladi.

Dialekt va lahja xilma-xilligi: O'zbek tilida turli dialektlar va lahjalar mavjud bo'lib, ular o'rtasida ba'zi farqlar mavjud. Bu dialektlar va lahjalarni tahlil qilishni qiyinlashtiradi.

Avtomatlashtirilgan tahlil va tarjima muammolari: O'zbek tilini boshqa tillardan avtomatlashtirilgan tarzda tarjima qilish yoki boshqa tillardan o'zbek tiliga tarjima qilish qiyinchilik tug'diradi, chunki til o'ziga xos murakkabliklarga ega.

Sintaktik tahlil va so'z tartibi: O'zbek tilida so'z tartibi erkin, bu esa sintaktik tahlil va gaplarni tushunishni qiyinlashtiradi.

Ushbu muammolarni hal qilish uchun o'zbek tili bo'yicha til resurslari yaratish, tadqiqotlarni rivojlantirish, va algoritmlar, morfologik analizatorlar ishlab chiqish kerak. Bundan tashqari mashina tilini rivojlantirish uchun til mutaxassislarini bevosita dasturlash va kompyuter savodxonligini oshirish masalalarini keltirib o'tishimiz mumkin. Bu muammolarni hal etishda kompyuter lingvistikasi fanining o'rni muhim hisoblanadi. Shuningdek, tilning rivojlanishi uchun bilimli, har tomonlama kuchli aql-idrokga ega kadrlarni yetishtirib chiqarish fanning asosiy vazifalaridan biri hisoblanadi.

Foydalanilgan adabiyotlar:

1. 2020-2030 йилларда ўзбек тилини ривожлантириш ва тил сиёсатини такомиллаштириш Концепцияси (ЎзР. Президентининг 2020 йил 20 октябрь ПФ-6084-сон Фармонига 1-илова)
2. Xolmanova Z. Kompyuter lingvistikasi. Toshkent, 2020.
3. Po'latov A., Muhamedova S. Kompyuter lingvistikasi Toshkent – 2009
4. Ismailov, A. S., & Abdurakhmonova, N. (2022). The development of Alisher stemmer for Uzbek Language. Science and Education, 3(4), 187-213.

5. Abdurakhmonova, N., Shakirovich, I. A., & O‘G‘Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. Science and Education, 3(9), 146-156.
6. Boltaboyev H. Adabiyot ensiklopediyasi. 1-2 jild. T., 2015, 2019.
7. Tursunaliyevich, F. N.., & Sharofiddinovich, S. S. (2023). Barqaror
8. Rivojlanishni Maktabgacha Ta’lim Yoshidagi Bolalarga
9. Singdirish. JOURNAL OF SCIENCE, RESEARCH AND TEACHING, 2(2), 26–29. Retrieved from https://jsrt.innovascience.uz/index.php/jsrt/article/view/45
10. O‘zbek tilining izohli lug‘ati (5 jildli). – Toshkent: O‘zbekiston milliy ensiklopediyasi, 2005–2008.
11. Abduraxmonova, N., & Abduvaxobov, G. I. (2021). O‘quv lug ‘atini tuzishning nazariy metodologik asoslari. СЎЗ САНЪАТИ ХАЛИҚАРО ЖУРНАЛИ, 103.
12. Sulevmanov, D., Gatiatullin, A., Prokopyev, N., & Abdurakhmonova, N. (2020, November). Turkic morpheme web portal as a platform for turkology research. In 2020 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-5). IEEE.

ЛИНГВИСТИЧЕСКИЙ ТРЕНАЖЕР ОБУЧАЮЩИЙ УЗБЕКСКОМУ ЯЗЫКУ КАК ИНОСТРАННОМУ НА БАЗЕ АНАЛИЗАТОРА РЕЧИ

Ибрагимова Сайёра
Старший научный сотрудник
Научно-исследовательский институт
Развития цифровых технологий и искусственного интеллекта.

Мирзаева Гулмира
PhD, старший преподаватель
Ташкентский университет информационных
технологий имени Мухаммада аль-Хоразмий.

Аннотация: Тема, представленная в данной статье, является довольно актуальной в современном мире. Существует множество различных информационных технологий, позволяющих усовершенствовать навыки аудирования, чтения и написания при изучении иностранных языков, но направлению улучшения навыков разговорной речи уделяется недостаточно внимания. В то же время для обучающегося важно научиться правильно произносить звуки, слова и целые предложения, так как это является основой для коммуникации в процессе деловых и производственных взаимоотношений. В современном темпе жизни, где ценится каждая сэкономленная минута, появляется потребность в обучении языку в комфортной и удобной среде: дома, в гостинице, в офисе. Повсеместное внедрение технологий распознавания позволяет улучшать навыки людей в различных областях знаний, среди которых особое место занимает обучение иностранным языкам. В данной статье представлен краткий обзор работа технологии преобразования речи в текст, определение важности распознавания речи в современных условиях, а также актуальности создания лингвистического тренажера для обучения узбекскому языку с учетом особенностей говорящего.

Ключевые слова: лингвистический тренажер, обучение узбекскому языку, речевые технологии, анализатор речи, интеллектуальный интерфейс.

Annotation: The topic presented in this article is quite relevant in the modern world. There are many different information technologies that allow you to improve listening, reading and writing skills when learning foreign languages, but not enough attention is paid to improving speaking skills. At the same time, it is important for the student to learn how to correctly pronounce sounds, words and entire sentences, as this is the basis for communication in the process of business and industrial relationships. In the modern pace of life, where every saved minute is valued, there is a need for language learning in a comfortable and convenient environment: at home, in a hotel, in the office. The widespread introduction of recognition technologies makes it possible to improve people's skills in various fields of knowledge, among which teaching foreign languages occupies a special place. This article provides a brief overview of the work of speech-to-text conversion technology, determining the importance of speech recognition in modern conditions, as well as the relevance of creating a linguistic simulator for teaching the Uzbek language, taking into account the characteristics of the speaker.

Keywords: linguistic simulator, teaching the Uzbek language, speech technologies, speech analyzer, intelligent interface.

Сегодня интерес к дистанционному обучению иностранным языкам чрезвычайно велик: платформа для изучения языков Duolingo имеет около 300 миллионов зарегистрированных пользователей; проект в сфере массового онлайн-образования Coursera предлагает сотни курсов английского, испанского, китайского, русского и других языков. На рынке появляется все больше компаний, разрабатывающих программное обеспечение для изучения языков, – таких, как Rosetta Stone, Transparent Language, Rocket Languages. Самостоятельное изучение языков без помощи учителя имеет много преимуществ: пользователь может выполнять упражнения в своем темпе, в удобное для себя время, не выходя из дома или по дороге на работу, не переплачивая при этом за языковые курсы или репетитора.

Лингвистические тренажеры для взрослых с нормальным слухом, изучающих иностранный язык, начали создаваться позже, в конце 80-х годов, в связи с развитием технологий по автоматическому распознаванию звучащей речи (АРР). Ранние программы по распознаванию речи были нацелены на мужские голоса: более высокие женские и детские голоса распознавались некачественно. В 1980-х годах разрабатываются технологии по нормализации голоса говорящего -– среди прочего, различные методы нормализации акустического влияния длины речевого тракта говорящего [Gerstman L., 1968: 78-80; Lobanov B., 1971: 606-608; Wakita H., 1977: 183-192], – позволившие перейти к независимым от диктора системам распознавания речи и расширить область их применения. В 1988 году в работе [Flege J., 1988: 353-381] высказывается идея о том, что технологии по автоматическому распознаванию речевых ошибок могут применяться при обучении иноязычному произношению иностранцев. Лингвистические тренажеры для иностранцев, изучающих английский язык, разрабатываются в это же время и Дж. Бернстейном [Bernstein J. et al., 1990: 1185-1188].

В структуре лингвистических тренажеров можно выделить следующие пять основных рабочих блоков: 1) автоматическое распознавание речи (АРР); 2) общая фонетическая оценка качества речи; 3) локализация произносительных ошибок; 4) диагноз ошибок; и 5) обратная связь с пользователем.

Задача блока автоматического распознавания речи осложняется тем, что технологии АРР, обученные на речи носителей языка, заметно хуже справляются с обработкой интерферированной речи иностранцев. Для некоторых языков уже существуют речевые корпуса, дикторы которых не являются носителями этих языков, однако они сравнительно немногочисленны. Качество распознавания речи крайне важно для последующей сегментации акустического материала записи и выделения конкретных ошибок в произношении пользователя: при неправильной сегментации речевого сигнала пользователя все последующие этапы оценки произношения последнего будут также неверны. В связи с этим, чаще всего в CAPT-системах упрощают задачу модуля АРР и предлагают пользователю произнести заданные слова или фразы. Отметим, что за последние тридцать лет был разработан широкий спектр лингвистических тренажеров для самых разных языков и целевых групп, со значительно отличающимся друг от друга

функционалом и принципами работы, но ни в одной из них нет обучения узбекскому языку, что и делает данную сферу актуальной [Ибрагимова С., 2023: 775-777].

Эффективное овладение узбекским языком на фонетическом уровне определяет эффективность протекания целого ряда процессов: распознавания звуковых сигналов и их идентификации, понимания речи, продуцирования разных форм речи – как устной, так и письменной.

Речевые навыки при обучении узбекскому языку как иностранному при использовании интеллектуальной системы вырабатываются параллельно с овладением словарным запасом, в том числе практика запоминания новых лексем сочетается с усвоением грамматических паттернов путем персонализированного обучения.

Функциональная схема лингвистического тренажера обучающий узбекскому языку представлена на рис. 1. Разрабатываемая интеллектуальная система рассчитана на работу без преподавателя. Упражнения и справочный грамматический материал, хранящиеся в базе данных тренажера, позволяют изучать следующие разделы фонетики:

1. Звуки: гласные, согласные, буквы алфавита;
2. Произношение: слогов, слов, фраз;
3. Аудирование: диктант, диалог с тренажером.



Рис. 1. Функциональная схема лингвистического тренажера

Текст упражнения подается на вход лингвистического процессора, который обрабатывает входной поток данных и передает решение в блок проверки упражнений, а также, при необходимости, формирует команды синтезатору речи, формируя корректный интонационный контур. Синтезатор речи озвучивает упражнение, чтобы показать, как правильно должно звучать то или иное слово и как изменяется интонация в процессе чтения данного текста. Одновременно с этим на экране монитора обучаемого тренажер отображает транскрипционную форму озвучиваемого им текста.

После прослушивания текста обучаемый имеет возможность произнести его в микрофон. Таким образом, в тренажере организована обратная связь, позволяющая оценить эффективность процесса обучения узбекскому языку. Речевой сигнал поступает в анализатор речи, который отображает на экране монитора транскрипцию произнесенных звуков для сравнения с эталонными характеристиками указанных параметров озвученного ранее синтезатором текста.

Модуль распознавания речевых ошибок автоматически осуществляет сравнение образца с реальным звучанием, регистрирует ошибку и передает информацию о ней в экспертную систему, где ошибка квалифицируется и через модель обучаемого корректируется стратегия дальнейшего его обучения.

Экспертная система отправляет отчет о статусе выполнения задания в базу данных результатов.

Интеллектуальный интерфейс пользователя выводит информацию по следующим пунктам:

1. Анализ транскрипции.
2. Анализ ошибок и фонетические правила, которые необходимо повторить.

Анализ транскрипции указывает на место в тексте, где обучаемый неправильно произнес звук. Анализ ошибок и фонетические правила, которые необходимо повторить, позволяют обучаемому освежить знания и для их закрепления выполнить дополнительные упражнения по данной теме.

Таким образом, представленный лингвистический тренажер позволит подойти к проблеме изучения узбекского языка как иностранного комплексно, охватывая все стороны обучения, что в дальнейшем послужит прочной основой для успешной межкультурной коммуникации.

Список литературы:

1. Gerstman L., “Classification of self-normalized vowels”. IEEE Trans. Audio Electroacoust. AU-16, 1968. – P. 78-80.
2. Lobanov B.M., “Classification of Russian vowels spoken by different speakers”. J. Acoust. Soc. Am.,1971, V. 49. – P. 606-608.
3. Wakita H., “Normalization of vowels by vocal-tract length and its application to vowel identification”. IEEE Trans. ASSP, 1977, V. 25. – P. 183-192.
4. Bernstein J., Cohen M., Murveit H., Rtischev D., Weintraub M., “Automatic evaluation and training in English pronunciation”. Proc. ICSLP-90: 1990 Internat. Conf. on Spoken Language Processing, Kobe, Japan, 1990. – P. 1185-1188.
5. Ибрагимова С.Н., Мирзаева Г.Р. “Применение современных технологий распознавания речи при обучении узбекскому языку как иностранному”. Республиканская научно-практическая конференция на тему «Вопросы применения инновационных технологий обучения в условиях цифровизации высшего образования (ICT Edu 2023)», Ташкент, 2023, 02-03 февраль. Стр. 775-777.

THE IMPORTANCE OF CORPORA IN TRANSLATION TEACHING ANALYSIS

Akhmedova Anorkhon Nasivali kizi
PhD, Associated professor of
Uzbekistan State World Languages University.

Annotation. Translation analysis involves examining and interpreting text in one language and accurately conveying its meaning in another language. To enhance the accuracy and efficiency of this process, many translators and linguists have turned to corpora – large collections of texts carefully selected and organized for linguistic research purposes. Incorporating corpora into translation analysis lessons can provide valuable insights and help improve translation quality. This article is devoted to give information with some examples of how corpora are utilized in translation analysis lessons.

Keywords: corpus analysis, Contextual Understanding, Terminology Research, Style and Register, Error Analysis, Quality Assurance, Cultural Considerations

Аннотация: Анализ перевода включает в себя изучение и интерпретацию текста на одном языке и точную передачу его значения на другом языке. Чтобы повысить точность и эффективность этого процесса, многие переводчики и лингвисты обратились к корпусам — большим коллекциям текстов, тщательно отобранных и систематизированных для целей лингвистических исследований. Включение корпусов в уроки по анализу перевода может дать ценную информацию и помочь улучшить качество перевода. Эта статья посвящена предоставлению информации с некоторыми примерами использования корпусов на уроках по анализу перевода.

Ключевые слова: корпусный анализ, понимание контекста, терминологические исследования, стиль и регистр, анализ ошибок, обеспечение качества, культурные аспекты.

In the world of language and translation, the use of corpora plays a crucial role in helping students and professionals refine their skills and enhance their understanding of different languages. A corpus refers to a large and structured collection of texts that are used to analyze language use and patterns. When it comes to teaching translation analysis, the importance of corpora cannot be overstated.

Using García-Izquierdo and Conde's (2012) words, "[i]n any event, regardless of their area of activity, most subjects feel the need for a specialised corpus combining formal, terminological-lexical, macrostructural and conceptual aspects, as well as contextual information" (p. 131). [García-Izquierdo, Conde, 2012; 131-156]

One of the key benefits of using a corpus in teaching translation analysis is the exposure it provides to a wide range of authentic texts. These texts come from various sources such as books, articles, websites, and more, reflecting the rich diversity of language usage. By working with real-world texts found in corpora, students can develop a deeper appreciation for the nuances and intricacies of language, thereby improving their translation skills.

Translations have sometimes been mentioned also in relation to a different type of corpus which includes two or more subcorpora. This has been called "multilingual" (Baker 1995) or

"comparable" (McEnery et al. 2008) corpus, and consists of two or more monolingual subcorpora in different languages, each collected according to similar design criteria. As with monolingual corpora it is often suggested that translations should not be included in multilingual comparable corpora. [Federico Zanettin, 2011; 14-23]

Moreover, corpora offer a wealth of linguistic data that can help students identify patterns, trends, and common structures in language. When students have access to a corpus, they can conduct detailed analyses of how words are used in different contexts, explore variations in language usage, and gain insights into idiomatic expressions and collocations. This deep dive into language data is invaluable for students aiming to become proficient translators.

Another significant advantage of incorporating corpora into translation analysis is the ability to enhance accuracy and consistency in translations. By consulting a corpus, students can verify their translation choices, check for potential errors, and ensure that their translations capture the intended meaning of the original text. This process not only helps students produce high-quality translations but also instills in them a sense of precision and attention to detail [Abdurakhmonova, N. 2019,2021,2022].

Furthermore, working with corpora in translation analysis allows students to stay up-to-date with evolving language trends and usage. Languages are dynamic and constantly evolving, with new words, phrases, and expressions emerging all the time. By engaging with a corpus, students can stay current with the latest developments in language, ensuring that their translations are relevant and reflective of contemporary language use.

The use of corpora in teaching translation analysis is paramount for students looking to hone their language skills and become proficient translators. From exposing students to authentic texts and linguistic patterns to enhancing accuracy and staying current with language trends, corpora offer a myriad of benefits that contribute to the overall growth and development of students in the field of translation. By leveraging the power of corpora, students can embark on a journey of exploration and discovery in the world of languages, paving the way for successful careers in translation and beyond [Abdurakhmonova, N. 2019,2021,2022].

Corpora are a crucial tool in the field of translation analysis. They are large collections of texts that are used to study language patterns and aid in the translation process. In translation analysis lessons, incorporating corpora can provide invaluable insights and improve the quality of translations. There are some examples of how corpora are integrated into translation analysis lessons:

1. Contextual Understanding. One-way corpora are used in translation analysis is to gain a better understanding of the context in which words and phrases are used. By analyzing how specific terms are used in different contexts within the corpora, translators can ensure that their translations accurately convey the intended meaning.

2. Terminology Research. Corpora are a treasure trove of specialized terminology. In translation analysis lessons, students can use corpora to research and identify the most appropriate terminology for specific subject areas. This helps ensure that translations are accurate and consistent, especially in technical or academic texts.

3. Style and Register. Corpora can also help students understand the nuances of style and register in different types of texts. By analyzing how certain words and structures are used across various genres and registers, students can fine-tune their translations to match the tone and style of the original text.

4. Error Analysis. When students are learning translation, corpora can be used to identify common errors and pitfalls in translation. By comparing their own translations to examples in the corpora, students can learn from mistakes and improve the accuracy of their translations over time.

5. Quality Assurance. Incorporating corpora into translation analysis lessons can also serve as a quality assurance tool. Students can use corpora to verify the correctness of their translations, identify inconsistencies, and check for adherence to language conventions and norms.

6. Cultural Considerations. Translations often involve cultural nuances and references. Corpora can help students understand how cultural elements are expressed in language and how they can be effectively translated to preserve the cultural context of the original text.

In conclusion, corpora play a vital role in translation analysis lessons by providing students with a wealth of linguistic data and insights. By incorporating corpora into their learning process, students can enhance their translation skills, deepen their understanding of language usage, and produce more accurate and culturally sensitive translations.

References:

1. García-Izquierdo, I., & Conde, T. (2012). Investigating specialized translators: corpus and documentary sources. Ibérica, 23, p. 131-156.
2. Federico Zanettin, (2011) Translation and corpus design. SYNAPS – A Journal of Professional Communication, p. 14-23
3. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 1790-1793). IEEE.
4. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
5. Gatiatullin, A., Suleymanov, D., Prokopyev, N., & Abdurakhmonova, N. (2022, November). "Turkic Morpheme": From the Portal to the Linguistic Platform. In World Conference Intelligent System for Industrial Automation (pp. 181-189). Cham: Springer Nature Switzerland.
6. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
7. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).

III SHO‘BA
NLPDA TIL VA NUTQ TAHLILI
(LINGVISTIK ANALIZATOR: MORFOLOGIK, SINTAKTIK VA SEMANTIK TAHLIL, NUTQ ANALIZI VA SINTEZI)

MAKING USEFUL TOOLS FOR LANGUAGE COMMUNITIES WITHOUT BIG CORPORA: A VERSATILE MORPHOLOGICAL ANALYSER AND SPELLCHECKER FOR SONGHAY BASED ON FREE SOFTWARE

Loïc Boizou
PhD, researcher
Department of Digital Humanities
University of Cologne.

Mohomodou Houssouba
PhD, researcher
Centre for African Studies
University of Basel.

Abstract: The current excitement over data-driven artificial intelligence makes headlines across the world, including in Africa. Accordingly, AI is supposed to enable an exponential growth in digital resources and level the playing field for less-resourced languages. However, this enthusiasm largely eludes the correlation between AI capacity and the availability of massive and well-curated datasets. In this regard, for the most part, African languages do not fulfill the basic requirements to participate in the new “revolution”. The Songhay language is a case in point, with slow and continuous work done over two decades to develop even a basic tool like a spellchecker. Still, even limited experience with localization and building corpora helps toward developing more advanced tools and content over time.

Keywords: free software - spellchecker - morphological tagging - morphology - Songhay language - under-resourced languages

Introduction

While big data and artificial intelligence make the headlines, the increasing digital rift is getting less attention in the media landscape. Only a small number of the world languages, roughly a big hundred, have access to the digital Graal while thousands are still practically or even totally left outside. Even for the languages that entered the digital space, there is a wide range of concrete situations, from AI-supported languages to languages with rough elementary resources. The situation is especially dire for African languages: practically no endogenous African language can be considered as well supported in the digital space. Furthermore, since cutting-edge technologies are highly data-intensive, most world languages, which are hardly written, let alone standardized, cannot really benefit from the development of these technologies. The big divide risks to widen further, unless a boom in speech-to-text technology can somehow mitigate the situation.

This article presents a Hunspell spellchecker and tagger designed for the Songhay language, a language spoken by several million speakers across borders, in Niger, Mali, Burkina

Faso, and Benin. Even if spellcheckers are a relatively old technology, they are still relevant for everyday users as a convenient help to avoid misspellings. As such, spellcheckers constitute a useful language-related technology, as well as some more complex tools such as machine translation systems, optical character recognition, speech-to-text systems or conversational agents. Traditional spellcheckers are quite easy to develop using standard mature tools such as Hunspell, on the basis of a lexicon and, if needed, a morphological model. Although not as advanced as predictive typing, which takes into account the context (the previous words), it avoids the need for significant corpora, which is still a missing resource for a huge number of languages.

The present contribution brings to the fore the weak integration of the African languages in the digital domain, to explain the rationale for developing a Songhay Hunspell module. It offers a general overview of the design of the Songhay morphological model, shows the two main usage cases, as spellchecker and morphological tagger and, finally, opens up perspectives for further enhancements of the tools for the language.

1. African languages in the digital space

As the internet opened on a global scale in 1996, most interfaces and contents were almost exclusively in English. Programmers had been mostly trained in English, and many wouldn't be able to work in their native languages, so even major international languages were sidelined at the outset. In most African countries, literacy and knowledge production were realized in European languages (English, French, Portuguese and Spanish) and Arabic. In many ways, the emergence of English as the nearly exclusive idiom of the new digital age prompted a feverish debate over cultural diversity, social equality and political participation. Paradoxically, a postcolonial framework like the Francophonie would become a vocal advocate for inclusiveness. Thus, African members of the Latin Union and Arab League played a major role in pushing the diversity agenda. Some participated in all three networks and related consortiums. The stakes were deemed high and existentially significant to major languages, not to mention the less-resourced ones spoken across Africa.

The International Francophone Network for Language Management (RIFAL) was a key instrument in coordinating multi-pronged initiatives at official and associative levels across countries. To be sure, while the institutional focus was on creating a French-speaking ecosystem in cyberspace, there were competing visions within the Francophonie. In many countries, French as an official language, though predominant in official business, was and is still spoken by only a small urban elite. The vast majority of the population communicates through local languages and creolized forms of French. Their inclusion into the digital age would require creating various kinds of resources on a sustainable basis. This imperative pushed language activists to aim beyond the OIF framework, by learning from different experiences and adapting the most accessible ones to their native environments.

It is worth noting that earlier projects tried, often in isolation, to produce tools for specific purposes. For instance, the Kamusi Project started in 1994 as a "living" Swahili dictionary, then evolved to become a multilingual platform linking over 40 languages across continents. Many initiators of similar single projects would come in contact once the internet enabled broad and continuous networking across campuses and regions. This was a pioneer era in which individuals and volunteer groups actually had the biggest impact in putting African languages on the map. By 2000, several networks emerged at the forefront. The PanAfrican Localization project (PAL)

quickly federated disparate efforts to translate software interfaces into African languages. It was a daunting task at a time of low-speed and costly dial-up connections. The African Languages Technology Initiative (Alt-I) was another effort to remove technical barriers by sharing know-how. Bisharat greatly contributed to aiming at the larger population, focusing on the development needs of rural communities – agriculture, animal and human health, or small trade [Osborn, 2010].

This loose coalition of actors also played a significant role in advocacy and lobbying work to effect policy change. After 2005, it coalesced into a younger and more structured platform, the African Network for Localization (ANLoc) based in South Africa, with a strong connection with Arabize in Egypt and technically savvy teams in Kenya, Ghana, Uganda, among others. With strong support by the Canadian International Development Research Centre (IDRC), the group carried out a coherent agenda to engage African languages with even a handful of volunteers willing to devote time and energy to the cause [Bailey, 2012].

The Songhay localization project, which started as a one-person initiative in 2004, really took off only in 2007, when a technically savvy volunteer joined and started to push translated user interfaces through the release pipelines. It joined the ANLoc network in 2009 and by 2012 registered a dozen released localizations, including Mozilla Firefox.

Apart from localization, access to language-related tools also remains scarce. Based on an overview considering two important tools for everyday users[15], spellcheckers and machine translation systems, the number of supported African languages is around 20 or sometimes even less.

Machine Translation systems for African languages

Google Translate + Facebook	Afrikaans, Amharic, Ganda, Hausa, Igbo, Lingala, Malagasy, Northern Sotho, Oromo, Somali, Swahili, Xhosa, Yoruba, Zulu
Google Translate only	Akan, Bambara, Chewa, Ewe, Kinyarwanda, Shona, Southern Sotho, Tigrinya, Tsonga
Facebook only	Pulaar-Fulfulde, Tswana, Swazi, Wolof

Hunspell modules for African languages (with number of words in the lexicon)[16]

Afrikaans (188,366), Akan (5007, full form), Amharic (182,552, full form), Bambara (26,012), Kituba (2171), Lingala (5194, some grammatical annotations), Malagasy (16,743, full form), Moore (105,696, full form), Sango (13,017), Tswana (61,598, full form), Shona (37,287), Swahili (Kenyan 48,317, full form, Tanzanian 67,900), Zulu (108,900, some grammatical annotations)[17]

[15] As far as virtual keyboards are concerned (also an important core technology), the number of Gboard configurations is far higher, but the predictive typing is often provided with almost no initial data (e.g. for Songhay).
[16] Theses modules were found in three locations: https://github.com/elastic/hunspell/tree/master/dicts, https://github.com/wooorm/dictionaries/tree/main/dictionaries, https://extensions.libreoffice.org/?q=Spellchecker&action_doExtensionSearch=Search
[17] There is also the very beginning of a Kinyarwanda Hunspell, but with 62 words only.

The wide scope of the size of the lexicons reveals the very diverse level of maturity of these Hunspell modules, yet the number of words in the lexicon alone can be partly misleading: for inflected languages, a relatively small lexicon with productive affixal patterns can be as expansive as a larger full-form lexicon. In certain mentioned Hunspell modules, e.g. Akan or Amharic, some patterns are defined in the affix files but are not used in the lexicon. No dictionary provides the structured set of grammar features (although the Lingala and the Zulu modules do include some pieces of grammatical information) that would allow the system to operate not only as a spellchecker but also as a morphological tagger.

The scarcity of the resources for the African languages, which is illustrated with both examples, is mainly the consequence of a sharp linguistic imbalance inherited from the colonial situation which left English, French and Portuguese (also Spanish to a far lesser extent) as strongly dominant languages [Ebongue and Hurst, 2017, more specifically Nakayiza and Ssentanda 2017, Kamwangamalu, 2019] as well as a pattern of state unilingualism [Ekkehard Wolff, 2017]. In some countries, Arabic (outside its African core territories) too imposes itself over local languages through religious prestige or official language policymaking. In almost all African countries, these mentioned languages are *de jure* or *de facto*[18] state languages (Ethiopia being an exception). When African languages are official or co-official, their role is most frequently purely symbolic (except for Somali and Swahili). Thus, exogenous languages dominate domains such as administration, education, big business, often religion, while African languages remain mainly vernacular or, at best, spoken vehicular and initial literacy mediums in primary schools. As a secondary consequence, there is a lack of available written accessible material for IA-based data-hungry technologies. Finally, given the administrative underuse of the local languages and the economic weakness of most African countries, the financial gains that could be expected from investing in resources for African languages are meager. The outcome might improve with the development of African middle classes, but only under the condition that this movement does not accelerate the relegation of endogenous languages. In such a grim context, each concrete tool for under-resourced African languages can have a slight positive outcome.

2.1 Rationale for Hunspell

Hunspell[19] is the most common tool for designing spellcheckers. It is free, mature and stable and well supported especially in the context of free software like Linux, which should be an important aspect to have in mind when fundings is scarce. It is the standard spellchecker for LibreOffice tools, for most browsers (Firefox, Chrome, Opera) and can be combined with the textual database ElasticSearch or used through Python[20] or R[21] libraries. The Hunspell files can be provided directly, but LibreOffice allows packing them and recording them as an extension on their extension bank. All these characteristics ensure Hunspell systems a good level of availability for any language community.

[18] Mali is a striking example of this situation. According to the Government Gazette of Mali (Décret N°2023-0401/PT-RM du 22 juillet 2023 portant promulgation de la constitution), French is not an official language, but is the working language. Nevertheless, as an obvious illustration of the *de facto* official status of French, this Government Gazette is published in French only.

[19] http://hunspell.github.io/

[20] https://pypi.org/project/hunspell/

[21] https://cran.r-project.org/web/packages/hunspell/vignettes/intro.html

Hunspell systems are a combination of lexicon and affix files encoded as Unicode plain texts. Therefore, Hunspell spellcheckers can be designed with simple unicode editors. As a drawback, it is easy to corrupt the data while developing or correcting Hunspell models, but this issue does not affect the end users.

Hunspell has a second potential benefit: if the model is properly designed, it can be used as a morphological tagger and lemmatizer. Morphological analysers are not a resource for everyday users but they are a traditional tool to provide annotated language data for linguists. Using Hunspell for both usage cases is an efficient way to avoid work reduplication for under-resourced languages. As for spellchecking, this task is carried out over isolated wordforms, without contextual disambiguation. It is an obvious limitation, but it a also what allows the tool to be designed without available corpus[22].

Previous Resources

Two pre-existing resources allowed developing the Songhay Hunspell Module very quickly.

The core element of Hunspell systems is a sufficiently comprehensive and cohesive lexicon. The development of the Songhay Hunspell module relies on previous work by the authors of the present contribution, as well as Abdoulbaki Seydou Cissé, who is the main developer of the website songhay.org, and the leading Songhay lexicographer Youssouf Mohamed Haïdara [Haïdara, 2010].

A draft morphological model was also ready. It was developed while designing a previous analyser for Songhay in Python, but a decision was made to switch to a widely available standard solution such as Hunspell in order to increase its availability. The initial model based on a traditional grammar (mainly the short introduction in Prost, 1956) was improved by generating all nominal inflected forms for the given lexicon and by checking the output manually. In this initial model, the nominal inflectional component was only developed for substantives, excluding adjectives, numerals and pronouns. The new enhanced model is described in the next section.

3. The designed Songhay Morphological Model

The present Hunspell module is designed for Mali's standardized Songhay, which is based on the Gao dialect, with minor adaptations needed to take into account localized specificities (e.g. -ante / -anta). Standardization is necessary to make a coherent spellchecker: allowing clearly different dialectal variations in the same text would risk enabling very artificial text and make the model far more complex, thus compromising the maintenance of the lexicon.

Hunspell has some limitations in the number of generation steps that impact the design for agglutinative languages[23] or, even worse, for polysynthetic languages. For Songhay, which possesses a relatively simple inflexional morphology and relatively stable stems, Hunspell fits perfectly.

As already mentioned, the model was designed in order to be usable not only as a spellchecker but also as a morphological analyser. For that purpose, the two most important aspects are to link all suppletive forms to the lemma through the Hunspell tag *st:* and to provide a full set of morphological tags for each form (either stem or derived).

[22] But Hunspell can also be a starting point in this direction [Pirininen and Lindén, 2010].
[23] It would require long compound endings, e.g. number + possession + case.

Parts of Speech (number of words)	Categories
noun (6069 common + 21 suppletive forms, 11 proper) verb (1925) adjective (201) adverb (75) interjection (56) numeral (24 + 9 suppletive forms) pronoun (15 + 11 suppletive forms) auxiliary (11 + 5 suppletive forms) preposition (11) postposition (10) conjunction (6) determiner (4) particles/adverbs[24] (4)	negative (aux.) definite/indefinite (noun, adj, num + rare pron) singular/plural (noun, adj, num, pron) person 1, person 2, person 3 (pron) emphatic (pron) complement (pron) proper (noun) ordinal (num) transitive (verb) infinitive (verb) participle (verb)

The categories include different levels of information, besides classical morphological categories (number, definiteness, polarity, person), there are single tags to mark some particular features, for propers nouns, emphatic and complement pronouns and transitive (or factitive) verbs. There are also tags to mark infinitives and participles as well as ordinal adjectives derived from numerals.

The lexicon contains 8468 words, but the real number of lexical units must omit the 46 suppletive forms; that is, 8422 lemmas. The remarkably low number of adjectives is related to a feature of the Songhay language whereby a significant part of words that would be semantically similar to Indo-European adjectives are verbs e.g. *faala* 'to be easy, to be cheap'.[25]

The lexical coverage is extended by the systematic derivation patterns: verbal adjectives / participles, verbal nouns / infinitives and ordinal adjectives. Given the affixal classes (14 suffix classes and 1 prefix class), the approximate number of word forms is about 30,000 nouns, 1000 adj, 20,000 verb forms, other about 300, probably about 55,000 forms in total.

The Hunspell model includes four rewriting rules, that allow adding the caron to some consonants before the letter *i* and *e* when missing: *si* → *ši*, *zi* → *ži*, *se* → *še*, *ze* → *že*. Usage and even norm are hesitant since in *s* and *z* (usually [s], [z]) are regularly palatalized before i and e as [ʃ] and [ʒ] in the Gao standard pronunciation, making the marking redundant (but needed in a strict phonographic approach) and slightly altering the stem in some cases (*bolši* / bolsi 'to boast' → *bol**s**andi*).

3.1 The Nominal Component

Songhay nominals have regularly four different forms, e.g. the word *gallu* 'town':

- gallu (singular indefinite)
- galloo (singular definite)

[24] No strong decision was made for these words.

[25] Languages with verb-like adjectives [Stassen, 2013] are not a rarity, Japanese being a widely cited example [Baker, 2003].

- galley (plural indefinite)
- galluyan (plural definite)

The Hunspell component describes nominals in three different ways.

3.1.1 Productive types

For usual nominals, the three suffixed forms are generated from the base form (singular indefinite). Songhay has three main models (the plural definite has only one form and therefore has no influence on the paradigms):

Songhay nominal endings

	paradigm 1	**paradigm 2**	**paradigm 3**
Singular definite	-oo	-aa	-aa
Plural indefinite	-ey	-ey	-awey
number of base words (excluding words with several combinations)	3417	756	2057

Some words show a certain variation. There are 55 words that can appear as 2nd or 3rd paradigms, two words of 1st or 2nd paradigms, and three words from the 3rd paradigm that can also appear with the *-oo* ending. As it is obvious from the previous table, the second model can be described as a mixed one, with singular similar to the third group and plural similar to the first group. Therefore, the Hunspell model was built about two singular suffixes classes and two plural suffix classes.

Regular nominal class used in the model

	paradigm 1	**paradigm 2**	**paradigm 3**
Singular definite	-oo (class O)	-aa (class A)	
Plural indefinite	-ey (class E)		-awey (class L, L for long)

It means that usual base words are defined by a combination of a singular class (O or A) and a plural one (E or L). If needed, the split between singular and plural would also allow to easily integrate singularia tantum and pluralia tantum in the lexicon.

Based on this approach, there are 3426 words within the class O, 2828 in the class A, 4246 in the class E and 2062 in the class L. In case of paradigmatic variability, a base word can belong to more than two classes.

These regular models imply the removing/substitution of occurring final vowels from the base form, e.g. *gallu* > *galloo* (the final *-u* is replaced by the definite *-oo* suffix). When the combination involves an extra consonant, it is generated through other models described in 3.1.2.

3.1.2 Minor paradigms

Minor paradigms describe the sets of lemmas that require an extra consonant between the stem and the ending. Such a consonant can be a final stem consonant lost in the base form or an extra *w* consonant to avoid a hiatus after a stem in stable vowel. It concerns a limited number of base words and their compounds.

The model defines the following classes, which combine singular and plural suffixes (contrary to the ones described in 3.1.1 for which singular and plural are split):

- y: suffixes -oo/-ey with *y* as a linking consonant (2 base words).
- w: suffixes -oo/-ey with *w* as a linking consonant (6 base words).
- w: suffixes -aa/-ey with *w* as a linking consonant (8 base words).
- g: suffixes -oo/-ey with *g* as a linking consonant (for the word *doo* 'place' and it compounds only).
- g: suffixes -aa/-ey with *g* as a linking consonant (for the word *baa* 'part' only).

The classes Y, W and g involve the shortening of long final vowels before *y*, *w* or *g*.

3.1.3 Idiosyncrasies

Idiosyncrasies are directly added in the base lexicon with a shared stem field (*st:* tag) for the morphological tagger. They may involve unexpected consonant or vowel alterations. It concerns the nouns *maa* 'nom, terme', *mee* 'mouth', *moo* 'eye', *alfa* 'peasant', *ban* 'good health' as well as some pronouns.

3.1.4 Numeral specifics

Numerals follow the nominal patterns previously described, but are extended by two specific classes. The o class enables to form ordinal adjectives from (cardinal) numerals. The i class, the only prefix class, generates the numeral form prefixed by *i-*, *woy* → *iwoy* (for numbers from two to nineteen). This prefixed variation appears in standalone number forms.

3.2 The Verbal Component

In Songhay tense, mood and polarity are expressed by verbal particles, therefore verbs have no inflection. As a consequence, the Hunspell verbal component is restricted to derivational patterns.

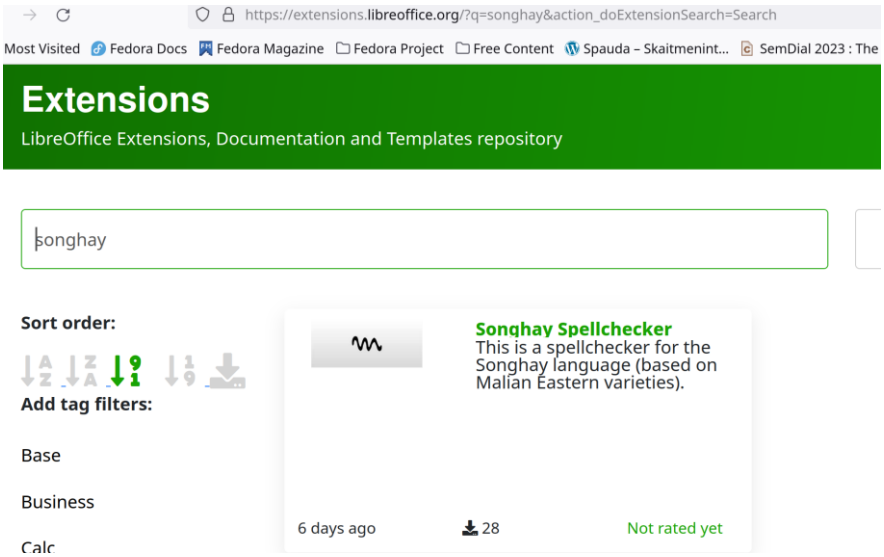
The verb affixes allow to generate the main verbo-nominal forms, the participle / verbal adjectives and the infinitive / verbal nouns. According to the nominal system, these deverbal forms are provided in singular and plural, in definite and indefinite. The verbo-nominal generation relies on the classes V and v, the latter involving a linking consonant *y* (after i) or *w* before endings, possibly with a long vowel shortening in the stem (*-uu* → *-u*, *-ii* → *-i*).

For some verbs, the generation model allows deriving the transitive/factitive suffixed counterpart in *-andi* (e.g. *naanay* 'be confident', *naanayandi* 'give confidence to sb') through classes D and d (the latter with a linking consonant as for v). This derivation pattern implies further derivation of the verbo-nominal forms for the derived transitive verb.

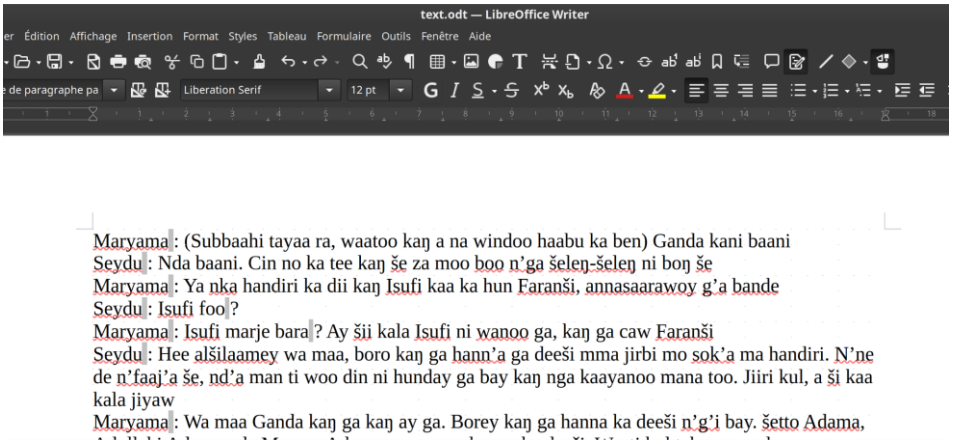
The next section illustrates the use of the Songhay Hunspell module, either packaged or not.

4. Use cases

The Songhay Hunspell Module is packed as a LibreOffice extension and can be downloaded from the dedicated webpage https://extensions.libreoffice.org/.



The spellchecker is directly usable to check the spelling of texts written using the tools provided by LibreOffice (Writer, Calc, Impress, etc.), as illustrated in the next screen capture.



The Songhay Hunspell files can also be used without the LibreOffice packaging, for browsers, for Elasticsearch or for the whole Linux system. Once installed on a Linux system, it is possible to use it for morphological analysis through the command line.

```
Maryama

Subbaahi  st:subbaahi po:noun ts:sing_indef

tayaa  st:tayo po:adj ts:sing_indef is:sing_def
tayaa  st:tayo po:noun ts:sing_indef is:sing_def
tayaa  st:taya po:adj ts:sing_indef is:sing_def
tayaa  st:taya po:noun ts:sing_indef is:sing_def

ra  st:ra po:postp

waatoo  st:waati po:noun ts:sing_indef is:sing_def

kaŋ  st:kaŋ po:pron
kaŋ  st:kaŋ po:noun ts:sing_indef
kaŋ  st:kaŋ po:verb

a  st:a po:pron is:3_sing

na  st:na po:aux
```

File sources and results files can be easily piped through Linux standard redirection operators. The output format is a verticalized text with each word form and its morphological interpretation with the lemma and the linguistic features and separated from the next word by an empty line. When the morphological analysis fails, the interpretation part remains empty (e.g. the proper name *Maryama* in the provided example). When a word form is morphologically ambiguous (e.g. *tayaa* or *kaŋ*), the word form is repeated for each possible interpretation.

5. Conclusion and Prospects

Given the lack of resources for African languages, it is important to try to find some paths to provide at least some basics for a greater number of languages with an emphasis on free tools, sharing and reusability. Hunspell module allows to fill a punctual gap for spellchecking and morphological analysis. Building on a few available language resources, namely the localization of some software interfaces and an internet lexicon on the website songhay.org, the Songhay Hunspell module is the first publicly available tool to automatically process the Songhay language. Except for some Songhay virtual keyboards, these handful of resources make up the single full digital backbone for Songhay to date. What could be the next steps?

Firstly, the Songhay Hunspell module can still be improved. The lexicon is in the process of being expanded. The analyser was run over all the texts at our disposal and the list of unknown words is currently under review. Indeed, this list contains incorrect words but also words missing in the Songhay Hunspell lexicon. After this list is thoroughly checked and purged, it is expected that a few thousands words will be added to the existing lexicon. Another expected task is to launch an initiative within the Songhay community to collect a significant list of proper nouns, mainly names, surnames and toponyms. Finally, some improvements are needed for shortened words with apostrophes, which are frequent in Songhay.

Beyond the goal of improving the Songhay Hunspell module, another step would be to provide a contextual tagger able to select one morphological interpretation for ambiguous wordforms. Together with a syntactic tagger, it would provide a very minimal set of automatic linguistic analysers, although such tools are not relevant for the everyday users of the language. But such analysers as well as other tools more directly useful for end users, like predictive typing, need at least basic task-oriented corpora. It means that in order to move further into developing tools for the Songhay community, corpus development cannot be avoided. That is why the incentive to produce written and spoken materials and to collect texts and recordings in Songhay is a fundamental task. It requires suitable tools but the challenge of mobilizing the whole language community is also of crucial importance, and it is not an easier issue than the technical one.

References:

1. Baker, M. ‘Verbal Adjectives as Adjectives without Phi-Features’. *Proceedings of the Fourth Tokyo Conference on Psycholinguistics.* Yukio Otsu (ed.), Keio University, 1-22, 2003.
2. Bailey, D. ‘Software Localization: Open Source as a Major Tool for Digital Multilingualism’. In Vannini, L., Le Crosnier, H. (eds.). NET.LANG: Toward the Multilingual Cyberspace. C&F Editions, 2012. http://net-lang.net
3. Ebongue, A. E., Hurst, E. (eds.). S*ociolinguistics in African Contexts: Perspectives and Challenges*, Switzerland: Springer International Publishing. 2017.

4. Ekkehard Wolff, H. Language ideologies and the politics of language in post-colonial Africa. *Stellenbosch Papers in Linguistics Plus*, Vol. 51, 2017, 1-22 doi: 10.5842/51-0 - 701
5. Haïdara, Y.M. *Dictionnaire soŋay-français. Kalima citaabu soŋay-annasaara senni*. Bamako: EDIS, 2010.
6. Kamwangamalu, N. M. 'Language ideologies and practices in Africa'. *Journal of Sociolinguistics*. 2019; 23: 543–554. https://doi.org/10.1111/josl.12327
7. Ògúnrẹ̀mí, T., Onyothi Nekoto, W., Samuel, S. *Decolonizing NLP for "Low-resource Languages"*, 2023.
8. Pirinen, T., Lindén, K. 'Creating and Weighting Hunspell Dictionaries as Finite-State Automata'. *Investigationes Linguisticae*. vol. XXI, 2010.
9. Osborn, D. *African Languages in a Digital Age: Challenges and Opportunities for Indigenous Language Computing*. HSRC & IDRC, 2010. Prost, A. *La langue soŋay et ses dialectes*. Dakar: IFAN, 1956.
10. Ssentanda, M., Nakayiza, J. "Without English There Is No Future": The Case of Language Attitudes and Ideologies in Uganda. In: *Sociolinguistics in African Contexts*, 2017, DOI: 10.1007/978-3-319-49611-5_7
11. Stassen, L. 'Predicative Adjectives'. In Dryer, M. S., Haspelmath, M. (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013. (http://wals.info/chapter/118, accessed on 2015-05-02)

ADVANCED NLP AI TOOLS IMPORTANCE IN MACHINE TRANSLATION FROM UZBEK TO ENGLISH

Nazirova Elmira
DSc, Head of the Department of Multimedia Technologies,
Tashkent University of Information Technologies
named after Muhammad Al-Khwarizmi.
E-mail: elmira_nazirova@mail.ru

Usmonova Kamola
PhD of the Department of Multimedia, Technologies,
Tashkent University of Information Technologies
named after Muhammad Al-Khwarizmi.
E-mail: kamolausmonova1234@gmail.com

Abstraction: This article explores the substantial influence of artificial intelligence (AI) on digital technology, particularly in the field of machine translation. It investigates the strides and obstacles in AI-powered machine translation, focusing specifically on Uzbek and English. It investigates the exciting developments in language translation technology, pushing the boundaries beyond the capabilities of current AI tools and Natural Language Processing. Recent strides in NLP, particularly in the domain of neural machine translation, have propelled us towards more accurate and contextually relevant translations. The integration of advanced NLP techniques, such as attention mechanisms and transformer architecture, has elevated the quality of translations, pushing the boundaries of what AI tools can achieve. Exciting developments in multilingual NLP models, fueled by AI-driven research, hold promise for overcoming existing limitations and unlocking new frontiers in cross-linguistic communication.

Keywords: Artificial Intelligence, Machine Translation, Uzbek, Russian, English, Natural Language Processing (NLP), Translation Technology, Linguistic Resources, Corpora.

Аннотация: В этой статье исследуется значительное влияние искусственного интеллекта (ИИ) на цифровые технологии, особенно в области машинного перевода. Она исследует прогрессы и препятствия в машинном переводе, основанном на ИИ, с особым акцентом на узбекском и английском языках. Также рассматриваются интересные нововведения в технологии языкового перевода, выходящие за рамки возможностей существующих инструментов ИИ и обработки естественного языка. Недавние достижения в области обработки естественного языка, особенно в сфере нейронного машинного перевода, подвигают нас к более точным и контекстно релевантным переводам. Интеграция передовых техник обработки естественного языка, таких как механизмы внимания и архитектура трансформера, повысила качество переводов, расширяя возможности инструментов ИИ. Захватывающие нововведения в мульти язычных моделях обработки естественного языка, стимулируемые исследованиями, основанными на ИИ, обещают преодолеть существующие ограничения и открыть новые горизонты в межъязыковой коммуникации.

Ключевые слова: Искусственный Интеллект, Машинный Перевод, Узбекский, Русский, Английский, Обработка Естественного Языка (NLP), Технология Перевода, Лингвистические Ресурсы, Корпусы.

Introduction

These days machine translation is very important for breaking down barrier between language. As Snell-Hornby (2006) mentioned that in the 1990s English became prevalent language for interacting as a technological lingua franca and allowed machine translation to localizing content. This shift was driven by the growing demand for rapid communication and globalization, highlighting the importance of natural language processing (NLP) techniques in understanding language structures while bridging cultural differences. Moreover, advancement of the evolution of technology designed to automatically interpret text or speech across diverse languages finds representation in machine translation—a fusion of artificial intelligence (AI) and language translation. As McKinsey (2018) pointed out that by digital productivity is expected to rise for the next decade by 60% in economic and social benefits. Enhancement and usage of AI and machine learning, with particularly focuse on NLP-driven machine translation. Yet, as Galati and Riediger (2017) stated that there appears some threat to jobs and population, thus machine translation will take over human role in the upcoming future. Despite these challenges, AI-powered translation tools offer undeniable benefits. By leveraging NLP algoritms, they can quickely process and translate large volumes of text with remarkable efficiency. According to Forcada (2017) Neural Machine Translation (NMT) is clearly out-lined as a “corpus-based machine translation’s new breed” and it demand extensive analysis of huge corpora of SL as well as TL sentences. As innovation computational approach, namely neural network is introduced. In addition, it is necessary for translation from Uzbek into English. While AI-powered translation tools have made remarkable strides, they face hurdles when dealing with intricate languages like Uzbek and English. These complexities have spurred an exciting journey towards a newer, more refined breed of translation technology that goes beyond mere word conversion.

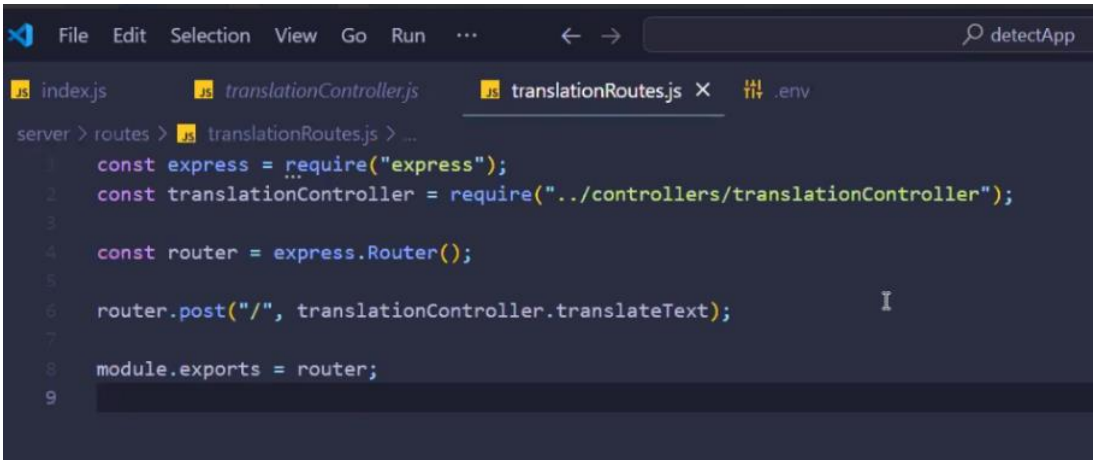
Challenges in Uzbek language translation

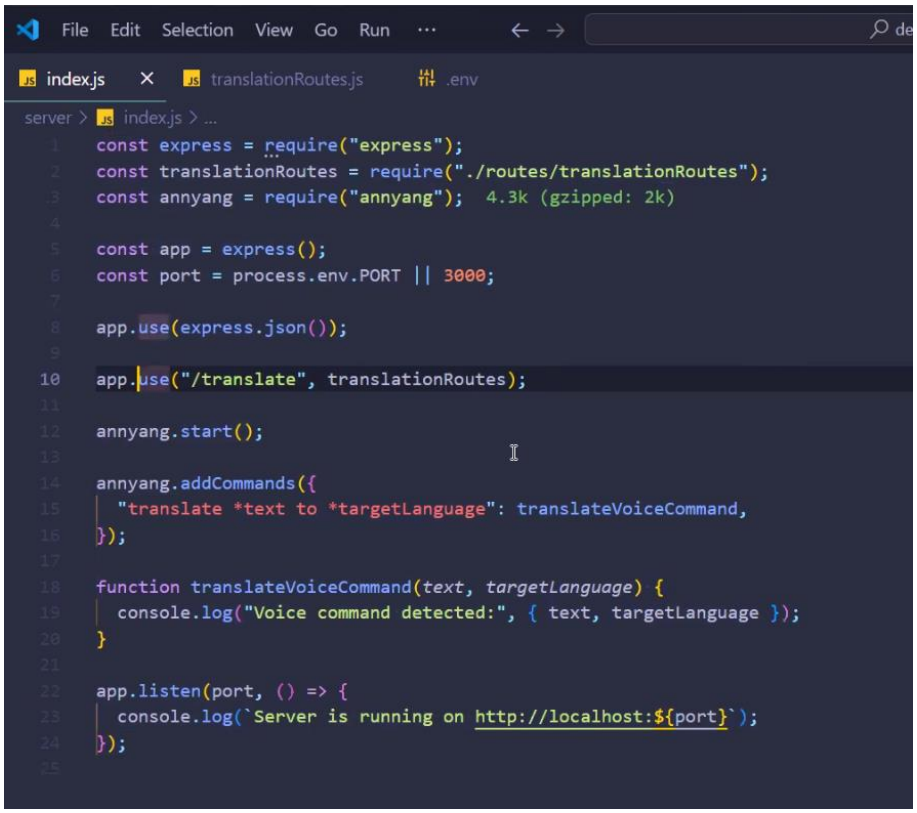
Unlike languages such as English, Uzbek lacks substantial linguistic assets like text corpora, annotated datasets, and lexicons, hindering the development of accurate NLP AI tools. Additionally, the complex morphology and syntax of Uzbek demand specific linguistic expertise and resources, further complicating the development of effective NLP AI tools. Except Uzbek, there is also some challenges in other language such as Russian translation. Languages rich in morphology, such as Russian, present complexities for machine translation systems. Vocabulary limitations fail to encompass Russian's intricate grammatical changes, affecting translation quality. Hence, it is necessary to use NLP models in order to train on large datasets can learn these patterns and provide contextually appropriate translation.

AI tools importance of Uzbek machine translation

As Munday (2012) pointed that translation has risen in sophistication with usage of technology as a tool for translation and it is explained scientific discipline that adduce both the theory and practice of translation. There are different tools for translating such as Google translator, Yandex and others. Yet, there are some challenges during translating using them. As Rudy and Christanta (2020) mentioned that through observation they find out that Google translate functions is not perfect translator in some cases as well as there are minor errors and diminished

precision. The reason of its inaccuracies, this translation tool indicates limited use. As a result, a mere alternation in translation meaning impact significantly the result's quality. Hence, the main barrier occurred when users use a translation tool, they do not get intended meaning because of conveying inaccurately. Moreover, In Uzbekistan, many people struggle translating some articles or different resources from Uzbek into English. The reason there is no trustworthy tool. This is because Uzbek and English structure very different with each other. Hence, there is some problems while translating using google or Yandex translator. Considering some translation tools such as Google translator, Yandex translator, UzbekVoice AI, MohirAI, Bing Microsoft translator, Taia, and others, we can notice inaccurate translation. For example, Google Translate struggles with conveying accurate meanings due to its inability to consider context. Complex texts and context-dependent nuances, uninterpretable without deep language knowledge, increase the likelihood of errors. Unlike professional translators who utilize various tools like glossaries, back translation, and reviewers to prevent literal and nonsensical translations, Google's direct translation often results in nonsensical outcomes. After translating, it is necessary to look through by someone and make context understandable. Hence, by harnessing the power of AI and NLP, researchers are pushing the boundaries of current capabilities, paving the way for more accurate, fluent, and culturally nuanced translation. The translation quality heavily relies on the specific language pair involved. Languages like Uzbek and English, more prevalent in Google's database, yield more accurate translations, whereas languages with limited database availability might suffer from inaccuracies. Grammatical errors are common in Google Translate as its system relies on language pair frequency rather than adhering to grammatical rules. This method leads to translations with significant grammatical flaws. Moreover, Google Translate lacks error correction systems. Users cannot report errors or proofread translations unless they are proficient in both the source and target languages. While creating machine translation tool we will get some words range from google and relook every word because there are some corrections needed. Yet, after deep analyses, we use our own algorithm in order to get more accurate output.





Enter the new breed of translation tech

It is time to deep analysis the morphology and syntax both language and create tool, which make translation clearer. After observation different platforms and tools, it was obvious there is no speech recognition from Uzbek to English and vice versa. Google provide this recognition for other languages such as Russia, German, however, there is no any microphone in order for speech recognition for Uzbek language. Similarly, Yandex translator also does not provide and detection. Hence, it is essential to do tools by using AI for Uzbek language. More precise, a tool that incorporates Uzbek cultural references, idiomatic expressions, and regional variations into the translation algorithms and speech recognition. This involves databases or modules specifically designed to comprehend cultural nuances, ensuring more culturally sensitive and accurate translations. It is developed for tailoring to Uzbek that prioritizes understanding context and nuances within sentences.

Conclusion

Striving for the advancement of AI translation tools designed for Uzbek signifies a significant leap toward facilitating seamless cross-cultural communication. Prioritizing the development of linguistic resources, contextual understanding, and cultural sensitivity within these tools is pivotal. By fostering collaborative platforms, utilizing advanced neural methodologies customized for Uzbek, and ensuring grammatical precision, these enhancements aim to elevate Uzbek translations to new levels of accuracy and cultural resonance. This concerted effort towards refining AI translation for Uzbek brings us closer to a future where language barriers diminish, fostering deeper connections and mutual understanding among Uzbek speakers and the global community.

References:

1. Abror, Sh. (2023). Recent advancements in Uzbek NPL tools. *Kompyuter lingvistikasining zamonaviy texnologiyalari- CTCL"2023".* Ilmiy amaliy anjuman. www.myscience.uz.
2. Bughin, Jaques, Manyika James. 2018. "The promise and challenge of the age of artificial intelligence". McKinsey Global Institute, May. https://www.mckinsey.it/idee/the-promise-and-challenge-of-the-age-ofartificial-intelligence 30/10. (Last accessed 13/10/2018).
3. Forcada, Mikel L., 2017. "Making sense of neural machine translation", in Translation Spaces (6)2 pp. 291–309 https://www.dlsi.ua.es//~mlf/docum/forcada17j2.pdf (Last accessed 29/10/2018).
4. Galati, Gabriele & Riediger Hellmut, 2017. "Traduttori Umani o Traduzione Automatica Neurale?". In Laboratorio Weaver, 24/06. http://www.fondazionemilano.eu/blogpress/weaver/2017/06/24/traduzioneautomatica-neurale/ (Last accessed 09/11/2018).
5. Kai, S. & Yue, Z. & Weihua, L (2018). Improved English to Russian translation by Neural suffix Prediction. https://dioi.org/1048550/arXiv 1801.03615
6. Munday, Jeremy. (2012). Introducing Translation Studies: Theories and Applications. ResearchGate. https://www.researchgate.net/publication/317451115_Jeremy_Munday_Introducing Translation_Studies_Theories_and_Applications.
7. Rudy, S. & Christanta, R. (2020). Problems in the application of google translate as a learning media in translation.384-389. https://dioi.org/10.30743/ii.v4i2.2893
8. Snell-Hornby, Mary. 2006. The Turns of Translation Studies - New paradigms or shifting viewpoints, Amsterdam; Philadelphia: John Benjamins Publishing Company.

THE DEVELOPMENT OF AN APPLICATION FOR COLLECTING AND FORMING CORPORA OF THE KAZAKH LANGUAGE

Karyukin Vladislav
PhD, Senior lecturer of
Al-Farabi Kazakh National University, Almaty, Kazakhstan.

Abdurakhmonova Nilufar
Doctor of science, Professor,
National University of Uzbekistan
named after Mirzo Ulugbek,Tashkent, Uzbekistan.

Annotation: Today, text processing and analysis tasks are crucial in developing the NLP field. Most of the tasks in this field, like text generation, sentiment analysis, and machine translation, require many language resources. While multi-resource languages, such as English, German, French, Chinese, and others, benefit from large, diverse corpora that facilitate the development of robust language models, low-resource languages face significant challenges related to data scarcity. This paper focuses on the lack of resources for the Kazakh language, which is one of the languages in the Turkic group, for high-quality model training in various NLP tasks. In order to increase the size of the available corpora, the research proposes the approach of parsing the Adilet legislative website, which includes a very large collection of well-structured and error-free texts. The dataset was gathered in two phases. In the first step, the links to the Adilet website were collected using the Content Downloader program. Then, the parser based on the Selenium WebDriver was utilized to extract data and form a dataset with *the title, the date of the law article, its text, and a URL link. The total corpora included 9575 texts.*

Keywords: data parsing, text processing, low-resource languages, Selenium WebDriver, Adilet website, Kazakh.

Аннотация: На сегодняшний день задачи обработки и анализа текста имеют решающее значение в развитии NLP области. Большинство задач в этой области, такие как генерация текста, анализ настроений и машинный перевод, требуют большого количества языковых ресурсов. В то время как многоресурсные языки, такие как английский, немецкий, французский, китайский и другие, получают развитие благодаря большим и разнообразным корпусам, которые способствуют разработке надежных языковых моделей, малоресурсные языки сталкиваются с серьезными проблемами, связанными с нехваткой данных. В данной статье основное внимание уделяется нехватке ресурсов казахского языка, который является одним из языков тюркской группы, для качественного обучения моделей различных NLP задач. Для увеличения размера доступных корпусов в исследовании предлагается подход к анализу законодательного сайта Адилет, который включает в себя очень большую коллекцию хорошо структурированных и не содержащих ошибки текстов. Из данного сайта датасет формировался в два этапа. На первом этапе ссылки сайта Адилет были собраны с помощью программы Content Downloader. Затем парсер на базе Selenium WebDriver был использован для извлечения данных и формирования датасета,

включающего заголовок, дату статьи, ее текст и URL адрес. Всего корпус включал 9575 текстов.

Ключевые слова: парсинг данных, предобработка данных, малоресурсные языки, Selenium WebDriver, веб-сайт Адилет, казахский язык.

Introduction

Currently, text data processing tasks play a significant role. The rapid development of machine learning [Mokhamed, 2024; 2] and large language models [Choi, 2024; 3] requires an increasing volume of corpora for text generation tasks [Jaesub, 2024; 5], sentiment analysis [Karyukin, 2022; 20], machine translation [Guo, 2023; 6], development of question-answer systems [Xu, 2024; 7] and other areas of data analysis. All these tasks require a large amount of language resources [Zhong, 2024; 10]. Multi-resource languages such as English, German, French, Spanish, Portuguese, and Chinese have extensive collections of text data, including scientific articles [Dagdelen, 2024; 7], news resources [Abd El-Mageed, 2024; 11], web pages [Modi, 2022; 15], literary works, and much more. Large corpora allow you to train more powerful and accurate language models. A wide range of available data allows developers to create custom NLP tools for various systems, including personalized recommendation and question-answer systems. For low-resource languages, this situation changes significantly and requires a serious challenge. Without enough data, machine learning algorithms cannot learn effectively, which hinders the creation of high-quality models for machine translation, sentiment analysis, speech recognition, and question-answering systems. This problem is especially relevant for systems processing the Kazakh language. There is a great lack of quality resources for the Kazakh language. The texts are only in Kazakh, Russian, and English on some sites. Such sites include the KazNU website (http://www.kaznu.kz), the Bolashak International Scholarship (http://www.bolashak.gov.kz), the Eurasian National University (http://www.enu.kz), Kazakhstan Post (http://www.kazpost.kz), news portals (http://inform.kz, http://tengrinews.kz) and others. In the work [Karyukin, 2023; 3], multilingual corpora from these sources were used to prepare a Kazakh-English machine translation system. This paper presents an expansion of the Kazakh language corpus by parsing the legislative website https://adilet.zan.kz/kaz/. This website contains a large number of laws of the Republic of Kazakhstan. The texts are of a formal business nature, are grammatically correct in structure, do not contain errors, and include many sections, such as Constitution, Constitutional law, Code, Law, Order, Decree, Agreement, Amendments, Article, Article of agreement, Changes, Charter, Classification, Concept, Conclusion, Conditions, Convention, Methods, Norms, etc.

Related works

In the field of data analytics, many works explore various aspects of data processing. The work [Wei, 2024; 5] analyzes the sentiment of comments in social networks of 22 different domains. The dataset includes more than 200 thousand reviews in Chinese. In the article [Afli, 2016; 10], the construction of statistical machine translation systems, which is demanding regarding the volume of corpora, is being studied. To train the system, texts in the source and target languages are used, obtained from Euronews websites and TED. The model's performance was assessed by analyzing the English-French translation, which showed significant improvements compared to the baseline. Multiclass classification of dialogue texts from the game Fallout labeled by emotional categories New Vegas presented in an article [Hämäläinen, 2022; 2]. Here, the texts are presented in 5 languages: English, Spanish, German, French, and Italian. During

the experiments, classification accuracy values of 54%-55% were obtained for these corpora using multilingual pre-trained BERT and XLMRoBERTa models.

A number of scientific articles also represent research on low-resource languages. The work [Allaberdiev, 2024; 3] presents the creation of a Uzbek-Kazakh corpus for machine translation. This corpus was created in several stages: at the first stage, corpora in Uzbek were collected from publicly available resources (books, websites), then the corpora were translated into Kazakh by a group of experts, and at the next stage, texts in two languages were combined into a parallel corpus. In work [Shymbayev, 2023; 3], a study of an extractive question-answer system using the BERT model is presented for the Kazakh language. Since the development of such a system requires tens of thousands of question-answer pairs, to solve this problem, a machine translation of the Stanford corpus was performed using Google Translate API. In the article [Tolegen, 2024; 4], an empirical study is presented that evaluates text generation models for the resource-poor and morphologically complex Kazakh language. This work trained a large transformer-based language model using a large text corpus of the Kazakh language from different areas. The second dataset contained question-answer problems in Kazakh and Russian languages.

Methodology

A Python web parser was developed to collect Kazakh legislative texts corpora. The first step was to collect links from Adilet (https://adilet.zan.kz/kaz/), the legislative website of the Republic of Kazakhstan. In order to implement this, we used the Content Downloader program. This program was designed to collect links and content from various websites. It is required to set the URL address of the website, the number of threads that determine the speed of data collection, and filtering parameters to gather links. The URL parameters were specified in the filtering window addresses, which must have indicated that a search was done for links to documents in the Kazakh language – *kaz/docs/*. Also excluded from the sample were links that did not contain documents, such as *search/*, *docs/search*, *index/docs*.

The window for collecting links and filtering in the Content Downloader program is presented in Fig. 1 and 2.

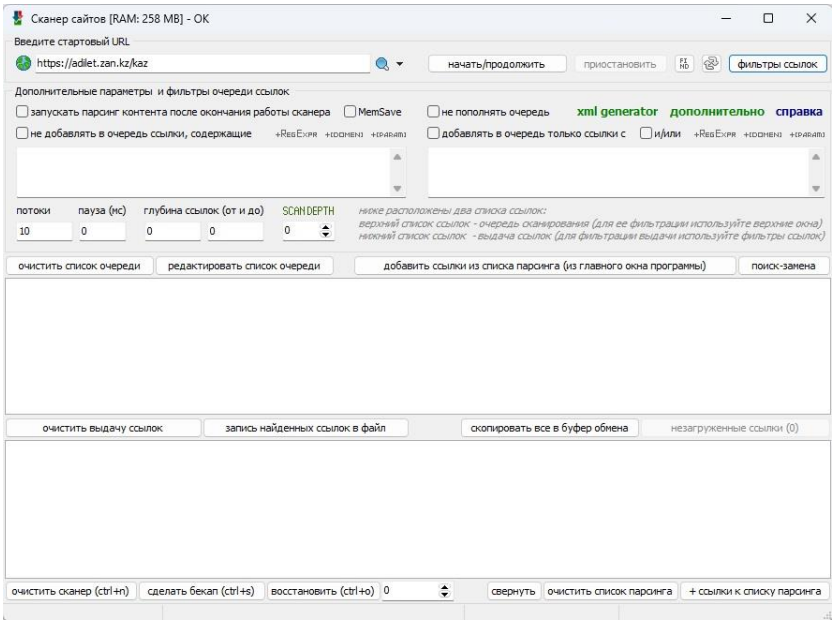


Figure 1 – The program windows of link collection

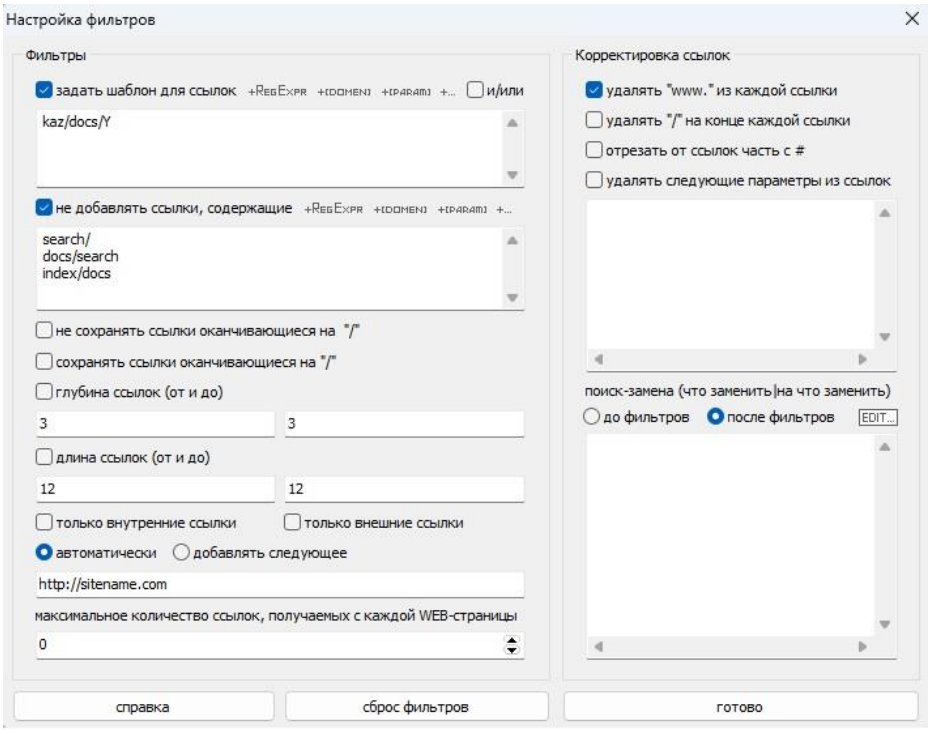


Figure 2 – The program windows of link filtering

Obtained using the Content Downloader program, links were saved in corresponding files. Next, the parser script adds the links extracted from the files to the list. The Selenium WebDriver was used to extract the content of web pages. This tool interacts with the browser at a lower level and uses browser-specific drivers that access it directly. This web driver receives information about the state of the web page and checks the presence of certain elements in it, their properties, and content. Thus, it is very effective for extracting data from web pages.

Results

The parser based on the Selenium WebDriver was utilized to extract the title, the date of the law article, and its text. This data, along with the URL address, was saved in JSON files. Then, the data was read from all JSON files, combined into one dataset, and saved in a CSV file. As a result, the dataset consisting of 9575 texts was obtained. It is shown in Fig. 3.

	header	date	text	source
0	Ветеринариялық бақылауға (қадағалауға) жататын...	2019 - 24 желтоқсан	2014 жылғы 29 мамырдағы Еуразиялық экономикалы...	https://adilet.zan.kz/kaz/docs/H19EK000237
1	Еуразиялық экономикалық одақтың кедендік аумағ...	2019 - 24 желтоқсан	Санитариялық, ветеринариялық-санитариялық және...	https://adilet.zan.kz/kaz/docs/H19EK000238
2	Дәрілік препараттың сапасы бойынша нормативтік...	2018 - 7 қыркүйек	2014 жылғы 29 мамырдағы Еуразиялық экономикалы...	https://adilet.zan.kz/kaz/docs/H18EK000151
3	Дәрілік заттарды сынақтан өткізудің аналитикал...	2018 - 17 шілде	2014 жылғы 29 мамырдағы Еуразиялық экономикалы...	https://adilet.zan.kz/kaz/docs/H18EK000113
4	Өсімдіктен алынатын дәрілік препараттардың сап...	2018 - 10 мамыр	Еуразиялық экономикалық комиссия Алқасы Еуразиялық...	https://adilet.zan.kz/kaz/docs/H18RK000006
...	...	...	...	...
9570	Қазақстан Республикасының кейбір заңнамалық ақ...	2017 - 3 шілде	РҚАО-ның ескертпесі!Қолданысқа енгізілу тәртібі...	https://adilet.zan.kz/kaz/docs/Z1700000086
9571	Қазақстан Республикасының кейбір заңнамалық ақ...	2021 - 3 қаңтардағы	1-бап. Қазақстан Республикасының мына заңнамал...	https://adilet.zan.kz/kaz/docs/Z2100000406
9572	Қазақстан Республикасының кейбір заңнамалық ақ...	2016 - 6 сәуір	1-бап. Қазақстан Республикасының мына заңнамал...	https://adilet.zan.kz/kaz/docs/Z1600000481
9573	Әділет органдары туралы	2002 - 18 наурыз	Ескерту. Орыс тіліндегі мәтінге өзгеріс енгізі...	https://adilet.zan.kz/kaz/docs/Z020000304_
9574	Өзбекстан Республикасының Еуразиялық экономика...	2006 - 4 шілде	РҚАО-ның ескертпесі!Келісім қолданысын тоқтатт...	https://adilet.zan.kz/kaz/docs/Z060000152_

9575 rows × 4 columns

Figure 3 – The dataset obtained with the use of the parser

Conclusion

This paper presents a study on creating text corpora for the low-resource Kazakh language, which is important for tasks in the field of NLP, such as machine translation, question-answer systems, text generation, etc. The main source of resources for quality texts in the Kazakh language was the legislative website Adilet, which is characterized by a high level of structure and grammatical correctness of the texts. An automated parsing script was used to obtain texts from this site, significantly simplifying and speeding up the process of collecting text data while maintaining high-quality information. The list of site links was generated with the use of the Content Downloader program, which returned a list of URLs for a given site. When setting up the program, the number of threads that determine the speed of data collection and filtering parameters was also set. The resulting links are added to the list and used in the parser program, where Selenium WebDriver is used to extract the content of web pages. During the experiment, a corpus was obtained that included 9575 legislative texts. In the future, the resulting corpus will be used to create a model of a question-answer system for the Kazakh language. It is also important to continue developing parsing methods to improve the accuracy and efficiency of data extraction, which will significantly contribute to the development of digital technologies in Kazakhstan.

References:

1. Mokhamed T., Harous S., Hussein N. *et al.* Comparative analysis of Deep Learning and Machine Learning algorithms for emoji prediction from Arabic text. *Social Network Analysis and Mining*, 14, 67, 2024. https://doi.org/10.1007/s13278-024-01217-w
2. Choi J., Lee B. Accelerating materials language processing with large language models. *Communication Materials*, 5, 13, 2024. https://doi.org/10.1038/s43246-024-00449-9
3. Jaesub Y., Jong-Seok L. Learning from class-imbalanced data using misclassification-focusing generative adversarial networks. *Expert Systems with Applications*, vol. 240, 2024. https://doi.org/10.1016/j.eswa.2023.122288
4. Karyukin V., Mutanov G., Mamykova Z. *et al.* On the development of an information system for monitoring user opinion and its role for the public. *Journal of Big Data*, 9, 110, 2022. https://doi.org/10.1186/s40537-022-00660-w
5. Guo S., Deng N., He Y. ISTIC's Neural Machine Translation Systems for CCMT' 2023. *Communications in Computer and Information Science*, vol. 1922, 2023. Springer, Singapore. https://doi.org/10.1007/978-981-99-7894-6_9
6. Xu L., Lu L., Liu M. *et al.* Nanjing Yunjin intelligent question-answering system based on knowledge graphs and retrieval augmented generation technology. *Heritage Science*, 12, 118, 2024. https://doi.org/10.1186/s40494-024-01231-3
7. Zhong Y., Goodfellow Sebastian D. Domain-specific language models pre-trained on construction management systems corpora. *Automation in Construction*, vol. 160, 105316, 2024. https://doi.org/10.1016/j.autcon.2024.105316
8. Dagdelen J., Dunn A., Lee S. *et al.* Structured information extraction from scientific text with large language models. *Nature Communications*, 15, 1418, 2024. https://doi.org/10.1038/s41467-024-45563-x
9. Abd El-Mageed A.A., Abohany A.A., Ali A.H. *et al.* An adaptive hybrid African vultures-aquila optimizer with Xgb-Tree algorithm for fake news detection. *Journal of Big Data*, 11, 41, 2024. https://doi.org/10.1186/s40537-024-00895-9

10. Modi A., Shah K., Shah S. *et al.* Sentiment Analysis of Twitter Feeds Using Flask Environment: A Superior Application of Data Analysis. *Annals of Data Science*, 11, 159–180, 2024. https://doi.org/10.1007/s40745-022-00445-1
11. Karyukin V, Rakhimova D, Karibayeva A, Turganbayeva A, Turarbek A. The neural machine translation models for the low-resource Kazakh–English language pair. PeerJ Computer Science 9: e1224, 2023. https://doi.org/10.7717/peerj-cs.1224
12. Wei Zh., Zhang Sh. A structured sentiment analysis dataset based on public comments from various domains. Data in Brief, vol. 53, 110232, 2024. https://doi.org/10.1016/j.dib.2024.110232
13. Afli H., Barrault L., Schwenk H. Building and using multimodal comparable corpora for machine translation, *Natural Language Engineering*, 22(4), pp. 603–625, 2016. https://doi.org/10.1017/S1351324916000152
14. Hämäläinen M., Alnajjar K. , Poibeau T. Video Games as a Corpus: Sentiment Analysis using Fallout New Vegas Dialog. In Proceedings of the 17th International Conference on the Foundations of Digital Games (FDG '22). Association for Computing Machinery, New York, NY, USA, Article 56, 1–4, 2022. https://doi.org/10.1145/3555858.3555930
15. Allaberdiev B., Matlatipov G., Kuriyozov E., Rakhmonov Z. Parallel texts dataset for Uzbek-Kazakh machine translation. Data in Brief, vol. 53, 110194, 2024. https://doi.org/10.1016/j.dib.2024.110194
16. Shymbayev M., Alimzhanov Y. Extractive Question Answering for Kazakh Language. *IEEE International Conference on Smart Information Systems and Technologies (SIST)*, Astana, Kazakhstan, pp. 401-405, 2023. https://doi.org/10.1109/SIST58284.2023.10223508
17. Tolegen G., Toleu A., Mussabayev R., Zhumazhanov B., Ziyatbekova G. Generative Pre-Trained Transformer for Kazakh Text Generation Tasks, *19th International Asian School-Seminar on Optimization Problems of Complex Systems (OPCS)*, Novosibirsk, Moscow, Russian Federation, pp. 144-118, 2023. https://doi.org/10.1109/OPCS59592.2023.10275765
18. Ismailov, A. S., & Abdurakhmonova, N. (2022). The development of Alisher stemmer for Uzbek Language. Science and Education, 3(4), 187-213.
19. Abdurakhmonova, N., & Tuliyev, U. (2018). Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education, 3, 68.
20. Abdurakhmonova, N., & Urdishev, K. (2019). Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL), 6(1-2019), 131-7.
21. Abduraxmonova, N. Z. (2018). Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref.

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ЯЗЫКОВОМ ПЕРЕВОДЕ

Назирова Э.Ш.
Ташкентский университет информационных технологий имени Мухаммада Ал-Хоразмий.

Усманова К.А.
Ташкентский университет информационных технологий имени Мухаммада Ал-Хоразмий.

Аннотация: Данная статья содержит в себе актуальность использования искусственного интеллекта при переводе и обучении иностранным языкам. Показана актуальность данной темы, преимущества применения ИИ. Также в статье кратко освещены работы учёных-лингвистов по данной теме. Почеркнуто именно на какие аспекты необходимо обратить внимание в момент их выражения на базе данных.

Особое внимание уделяется двум аспектам лексикографии, условно связанных друг с другом: научно-теоретическому и практическому.

Показано как использование искусственного интеллекта при машинном переводе открывает новые возможности для развития и совершенствования языковой коммуникации.

Ключевые слова: развитие, искусственный интеллект, факт, машинный перевод, лингвистика, аспект, алгоритм.

Abstract: This article contains the relevance of using artificial intelligence in translation and teaching foreign languages. The relevance of this topic and the advantages of using AI are shown. The article also briefly highlights the work of linguists on this topic. It is highlighted exactly what aspects need to be paid attention to at the time of their expression on the database.

Particular attention is paid to two aspects of lexicography, conditionally related to each other: scientific-theoretical and practical.

It is shown how the use of artificial intelligence in machine translation opens up new opportunities for the development and improvement of language communication.

Keywords: development, artificial intelligence, fact, machine translation, linguistics, aspect, algorithm.

Каждая эпоха сопровождается своими особенностями и историческими событиями. Если XX век у нас ассоциируется с полётом человека в космос и покорением Эвереста, то одним из неотъемлемых фактов, присущим нашему веку является открытие компьютера и развитие нанотехнологий, развитие интернета и искусственного интеллекта. Искусственный интеллект (ИИ) стал неотъемлемой частью многих сфер нашей жизни, включая языковой перевод. С развитием технологий и появлением новых алгоритмов машино-обучения, использование ИИ при машинном переводе стало все более распространенным и эффективным.

Одним из ключевых преимуществ использования искусственного интеллекта при машинном переводе является скорость и точность перевода. Благодаря мощным

алгоритмам и нейронным сетям, компьютеры могут обрабатывать большое количество текстов и выбирать наиболее подходящий вариант перевода за считанные секунды. Это позволяет сократить время на перевод текстов и повысить производительность работы переводчиков и педагогов в обучении иностранным языкам.

Еще одним важным преимуществом использования ИИ при машинном переводе является его способность к адаптации и улучшению качества перевода. Системы машинного перевода на базе ИИ могут обучаться на больших объемах текстов и корректировать свои алгоритмы в зависимости от обратной связи от пользователей. Это позволяет системам стать все более точными и эффективными с течением времени.

Кроме того, использование искусственного интеллекта при машинном переводе позволяет решать множество задач, которые для человека были бы труднопонимаемыми или слишком монотонными. Нейронные сети способны распознавать сленг, идиомы, а также адаптироваться к специфике языка и контексту текста, что делает перевод более точным и понятным.

Однако несмотря на все преимущества, использование искусственного интеллекта при машинном переводе имеет и некоторые недостатки. Например, системы машинного перевода могут допускать ошибки в интерпретации контекста или неверно переводить сложные фразы, требующие понимания культурных особенностей или идиоматики. Также важно помнить, что машины-переводчики могут не всегда передать точный смысл текста и не способны адаптироваться к новым языковым явлениям и изменениям. В качестве примера можем привести исследования учёных лингвистов. Иследования по словарной типологии проводили такие учёные, как Л. В. Щерба, В. В. Дубичинский, Л. В. Варпахович, П. Н. Денисов. В XXI веке исследования по созданию фразеологических словарей для школьников в учебной лексикографии оказались в центре русской лексикографии [2]. В этом смысле проблема учебной лексики вызвала дискуссию во мнениях, высказанных специалистами в различных областях. В частности, Ю. И. Рогалева в качестве важнейшего принципа учебной лексики выделяет постепенность, адаптированную к возрасту и уровню знания изучающих язык, в подаче статей.

В. М. Морковкин сообщает о «существовании теоретического и практического лингвометодического поля лексики в учебных целях». В связи с этим ученый делит учебную лексику на следующие группы [3]: 1) теория и практика создания учебной лексики; 2) теория и практика создания лексического минимума; 3) теория и практика учебной лексикостатики; 4) теория и практика лексикографического учебного пособия по лексике; 5) описание лексики словарей в учебном процессе и теория и практика их предъявления. В 20 в. научные исследования лексики английского языка как иностранного изучались Г. Палмером, А. Хорнби, К. Огденом, М. Уэстом как особый объект исследования в области педагогики и методологии.

Можно увидеть два аспекта лексикографии, условно связанных друг с другом: научно-теоретический и практический. LP Ступин выделяет лексикографию как область в трех аспектах: 1) особую отрасль языкознания, изучающую принципы создания различных словарей; 2) практическая деятельность по работе со лексикой; 3) словарь конкретного языка [20]. Учебный словарь – это словарь особого типа, приспособленный для изучения и преподавания определенного языка. Его отличие от других словарей в том, что лексическая единица, используемая при обучении пользователей, предназначена для образовательного

процесса. Материал, отобранный для содержания таких словарей, направлен с дидактической целью. Словари служат инструментом обучения в обучении. Его образовательный характер состоит в том, что содержание словаря, расположение статей, формы подачи языковой информации, способ изложения понятого смысла или содержания, размер и т. д. направлены на обучение или изучение языка. [1]. Одной из особенностей этих словарей является минимизация всех элементов языка в пределах определенного уровня. Единицы, включенные в словарь, основаны на принципе лексического минимума. Также социолингвистический характер слова и объяснение содержания лексики, отобранной для одноязычных словарей, показывают, что оно ориентировано на образовательную цель.

ИИ на сегодняшний день переводя данные словари, тексты, задания позволяет содержать фразеологические, орфографические аспекты, словосочетания, антонимы, синонимы и искусственные слова. Если в двуязычных учебных словарях наряду с указанными выше полями предусмотрены также перевод и объяснение слов, то содержание словаря при машинном переводе может в определенной степени различаться в зависимости от того, предназначены ли учебные словари для использования в образовании языка как родного, второго или иностранного. Словари также отличаются друг от друга тем, представлены ли статьи в алфавитном порядке, плоско или в виде сетки. По мнению аудитории пользователей, он адаптирован к общим потребностям и интересам дошкольников, школьников, лиц с высшим образованием или изучающих язык. Подводя итог вышеизложенным различиям в создании учебных словарей, следует отметить, что словари данного типа должны служить средством общения изучающих язык. Также содержать помимо перевода с одного языка на другой предоставлять наиболее точную информацию по синонимичному, омонимичному и другим рядам, а при обучении определённой возрастной категории предоставлять, если нужно визуальный картиночное представление обучающемуся.

Развитие информационных возможностей, интенсивное обучение и прогресс, существующие сегодня в мире, требуют от педагогов и переводчиков следования в ногу со временем. В этом им помогает ИИ. Использование искусственного интеллекта при машинном переводе открывает новые возможности для развития и совершенствования языковой коммуникации. Современные технологии и алгоритмы продолжат совершенствоваться, делая машинный перевод все более точным и доступным для широкой аудитории. Важно помнить, что современные технологии могут быть мощным инструментом помощи переводчикам, но не заменять их креативный и интеллектуальный вклад в процесс перевода.

Использованная литература:

1. Ступин Л.П. Лексикография английского языка М.: Высшая школа, 1985. –С. 119.
2. ТрифоновА.С. ОСНОВНЫЕ ПАРАМЕТРЫ ТИПОЛОГИЗАЦИИ УЧЕБНЫХ СЛОВАРЕЙ /Теоретическая и прикладная лингвистика, 201
3. Расулова У.Р. Особенности разработки узбекско-английского двуязычного програмного комплекса на основе моделирования грамматических категорий глаголов и морфологического анализа при машинном переводе. Учительский журнал.ISSN 2782-4200 Всероссийское педагогическое издание. Министерство науки и высшего образования

4. 4.Абдурахмонова, Н., & Абдувахобов, Г. (2021). O ‘QUV LUG ‘ATINI TUZISHNING NAZARIY METODOLOGIK ASOSLARI. МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА, 4(6).
5. 5. Abdurakhmonova, N., Alisher, I., & Sayfulleyeva, R. (2022, September). MorphUz: Morphological Analyzer for the Uzbek Language. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 61-66). IEEE.
6. 6.Abdurakhmonova, N., Alisher, I., & Sayfulleyeva, R. (2022, September). MorphUz: Morphological Analyzer for the Uzbek Language. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 61-66). IEEE.

SEMANTIK TEGLASHNING FE'L SO'Z TURKUMIGA OID GURUHLAR TASNIFI

Ismadiyorov Zafar
Mirzo Ulug'bek nomidagi O'zbekiston
Milliy universiteti magistranti.
E-mail: zafarismadiyorov1@gmail.com

Annotatsiya: Mazkur maqolada matnni semantik teglash jarayoni, usullari, uning amaliy ahamiyati,o'ziga xos xususiyatlari va fe'l so'z turkumi doirasidagi fe'llarni o'zgarishlarining sabablari haqida bayon qilinadi.

Kalit so'zlar: teglash, ehtimollik nazariyasi, metriki, CLAWS, lemmatayzer, terminologiya, sintaktik tahlil, morfologik stemmerlar.

Abstract: This article describes the process of semantic tagging of the text, its methods, its practical importance, specific features, and the reasons for changing verbs within the verb phrase.

Keywords: tagging, probability theory, metric, CLAWS, lemmatizer, terminology, syntactic analysis, morphological stemmers.

Аннотация: В данной статье описан процесс семантической разметки текста, его методы, практическое значение, особенности и причины изменения глаголов внутри глагольного словосочетания.

Ключевые слова: тегирование, теория вероятностей, метрика, CLAWS, лемматизатор, терминология, синтаксический анализ, морфологические стеммеры.

Semantik teglash - bu so'zlar, iboralarga yoki ma'nolarga ta'sir qiluvchi o'zgaruvchilarni va ularning munosabatlari, o'zgarishlari va xususiyatlari to'g'risida o'rganish, tushuntirish va ta'riflashning ilmiy yo'nalishi hisoblanadi. Bu yo'nalish so'zlar, gaplarning va butun qatori ma'lumotlar turlari orasidagi ma'naviy aloqalarni o'rganadi. Matnlarni teglash uch usulda bajariladi: 1) *qoidaga asoslangan teglash* - bunda teglash tizimi leksikonga asoslanadi, bioq o'zi uchun noma'lum bo'lgan konstruksiyaga tegishli teglarni aniqlay olmasligi mumkin; 2) *ehtimolga asoslangan teglash* - ehtimollik nazariyasi metrikasi yordamida amalga oshiriladi. Agar so'z noma'lum grammatik sinfga tegishli bo'lsa , u qo'shimcha ma'lumotlar bazasidan qidiriladi ; 3) *gibrid asosli teglash* (CLAWS texnologiyasi yordamida) kiradi. Bunda ham ma'lumotlar bazasidan, ham ehtimollik vaziyatlariga qarab teglar tanlanadi [N.Abduraxmonova. 202, 75].

Hozirda bu kabi teggerlarning namunalarini internet tizimida qo'llash mumkin. Penn sintaktik korpuslaring natijasi sifatida sintaktik tahlil uchun Fidditch teggeri, tobelanish analizi uchun uzviy grammatika tizimi, Britaniya milliy korpusi uchun CLAWS morfologik teggeri, ko'plab korpuslarga amal qilinadigan USAS semantik teggeri va sistemalar tarkibida yaratilgan lemmatayzer yoki morfologik stemmerlar shular jumlasidandir [T. McEnery, A. Hardie. 2012:29].

Matnlarni semantik teglashning ahamiyati juda katta, chunki bu jarayon, matnlarning ma'nolarini aniqlash va tushuntirishga yordam beradi. Semantik teglash matnning tahlilini qilish, undagi asosiy mavzuni aniqlash, muhim ma'nolarni ajratib chiqish va ularning mantiqiy aloqalarini tushuntirishda juda foydali bo'ladi [Abdurakhmonova, N. 2019,2021,2022,2023].

Misol uchun, so‘zlar, gaplar yoki to‘plam ma’lumotlari o‘zining konteksti ichida qanday ma’nolarni anglatadi va ularga qanday munosabatda qo‘llaniladi, semantik teglash orqali aniqlanishi mumkin. Bu matnlarni yaxshi tushunish, ulardan foydalanish va ma’lumotlarni manbalar orqali muvofiqlashda yordam beradi. Semantik teglashning amaliyoti odatda quyidagi bosqichlarga bo‘linadi:

-Dastlab, matn yoki ma’lumotlar to‘planadi va ularni ma’nolari bo‘yicha tahlil qilinib, matnning qanday ma’lumotlar o‘z ichiga olganligini tushuntirish uchun kerakli ma’lumotlarni topish, to‘plash va tahlil qilishni o‘z ichiga oladi.

-Matnning asosiy mavzusini aniqlash va matnni qanday kontekstda yoki muhitda qo‘llab-quvvatlashi matnning ma’nosi va uning ahamiyatini tushuntirishga yordam beradi.

-Har bir so‘z, ifoda yoki qism ma’nolari o‘rganiladi va tahlil qilinib, har bir so‘zning o‘ziga xos ma’nolari va ularga qo‘llanilishi kerak bo‘lgan o‘rinlari keltiriladi.

-Keyinchalik, matnlarning ichidagi so‘zlar va ifodalar o‘rtasidagi munosabatlarni aniqlashga alohida e’tibor qaratilib, so‘zlar, gaplar va ifodalarning bir-biriga qanday ta’sir qilishini va ulardagi aloqalarni tushuntirishda yordam beradi.

-Oxirgi bosqichda asosiy tahlil va ta’riflash jarayonlarini o‘z ichiga olib, ma’lumotlar o‘zgaruvchilarini, kontekstlarni va munosabatlarni tahlil qilishni va matnlarni tushuntirishni o‘z ichiga oladi.

Aslida, semantik teglash barcha turkumlarda amalga oshirilishi mumkin, chunki uni matnning ma’nolari, tushunchalari va munosabatlari haqida tahlil qilish uchun qo‘llash mumkin. So‘z turkumlarini teglashning dastlabki shakli Lankaster universitetining tadqiqot markazida (UCREL) o‘tgan asrning 80- yillarida Jeorey Lich rahbarligida ishlab chiqilgan. Ingliz tili uchun CLAWS (the Constituent Likelihood Automatic Word- tagging System) avtomat tegi 1980-yilda paydo bo‘lgan. O‘sha vaqtdagi taggerning imkoniyatlari muayyan darajada o‘zgartirilgan bo‘lsa-da, hanuz uch bosqichini saqlab qolgan: tahriroldi, avtomatik teglash, natijani tahrirlash. CLAWS yillar davomida turli teglardan foydalangan bo‘lsa - da, hozirda uning 132 asosiy tegi mavjud[Abdurakhmonova, N. 2019,2021,2022,2023]. Braun korpusidagi aksariyat teglar yuqorida nomi qayd qilingan korpus teglari bilan bir xil. Ushbu teglash tizimi matnni teglash uchun beshta alohida bosqichni amalga oshiradi [R. Garside. 33]: 1) tahrir oldi fazasi - ushbu bosqich uchun yarim avtomat , yarmi qo‘l mehnati orqali matnni teglash tizimiga moslashtiriladi; 2) tegni belgilash - matnga kiritilgan har bir so‘z bir yoki bir nechta tegga ega bo‘ladi. Ushbu bosqich so‘zining matndagi holatini emas, balki ixtiyoriy matnda qo'llanilshi mumkin bo‘lgan holat uchun tegni aniqlaydi; 3) idiomalarni teglash bosqichda muayyan tegga mos keluvchi kontekstdagi so‘zlarning teglari belgilanadi; 4) teglarni aniqlashtirish mazkur bosqichda kontekstda uchragan so‘z birdan ortiq tegga to‘g‘ri kelib qolganda eng munosib varianti aniqlanadi; 5) tahrirdan so‘ng kompyuterda yo‘l qo‘yilgan ortiqcha ma’lumotlarlarga ishlov beriladi.

Shuningdek, o‘zbek tili korpuslari uchun semantik teglash tizimini ishlab chiqish uchun o‘zbek tilining izohli lug‘atidagi fe’l so‘z turkumiga kiruvchi ma’lum guruhga kiradigan fe’llar uchun atributlar tizimini ishlab chiqish, ushbu teg guruhlarining ichiga shu ma’no-mazmunga ega bo‘lgan so‘zlarni joylashtirib chiqish va natijada dasturiy ta’minot yaratish dolzarb masala hisobladi[Abdurakhmonova, N. 2019,2021,2022,2023].

Jumladan, O‘zbek tilining izohli lig‘atida [Ўзбек тилининг изоҳли луғати 5 жилдлик, 2000-2006: 23] yur fe’lini quyidagicha ma’nolari borligini keltiradi: 1.Qadam tashlab olg‘a bosmoq, qadamlab jilmoq. 2. Ma’lum yo‘nalishda harakat qilmoq, bir joydan boshqa joyga jilmoq,

bormoq yoki ketmoq (asosan transport vositasi haqida). 3. Umuman, bir joydan jilmoq, siljimoq (suzmoq, sudralmoq, o‘rmalamoq va shu kabilar). 4. Kezish, aylanish, sayohat va shu kabilarda bo‘lmoq. 5. Hayot kechirmoq, yashamoq. 6. Ish-faoliyat va shu kabilarda birga bo‘lmoq. 7. Qatnamoq. 8. Davom etmoq (o‘zi birikib kelgan so‘z bildirgan holatda). 9. Shaxmat, qarta, domino kabi o‘yinlarda o‘z navbatida, navbati kelganda harakat qilmoq, o‘ynamoq; yurish qilmoq. 10. Harakatga kelmoq, ishlab turmoq, ishlamoq (mexanizmlar haqida). 11. Suygan yor bilan vaqt o‘tkazmoq, yaqin munosabatda bo‘lmoq. 12. Rozi bo‘lmoq, ko‘nmoq, unamoq. 13. Yoyilib ketmoq, yoyilmoq; tarqalmoq (gap-so‘z, xabar haqida).

Xulosa o‘rnida, fe’l so‘z turkumida semantik teglash jarayoni fe’llarning ma’nolari, tushunchalari va ularga qo’llanilish usullari haqida tahlil qilishni o‘z ichiga oladi. Bu fe’llarning turkumning barcha jabhalari bo‘yicha tahlil qilinishini va ulardagi munosabatlarni tushuntirishni o‘z ichiga oladi. Semantik teglash fe’llar turkumida quyidagi bosqichlarga asoslanadi:

-turkumning ichidagi har bir fe’lning o‘ziga xos asosiy ma’nosini aniqlash;

-fe’llarining turkum ichidagi ma’nolari va ularga qo‘llanilishi tahlil qilish;

-bu fe’llarning qanday kontekstda yoki munosabatda qo‘llanilishi va ulardagi farqlar tushuntiriladi. Ba’zi fe’llarning ma’nosida o’zgaruvchiliklar mavjud bo‘lishi mumkin. Semantik teglashda bu o‘zgaruvchiliklarni aniqlash va ulardagi farqlarni tahlil qilish juda muhimdir. Fe’llar turkumidagi kontekst va uslub tahlili ham semantik teglashning muhim qismlaridan biridir. Fe’llar o‘zining turkum ichidagi joyi va ularga qo‘llanish usullari o‘zaro bog‘liq bo’lishi mumkin.

Foydalanilgan adabiyotlar:

1. N. Abduraxmonova. Mashina tarjimasining lingvistik ta’minoti(monografiya). - Toshkent:Muharrir, 2018. 165b.
2. N. Abduraxmonova. Kompyuter lingvistikasi (darslik) - Toshkent, 2021. 400b.
3. T. McEnery, A. Hardie. Corpus linguistics: method, theory and practice, Cambridge. 2012, -P, 29
4. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 1790-1793). IEEE
5. Абдурахмонова, Н., & Бойсариева, С. (2023). Tabiiy tilni qayta ishlashda (nlp) okkazionalizmlarning morfem tahlili. Международный журнал искусство слова, 6(3)
6. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240)
7. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
8. R. Garside. The CLAWS Work-tagging System. In: R. Garside, G. Leech and G. Sampson(eds), The Computational Analysis of English: A Corpus-based Approach. London: Longman, P. 3.
9. R. Delmonte. Computational Linguistics Text Processing: Lexicon, grammar, Parsing and Anaphora Resolution, Nova Science Publishers, Inc. New York, 2008.

SINTAKTIK SINONIMIYA VA UNI ANIQLASH XUSUSIDA

Uzoqova Mohiyaxon Tuyg‘un qizi
Kompyuter lingvistikasi mutaxassisligi magistranti
Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti.

Annotatsiya: Sinonimlar turli belgilariga ko‘ra, shu jumladan, mazmuni va shakliga ko‘ra tasniflanishi mumkin. Shakliy sinonimlar qatoriga sintaktik sinonimlarni ham kiritish mumkin. Ushbu maqolada sintaktik sinonimiya hodisasiga va uning shakllanish usullariga to‘xtab o‘tilgan. Shuningdek, sintaktik sinonimiyani faqat algoritmik yo‘l bilan avtomatik aniqlash murakkabligi bois, jarayonni soddalashtiruvchi va muallif tomonidan samaraliroq deb baholanayotgan til modellaridan biri – T5 modeliga ham qisqacha to‘xtalib o‘tilgan.

Kalit so‘zlar: Sinonimlar tasnifi, sintaktik sinonimiya, til modellari, T5 modeli, parafrazalash.

Abstract: This paper explores the complex realm of synonymy, focusing on its various classifications based on meaning and form, including syntactic synonymy. It delves into the intricate methods of shaping syntactic synonymy, acknowledging the challenges in automatic identification using algorithms alone. The paper proposes the use of advanced AI models, specifically the T5 language model, as a more effective solution for this linguistic task. The study aims to enhance our understanding of synonymy and contribute to the development of more sophisticated language processing tools.

Keywords: classification of synonyms, syntactic synonomy, language models, T5 model, parapharsing.

Аннотация: В данной статье исследуется сложный мир синонимии, уделяя особое внимание ее различным классификациям, основанным на значении и форме, включая синтаксическую синонимию. Он углубляется в сложные методы формирования синтаксической синонимии, признавая проблемы автоматической идентификации с использованием одних только алгоритмов. В статье предлагается использовать передовые модели искусственного интеллекта, в частности языковую модель T5, как более эффективное решение этой лингвистической задачи. Исследование направлено на улучшение нашего понимания синонимии и содействие разработке более сложных инструментов языковой обработки.

Ключевые слова: классификация синонимов, синтаксический синоним, модели языка, модель T5, перефразирование.

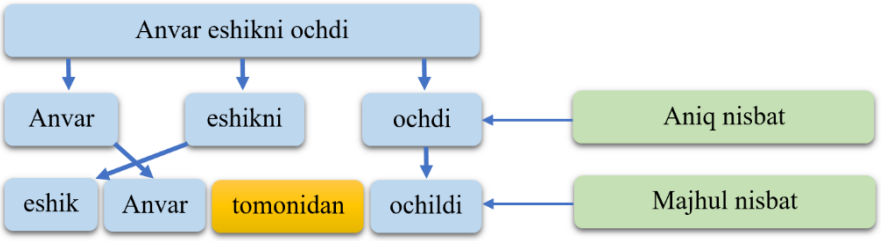
Sinonimlarni tasniflashda turli yondashuvlar mavjud. Umumiy holatda ularni mazmuniy va shakliy sinonimlar kabi yirik guruhlarga bo‘lish mumkin. Mavjud tasniflarni umumlashtirgan holda biz shakliga ko‘ra tasniflanuvchi sinonimlar sifatida leksik, morfologik, paremeologik va sintaktik sinonimlar taklif qilamiz. Sintaktik sinonimlarning mavjudligi haqida o‘zbek tilshunosligida Pinxasov D. [Пинхасов, 1969; 16] va Polvonovaning B. [Polvonova, 2021; 891] ham qayd etib o‘tgan.

Sintaktik sinonimlar – bu gapning mazmuni saqlangan, ammo sintaktik o‘zgarish kuzatilgani sababli gapning shakli o‘zgargan sinonimlardir. Sintaktik sinonimlikni ba’zi olimlar *sintaktik variatsiyalar* deb ham atashadi [Rashidli, 2022; 74].

Sintaktik sinonimlikni hosil qilishning bir qancha usullari mavjud:

1. Gapning nisbatini o‘zgartirish orqali.

Gapning nisbatini o‘zgartirish orqali uning mazmunini saqlagan holda umumiy shaklini o‘zgartirish mumkin. Bunday holatda sintaktik sinonimlik ko‘pincha aniq nisbat – majhul nisbat o‘rtasida kuzatiladi (Qarang: 1 -rasm).



1-rasm. Gapning nisbatini o‘zgartirish orqali sintaktik sinonimlikni hosil qilish

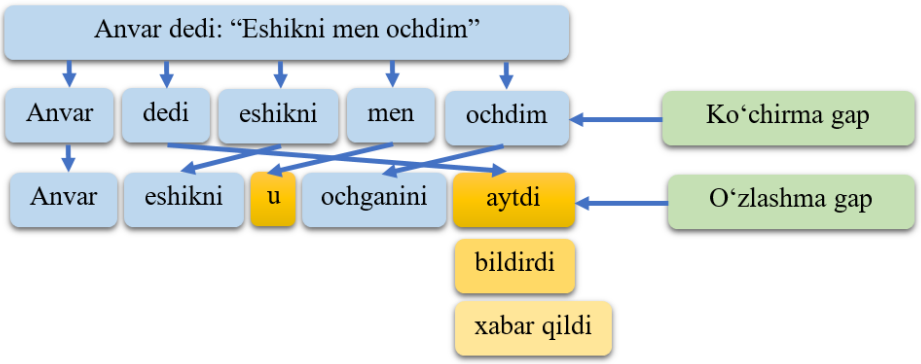
2. Ko‘chirma gapni o‘zlashtirma gapga o‘zgartirish orqali.

Ko‘chirma gapda o‘zga gapining asosiy mazmuni aks etadi, muallif gapida esa ko‘chirma gap kimga tegishli ekanligi, unga xos xususiyat, qanday maqsad va tarzda aytilgani va boshqa qo‘shimcha tavsifiy ma’lumotlar beriladi. Sintaktik sinonimlikni ayan ko‘chirma gapni o‘zlashtirma gapga aylantirish yoki aksini qilish orqali ham hosil qilish mumkin. Bunda o‘zgalar gapining mazmuni saqlanib, grammatik shakli yoki gap qurilishi o‘zgartirib yuboriladi (Qarang: 2 -rasm):

– Ko‘chirma gapni o‘zlashtirma gapga aylantirishda ko‘chirma gapdagi faol qo‘llanuvchi demoq fe’li aytmoq, gapirmoq, ta’kidlamoq fe’llariga almashtiriladi;

– Ko‘chirma gap o‘zlastirma gapga aylantirilganda, ba’zan uning uslubiy xususiyatlari ham o‘zgaradi. Ayniqsa, so‘zlovchining nutq jarayoniga turli munosabatini ifodalovchi ayrim kirish so‘zlar, undovlar; *axir, xo‘sh, ha, yo‘q, xo‘p, mayli* kabi gap bilan bog‘lanmagan elementlar, so‘zgaplar ko‘pincha tushib qoladi.

– Undov gap, atov gap, ritorik so‘roq gaplar o‘zlashtirma gapgdan tushirib qoldiriladi.



2-rasm. Gapning nisbatini o‘zgartirish orqali sintaktik sinonimlikni hosil qilish

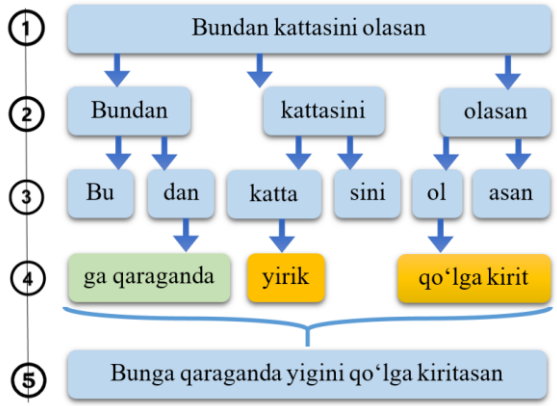
Biroq ko‘chirma gapni o‘zlashtirma gapga aylantirish orqali sintaktik sinonimlikni hosil qilish, bizningcha, sintaktik sinonimizatsiyadagi eng qiyin jarayonlardandir. Chunki ko‘chirma gap qanchalik murakkab tuzilgan bo‘lsa, uni o‘zlashtirma gapga aylantirish shunchalik qiyinlashadi. Bu murakkab sintaktik parsing qilish jarayonini talab qiladi. Bizning ilmiy tadqiqotimiz doirasida matnni sintaktik analiz qilish vazifasi qo‘yilmagani bois bunday sinonimlashni o‘zbekcha sintaktik analizator ishlab chiqilgach unga integratsiya qilish orqali amalga oshirishni istiqbolli reja sifatida belgilab oldik.

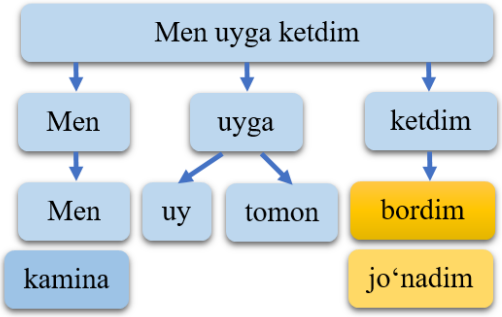
3. So‘z yuklamalarning shakl yasovchi qo‘shimchalar bilan o‘zgarishi.

Berilgan gap so‘zlar tarkibidagi ayrim shakl yasovchi qo‘shimchalarning “qo‘shimcha + so‘z yuklama”larga almashishi (Qarang: 3-rasm) yoki so‘z yuklamalarga (Qarang: 4-rasm) ham shakliy o‘zgarishga dalolat qiladi.

4. Tasviriy ifoda yoki izohlovchilar yordamida gapni o‘zgartirish orqali.

Tasviriy idoda yoki izohlovchilar ham o‘z doirasida (tasviriy ifoda – tasviriy ifoda; izohlovchi – izohlovchi), ham o‘zaro perependikulyar (tasviriy ifoda – izohlovchi; izohlovchi – tasviriy ifoda), ham leksema bilan almashib kelishi mumkin (Qarang: 5-rasm). Bunda gapning sezilarli qismi shaklan o‘zgarishi hisobiga gap qurilishi ham o‘zgarishi mumkin.





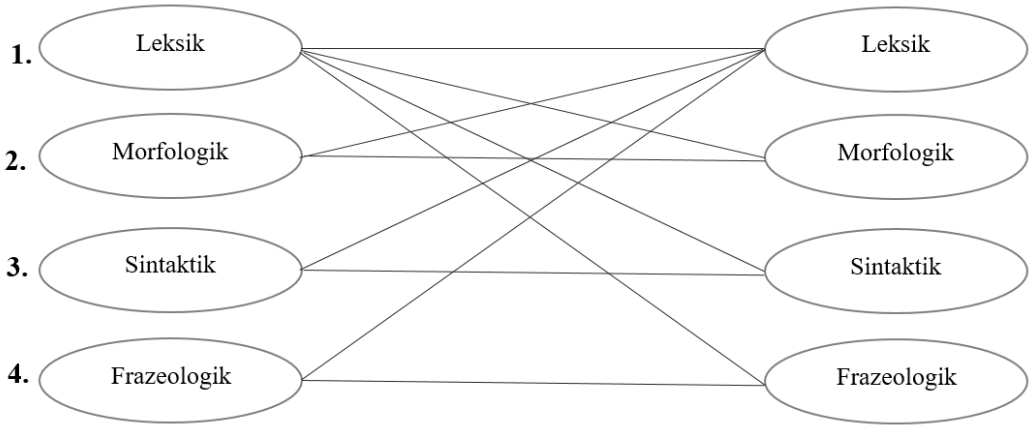
3-rasm. Gapning nisbatini o‘zgartirish orqali sintaktik sinonimlikni hosil qilish

4-rasm. Gapning nisbatini o‘zgartirish orqali sintaktik sinonimlikni hosil qilish

Sintaktik sinonimlar, o‘z navbatida, matn ichida boshqa sinonimik birliklar bilan kesishishi mumkin.

Sintaktik sinonimlarning boshqa turdagi sinonimlar bilan kesishishi:

1. Sintaktik sinonim – leksik sinonim: Toshkent – O‘zbekistonning poytaxti.
2. Sintaktik sinonim – sintaktik sinonim: Aziz eshikni ochdi – Eshik Aziz tomonidan ochildi.
3. Sintaktik sinonim – sintaktik sinonim: “Darsga kim javob beradi?” – dedi ustoz – Ustoz darsga kim javob berishini so‘radi.
4. Sintaktik sinonim – sintaktik sinonim: *O‘zbekistonning poytaxti*ga xush kelibsiz – *Yurtimizning yuragi*ga xush kelibsiz.



5-rasm. Sinov uchun berilgan gap bo‘yicha asosida modelning kichik dataset bilan mashq qildirilgan dastlabki versiyasidan olingan eng yuqori beshta natija

Sintaktik sinonimlarning bunday turli sinonimlar bilan ham, o‘z turida ham matnda sinonimlasha olishi algortmik yo‘l bilan avtomatik tarzda sinatktik sinonimizatsiyani amalga oshirishni murakkablashtiradi. Buning uchun yoki kuchli lingvistik va texnik tizim asosida ishlovchi morfoanalizator va sintaktik analizatorlarni sinonimayzer dasturiga integratsiya qilish lozim yoki matnni parafaraza qilishga yo‘naltirilgan sun’iy intellekt modellariga murojaat qilish lozim.

Ana shunday modellardan biri sifatida biz keyinchalik tadqiqotimizga joriy qilishni ko‘zlayotgan T5 modelini taklif qilmoqchimiz [Huggingface, 4]. T5 yoki Text-to-Text Transfer Transformer universal til modeli NLPning juda ko‘p masalalarini hal qilishda: mashina tarjimasida [Huggingface, 5], matnni klassifikatsiya qilishda [Huggingface, 6], NER (named entity recognition) aniqlashda [Huggingface, 7], matnni avtomatik generatsiya qilishda [Huggingface, 8], matnni xulosalashda [Huggingface, 9], o‘rindosh taklif qilishda (Fill-Mask) [Huggingface, 10], savol-javob tizimlarini qurishda [Huggingface, 11], shuningdek, matnni parafraza qilishda foydalaniladi. Demak, bu modeldan gapni avtomatik parafrazalashda ham foydalanish mumkin.

Buning uchun quyidagi ketma-ketlikda ish bajarish lozim:

1. Kerakli datani mos formatda to‘plash.
2. T5 modelini sozlash yoki avvaldan mashq qildirilgan hamda sozlangan T5 modelini topish.
3. Kiruvchi va chiquvchi ma’lumotlarni taqdim etish.
4. Natijani sifat nazoratidan o‘tkazish va zarur bo‘lsa, modelni mukammalashtirish.

Modelni mashq qildirish uchun o‘zbek tilidagi kerakli formatdagi ma’lumotlar korpusini to‘plash quyidagicha ko‘rinishda bo‘lishi lozim:

1. **Data to‘plash boqichi.** Buning uchun qulay kengaytma bilan saqlangan (masalan, .xsl) “Almashuvchi” va “almashtiriluvchi” qatorlar jufligini hosil qilish kerak (Qarang: 1- jadval).

Almashuvchi gap (“Input”)	Almashtiriluvchi gap (“Target”)
Odatda o‘quvchilarimga shu kundaligimni namuna qilib ko‘rsataman.	Ko‘p hollarda bolalarga shu kundalik namuna qilib ko‘rsatilar edi.
Kundalikni qo‘lma-qo‘l olib tomosha qila boshlashdi.	Kundalikni qo‘llaridan qo‘ymay tomosha qilishdi.

1-jadval. Sinov uchun berilgan gap bo‘yicha asosida modelning kichik dataset bilan mashq qildirilgan dastlabki versiyasidan olingan eng yuqori beshta natija

2. To‘plangan datani tayyorlash bosqichi. Bu bosqichda tayyorlangan juftliklarni tozalash, tokenizatsiya qilish va model tushunishi mumkin bo‘lgan formatga solish lozim.

3. Tayyor datani qismlarga bo‘lish bosqichi. Bu bosqichda tayyor datani 3 qismga: mashq qilish uchun, sozlamarni to‘g‘rilash uchun validatsiyaga va testlash uchun bo‘lish lozim.

4. Tayyor datani raqamli ko‘rinishga o‘tkazish (enkodlash) bosqichi.

Shunday qilib, tayyor data asosida modelni mashq qildirish orqali parafrazalash jarayonini yengillashtirish, mukammalashtirish mumkin. Chunki sintaktik sinonimiyaning har bir turi uchun alohida algoritm ishlab chiqish va uni integratsiya qilish katta resurs talab qilish bilan birga, sinonimizatsiya jarayonini sekinlashtirishi mumkin. Parafrazalash modellaridan foydalanish orqali esa bu muammo tezroq, samaraliroq va sifatliroq yechilishi mumkin degan deb hisoblaymiz.

Foydalanilgan adabiyotlar:

1. Пинхасов Д. Ҳозирги ўзбек адабий тили. Лексикология ва фразеология. –Тошкент: Ўқитувчи, 1969. –Б. 16.
2. Polvonova B. O‘zbek tilida sinonimiya talqini. Oriental Renaissance: Innovative, educational, natural and social sciences. VOLUME 1 | ISSUE 11 ISSN 2181-1784, 2021. – B. 891.
3. Vafa Ahmed gizi Rashidli. Syntactic synonymy, syntactic variability, syntactic transformation, syntactic equivalence // “Colloquium-journal” / Philological sciences / 17 (140), 2022, pp. 74-77. http://doi.org/10.24412/2520-6990-2022-17140-73-77
4. https://huggingface.co/docs/transformers/model_doc/t5
5. https://huggingface.co/docs/transformers/tasks/translation
6. https://colab.research.google.com/github/patil-suraj/exploring-T5/blob/master/t5_fine_tuning.ipynb
7. https://colab.research.google.com/drive/1obr78FY_cBmWY5ODViCmzdY6O1KB65Vc?usp=sharing
8. https://colab.research.google.com/github/NielsRogge/Transformers-Tutorials/blob/master/T5/Fine_tune_CodeT5_for_generating_docstrings_from_Ruby_code.ipynb
9. https://huggingface.co/docs/transformers/tasks/summarization
10. https://github.com/huggingface/transformers/tree/main/examples/flax/language-modeling#t5-like-span-masked-language-modeling
11. https://colab.research.google.com/github/snapthat/TF-T5-text-to-text/blob/master/snapthatT5/notebooks/TF-T5-Datasets%20Training.ipynb

NATURAL LANGUAGE PROCESSING (NLP) APPLICATIONS IN LANGUAGE TEACHING

Otajonova Shakhnoza
Senior teacher of National University of Uzbekistan.

Inogamova Nargiza
Senior teacher of National University of Uzbekistan.

Annotation: This article explores the applications of Natural Language Processing (NLP) in language teaching, examining its role in automated language assessment, intelligent tutoring systems, language learning applications, and text-to-speech/speech recognition systems. Through in-depth analysis, it highlights NLP's benefits such as enhanced personalization, immediate feedback, and adaptability, addressing challenges in language teaching like scalability and engagement. The integration of NLP with educational methods fosters dynamic learning environments, nurturing linguistic proficiency and cultural understanding. Future strides in NLP, coupled with virtual reality integration, promise transformative shifts in language education, offering immersive, interactive, and personalized learning experiences. However, ethical considerations such as bias mitigation and data privacy require attention. Collaborative efforts among educators, policymakers, and developers are crucial for responsible and inclusive NLP deployment in language teaching, empowering learners with global communication skills and cultural fluency in an interconnected world.

Keywords: Natural Language Processing (NLP), Intelligent Tutoring Systems (ITS), language learning applications, Text-to-Speech (TTS), Speech Recognition Systems (SRS), Virtual Reality (VR), Augmented Reality (AR), digital literacy.

Аннотация: В данной статье исследуются применения обработки естественного языка (NLP) в преподавании языка, рассматривается ее роль в автоматизированной оценке языка, интеллектуальных системах обучения, приложениях для изучения языка и системах распознавания текста и речи. Благодаря углубленному анализу в статье подчеркиваются преимущества NLP, такие как улучшенная персонализация, моментальная обратная связь и адаптируемость, что позволяет решить проблемы преподавания языка, такие как масштабируемость и вовлеченность. Интеграция NLP с образовательными методами способствует созданию динамичной среды обучения, развивая языковую грамотность и понимание культурных особенностей. Будущие шаги в NLP, вместе с интеграцией с виртуальной реальностью, обещают революционные изменения в языковом образовании, предлагая захватывающий, интерактивный и персонализированный опыт обучения. Однако важно учитывать этические аспекты, такие как устранение предвзятости и конфиденциальность данных. Совместные усилия преподавателей, законодателей и разработчиков необходимы для ответственного и инклюзивного внедрения NLP в преподавание языков, предоставления учащимся навыки глобальной коммуникации и культурной грамотности во взаимосвязанном мире.

Ключевые слова: Обработка естественного языка (NLP), Интеллектуальные системы обучения (ITS), приложения для изучения языка, Преобразование текста в речь

(TTS), Системы распознавания речи (SRS), Виртуальная реальность (VR), Дополненная реальность (AR), цифровая грамотность.

Introduction

Natural Language Processing (NLP) refers to the branch of artificial intelligence that focuses on the interaction between computers and human languages [Meera and Geerthik, 2022]. It involves the development of algorithms and models that enable computers to understand, interpret, and generate human language. NLP encompasses a wide range of tasks, including but not limited to text analysis, speech recognition, machine translation, and sentiment analysis. Researchers in NLP aim to bridge the gap between human communication and computational systems, allowing for more natural and effective interactions between humans and machines. Language teaching plays a crucial role in facilitating communication and fostering cultural understanding in a globalized world [Liddicoat and Scarino, 2013]. It equips individuals with the skills to express thoughts, ideas, and emotions effectively, thereby enabling meaningful interactions across linguistic and cultural boundaries. The use of technology, particularly NLP, in language teaching has the potential to enhance the learning experience by providing personalized and adaptive resources, immediate feedback, and engaging interactive platforms. By leveraging NLP tools and techniques, educators can create immersive and dynamic learning environments that cater to the diverse needs and preferences of learners.

This article explores the applications of Natural Language Processing (NLP) in the field of language teaching. It delves into the various ways in which NLP technologies can be integrated into language learning environments to improve teaching effectiveness, enhance learner engagement, and provide personalized learning experiences. The article begins by defining NLP and discussing its relevance in the context of language education. It then examines specific applications of NLP in automated language assessment, intelligent tutoring systems, language learning applications, and text-to-speech/speech recognition systems. Furthermore, the article explores the benefits and challenges associated with the use of NLP in language teaching, highlighting its potential to revolutionize the way languages are taught and learned in the digital age.

NLP Applications in Language Teaching

Automated Language Assessment (ALA) refers to the use of NLP technologies to evaluate and analyze language proficiency levels without human intervention. ALA systems leverage machine learning algorithms to assess various linguistic aspects, including grammar, vocabulary, fluency, and pronunciation. These systems can generate automated feedback and performance reports, providing learners with valuable insights into their strengths and areas for improvement [Evanini, Xie and Zechner, 2013]. By integrating ALA tools into language teaching platforms, educators can streamline assessment processes, track learner progress more effectively, and tailor instruction to meet individual learning needs.

Intelligent Tutoring Systems (ITS) utilize NLP techniques to create interactive and adaptive learning environments that mimic the role of a human tutor. These systems employ natural language understanding and generation algorithms to engage learners in meaningful conversations, provide personalized feedback, and deliver targeted instructional content [VanLEHN, 2011]. ITS can analyze learner responses in real time, identify misconceptions or

difficulties, and offer customized learning pathways to optimize learning outcomes. By harnessing the power of NLP, ITS revolutionizes traditional tutoring approaches by offering scalable, accessible, and personalized learning experiences.

NLP-driven **Language Learning Applications** encompass a wide range of tools and resources designed to support language acquisition and practice. These applications leverage NLP technologies such as machine translation, sentiment analysis, and natural language generation to facilitate immersive language learning experiences [Heil et al., 2016]. Language learning apps can provide interactive lessons, vocabulary drills, conversational simulations, and cultural context insights, catering to learners of different proficiency levels and learning styles. NLP-powered applications enhance engagement, motivation, and retention among language learners, making language learning more accessible and enjoyable.

Text-to-Speech (TTS) and Speech Recognition Systems (SRS) are integral components of NLP that play a crucial role in language teaching and learning. TTS systems convert written text into spoken language, allowing learners to listen to and practice pronunciation, intonation, and fluency [Kamath, Liu and Whitaker, 2019]. SRS, on the other hand, enables computers to recognize and interpret spoken language, facilitating interactive dialogue-based learning activities. By integrating TTS and SRS technologies into language teaching platforms, educators can create immersive and interactive language learning experiences that promote oral proficiency and communication skills.

Beyond automated assessment, **NLP contributes significantly to enhancing assessment and feedback mechanisms** in language teaching. NLP-powered systems can analyze learner-generated content, such as essays or spoken responses, to provide detailed feedback on grammar, vocabulary usage, coherence, and overall language proficiency [Shah, 2020]. This granular feedback not only helps learners identify specific areas for improvement but also fosters self-correction and autonomous learning skills. Moreover, NLP enables educators to track learner progress longitudinally, identifying patterns, trends, and areas of persistent difficulty for targeted intervention and support. Integrating NLP-enhanced assessment and feedback mechanisms into language teaching strategies enriches the learning experience, promotes continuous improvement, and empowers learners to achieve higher levels of linguistic competence.

Benefits and Challenges of NLP in Language Teaching

One of the key benefits of integrating NLP into language teaching is the **enhanced personalization and adaptability** it offers [Keezhatta, 2019]. NLP-powered systems can analyze learner data, preferences, and performance patterns to create personalized learning pathways and content recommendations. This level of customization ensures that learners receive tailored instruction that aligns with their individual learning styles, pace, and goals. Furthermore, NLP-driven platforms can adapt content and difficulty levels based on learner progress, maximizing learning outcomes and engagement.

NLP technologies enable the provision of **immediate feedback and error correction**, which are critical components of effective language learning [Keezhatta, 2019]. Automated language assessment systems can identify and highlight grammatical errors, pronunciation issues, and vocabulary gaps in real time, allowing learners to receive instant feedback on their performance. This immediate feedback not only facilitates timely error correction but also promotes self-correction and self-monitoring skills among learners. NLP-driven feedback

mechanisms enhance the efficiency and effectiveness of language instruction by addressing errors promptly and facilitating continuous improvement.

Despite its numerous benefits, the integration of NLP in language teaching also presents certain **limitations and ethical considerations** [Parra Escartín et al., 2017]. One potential limitation is the reliance on technology, which may result in overdependence and reduced critical thinking or communication skills among learners. Additionally, NLP algorithms may exhibit biases or inaccuracies, leading to skewed assessments or content recommendations. Ethical considerations include data privacy concerns, algorithm transparency, and equitable access to NLP-driven resources for all learners. Educators and developers must address these challenges by promoting digital literacy, ensuring algorithm fairness and transparency, and prioritizing ethical guidelines in NLP applications for language education.

Moreover, another challenge in integrating NLP into language teaching is the **need for continuous training and professional development for educators**. NLP technologies evolve rapidly, requiring educators to stay updated with the latest advancements, tools, and methodologies. Professional development programs should focus on enhancing educators' digital literacy, NLP proficiency, and pedagogical strategies for effective integration of NLP-powered resources in language teaching. Additionally, collaboration between educators, technologists, and curriculum designers is crucial to co-create innovative NLP-based learning solutions that address diverse learner needs and enhance educational outcomes. By investing in ongoing training and collaboration, educators can harness the full potential of NLP in language teaching while mitigating implementation challenges and maximizing benefits for learners.

Future Directions

The rapid **advancements in Natural Language Processing (NLP) technology** hold immense potential for further innovation in language teaching [Shah, 2020]. Future developments in NLP are expected to focus on improving accuracy, scalability, and efficiency in language processing tasks. Enhanced machine learning algorithms, deep learning models, and neural network architectures will enable NLP systems to achieve higher levels of language understanding, generation, and interaction. Additionally, advancements in data annotation, linguistic modeling, and cross-lingual transfer learning will contribute to more robust and versatile NLP applications for language education.

The integration of NLP with **Virtual Reality (VR) and Augmented Reality (AR)** technologies represents a promising direction for immersive language learning experiences [Lampropoulos et al., 2022]. By combining NLP-driven language processing capabilities with VR/AR environments, educators can create simulated real-world scenarios, interactive language exercises, and virtual language immersion experiences. Learners can engage in role-playing activities, cultural simulations, and communicative tasks within virtual environments, enhancing their language skills, cultural competence, and global awareness. The synergy between NLP, VR, and AR offers new opportunities to transform language teaching methodologies and provide enriching learning experiences.

Moreover, the future of NLP in language education also includes **advancements in natural language understanding (NLU) and generation (NLG) capabilities**. NLU improvements will enable NLP systems to interpret and respond to learner queries, prompts, and commands more accurately and contextually. NLG advancements, on the other hand, will enhance

the ability of NLP systems to generate coherent and contextually relevant language outputs, such as explanations, summaries, and instructional content. These developments will foster more natural and engaging interactions between learners and NLP-driven educational resources, promoting deeper learning and knowledge retention. Furthermore, future NLP innovations will focus on **enhancing multimodal language processing capabilities**. Multimodal NLP combines text, speech, images, and gestures to facilitate more comprehensive and interactive language learning experiences. By incorporating multimodal NLP techniques, educators can create diverse and engaging learning materials that cater to different learning styles and preferences. For example, interactive language learning platforms may integrate text-based exercises with audiovisual content, interactive simulations, and real-time feedback mechanisms, providing learners with immersive and effective learning experiences.

Looking ahead, future advancements in NLP technology, coupled with integration with Virtual Reality and Augmented Reality, will further revolutionize language education by offering immersive, interactive, and personalized learning environments. Educators, policymakers, and developers must collaborate to address ethical concerns, ensure equitable access to NLP-driven resources, and harness the full potential of NLP in promoting language proficiency, cultural understanding, and global communication skills among learners.

Conclusion

The integration of Natural Language Processing (NLP) in language teaching marks a pivotal advancement in educational methodologies. This article has examined various NLP applications, highlighting benefits like personalized learning, instant feedback, and adaptable instruction. These technologies address longstanding teaching challenges, enhancing scalability and engagement. NLP's synergy with educational methods fosters dynamic learning environments, nurturing linguistic proficiency and cultural understanding. Future strides in NLP, alongside virtual reality integration, promise transformative shifts in language education. While these developments offer tremendous potential, ethical considerations, including bias mitigation and data privacy, require careful attention. Collaborative efforts among educators, policymakers, and developers are essential to ensure responsible and inclusive deployment of NLP resources in language teaching. In essence, NLP's adoption signifies a journey toward tailored, adaptive, and immersive learning, empowering learners with global communication skills and cultural fluency.

References:

1. K. Evanini, S. Xie, and K. Zechner, “Prompt-based Content Scoring for Automated Spoken Language Assessment,” 2013. Accessed: Apr. 12, 2024. [Online]. Available: https://aclanthology.org/W13-1721.pdf
2. K. VanLEHN, “The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems,” Educational Psychologist, vol. 46, no. 4, pp. 197–221, Oct. 2011, doi: https://doi.org/10.1080/00461520.2011.611369
3. C. R. Heil, J. S. Wu, J. J. Lee, and T. Schmidt, “A Review of Mobile Language Learning Applications: Trends, Challenges, and Opportunities,” The EuroCALL Review, vol. 24, no. 2, p. 32, Sep. 2016, doi: https://doi.org/10.4995/eurocall.2016.6402

4. M. S. Keezhatta, “The Impact of Neuro-Linguistic Programming on English Language Teaching: Perceptions of NLP-Trained English Teachers,” International Journal of English Linguistics, vol. 9, no. 6, p. 454, Nov. 2019, doi: https://doi.org/10.5539/ijel.v9n6p454
5. C. Parra Escartín, W. Reijers, T. Lynn, J. Moorkens, A. Way, and C.-H. Liu, “Ethical considerations in NLP shared tasks,” doras.dcu.ie, Apr. 04, 2017. https://doras.dcu.ie/23231/
6. U. Kamath, J. Liu, and J. Whitaker, Deep Learning for NLP and Speech Recognition. Cham: Springer International Publishing, 2019. doi: https://doi.org/10.1007/978-3-030-14596-5
7. V. Shah, “Advancements in Deep Learning for Natural Language Processing in Software Applications,” INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY, vol. 4, no. 3, pp. 45–56, Sep. 2020, Available: http://ijcst.com.pk/index.php/IJCST/article/view/373
8. G. Lampropoulos, E. Keramopoulos, K. Diamantaras, and G. Evangelidis, “Augmented Reality and Virtual Reality in Education: Public Perspectives, Sentiments, Attitudes, and Discourses,” Education Sciences, vol. 12, no. 11, p. 798, Nov. 2022, doi: https://doi.org/10.3390/educsci12110798
9. S. Meera and S. Geerthik, “Natural Language Processing,” Artificial Intelligent Techniques for Wireless Communication and Networking, pp. 139–153, Feb. 2022, doi: https://doi.org/10.1002/9781119821809.ch10
10. J. Liddicoat and A. Scarino, Intercultural Language Teaching and Learning. John Wiley & Sons, 2013. Accessed: Apr. 12, 2024. [Online]. Available: https://books.google.com/books?hl=ru&lr=&id=cb6WbN-le-AC&oi=fnd&pg=PR8&dq=Language+teaching+plays+a+crucial+role+in+facilitating+communication+and+fostering+cultural+understanding+in+a+globalized+world+&ots=7ee_j4EP4u&sig=DiipBMFs4m5_FkEFg8jXVOiWmMc

STRUKTUR TAHLIL METODLARI ASOSIDA YARATILGAN SINTAKTIK NAZARIYALAR

Jo‘raqulova Madina Subhonovna
O‘zMU O‘zbek filologiyasi fakulteti
Kompyuter lingvistikasi va amaliy tilshunoslik
kafedrasi o‘qituvchisi.

Qodirova Zebo Gulboyevna
O‘zMU O‘zbek filologiyasi fakulteti
Kompyuter lingvistikasi va amaliy tilshunoslik
kafedrasi o‘qituvchisi.

Annotatsiya: Ushbu maqolada grammatik aniqlikni ta’minlashda sintaktik tahlilning o‘rni haqida fikr yuritiladi. Xususan, matndagi muayyan sintaktik birliklarni va strukturalarni avtomatik tarzda analiz qila oladigan sintaktik analiz (parsing) uchun olimlar tomonidan ishlab chiqilgan struktur tahlil metodlari va ular asosida yaratilgan nazariyalar tahlil qilinadi. Ushbu maqolaning asosiy maqsadi, gaplarni sintaktik tahlil qilishda struktur tahlil metodlarining vazifalarini yoritib berishga qaratilgan bo‘lib, tillarning xususiyatidan kelib chiqib, yaratilgan turli grammatik nazariyalarning umumiy va farqli jihatlari yoritib berilgan.

Kalit so‘zlar: Sintaktik parsing algoritmi, tabiiy tilning matematik modeli, transformatsion grammatika, generativ grammatika, formallashtirish, verbosentrizm

Annotation: This article discusses the role of syntactic analysis in ensuring grammatical accuracy. In particular, structural analysis methods developed by scientists and theories based on them are analyzed for syntactic analysis (parsing), which can automatically analyze certain syntactic units and structures in the text. The main goal of this article is to clarify the tasks of structural analysis methods in the syntactic analysis of sentences, and the common and different aspects of various grammatical theories created based on the characteristics of languages are highlighted.

Keywords: Syntactic parser algorithm, mathematical model of natural languages, transformational grammar, generative grammar, formalization, verbocentrism

Аннотация: В данной статье рассматривается роль синтаксического анализа в обеспечении грамматической точности. В частности, разработанные учеными методы структурного анализа и основанные на них теории анализируются для синтаксического анализа (парсинга), который позволяет автоматически анализировать определенные синтаксические единицы и структуры в тексте. Основная цель данной статьи - уточнить задачи методов структурного анализа при синтаксическом анализе предложений, а также выделить общие и различные аспекты различных грамматических теорий, созданных на основе особенностей языков.

Ключевые слова: алгоритм синтаксического анализатора, математическая модель естественных языков, трансформационная грамматика, порождающая грамматика, формализация, вербоцентризм

Kompyuter lingvistikasi - lingvistik qonuniyatlarni tavsiflash uchun matematik modellardan foydalanishni amalga oshiradi. Kompyuter lingvistikasini ikkita katta qismga bo‘lish mumkin. Birinchi qism lingvistik tadqiqotlarda kompyuter texnologiyalaridan foydalanish usullarini aniqlash uchun ma’lum matematik usullardan foydalanishni o‘rganadi. Ikkinchi qism lingvistik muammolarni hal qilishning matematik modellarini yaratish va ushbu modellar asosida ishlaydigan dasturlarni ishlab chiqishda foydalaniladi. Kompyuter lingvistikasining bu qismi tabiiy til matnlarini qayta ishlash tizimlarini ishlab chiqish bilan shug‘ullanadigan sun’iy intellekt bo‘limi bilan chambarchas bog‘liq. Tabiiy tilni qayta ishlash – bu nutq yoki matn orqali tabiiy tilni o‘rganish va uni dasturiy ta’minot va algoritmlar yordamida qanday boshqarish yoki tushunish jarayoni hisoblanadi. NLPni o‘rganish uzoq vaqtdan (50 yildan ortiq) beri mavjud bo‘lib, matnni qayta ishlash va tushunish uchun qanday lingvistik modellar yaratish mumkinligi bilan qiziqish uyg‘otdi. NLP tilni qayta ishlash uchun statistik ma’lumotlar va mumtoz tilshunoslik modellaridan foydalanishga e’tibor qaratgan bo‘lib, Kompyuter lingvistikasida tabiiy tilni modellashtirish inson faoliyati davomida axborot almashinuvida foydalaniladigan va ushbu faoliyatga asosan doimiy o‘zgarib boradigan ko‘p bosqichli belgilar tizimi bilan bog‘liq jarayondir. Ma’lumotlarni tahlil qilish uchun an’anaviy vositalar, metodologiyalar va ko‘nikmalardan foydalanish uchun berilgan matnni strukturlangan formatga o‘tkazish lozim bo‘ladi.

Tabiiy tilni qayta ishlash jarayonining ikkita asosiy usuli mavjud. Bular: sintaktik va semantik tahlil. Sintaksik tahlil grammatik aniqlikni ta’minlash uchun so‘zlarni joylashtirishga imkon beradi. Aniq va mukammal tushunish uchun tabiiy tilning grammatik qoidalarga muvofiqligini baholaydi. Sintaktik tahlil yordamida so‘z turkumlari ajratib olinadi va grammatik qoidalar orqali ularning ma’nosi aniqlanadi. Til sintaksisi generativ matn uchun asos bo‘lib, nutq qismlari va tahlil iyerarxiyasi uchun asos bo‘lib xizmat qiladi. [Abduraxmonova N. 2021: 154]

Har bir fan kabi tilshunoslik fani ham tarixiy taraqqiyoti davomida lingvistik hodisalarni o‘rganishda, tadqiqot obyektini ilmiy jihatdan tahlil etib, uning ichki strukturasini, o‘ziga xos jihatlarini ochishda, tabiatini aniqlashda, ilmiy xulosalar chiqarishda o‘ziga xos metodlar asosida ish olib boradi. Tilning qonuniyatlarini sinxron nuqtayi nazardan tekshiradigan *distributiv, transformatsion, bevosita ishtirokchilar, derivatsion* kabi struktur tahlil metodlari mavjud bo‘lib, ushbu maqolaning asosiy maqsadi gaplarni sintaktik tahlil qilishda struktur tahlil metodlarining vazifalarini yoritib berishga qaratilgan bo‘lib, bu haqida quyida to‘xtalamiz. Sintaktik strukturalarni o‘rganish tillarning xususiyatidan kelib chiqib, turli nazariyalarning yaratilishiga sabab bo‘lgan. Tilshunoslik fani o‘zining bugungi taraqqiyoti davrigacha erishgan ulkan yutuqlarida xorijiy mamlakatlar olimlarining ham salmoqli ulushlari bor, albatta. Ayniqsa, amerikalik strukturalistlarining, xususan, deskriptiv tilshunoslik namoyandalarining, Praga tilshunoslik maktabi a’zolarining, Kopengagen glossematik yo‘nalishi vakillarining bu sohadagi xizmatlari tahsinga sazovordir.

Tilshunoslik fanida “struktura” va “struktur tilshunoslik” terminlari 1939-yilda Kopengagen tilshunoslik maktabining yirik namoyandalaridan biri V.Bryondalning “Struktur tilshunoslik” nomli maqolasidan keyin qo‘llanila boshlandi. Kopengagen strukturalizm maktabi jahon tilshunosligida *glossematika* nomi bilan mashhur bo‘ldi. Uning asosiy tashkilotchilaridan biri Lui Yelmslevdir.

Lui Yelmslev tilshunoslikda birinchilardan bo‘lib, til materiali talqiniga matematik qoida va metodlarni tadbiq etdi. [Turniyozov N. 2009: 6] L.Yelmslev til unsurlarining paradigmatik va

sintagmatik munosabatlari xususida ham ilmiy mulohaza yuritadi. Paradigmatik munosabat tilga, sintagmatik munosabat nutqqa taalluqli ekanligini aytadi.

Shuni alohida ta'kidlash lozimki, yuqorida eslatib o'tilgan tilshunoslik maktablari namoyandalarining ko'pchiligi u yoki bu ma'noda V.fon Gumboldt va F.de Sossyurni o'zlarining ustozlari, deb biladilar. Bundan tashqari, L.Blumfildning lingvistik qarashlari talqiniga ham alohida to'xtash lozim, zero, u deskriptiv tilshunoslikning asoschisidir[Abdurakhmonova, N. 2019,2021,2022,2023].. XX asrning 50-yillarida shakllanib, rivojlanish bosqichiga o'tgan struktur tilshunoslikning deskriptiv maktabi vakillari an'anaviy tilshunoslikning aniq lingvistik tadqiqot tamoyillari yo'q ekanligini tanqid qilgan holda, o'z oldilariga tilni o'rganishning formallashtirilgan tahlil tizimini ishlab chiqishni maqsad qilib qo'ydilar. Ana shunday tahlil asosida tadqiq etilayotgan obyektni o'rganishda har qanday subyektivlikdan holi bo'lgan obyektiv xulosaga kelish mumkinligini ta'kidlaydilar. Lekin struktur tilshunoslikning glossematika yo'nalishida substansiyani shakldan ajratish shu darajaga yetdiki, natijada shakl o'zining substansiyasidan ajralib qoldi.

Bu esa struktur tilshunoslik bag'ridan yangi yo'nalish-generativ lingvistikaning o'sib chiqishiga olib keldi.

Amerika deskriptiv (tavsifiy) tilshunoslik maktabi Leonard Blumfildning lingvistik konsepsiyasi asosida taraqqiy etdi. Buni strukturalizmning turli yo'nalishlarida tadqiqot ishlari olib borgan Z.Xarris, Ch.Xokket, Ch.Friz, N.Xomskiy, Y. Nayda kabi tilshunoslarning tadqiqot ishlarida kuzatish mumkin. Deskriptiv tilshunoslikning asosiy tahlil metodlaridan biri distributiv metoddir. Mazkur metodga aslida L.Blumfild asos solgan edi. Biroq uning asosiy tamoyillari Z.Xarrisning "Struktural tilshunoslik metodlari" deb nomlanuvchi asarida mukammal yoritib berildi. "Distributsiya" termini inglizcha so'z bo'lib, "qurshov" degan ma'noni anglatadi. Shuning uchun distributiv tahlilning asosiy maqsadi til birliklarining, xususan, fonema va morfemalarning real qo'llanishi qurshovini o'rganishdan iboratdir. Boshqacha aytganda, distributiv tahlil metodi til birliklarining bir-birlari bilan qanday qurshovda munosabatga kirisha olishi masalasi tavsifi bilan shug'ullanadi. Distributiv tahlilda til birliklarining sintagmatik qatorda qay darajada o'rinlashishi xususida qimmatli ma'lumotlar olish mumkin. Lekin mazkur tahlilning zaif tomonlari ham bor, albatta. Masalan, ushbu metod orqali sintagma yoki so'z birikmalarining o'zaro qurshov munosabatlarini tahlil qilish qiyin. Boshqacha aytganda, sintaktik strukturalarga qo'llab bo'lmaydi. Shu munosabat bilan deskriptiv tilshunoslikda yangi tahlil metodi – bevosita ishtirokchilar metodini joriy etishga ehtiyoj tug'ildi.

Gapning bevosita ishtirokchilarga bo'linishi metodi lingvistik adabiyotlarda *bevosita ishtirokchilar* **(BI)** deb yuritiladi. Amerika struktural tilshunosligida Z.Xarris tomonidan ilmiy asoslangan distributiv tahlil usulining ba'zi nuqsonlari mavjudligi, masalan, gapni segmentlarga ajratib tahlil qilishda seziladigan zaiflik tomonlari, tilshunoslarni yangi tahlil metodlarini joriy etishga undadi. Bunday tahlil metodlaridan biri BI tahlili metodidir. Mazkur metod haqida dastlabki ma'lumot L.Blumfild tomonidan "Til" deb nomlangan asarida berilgan edi. Keyinchalik bu tahlilning mukammal tavsifi amerikalik tilshunoslar R.Ueellz va E.Naydalarning tadqiqot ishlarida yoritildi.

BI metodi vositasida berilgan gap dastlab ikki segmentga (bo'lakka) ajratiladi. Bu segmentlarning har biri, ikkinchisi bilan bog'liq bo'lishidan tashqari, o'z sathida bir necha bevosita ishtirokchi unsurni birlashtirgan bo'ladi. Tahlilning ikkinchi bosqichida avval ajratilgan BI

unsurlar yana ikkiga bo‘linadi va ana shu tartibda ajralmaydigan qism qolmaguncha tahlil davom etaveradi. Masalan:

1. *Cho‘lponning so‘zi Jo‘rabekni hushyor torttirdi*

Birinchidan, gapning BI metodiga ko‘ra tahlilini gorizontal chiziq asosida ko‘rib o‘tamiz:

Cho‘lponning || so‘zi | Jo‘rabekni || hushyor torttirdi.

Bu munosabatni quyidagi shartli belgilar yordamida ham ko‘rsatish mumkin:

S=NP+VP; NP=N1+N2; VP=N+V.

Shartli belgilar quyidagilarni bildiradi: **S** – gap, **NP** – ot tasarrufidagi guruh, **VP** – fe’l tasarrufidagi guruh, **N** – ot yoki uning o‘rnida kelayotgan biror so‘z, **V** – fe’l yoki uning o‘rnida kelayotgan biror so‘z, **D** – har qanday aniqlovchi yoki yordamchi so‘z, raqamlar esa guruh tarkibidagi bir turkumdagi birliklarning miqdorini bildiradi.

Bundan tashqari, BI tahlili vertikal chiziq asosida ifodalanishi ham mumkin.

Gapning BI birliklarga bo‘linishi yuqoridagi kabi berilsa, uni BI analizi deb ataymiz. Ammo zarurat paydo bo‘lganda, tahlilni eng kichik BI unsurlarining o‘zaro birikib, katta BI guruhlarni hosil qilishdan boshlash mumkin. Bunday tahlil BI sintezi deb ataladi[Abdurakhmonova, N. 2019,2021,2022,2023].

Shunday qilib, BI tahlili metodi an’anaviy sintaktik tahlil usulidan keskin farqlanishini va gap strukturasining grammatik shakllanishini mukammal darajada aniq ifodalashini ko‘rish mumkin.

BI tahlili usulining ham ba’zi nuqsonlari mavjud bo‘lib, mazkur tahlil usulida bir yo‘la ikki va undan ortiq mustaqil gaplarni tahlil etib bo‘lmasligini kiritish mumkin. Biroq ushbu kamchiliklar transformatsion tahlilda barham topadi.

BI tahlili metodining quyidagi nuqsonlari yangi metodning yaratilishi uchun zarurat tug‘dirdi.

1. BI tahlil metodi gaplarning ma’no jihatiga kam e’tibor beradi, zotan, u tahlil uchun berilgan jumlaning ishtirokchi unsurlari o‘rtasidagi sintaktik aloqalarni ko‘rsatib beradi, xolos.

2. BI metodiga ko‘ra, berilgan gap tayyorligicha tahlil qilinadi. Masalan, aniq nisbat shaklidagi gap berilgan bo‘lsa, uning majhul nisbat shaklidagi qurilma bilan aloqasini ochib bera olmaydi.

3. BI tahlili metodi sintaktik nuqtayi nazardan ichki tub strukturani topish va uning ahamiyatini ko‘rsatib berish imkoniyatiga ega emas.

Ana shu sabablarga ko‘ra, gaplarning sintaktik modellarini aniqroq ifodalab bera oladigan transformatsion metod oriy etildi. Uning mualliflari Z.Xarris va N.Xomskiy hisoblanadi. Transformatsion metod (TM) distributiv hamda BI tahlili metodlarining barcha yutuqlarini o‘zida mujassamlashtiradi va shu bilan birga, quyidagi shartlari mavjudligi bilan muhim ahamiyat kasb etadi:

1. Muayyan tayanch strukturaga asoslanadi.
2. Berilgan gapning ma’nosini to‘liq saqlagani holda sintaktik shaklini o‘zgartiradi.
3. Tahlil jarayonida jumlaning ma’nosiga ta’sir etmaydigan leksik o‘zgarishlar yasay oladi.

Transformatsiyaning turli xil turlari mavjud bo‘lib, ulardan *nominalizatsiya, aktiv qurilmaning passiv qurilmaga aylanishi, almashinuv, kontaminatsiya, konversiya* transformatsiyalari ko‘p qo‘llanadi [Turniyozov N. 2009: 35]. Transformatsiya metodi gapning ichki va tashqi strukturasi va bu struktura birliklarining o‘zaro munosabatiga tayanadi. Agar distributiv tahlil va BI metodlari jumla qanday mayda qismlardan tashkil topadi, degan savolga

javob berishga harakat qilgan bo‘lsa, TM bu jumla qanday gapning transformatsion qayta shakllanishidan hosil bo‘lgan, degan savolga javob beradi. F.M.Berezin to‘g‘ri ta’kidlaganidek, TMning ojiz tomoni ana shunda namoyon bo‘ladi. Chunki Transformatsion tahlil bevosita lingvistik materialdan kelib chiqmaydi, balki dastlab sof empirik asosga quriladi, so‘ngra til materiali asosida tekshiriladi. Bundan tashqari, transformatsiya metodining qo‘llanilish doirasi va turli tillardagi yadro gaplar tiplarining miqdori aniq belgilanmagan.

Transformatsiya metodidagi ana shu cheklangan tomon e’tiborga olinib, sintaktik transformatsiya bilan sintaktik derivatsiyani bir-biridan farqlash lozimligi ta’kidlanadi.

S.D.Kasnelson gaplar o‘zaro faqat transformatsiya munosabatida emas, balki derivatsiya munosabatida ham bo‘lishini bayon qilgan edi. [Кацнельсон С.Д. 1967: 54] Derivatsiyaning tarkibiy qismlari ham transformatsiyaning tarkibiy qismlariga analogiya yo‘li bilan quyidagicha belgilanadi: asos gap uchun *operand,* hosila gap uchun *derivat,* yasovchi vosita uchun esa *derivatsiya operatori* atamalarini qo‘llaydi.

Amerikalik tilshunos Noam Xomskiy (Noam Chomsky) transformatsion grammatika (yoki transformatsion-generativ grammatika), deb nomlanuvchi nazariyani taklif qildi. N.Xomskiyning fikricha, “til – til birliklari va ularning sinflari majmuyi emas, balki to‘g‘ri jumlalar tarkib toptiradigan mexanizm”. Sintaksisni Xomskiy jumlalar tuzish tamoyillari va usullari haqidagi ta’limot sifatida belgiladi[Abdurakhmonova, N. 2019,2021,2022,2023].

N.Xomskiyning 1957-yilda nashr qilingan “Sintaktik strukturalar” asari jahon tilshunosligida yangi yo‘nalishga asos soldi. Mazkur tadqiqotda Xomskiy “tilning barcha gaplarini bilish mumkin emas, til asosida ayrim cheklangan qoidalar tizimi bo‘lishi kerak”, degan fikrni ilgari suradi. [Noam Chomskiy. 1957: 13] “Bu qoidalar tizimi esa tilning grammatikasi” demakdir. So‘zlovchi ham tinglovchi ham muloqot jarayonida grammatikani uning vositasida “to‘g‘ri” ifoda qurish, yoki “to‘g‘ri” qurilgan ifodani tushunish uchun vujudga keltiradi”:

1)pie little blue mud make eye girl was
non kichkina moviy qum yasamoq ko‘zli qizcha

2) The little pie with mud eyes was making a blue girl. Kichkina non qum ko‘zli moviy qizcha yasar edi.

3) The little girl with blue eyes was making a mud pie. Moviy ko‘zli kichkina qizcha qumdan non yasar edi.

Eshitilgan tovushlar tizimi mazkur til grammatikasini bilgandagina ma’no kasb etadi.

Xomskiyning ta’limotiga ko‘ra, har bir tilda mavjud bo‘lgan son-sanoqsiz gaplar, nutq so‘zlovchisining yangi-yangi gaplarni cheklanmagan darajada yarata olish qobiliyati bilan bog‘liqdir. Bu o‘rinda ma’no anglatmagan jumlalar ham grammatik jihatdan to‘g‘ri bo‘lishi mumkin. Boshqacha qilib aytganda, har bir kishi o‘zining nutqiy faoliyati jarayonida shu paytgacha eshitilmagan yangidan-yangi gaplarni yaratishi mumkin va bu yangi gap tinglovchi tomonidan to‘g‘ri tushuniladi.

N.Xomskiy nazariyasiga ko‘ra, transformatsion tahlil sintaktik strukturalarning *zohiriy*dan *botiniy*ga aylanish yo‘li bilan tahlil qilishdir. Haqiqatdan ham, nutq va tafakkur jarayonlarida *botiniy* strukturaning o‘rni mavjud. Biroq ushbu “*botiniylik*”ning qanday va qayerdan paydo bo‘lganligiga N.Xomskiyning dastlabki tadqiqot ishlarida ahamiyat berilmaydi. Uning transformatsion modeli tabiiy til ifodalarini avtomatik sxemalarga solish imkoniyatinigina yaratgandi. Keyinroq N.Xomskiy asta-sekin insonning nutqiy faoliyati muammolari bilan qiziqa boshladi.

N. Xomskiy F.de Sossyurning til-nutq dixatomiyasiga muqobil bo‘lgan *lisoniy qobiliyat* (*linguistie competence*) va *lisoniy faollik* (*linguistie performance*) tushunchalarini rivojlantirdi. U lisoniy qobiliyatni tilshunoslik sohasiga, lisoniy faollikni psixologiya sohasiga aloqador va lisoniy qobiliyatni mazkur tilda so‘zlashish qobiliyatini tashkil qiluvchi, lisoniy faollikni esa til egasi tomonidan yaratiladigan ifodalar sifatida izohladi[Abdurakhmonova, N. 2019,2021,2022,2023].

“Xomskiy inqilobi” lingvistning yo‘nalishini, tilga taksonomik yondashuvdan generativ yondashuvga o‘zgartirdi. Bungacha AQShda distributiv tahlilga e’tibor qaratilgan edi. Endi oldingi o‘ringa N.Xomskiy yaratgan formal qoidalar chiqdi. Grammatik dalillar (grammatik moslashuv, ellipsis kabilar)ni tavsiflashda formal qoidalarga tayanish masalasi markaziy masalaga aylandi. Grammatik nazariyalarni formallashtirishda yangi metatil yaratildi.

Vujudga keltirish grammatikasi termini ilmiy muomalaga Noam Xomskiy ishlarida 1950-yillar oxirida keltirilgan. Manbalarda N.Xomskiyning bundan avval transformatsion grammatika terminini qo‘llaganligi (ingl. *transformational grammar*) qayd etiladi. N.Xomskiyning fikricha, vujudga keltirish grammatikasining xususiyatlari universal grammatika asosida keltirib chiqarilgan.

Vujudga keltirish grammatikasi tarafdorlari grammatik qurilish kommunikativ faoliyat natijasi emas va inson tomonidan o‘zlashtirilmaydi, deb uqtiradilar. Bu jihatdan vujudga keltirish grammatikasi boshqa yo‘nalishlardan farq qiladi.

Vujudga keltirish grammatikasining qoidalari algoritm kabi funksionallashib, unda gapning grammatik jihatdan to‘g‘ri yoki noto‘g‘riligini aniqlashga qaratiladi.

Generativ lingvistika (generativizm, vujudga keltirish grammatikasi, transformatsion vujudga keltirish grammatikasi) quyidagi qoidalarga asoslanadi:

1.Til grammatikasi formal sistema sifatida qaraladi. Ilmdagi bu burilish nafaqat tilni o‘rganish, balki til grammatikasini formal obyekt sifatida belgilashga olib keldi.

2.Transformatsion vujudga keltirish grammatikasining asosiy xulosalaridan biri shundaki, unda gap mazmun bilan bog‘langan ichki va talaffuz hamda yozuvda o‘z ifodasini topuvchi tashqi strukturaga ega, u insonga tilga oid universal xususiyatlar tug‘ilganidayoq taqdim etilgan. Insonning nutqi ustida tajriba o‘tkazilishi bilan til bilimi oshib boradi. “Umuman til” universaliyasi konkret til detallari bilan boyib boradi. “Universal grammatika”ning eng umumiy mexanizmlari konkret til mexanizmlari bilan boyiydi. O‘z ona tilisini o‘zlashtirar ekan, bola mavjud barcha xatolarga emas, balki ayrim xatolarga yo‘l qo‘yadi. Shu bois, inson ona tilisini o‘zlashtirishi uchun nutq materiallari bilan bir qancha tajribalar o‘tkazsa va gap qurilishidagi zarur parametrlarni bilsa, bo‘lgani. Masalan, gapda eganing o‘rni qayerda, moslashuvli bog‘lanish bormi, olmosh bilan ifodalangan ega ellipsisiga uchraganmi va hokazo. Noam Chomskiyning ilmiy qarashlari tilning formal tavsifini yozishda asos vazifasini o‘tadi. U lingvistik modelning ikki turini keltiradi: 1-minimal lingvistik nazariyaga moslashgan Markov jarayoni sifatida til konsepsiyasiga asoslangan nazariy model hamda tarkib toptiruvchi analiziga asoslangan frazali strukturaviy model. N.Xomskiy strukturalarni analiz qilishning ikki yondashuvi mavjud ekanligi, bu o‘z navbatida har bir sintaktik qolipning tuzilishidan kelib chiqishini ta’kidlaydi: *chuqur struktura (deep structure)* va *yuzaki struktura (surface (shallow) structure.* Chuqur (ichki) strukura bu semantik munosabatga qurilgan tuzilma, yuzaki struktura esa uning sintaktik (shakliy) ifodasi. [Noam Chomskiy. 1957: 15]

Har bir tabiiy til matematik jihatdan tuzilmagan va formallashtirilmagan tarkiblardan tashkil topgan murakkab tizimdir. Biroq tabiiy til bo‘yicha olib borilgan tadqiqotlar shuni

ko‘rsatadiki, tuzilmagan va rasmiylashtirilmagan tabiiy tilni chiziqli usul yordamida tuzilgan va rasmiylashtirilgan shaklga keltirish mumkin[Abdurakhmonova, N. 2019,2021,2022,2023]. Bu so‘z, so‘z va gaplarning turlari bo‘yicha mantiqiy-lingvistik modellarni qayta qurish, so‘ngra matematik modellarni yaratish bilan amalga oshiriladi. Bunda grammatika lingvistik generator sifatida, ba’zan “rekognayzer” - berilgan chiziqlarning tilga tegishli (to‘g‘ri yoki noto‘g‘ri) ekanligini aniqlash me’yori sifatida qo‘llanadi. Parsing-tabiiy tildagi chiziqli munosabatlarni qismlarga ajratish va simvollarga biriktirish, har birini til grammatik qoidalariga solishtirib ko‘rish jarayonidir. Ushbu metodologiyani tilni formallashtirish darajasi sifatida aniqlash mumkin. Formallashtirish darajasi, o‘z navbatida, tabiiy til semantikasining qoidalarga solinganlik darajasini va algoritmning aniqligini belgilaydi.

Struktur tilshunoslik haqida mulohaza yuritilganda, fransuz tilshunosi Lusyen Tenyerning bu sohadagi tadqiqot ishlarini eslamay o‘tib bo‘lmaydi. L.Tenyer tilshunoslikda “Struktur sintaksis asoslari” deb nomlanuvchi asari bilan shuhrat qozondi. Mazkur asar 1959-yilda olimning vafotidan keyin chop etildi. Asarni Tenyerning qo‘lyozmasi asosida uning safdoshlari Furkye hamda Domalar nashrga tayyorladilar. Olimning bu asari struktur tilshunoslik masalalari talqiniga bag‘ishlangan bo‘lsa-da, u odatdagi struktur tilshunoslik doirasida yozilgan ishlarga mutlaqo o‘xshamas edi (N.Xomskiy kabi amerikalik olimlarning struktur tilshunoslik doirasida yozilgan asarlari ko‘zda tutilmoqda). Darhaqiqat, L.Tenyerning til materialini tahlil qilish usuli o‘ziga xosdir va u deskriptiv tilshunoslikning tahlil usullaridan farq qiladi[Abdurakhmonova, N. 2019,2021,2022,2023]. L.Tenyerning sintaktik nazariyasi “verbosentrik” tamoyilga asoslangan bo‘lib, uning asosiy mazmunini fe’l valentligi masalasi tashkil etadi. Olim sintaktik struktura, shu jumladan, gap tarkibida qatnashayotgan turli turkumlardagi so‘zlar ichida fe’lning muhim ahamiyat kasb etishini, boshqa turkumlardagi so‘zlar esa u orqali boshqarilishini alohida ta’kidlaydi. L.Tenyer jumla tarkibida fe’l ohanrabo vazifasini bajarishini va shu bois jumlaning boshqa komponentlari u orqali o‘z o‘rinlarini topishini ko‘rsatib beradi. [Гак В. 1988: 8]

Olimning til materiali tahlil usulining o‘ziga xosligi shundaki, u jumlani teatr sahnasiga o‘xshatadi va bunda sahnadagi harakatni jumlada fe’l orqali bajariladigan harakat bilan qiyoslaydi. Ma’lumki, sahnadagi harakat aktyorlar tomonidan bajariladi. Gapda (jumlada) esa ish-harakatning bajarilishi aktantlar zimmasiga tushadi[Abdurakhmonova, N. 2019,2021,2022,2023]. Olim bu o‘rinda ham aktyor va aktant tushunchalarini qiyoslaydi. Yuqoridagilardan tashqari, teatr sahnasida harakatning bajarilishi ma’lum bir muhit bilan bog‘liqdir. L.Tenyer bunday holatni jumlada holat bildiruvchi so‘zlar (sirkonstantlar) bajarishini izohlab beradi. L.Tenyer jumla tarkibida fe’l orqali boshqariladigan va ish-harakatning voqelanishida faol ishtirok etadigan unsurlarni aktant deb ataydi. Uning “verbosentrik” yoki “bir cho‘qqili” nazariyasida aktantlar uchga bo‘linadi: *1) birinchi darajali aktant; 2) ikkinchi darajali aktant; 3) uchinchi darajali aktant.*

Birinchi darajali aktant ish-harakatning ijrochisi hisoblanadi va u qolgan ikki aktantga nisbatan o‘ta faol sanaladi:

1. *Hamlet explained it to me* .

(Hamlet buni menga tushuntirdi.)

2. *Saida qo‘llarini sochiqqa artdi.*

Mazkur gaplarda **Hamlet, Saida** so‘zlari birinchi darajali aktant vazifasini bajarmoqda. Zotan, bunda **Hamlet, Saida** so‘zlari ish-harakatning ijrochisidir. Bu jihatdan birinchi aktant an’anaviy tilshunosligimizdagi gapning bosh bo‘lagi – egaga to‘g‘ri kelmoqda.

Ikkinchi darajali aktant ish-harakatning mazmunini o‘ziga to‘liq singdirib oladi. Bu jihatdan uni odatdagi vositasiz to‘ldiruvchi bilan qiyoslash mumkin. Yuqorida berilgan gaplarda **it, qo‘llarini** so‘zlari ana shu vazifada kelmoqda.

Uchinchi darajali aktant ish-harakatning ijrosi bilan bilvosita aloqador bo‘ladi. Bu jihatdan uni odatdagi vositali to‘ldiruvchi bilan qiyoslash mumkin bo‘ladi. Aynan ana shu gaplarda qatnashayotgan **to me, sochiqqa** so‘zlari esa uchinshi darajali aktantdir.

Ko‘rinadiki, aktant vazifasida har doim ot turkumidagi so‘z keladi. Ammo buni keng ma’noda tushunmoq lozim. Chunki ot vazifasini olmosh hamda otlashgan so‘zlar ham bajara oladi.

Albatta, harakat ijrosi qurshovida nafaqat aktantlar, balki sirkonstantlar ham, determinativ so‘zlar (aniqlovchi) ham kelishi mumkin. L.Tenyer determinativ so‘zlarni maxsus nom bilan atamaydi va ularni aktantlar tasarrufida kelib, ularga izoh beruvchi so‘zlar deb ko‘rsatadi:

1. *My brother’s family is large* (Akamning oilasi katta.)

2.*O‘qishning foydasi ko‘p bo‘ladi.*

Ushbu misollarda qatnashayotgan **brother’s, o‘qishning** so‘zlari birinchi darajali aktant **family, foydasi** so‘zlarining tasarrufiga kiradi.

Holat bildiruvchi so‘zlarni olim sirkonstant termini orqali ataydi. Masalan:

1. *He may return soon* (U tezda qaytishi mumkin.)

2. *Ular juda sekin yurishdi.*

Berilgan misollarda **soon, sekin** so‘zlari sirkonstant sanaladi. Boshqacha aytganda, an’anaga ko‘ra hol deb ataluvchi so‘zlar L.Tenyer nazariyasida sirkonstant nomi bilan ataladi.

Shuni ham aytish kerakki, jumla tarkibidagi aktantlar soni fe’lning valentlik darajasini belgilaydi. Gapda aktantlar soni qancha bo‘lsa, fe’lning sintaktik valentligi darajasi ham shuncha bo‘ladi. Biroq olim ba’zi fe’llar aktantsiz bo‘lishini, ayrimlari bir aktantli, yana boshqa birlari esa ikki va undan ortiq aktantlarni boshqarishi mumkinligini aytadi. Masalan, ingliz tilida **Read** (O‘qi) deb aytilgan buyruq maylidagi faqatgina kesimdan iborat gapda aktant kuzatilmaydi. Demak, ayni paytda, fe’lning valentlik darajasi nolga teng. Agar **Sobir Komilning kitobini Akbarga berdi,** deb gap tuzsak, u holda fe’lning valentlik darajasi uchga teng bo‘ladi. Chunki bu gapda uchta aktant ishtirok etmoqda: Sobir – 1-darajali, kitobini – 2-darajali, Akbarga – 3-darajali aktantlardir.

Shunday qilib, L.Tenyerning “verbosentrik” nazariyasi o‘ziga xosligi bilan Praga tilshunoslik maktabi hamda amerikalik olimlar tomonidan asos solingan struktur tilshunoslik tamoyillaridan tubdan farqlanishini ko‘ramiz. L.Tenyer til materiali talqiniga mantiqiy tushunchalarni aralashtirmaslikka harakat qiladi. Formallashtirish - bu bilimlar majmuasini, mazmunini muayyan shakllar yordamida sun’iy til belgilari orqali ifodalashdir. Grammatikani formallashtirish tendensiyasi XX asrning o‘rtalarida kuchaydi. Bu bevosita tilning strukturasini yanada aniq parametrlarda ifodalashga intilish, uni kompyuterga moslashtirishga urinish bilan bog‘liqdir. Yuqorida sanab o‘tilgan sintaktik tahlil metodlarining barchasi gapni segmentlarga bo‘lgan, faqat segmentlarga bo‘lishda yondashuvi bir-biridan farq qiladi.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. Kompyuter lingvistikasi (darslik), Toshkent: Nodirabegim, 2021. - B. 400
2. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. InProceedings of the International Conference on Language Technologies for All (LT4All) 2019.

3. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 1790-1793). IEEE.
4. Абдурахмонова, Н., & Бойсариева, С. (2023). Tabiiy tilni qayta ishlashda (nlp) okkazionalizmlarning morfem tahlili. Международный журнал искусство слова, 6(3).
5. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
6. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
7. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
8. Гак В. Л.Теньер и его структурный синтаксис // Вступительная статья: Теньер Л. Основы структурного синтаксиса. –М., 1988.-С.8.
9. Кацнельсон С.Д. Порождающая грамматика и принцип деривации // Проблемы языкознания, - М., 1967.
10. Noam Chomskiy. Syntactic structures.–Paris: Mouton, 1957.–P.13
11. Теньер Л. Основы структурного синтаксиса. –М.: Прогресс, 1988.
12. Turniyozov N., Turniyozova K., Xayrullayev X. Struktur sintaksis asoslari. -Toshkent: Fan, 2009. -B 150

НЕЧЕТКАЯ ЛОГИКА НА БАЗЕ ЯЗЫКОВЫХ МОДЕЛЕЙ: ПОСТРОЕНИЕ ГИБРИДНОЙ СИСТЕМЫ КЛАССИФИКАЦИИ ДИАЛЕКТОВ И ОПРЕДЕЛЕНИЯ ДИАЛЕКТИЗМОВ

Миндубаев Артур
Аспирант Академии Наук Республики Татарстан,
Институт прикладной семиотики.
E-mail: a.mindubaev@bk.ru

Аннотация: В данной работе поставлена цель создание информационной системы классификации диалектов и определения диалектизмов с использованием нейросетевой модели и механизма логического вывода на базе правил нечёткой логики. Данная система не имеет прямых аналогов для русского языка, что подчёркивает актуальность данной работы.

Для решения поставленной задачи использовался язык Python с библиотеками для обработки естественного языка, обучения нейронной сети, работы с трансферной моделью Google Bert и создания базы правил нечёткой логики для модели Мамдани.

В научно-исследовательской работе был выполнен анализ диалектологической предметной области и поиск методов решения задач диалектологии с помощью методов обработки естественного языка и машинного обучения. Автором была предложена архитектура нейросетевой модели для классификации диалектов и сформированы основные концепции алгоритма определения диалектизмов. Следующим шагом было произведено обучение нейросетевой модели на основе Bert и реализован экспериментальный алгоритм для поиска диалектизмов.

Ключевые слова: классификация диалектов, нейронные сети, LM, Bert, нечёткая логика, модель Мамдани.

Abstract: This work aims to create an information system for classifying dialects and identifying dialectics using a neural network model and a logical inference mechanism based on the rules of fuzzy logic. This system has no explicit analogues for the Russian language, which emphasizes the relevance of this work.

To solve the problem, the Python programming language was used with libraries for natural language processing, neural network training, Google Bert model usage and base of fuzzy logic rules creation for the Mamdani model.

The research work involved an analysis of dialectology field and a search for methods using natural language processing and machine learning methods for dialect tasks. The author offered the architecture of the neural network model for the dialect classification and formed the basic concepts of the dialethism's determination algorithm. The next step was a neural network model training on Bert outputs and dialethism's determination experimental algorithm implementation.

Keywords: dialect classification, neural network, LM, Bert, fuzzy logic, Mamdani model.

Вступление

Современная форма распространённых естественных языков зачастую имеет установленный набор норм и правил. Такая форма языка называется литературной или

стандартной. Однако стандартизация естественных языков – это достижение нескольких последних столетий. В то же время территориальная, экономическая или социальная обособленность групп людей, говорящих на одном и том же языке, неизбежно приводила к характерным отклонениям и порождала диалекты. Степень отличия диалектов сильно варьируется в зависимости от условий, из-за чего носители диалектов могут, как испытывать незначительные трудности при общении, так и полностью потерять возможность понимать друг друга. Поэтому диалектическая экспертиза текстов может быть крайне сложной и кропотливой работой.

Несмотря на это, в результате анализа предметной сферы и поиска готовых коммерческих решений сложилось мнение, что набор инструментов для экспертов в диалектологии скуден, и всё становится ещё хуже, если речь идёт об экспертных системах для первичной экспертизы текста. И это не удивительно, так как для построения эффективной экспертной системы придётся использовать самые современные методы обработки естественных языков (NLP), связанные, в первую очередь, с машинным обучением, и, в частности, с нейронными сетями и трансферным обучением. Данная же работа посвящена созданию нейросетевой модели для русского языка.

Анализ существующих методов и подходов

Задачу определения диалекта и поиска диалектизмов можно отнести к области понимания естественного языка. Так, для задачи определения диалекта системе потребуется определить признаки текста, присущие определённому диалекту, то есть классифицировать текст.

Аналогичным образом можно поступить с задачей поиска диалектизмов. Для её решения потребуется выделить из текста всех потенциальных кандидатов и произвести их классификацию.

Существующие методы определения диалектов на основе языковых моделей

Большинство систем идентификации языка работают с речью [A. Etman, A. A. L. Beex, 2015; 220]. Так, при фонотактическом моделировании используют два популярных подхода распознавание звуков на основе языковой модели (PRLM) и параллельное распознавание звуков на основе языковой модели (PPRLM). В обоих подходах языковые модели оценивают речевую транскрипцию на принадлежность к определённому диалекту, то есть производят классификацию.

Задачу определения диалекта в тексте зачастую сводят к задаче определения языка. В работе [Jauhiainen T, Lindén K, Jauhiainen H., 2019; 561] рассматривается метод HeLI и HeLI 2.0. HeLI – это метод обучения с учителем, использующий языковые модели и схожий с методом наивного Байеса. В статье предлагался новый способ улучшения метода HeLI 2.0 с помощью адаптации языковой модели. Адаптация выполняется итеративно, путём пополнения словаря модели с наивысшей оценкой уверенности n-граммами из тестовых данных.

Языковые модели для задачи классификации

Самым быстроразвивающимся и многообещающим подходом обработки языков являются языковые модели, обучаемые на больших корпусах текстов [Mitchell, Melanie, David C. Krakauer, 2024]. Одной из таких моделей в открытом доступе является Google Bert. Данный трансформер по оценкам GLUE достигает, а местами и превосходит человеческие показатели. Bert способен решать разные задачи классификации текста [Sun, C., Qiu, X., Xu,

Y., Huang, X., 2019]: определение грамматической корректности (CoLa), определение истинности логического вывода (MNLI), определение парафразов (MRPC), определение ответа на вопрос (QNLI) и определение тождественных вопросов (QQP). Для адаптации модели под конкретную задачу классификации обычно используют дополнительный слой нейронной сети прямого распространения.

Модели нечёткой логики

Нечёткая логика является мощным инструментом реализации систем принятия решения на базе правил с 1970-х годов [Zimmermann, H-J., 2010; 317]. Заде в своей работе [Zadeh, 1973; 28] впервые предложил концепцию контроллера на базе логических правил.

Фактически финальный этап дефазификации можно рассматривать как классификацию предлагаемого решения на основе входных значений переменных и системы логических правил. Примером системы классификации на основе системы правил нечёткой логики является [McBratney, Alex B., Adrian W. Moore, 1985; 165]. В данной работе авторы пытались реализовать климатическую классификацию.

В основе контроллера нечёткой логики лежит механизм логического вывода. Самым популярным механизмом логического вывода является модель Мамдани [Iancu, Ion, 2012; 325]. Модель Мамдани проста для вычисления, так как использует функции минимума и максимума для агрегации правил. Всё это позволяет создавать большие системы правил для принятия решений.

Классификация диалектов

Основой нейросетевой модели системы классификации диалектов будет модель обработки естественных языков Google Bert модификации Large, обученной компанией SberDevices для своей сисемы «Салют» преимущественно на русском корпусе текстов [Д. Антюхов, 2020]. Таким образом, словарь модели составляет 120138 уникальных токенов.

Для задачи классификации понадобится вектор скрытого состояния токена [CLS] размерности 1024. Он пропускается через линейный слой нейронной сети (pooler) с tanh функцией активации. Данный слой использовался для задачи предсказания следующего предложения во время первичного обучения модели. Именно выход pooler-слоя размерностью 1024 и станет входом для последнего слоя нейронной сети (head) для тонкой настройки под задачу классификации диалектов. Данный слой тонкой настройки был обучен на подготовленных данных подробнее о которых будет рассказано в разделе 5

В качестве финального слоя классификатора нейронной сети будет однослойная нейронная сеть прямого распространения размерностью (1024,3) с функцией активации Relu. Так как требуется провести классификацию по трём диалектам (северный, среднерусский и южный) будет использовала функция потерь кросс-энтропии. В качестве метода оптимизации был выбран Adam (Adaptive moment estimation).

Однако учитывая небольшую выборку обучающих примеров для классификации диалектов, велика вероятность переобучения. Поэтому было принято решение использовать метод случайного исключения активных нейронов (dropout). Итоговая архитектура модели классификации изображена на рисунке 1. Экспериментальным образом были подобраны параметры обучения слоя классификации: скорость обучения - 10^{-6}; количество эпох – 5; вероятность исключения нейрона – 0,5; размер пакета обучения в рамках итерации - 2. Результаты обучения можно увидеть на графиках рисунков 2 и 3.

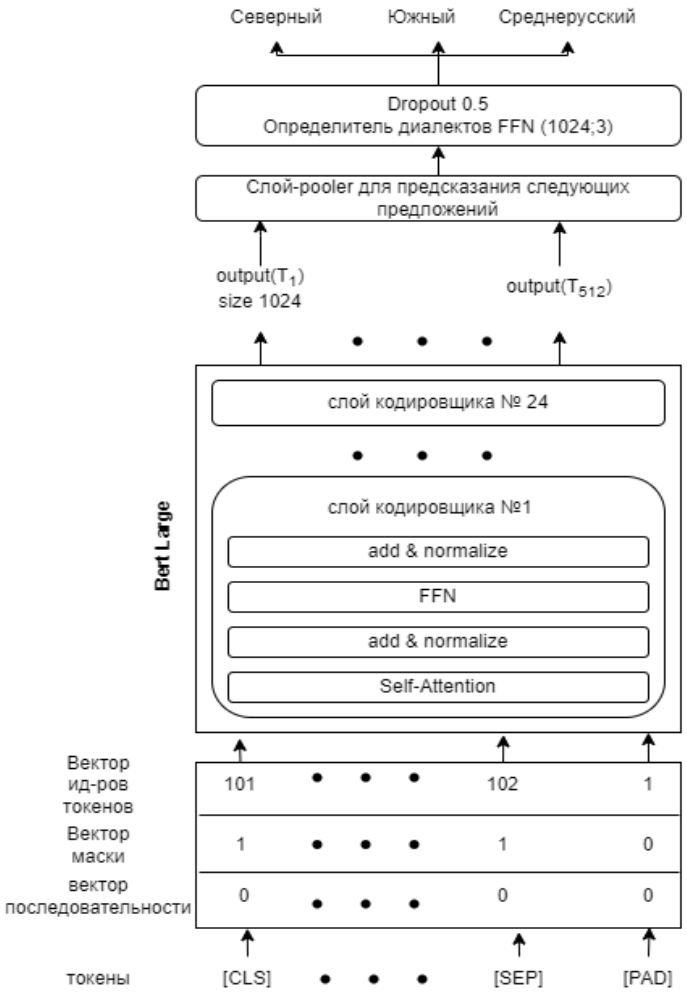


Рисунок 1. Архитектура нейросетевой модели классификации диалектов

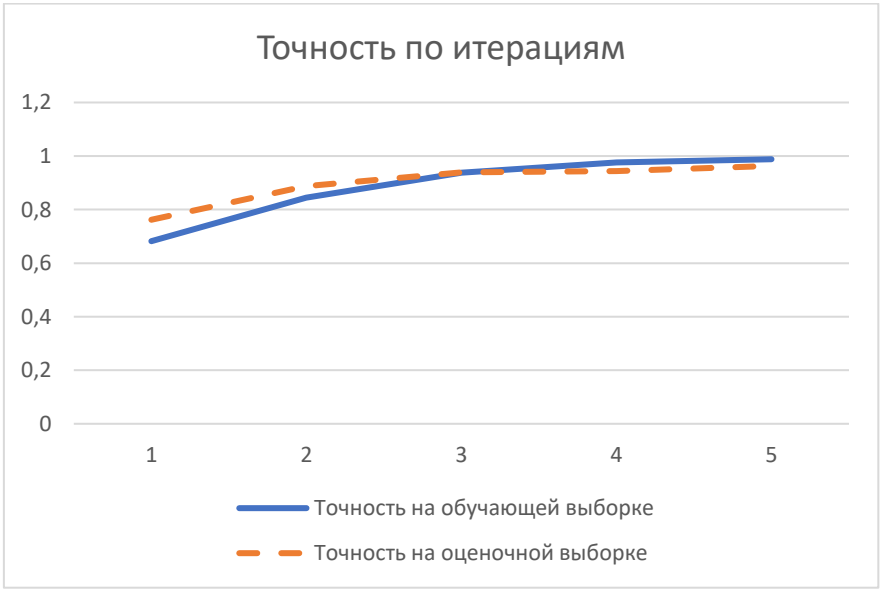


Рисунок 2. Диаграмма точности классификации диалекта во время обучения

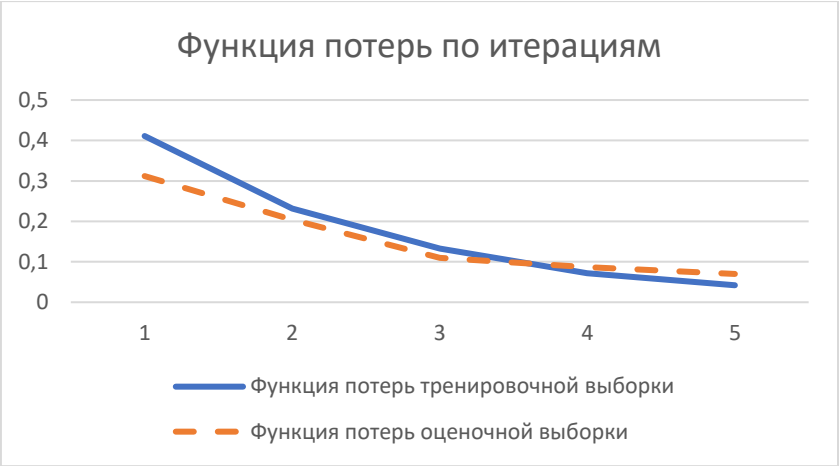


Рисунок 3. Диаграмма функции потерь во время обучения

Определение диалектизмов

Ключевой проблемой обучения нейросетевой модели для поиска диалектизмов является недостаточное количество данных о диалектизмах в контексте. Поэтому было принято решение использовать гибридный подход, основанный на нечёткой логике и метриках модели Bert.

Нечёткая логика позволяет строить системы правил, основанные на нечётких переменных и их функциях принадлежности к множеству. В качестве ответа такой системы выступает не явное решение задачи классификации, а вектор действительных чисел степени принадлежности к тому или иному множеству классифицируемых объектов.

$$F_t(x,a,b,c,d) = \begin{cases} 0, \text{ если } x < a \\ \frac{x-a}{b-a}, \text{если } a \le x \le b \\ 1, \text{ если } b \le x \le c \\ \frac{d-x}{d-c}, \text{если c} \le x \le d \\ 0, \text{ если } x > d \end{cases} \quad (1)$$

Было принято решение использовать кусочно-линейную трапециевидную функцию, так как данная функция имеет простое построение и идеально подходит для прототипирования системы правил. Трапециевидная функция (1) задаётся четырьмя параметрами: a, b, c, d. Параметры функции подбирались экспериментально. Остальные классы функций подходят для более тонкой настройки системы экспертами предметных областей. Их стоит использовать, если известно точное распределение классифицируемых объектов.

Для построения системы правил были сформулированы 4 лингвистические переменные, которые будут далее описаны набором (β, T, X), где β – название лингвистический переменной, T – набор термов или нечётких переменных, X – область допустимых значений (универсальное множество): частота употребления в контексте (β_1); относительная частота появления в диалекте i ($\beta_{2i}\ \ i = \overline{1..3}$); косинусное сходство вложения текста со средним вектором вложения диалекта i ($\beta_{3i}\ \ i = \overline{1..3}$); принадлежность слова к диалекту β_{4i}.

Лингвистическая переменная «Частота употребления в контексте»

Данная лингвистическая переменная будет обозначаться в системе правил как β_1. Смысл данной переменной заключается в том, чтобы описать шанс появления в контексте предложения через предсказание замаскированного слова с помощью модели Bert. Финальный слой нейросети производит моделирование языка, то есть производит классификацию замаскированного слова по токенам словаря. Таким образом, выход нейроной сети обрабатывают функцией softmax, чтобы смоделировать вероятности появления токена.

Тем не менее, использование вероятностей в явном виде приводит к оперированию маленькими действительными числами. Более того нижняя граница вероятностной оценки модели Bert, когда слово считается «не использующимся в контексте» будет зависеть от размера словаря, что крайне важно для функции принадлежности данной нечёткой переменной. Чтобы компенсировать это потребуется провести нормализацию относительно максимального элемента: $normalize(Output)_i = \frac{Output_i - \min(Output)}{\max(Output) - \min(Output)}$.

Однако важно помнить, слово может состоять из нескольких токенов. В таком случае определим частоту встречи слова ($freq_w$) в контексте как среднее гармоническое от всех частот токенов ($freq_{t_i},\ t_i \in T_w, i = \overline{1..n}$) этого слова вычисляется по формуле: $freq_w = \frac{n}{\sum_{i=1}^{n} \frac{1}{freq_{t_i}}}$.

Таким образом, область допустимых значений лексической переменной «Частота употребления в контексте» - $X \in [0; 1]$. Данная область разделяется между следующими термами (нечёткими переменными) с трапециевидными функциями принадлежности: «Не использующееся в контексте» α_{11}, f_t с параметрами: $a = 0,\ b = 0,\ c = 0,\ d = 5 * 10^{-10}$; «Редко использующееся в контексте» α_{12}, f_t с параметрами: $a = 4 * 10^{-10},\ b = 0{,}1,\ c = 0{,}2,\ d = 0{,}35$; «Часто используется в контексте» α_{13}, f_t с параметрами: $a = 0{,}3,\ b = 0{,}8,\ c = 1,\ d = 1$.

Лингвистическая переменная «Относительная частота появления в диалекте»

Лингвистическая переменная «Относительная частота появления в диалекте» считается для каждого диалекта и обозначается в системе правил как $\beta_{2i}, i = \overline{1..3}$. Смысл данной лингвистической переменной описать насколько часто диалектизм встречается в диалекте i относительно общего числа встреч во всех диалектах. Тогда формула вычисления метрики: $relativeFreq_i(w) = \begin{cases} \frac{count_i(w)}{\sum_{k=1}^{3} count_i(W)}, \text{если } \sum_{k=1}^{3} count_i(W) > 0 \\ 0, \text{если } \sum_{k=1}^{3} count_i(W) = 0 \end{cases}$,

где $count_i(w)$ – зафиксированное количество классификаций слова w как диалектизма диалекта i. В случае если у системы нет информации о слове w как о диалектизме: $count_i(w) = 0$. Данная статистика будет собираться при обработке обучающей и тренировочной выборке. Так же система будет давать возможность экспертам внести новые данные для обновления статистики диалектизмов. Всё это поможет экспертам напрямую влиять на результаты работы алгоритма, построенного на нечёткой логике.

Исходя из формулы метрики очевидно, что область допустимых значений лингвистической переменной «Относительная частота появления в диалекте» - $X \in [0; 1]$. Данная переменная будет иметь следующие термы с трапециевидными функциями

принадлежности на своём универсуме: «Не используется в диалекте i» α_{21i}, f_t с параметрами: $a = 0,\ b = 0,\ c = 0,\ d = 0{,}05$; «Используется в диалекте i» α_{22i}, f_t с параметрами: $a = 0,\ b = 0{,}2,\ c = 0{,}3,\ d = 0{,}5$; «Часто используется в диалекте i» α_{23i}, f_t с параметрами: $a = 0{,}4,\ b = 0{,}5,\ c = 1,\ d = 1$.

Лингвистическая переменная «Косинусное сходство вложения текста со средним вектором вложения диалекта»

Лингвистическая переменная «Косинусное сходство вложения текста со средним вектором вложения диалекта» считается для каждого диалекта так же, как и «Относительная частота появления в диалекте». В системе нечётких правил данная переменная будет обозначаться как β_{3i}.

Метрика данной лингвистической переменной считается по формуле косинусного сходства векторов: $simularity(D^i, S) = \frac{D^i * S}{\|D^i\| * \|S\|} = \frac{\sum_{j=1}^{n} D_j^i * S_j}{\sum_{j=1}^{n} \sqrt[2]{D_j^i} * \sum_{j=1}^{n} \sqrt[2]{S_j}}$, где D_i – средний вектор вложений предложений, относящихся к диалекту i; S – вектор вложения предложения, которому принадлежит замаскированное слово. Первым важным примечанием является то, что вложение формируется по исходному предложению до использования процедуры маскирования, чтобы вложение содержало полную информацию о контексте. Второе примечание связано с тем как именно формируется вложение предложения для формирования среднего вложения по диалекту: $D^i = \frac{\sum_{j=1}^{n} S_j^i}{n}$, где S_j^i – вложение j предложения в обучающих данных принадлежащее к i диалекту, а n – общее число предложений обучающей выборки, принадлежащих диалекту i) и участия в вычислении косинусного сходства (S).

Для подсчёта метрики косинусного сходства будет использоваться подход взвешенного среднего. Роль весов играет вектор маски токенов, который является частью входа модели Bert. Таким образом, формула вложения предложения будет иметь вид:

$s_i = \frac{\sum_{j=1}^{512} Output_{ji} * M_j}{\sum_{j=1}^{512} M_j}$, где $i = \overline{1..1024}$; M – вектор масок токенов, состоящий из 0 и 1; Output – матрица выхода модели Bert.

Так как данная метрика ограничена значениями косинуса, область значения лингвистической переменной «Косинусное сходство вложения текста со средним вектором вложения диалекта» - $X \in [0; 1]$. Данная переменная будет иметь следующие термы с трапециевидными функциями принадлежности на своём универсуме: «Отлично от среднего вложения диалекта i» α_{31i}, f_t с параметрами: $a = 0,\ b = 0,\ c = 0,\ d = 0{,}3$; «Незначительное сходство со средним вложением диалекта i» α_{32i}, f_t с параметрами: $a = 0{,}25,\ b = 0{,}4,\ c = 0{,}45,\ d = 0{,}55$; «Схоже со средним вложением диалекта i» α_{33i}, f_t с параметрами: $a = 0{,}5,\ b = 0{,}8,\ c = 1,\ d = 1$.

Лингвистическая переменная «Принадлежность слова к диалектам»

Лингвистическая переменная «Принадлежность к диалектам» обозначается γ. Данная переменная служит логическим выводом системы правил нечёткой логики модели Мамдани. Универсум данной переменной - $X \in [0; 1]$. Универсум разделяется следующими нечёткими переменными: «Не принадлежит к диалектам» λ_1, f_t с параметрами: $a = 0,\ b = 0,\ c = 0{,}2,\ d = 0{,}3$; «Возможно принадлежит к диалектам» λ_2, f_t с параметрами: $a = 0{,}2,$

$b = 0{,}4,\ c = 0{,}5,\ d = 0{,}7$; «Принадлежит к диалектам» λ_3, f_t c параметрами: $a = 0{,}6,\ b = 0{,}8,\ c = 1,\ d = 1$.

Система правил логического вывода для модели Мамдани

Основой логического вывода в нечёткой логике является система правил вывода. Причём правила составления данной системы правил зависит от самого механизма нечёткой логики. В данной работе будет использована модель Мамдани.

Логика построения системы правил базируется на нескольких предположениях:

1) Приоритет имеет переменная «Относительная частота появления в диалекте» (β_{2i}), так как это точная статистическая метрика. Более того, эксперты смогут добавлять в системы новые диалектизмы, тем самым напрямую влияя на статистику слова, а, следовательно, и на логический вывод модели Мамдани.
2) Если слово не встречается в диалекте, то упор идёт на метрики полученные с помощью модели Bert (β_1 и β_{3i}), которые безусловно являются менее надёжными, но позволяют сделать логический вывод в ситуации неизвестности. Причём переменная «Частота употребления в контексте» (β_1) имеет приоритет над переменной «Косинусное сходство вложения текста со средним вектором вложения диалекта» (β_{3i}), так как задача предсказания является первичной для модели Bert.
3) В случае встречи слова, которое предсказывается моделью Bert с крайне низкой вероятностью или токена которого вовсе нет в словаре модели, и при этом отсутствует статистика по встречаемости данного слова в диалектах, и вектор вложения предложения отличается от среднего вектора вложений диалектов, тогда система находится в ситуации полной неизвестности. Было принято решения трактовать подобное слово как потенциальный диалектизм (λ_2).

Согласно предположениям, была сформирована следующая система правил для механизма вывода нечёткой логики:

$$i = \overline{1..3}$$
$$\beta_1 \in \alpha_{11} \cap \beta_{2i} \in \alpha_{21i} \cap \beta_{3i} \in \alpha_{31i} \to \gamma \in \lambda_1$$
$$\beta_1 \in \alpha_{11} \cap \beta_{2i} \in \alpha_{21i} \cap \beta_{3i} \in \alpha_{32i} \to \gamma \in \lambda_2$$
$$\beta_1 \in \alpha_{11} \cap \beta_{2i} \in \alpha_{21i} \cap \beta_{3i} \in \alpha_{33i} \to \gamma \in \lambda_3$$
$$\beta_1 \in \alpha_{11} \cap \beta_{2i} \in \alpha_{22i} \cap \beta_{3i} \in \alpha_{31i} \to \gamma \in \lambda_2$$
$$\beta_1 \in \alpha_{11} \cap \beta_{2i} \in \alpha_{22i} \cap \beta_{3i} \in \alpha_{32i} \to \gamma \in \lambda_3$$
$$\beta_1 \in \alpha_{11} \cap \beta_{2i} \in \alpha_{22i} \cap \beta_{3i} \in \alpha_{33i} \to \gamma \in \lambda_3$$
$$\beta_{2i} \in \alpha_{23i} \to \gamma \in \lambda_3$$
$$\beta_1 \in \alpha_{12} \cap \beta_{2i} \in \alpha_{21i} \to \gamma \in \lambda_1$$
$$\beta_1 \in \alpha_{12} \cap \beta_{2i} \in \alpha_{22i} \to \gamma \in \lambda_3$$
$$\beta_1 \in \alpha_{13} \cap \beta_{2i} \in \alpha_{21i} \to \gamma \in \lambda_1$$
$$\beta_1 \in \alpha_{13} \cap \beta_{2i} \in \alpha_{22i} \to \gamma \in \lambda_4$$
$$\beta_1 \in \alpha_{11} \cap \beta_{21} \in \alpha_{211} \cap \beta_{31} \in \alpha_{311} \cap \beta_{22} \in \alpha_{212} \cap \beta_{32} \in \alpha_{312} \cap \beta_{23} \in \alpha_{213} \cap \beta_{33} \in \alpha_{313} \to \gamma \in \lambda_2$$

Набор данных

Набор данных для обучения и тестирования нейросетевой модели был взят из диалектического корпуса НКРЯ. Доступный набор данных был не сбалансирован (рисунок 4) и состоял из 2142 примерах. Для поиска диалектизмов использовался поиск по тегам

dialex, dialtem, dialsfx и dialform. Первые три записи загруженных данных можно увидеть в таблице 1.

В первоначальных данных нет сведений о диалектах, однако колонка «Title» помимо названия содержит сведения о месте публикации или нахождения текста. Соответственно, с помощью регулярных выражений из колонки «Title» было выделено название субъекта РФ. Далее, используя диалектологическую карту 1965 [Захарова, К. Федоровна, В. Г. Орлова, 2004], можно примерно отнести субъекты РФ по географическому расположению к тому или иному диалекту.

Всего загруженная выборка содержала данные по 7 субъектам РФ, которые были разбиты по 3 диалектам следующим образом:

1) Северный диалект: Вологодская область, Архангельская область;
2) Среднерусский диалект: Псковская область, Пермский край;
3) Южный диалект: Смоленская область, Ставропольский край, Самарская область.

Предобработанные данные были разделены на обучающую, оценочную и тестовую в соотношении 8:1:1. Оценочная выборка была выделена для проверки переобучения во время эпох обучения.

index	Full context	Center	Title
0	развяжут это фсе, повернёш на другу сторону,	другу	О выделке льна (деревня Усть-Поча, Плесецкий р...
1	высохнё,	высохнё	О выделке льна (деревня Усть-Поча, Плесецкий р...
2	потом начьнут в овине или на пецьке, ну в ови...	поставя	О выделке льна (деревня Усть-Поча, Плесецкий р...
3	И мялка была така, на ношках,	така	О выделке льна (деревня Усть-Поча, Плесецкий р...

Таблица 1. Структура данных загруженных из НКРЯ

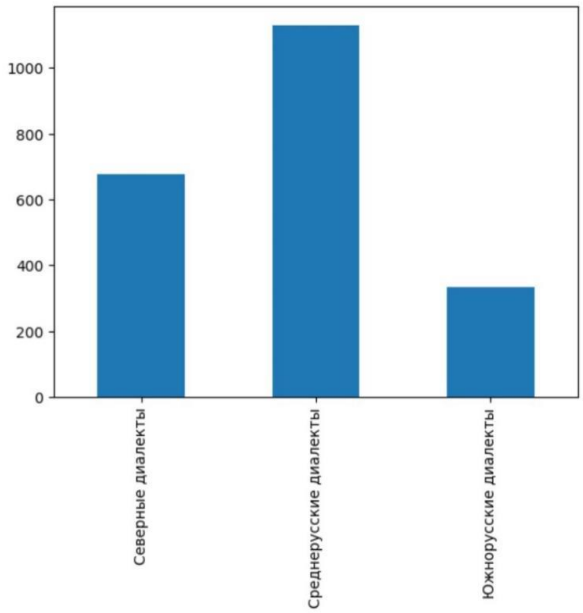


Рисунок 4. Распределение диалектов в обработанном наборе данных.

Оценка результатов

Настроенная на задачу классификации диалектов модель Bert показала точность 95,3% на тестовой выборке, состоящей из 215 примеров. Полученный результат нельзя назвать объективным из-за малого набора данных, что зачастую приводит к переобучению. Для более точной оценки эффективности модели в будущем потребуется расширить тестовый набор данных.

Вторым этапом тестирования является оценка эффективности определения диалектизмов. Для проведения тестирования было принято решения использовать весь набор пред обработанных данных, который был использован для обучения и тестирования тонкой настройки модели Bert под задачу классификации диалектов.

Из каждой записи набора данных (таблица 1) составлялись тестовые случаи. Если текст записи состоял только из диалектизма, то составлялся один позитивный пример. Если был предоставлен полноценный контекст диалектизма, то составлялся позитивный и негативный пример. Так, положительный случай состоит из диалектизма из колонки «Center» и текста («Full context»), негативный - из случайного слова из контекста, исключая диалектизм, и текста. Проводился строгий подсчёт верных классификаций, поэтому класс «Возможно принадлежит к диалектам» считался заведомо не верным. Точность классификации (accuracy) на 4279 тестовых примерах составила 0,4162.

На полученную точность напрямую влияют подобранные параметры функции принадлежности к множеству. Можно попробовать улучшить результат проведя больше экспериментов с ними.

Вторым важным фактором является наличие класса «Возможно принадлежит к диалектам», который ухудшает результат строгой классификации, но помогает в качестве рекомендаций выделять слова из текста.

Заключение

Результатом данной работы является экспериментальная реализация системы классификации диалектов и определения диалектизмов. Созданная система способна классифицировать диалекты русского языка введённого текста и в рекомендательном порядке показывать список найденных диалектизмов.

Экспериментальность данной системы заключается в гибридном подходе к определению диалектизмов, где система правил нечёткой логики опирается на статистические метрики и метрики, полученные с помощью языковых моделей.

Преимуществом предложенной архитектуры системы является адаптируемость под другой язык, в частности, полученные наработки могут быть использованы в системе диалектологического атласа тюркских языков, разрабатываемой в Институте прикладной семиотики АН РТ.

В дальнейшем планируется более точно настроить функции принадлежности термов лингвистических переменных для системы правил нечёткой логики. Для этого потребуется много экспериментов и помощь экспертов в данной сфере. Так же планируется дополнить набор данных для обучения и тестирования.

Использованная литература:

1. A. Etman and A. A. L. Beex, Language and Dialect Identification: A survey, 2015 SAI Intelligent Systems Conference (IntelliSys), London, UK, 2015, pp. 220-231, https://doi.org/10.1109/IntelliSys.2015.7361147
2. Iancu, Ion. A Mamdani type fuzzy logic controller. Fuzzy logic-controls, concepts, theories and applications 15.2 (2012): 325-350.
3. Jauhiainen T, Lindén K, Jauhiainen H. Language model adaptation for language and dialect identification of text. Natural Language Engineering. 2019;25(5):561-583. https://doi.org/10.1017/S135132491900038X
4. McBratney, Alex B., and Adrian W. Moore. Application of fuzzy sets to climatic classification. Agricultural and forest meteorology 35.1-4 (1985): 165-185.
5. Mitchell, Melanie, and David C. Krakauer. The debate over understanding in AI's large language models. Proceedings of the National Academy of Sciences 120.13 (2023): e2215907120.
6. Sun, C., Qiu, X., Xu, Y., Huang, X. (2019). How to Fine-Tune BERT for Text Classification? In: Sun, M., Huang, X., Ji, H., Liu, Z., Liu, Y. (eds) Chinese Computational Linguistics. CCL 2019. Lecture Notes in Computer Science (), vol 11856. Springer, Cham. https://doi.org/10.1007/978-3-030-32381-3_16
7. Zadeh LA. Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans Syst Man Cybernet 1973, 3:28–44.
8. Zimmermann, H-J. Fuzzy set theory. Wiley interdisciplinary reviews: computational statistics 2.3 (2010): 317-332.
9. Д. Антюхов. Обучение модели естественного языка с BERT, блог компании SberDevices, 2020. Habr. https://habr.com/ru/company/sberdevices/blog/527576/
10. Захарова, Капитолина Федоровна, and Варвара Георгиевна Орлова. Диалектное членение русского языка. УРСС, 2004.

NATURAL LANGUAGE PROCESSING AND ITS MAIN APPROACHES

Suyunova Mohinur
PhD student of National University of Uzbekistan.

Abstract: The article explores the relevance of NLP (or Computational Linguistics) to language learning, namely written and spoken communication. In addition, it further focuses on characterizing the techniques of NLP and uses of it for language learning. Together with presenting the background necessary to explain main principles of NLP, the paper also investigates the use of NLP in detecting language faults by students when they are learning the language as a second language.

Keywords: Statistical approaches, classical approaches, deep learning approaches, intelligent language tutoring systems, NLP algorithms, mal-rule technique, meta-rule

Аннотация: В статье исследуется значимость НЛП (или компьютерной лингвистики) для изучения языка, а именно письменного и устного общения. Кроме того, основное внимание уделяется характеристике техник НЛП и их использованию для изучения языка. Наряду с представлением исходной информации, необходимой для объяснения основных принципов НЛП, в статье также исследуется использование НЛП для выявления языковых ошибок у студентов, когда они изучают язык как второй язык.

Ключевые слова: статистические подходы, классические подходы, подходы глубокого обучения, интеллектуальные системы обучения языку, алгоритмы НЛП, техника неправильных правил, метаправило.

The automatic processing of human languages is known as natural language processing, or NLP (also known as computational linguistics). Since NLP is a broad, multidisciplinary field that is still relatively young, there are numerous definitions in use that are used by various practitioners. Any competent individual would include the following definition in their list:

In order to achieve human-like language processing for a variety of activities or applications, a theoretically motivated spectrum of computing techniques known as "natural language processing" is used to analyze and represent naturally occurring texts at one or more levels of linguistic analysis.

NLP, then, is a branch of computer science and linguistics that studies how computers and human (natural) languages interact. Furthermore, it is heavily influenced by developments in artificial intelligence (AI) and machine learning (ML). The natural language processing (NLP) approaches are designed to enable computers to comprehend and respond to commands in natural language. It should be mentioned that spoken and written language include the two categories of natural language processing.

Speech is the primary mode of human communication; hence written languages serve a secondary role in most activities. Written language is easier to understand than spoken language, which often contains noise and ambiguity. Natural language processing (NLP) is considered a challenging computer science topic due of its ambiguity [Abdurakhmonova, N. 2019,2021,2022,2023].

Natural language processing research dates back to the late 1940s. Machine translation (MT) was one of the first computer-based applications for natural language. According to Cambria and White (2014), NLP research has progressed from the time of punch cards and batch processing, when analyzing a sentence may take up to 7 minutes, to the present day, where millions of webpages can be analyzed in under a second.

The 'levels of language' technique is the most informative way to convey what actually happens within a Natural Language Processing system. People employ these levels to obtain meaning from written or spoken words. This is because language processing is mostly based on formal models or representations of knowledge relevant to various levels. Furthermore, language processing applications differ from data processing systems in that they make use of linguistic expertise. Natural language processing analysis is divided into the following levels: phonology, morphology, lexicon, syntactic, semantic, discourse, and pragmatics. The appendix contains the meanings of all levels.

There are several applications available that include both theory and implementations. Natural language processing can be applied to any text-based application. The appendix includes an overview of the most commonly used NLP applications. The majority of these challenges can be codified into five key tasks: classification, matching, translation, structured prediction, and sequential decision making.

According to Liddy (2001), natural language processing techniques can be classified into four categories: Symbolic approaches conduct in-depth analyses of linguistic phenomena and are based on the explicit description of language facts using well-understood knowledge representation schemes and algorithms. Statistical approaches use mathematical tools and vast text corpora to create approximate models of linguistic processes. These models are based on actual examples and do not require additional linguistic or world information.

Connectionist and statistical techniques both construct broad models of linguistic phenomena. Connectionism, also known as "parallel distributed processing," "neural networks," or "neuro-computing," differs from other statistical methods in that it combines statistical learning with representational theories, allowing for transformation, inference, and manipulation of logic formulae.

A rising number of researchers use a combination of data-driven and knowledge-driven methodologies, known as hybrid approaches. The following demonstrates that the approaches have both similarities and differences. For example, each technique has distinct assumptions, philosophical grounds, and sources of proof. There are two sorts of text classification methods: classical and deep learning.

Classical approaches rely on manual feature engineering and rules in combination with statistical algorithms. The manually designing features of data instances into feature vectors can be done in several ways. Studies have shown that the most effective surface features in hate speech detection are bag of words, word and character n-grams. In terms of classifiers, the most popular algorithm used, is the Support Vector Machine. Algorithms like Naive Bayes, Logistic Regression and Random Forest are also used for classification task [Abdurakhmonova, N. 2019,2021,2022,2023].

Deep learning approaches use neural networks to automatically learn several layers of features from input data. By the early 2000s, developments in computer hardware and optimization and training techniques have enabled the construction of ever larger and deeper networks, giving

rise to the modern term deep learning. Traditionally, NLP techniques rely on linear models trained on sparse feature vectors. However, non-linear neural networks with dense inputs have recently demonstrated success. The most commonly utilized networks are Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), sometimes known as Long Short-Term Memory networks (LSTM).

Intelligent language tutoring systems (ILTSs) use natural language processing (NLP) to provide personalized feedback to learners through workbook-style exercises. Examples include the E-Tutor (Heift, 2010), Robo-Sensei (Nagata, 2009), TAGARELA (Amaral & Meurers, 2011), i-tutor (Choi, 2016), and FeedBook (Meurers et al., in press). NLP analysis can help optimize material sequencing and update learner models (Schulze, 2011). The analysis typically focuses on linguistic faults committed by the learner, although feedback can also highlight correctly employed forms, meaning, or appropriateness of a learner's response to an exercise.

What drives the adoption of NLP in tutoring systems? To provide feedback and track students' abilities in a learner model, an ILTS must first gather information about their abilities. The approach depends on the activities and learner reactions, including language production and system interaction. Language assessment (Bachman & Palmer, 1996) and task-based language teaching (Ellis, 2009) both emphasize the importance of the interaction between the activity and the learner's response. It is also crucial for determining the system analysis requirements of various activity types (Quixal & Meurers, 2016).

For many types of language learning activities, however, extensionally describing a direct and complete mapping between potential learner input and desired feedback is not possible. Nagata (2009, pp. 563-4) illustrates this with an exercise from her Japanese instructor ROBO-SENSEI. The learner reads a short communicative context and is instructed to write a sentence in Japanese based on the system's English translation. The learner response in this activity is directly dependent on the input provided by the exercise (a direct response in the framework of Bachman & Palmer, 1996), so a short, seven-word statement can be defined as the goal response.

After considering lexical, orthographic, and word order alternatives, there are 6,048 correct phrases that the learner can enter. Even limiting ill-formed patterns to erroneous particle and conjugation choices results in about a million phrases [Abdurakhmonova, N. 2019,2021,2022,2023]. Explicitly establishing a mapping between a million predicted replies and their related feedback is clearly impossible. Note that the explosion of viable learner answers depicted by Nagata is already a problem for direct response in a limited activity, where the meaning to be represented was fixed and only form variety was expected. Beyond analyzing language as a system (e.g., system-referenced tests in language assessment; Baker, 1989), performance-referenced analysis examines the ability to use language appropriately to complete a task.

To analyze learner input for various language activities that require significant variation in form, meaning, or task-appropriateness, NLP algorithms and resources should be used to abstract away from specific strings to more general properties. NLP analysis can provide feedback, update learner models, and sequence instruction based on a small set of language features and categories, rather than the enormous number of string instances.

Errors in native language texts have led to a shift from theory-driven, rule-based NLP to data-driven, statistical, and machine-learning approaches in the 1980s and 1990s. This shift is driven by the need to develop robust NLP that works in suboptimal conditions, such as noise and unknown forms and patterns. The purpose of using NLP in ILTS differs from other NLP domains.

NLP is designed to handle errors and unexpected input to produce results, such as syntactic analysis from a parser or translation from a machine translation system.

The basic purpose of NLP in ILTS is to provide feedback to learners by identifying language and response characteristics that differ from intended targets. The NLP abstraction aims to characterize faults, rather than ignoring them due to processing robustness [Abdurakhmonova, N. 2019,2021,2022,2023].

Writers' aids, such as standard spell and grammar checkers (Dickinson, 2006), share the ILTS focus on error detection, but they are based on assumptions about common errors committed by native speakers, which do not apply to language learners. Rimrott and Heift (2008) found that many L2 misspellings are multiple-edit errors and cannot be repaired by spell checkers meant for native writers. Tschichold (1999) also points out that typical writers' aids are not always useful for language learners because they require more scaffolding than a list of options from which to pick.

Language licensing systems rely on formal grammars of the language to be licensed, which can be articulated in one of two ways (Johnson, 1994). In a validity-based approach, recognizing a string involves identifying valid derivations from a collection of rules known as a grammar. Adding more rules to the language allows for licensing of more string kinds, whereas removing rules prevents licensing of any type. In a satisfiability-based setup, a grammar is a set of constraints, and a string is considered licensed if its model meets all of the constraints in the grammar.

There are two methods for examining a string to diagnose learner faults, similar to the two forms of formal grammars. The mal-rule technique takes the validity-based approach and using typical parsing algorithms. Starting with a conventional native language grammar, rules are added to license strings used by language learners but not in the native language. These rules are known as mal-rules and are used to license learner errors (Sleeman, 1982; Matthews, 1992). Meta-rules can capture generalizations across rules, such as errors in subject-verb agreement, which can occur in several rules involving subjects and finite verbs (Weischedel & Sondheimer, 1983). The mal-rule strategy is effective when errors correspond to the local tree of a single grammar rule. Otherwise, additional rules must be considered, making it substantially more difficult to detect errors and govern the interaction of mal-rules with ordinary rules. To restrict the search space created by rule interaction, the use of mal-rules might be limited. In basic cases, mal-rules are created after ordinary grammar analysis fails. However, this just limits the search space for well-formed strings. If parsing fails, it remains unclear which mal-rules should be added [Abdurakhmonova, N. 2019,2021,2022,2023].

Recent research on grammatical error correction (GEC) focuses on recognizing and repairing faults in written language as a translation problem, as opposed to statistical approaches that capture insights based on rules. The goal is to use statistical machine translation (Junczys-Dowmunt & Grundkiewicz, 2016) or neural network approaches (Chollampatt & Ng, 2018) to convert ill-formed text to well-formed text. However, these methods have limited relevance for research and applications that rely on linguistic rules and conceptualization.

Constraint relaxation, also known as the second group of language licensing approaches (Kwasny & Sondheimer, 1981), is an option for satisfiability-based grammars or rule-based grammars with complex categories that can be relaxed. Boyd (2012) suggests that a conflict detection system can diagnose learner errors when parsing is seen as a broad constraint satisfaction

problem. Constraint relaxation removes certain requirements from grammar, such as enforcing agreement. This allows for greater string licensing.

This presupposes that an error may be mapped to a specific constraint to be relaxed—that is, the domains of the learner error and the constraint in the grammar must be tightly related. Constraints can be associated with weights to control the likelihood of analysis (Foth, Menzel, & Schröder, 2005). This raises the question of how flexible control can be informed by the ranking of errors likely to occur for a specific learner given a task. Other proposals incorporate constraint relaxation and parts of mal-rules [Abdurakhmonova, N. 2019,2021,2022,2023].

Reuer (2003) combines a constraint relaxation technique with a modified parsing algorithm to license strings with inserted or omitted words. This approach incorporates meta-rule-like generalizations into the parsing process.

NLP contributes to language acquisition by exposing learners to actual native language and providing possibilities for interaction. This area of research involves work on finding and improving authentic texts for learners to read, as well as the automatic development of exercises and examinations based on such materials.

What are the specific learner language properties that need to be identified? Historically, readability has been the primary factor in text selection, with several algorithms devised for this purpose (DuBay, 2004). Current machine learning methods, informed by a broader range of linguistic characteristics, are significantly more accurate than traditional measures based on shallow features such as average sentence and word lengths (Xia, Kochmar, & Briscoe, 2016; Crossley, Skalicky, Dascalu, McNamara, & Kyle, 2017; Weiss & Meurers, 2018), including some commercial systems (Nelson, Perfetti, Liben, & Liben, 2012).

Beyond readability, SLA research has revealed that awareness of language categories and forms is an important component of successful second language acquisition (see Lightbown & Spada, 1999), and a variety of linguistic properties, including morphological, syntactic, semantic, and pragmatic information, have been identified as relevant for language awareness (see Schmidt, 1995, p. 30). In answer to this demand, the FLAIR system (Chinkina & Meurers, 2016) provides linguistically aware Web searches, allowing language learners' input to be systematically enriched with the types of language patterns to be learned next.

In conclusion, although NLP has had limited impact on real-life language teaching and SLA, it offers numerous opportunities for developing applications to support language learning and research. Interdisciplinary collaboration between SLA and NLP is critical for establishing accurate annotation systems and analysis algorithms that uncover key properties for analyzing learner language.

References:

1. Chinkina, M., & Meurers, D. Linguistically aware information retrieval: Providing input enrichment for second language learners. In *Proceedings of the 11thWorkshop on Innovative Use of NLP for Building Educational Applications (BEA)* (pp. 188–98). San Diego, CA: Association for Computational Linguistics. Retrieved April 4, 2019 from https://www.semanticscholar.org/paper/Linguistically-Aware-Information-Retrieval%3A-Input-Chinkina-Meurers/d1e6e66b181b5912c101a19d345e56a0c5c28bba. ,2016.

2. Chollampatt, S., & Ng, H. T. A multilayer convolutional encoder–decoder neural network for grammatical error correction. In *Thirty-Second AAAI Conference on Arti!cial Intelligence (AAAI)* (pp. 5755–62). New Orleans, LA: AAAI, 2018.
3. DuBay, W. H. *The principles of readability*. CostaMesa, CA: Impact Information. Retrieved April 4, 2019 from http://www.impact-information.com/impactinfo/readability02.pdf, 2004.
4. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
5. Абдурахмонова, Н., & Бойсариева, С. (2023). TABIIY TILNI QAYTA ISHLASHDA (NLP) OKKAZIONALIZMLARNING MORFEM TAHLILI. *МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА*, *6*(3).
6. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In *2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)* (pp. 1790-1793). IEEE.
7. Mengliev, D. B., Abdurakhmonova, N., Hayitbayeva, D., & Barakhnin, V. B. (2023, November). Automating the transition from dialectal to literary forms in Uzbek language texts: an algorithmic perspective. In *2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE)* (pp. 1440-1443). IEEE.
8. San'atbek, M., Mirsaid, A., & Nilufar, A. (2018). Modeling WordNet type thesaurus for Uzbek language semantic dictionary. *International Journal of Systems Engineering*, *2*(1), 26.
9. Foth, K., Menzel, W., & Schröder. Robust parsing with weighted constraints. *Natural Language Engineering*, *11*(1), 1–25, 2005.
10. Johnson,M. Two ways of formalizing grammars. *Linguistics and Philosophy*, *17*(3), 221–48, 1994.
11. Lightbown, P. M., & Spada, N. *How languages are learned*. Oxford, England: Oxford University Press, 1999.
12. Reuer,V. Error recognition and feedback with lexical functional grammar. *CALICO Journal*, *20*(3), 497–512, 2003.
13. Schmidt, R. Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu: University of Hawai'i Press, 1995.

BIR SO‘Z TURKUMIDAGI OMONIM SO‘ZLARNI ANIQLASH USULLARI

Malikov Elbek
Kompyuter lingvistikasi yo’nalishi magistaranti
Toshkent davlat o’zbek tili va adabiyoti universiteti.

Annotatsiya: Tabiiy tilga ishlov berish sohasidagi muhim vazifalardan biri bu omonimiya masalasidir. Bu masala turli tabiiy tillarda mavjud bo‘lib, har bir tabiiy tilda turlicha usullar taklif qilinib kelinmoqda. Omonim so‘zlarni semantik farqlashda turli usullardan foydalanish mumkin. O‘zbek tilidagi omonim so‘zlarni semantic farqlashda ularni turli so‘z turkumlari va bir so‘z turkumi doirasida uchrashiga ko‘ra turli usullardan foydalanish mumkin. Ushbu maqolada aynan bir so‘z turkumi doirasidagi omonim so‘zlarni semantic farqlash usullari haqida so‘z olib boriladi. Bir so‘z turkumi doirasidagi omonimiyani aniqlashda Mashinali o‘qitish yondashuvining usullaridan foydalanishning jahon tajribalari o‘rganildi.

Kalit so‘zlar: tabiiy tilga ishlov berish (NLP natural language processing), omonimiya, Mashinali o‘qitish yondoshuvi, kontekst, so‘z ma’nosini aniqlash (WSD-word sense disambiguation)

Annotation: One of the important tasks in the field of natural language processing is the issue of homonymy. This problem exists in different natural languages, and different methods are proposed in each natural language. Different methods can be used for semantic differentiation of homonyms. In the semantic differentiation of homonymous words in the Uzbek language, different methods can be used depending on whether they meet in different word groups and within the same word group. This article discusses the methods of semantic differentiation of homonyms within the same word group. The world experiences of using the methods of the Machine Learning approach in determining homonymy within a word group were studied.

Keywords: NLP (natural language processing), homonymy, Machine learning approach, context, word sense disambiguation (WSD-word sense disambiguation)

Аннотация: Омонимия имеет важное значение в обработке естественного языка. Этот вопрос существует в разных естественных языках. Для семантической дифференциации омонимов можно использовать различные способы. При семантической дифференциации омонимов в узбекском языке могут применяться различные методы в зависимости от того, встречаются ли они в разных частях речи и в рамках одной части речи. В данной статье рассмотрены способы семантической дифференциации омонимов внутри одной части речи. Изучен мировой опыт использования методов подхода машинного обучения при определении омонимии внутри одной части речи.

Ключевые слова: обработка естественного языка (NLP), омонимия, подход машинного обучения, контекст, определение смысловой неоднозначности

Tabiiy tilda o‘qilishi bir xil, lekin turli kontesktlarda ma’nosi har xil bo‘lgan so‘zlar uchraydi. Bu so‘zlar omonim so‘zlar hisoblanadi. Inson omonim so‘zni oson farqlab, ma’nosini tushunadi. Lekin bu mashina uchun qiyin vazifa, uzoq yillardan buyon odamlarga o‘xshab

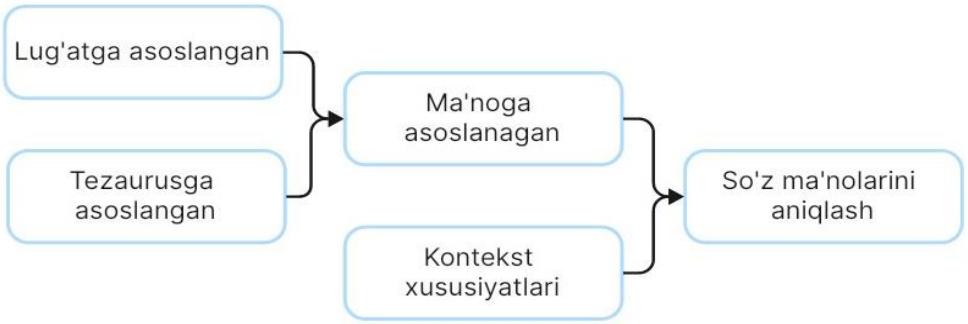
so‘zning ma’nosini avtomatik aniqlaydigan tizim ishlab chiqish masalasi o‘zining dolzarbligini saqlab kelmoqda.

So‘z manolarini aniqlash usullari omonim so‘zning to‘g‘ri manosini topish va uni to‘g‘ri o‘ringa avtomatik joylashtirishdir vazifasidir. So‘z manolarini aniqlash tabiiy tilni qayta ishlash sohasining muhim, murakkab masalasidir. Bu mashina tarjimasi, semantik xaritalash, semantik izohlash va ontologiyani o‘rganish kabi ko‘plab real dunyo ilovalari uchun zarur. Shuningdek, u ko‘plab axborot qidirish, ma'lumot olish va nutqni aniqlash kabi ilovalarning ish faoliyatini yaxshilashda foydali bo‘ladi. Ingliz [8], hind [9], fransuz [10], ispan [11], xitoy [12] va boshqa ko‘plab tabiiy tillar omonim so‘zlarga ega tillar hisoblanadi. O‘zbek tilidagi omonim so‘zlarni semantik farqlashda qator ilmiy izlanishlar olib borilmoqda. Omonimiyani aniqlashda ularni ikkita: *bir so‘z turkumi va turli so‘z turkumlari* doirasidagi omonim so‘zlarga ajratilgan holda aniqlash bo‘yicha ilmiy izlanishlar mavjud. Jumladan tadqiqotchi X.Axmedova hamda texnika fanlari falsafa doktori B.B.Elov tomonidan turli so‘z turkumlari doirasidagi omonim so‘zlarni so‘z turkumlari doirasida uchrashiga ko‘ra *ikki so‘z turkumi, uch so‘z turkumi va to‘rtta so‘z turkumi doirasidagi omonimlar* kabi guruhlarga ajratgan holda semantik farqlash bo‘yicha amaliy natijalarga erishgan [1], [2], [3], [4]. Turli so‘z turkumlari orasidagi omonimiyani aniqlash masalasi jumlalarni POS (Part of speechs tagging) teglash jarayonida o‘z yechimini topadi. So‘z turkumlarini aniqlashda Yashirin Markov modelidan foydalanish bo‘yicha ko‘plab ilmiy maqolalarni ko‘rish mumkin [5]. O‘zbek tilida shunday omonim so‘zlar borki, turli so‘z turkumlari ichida ham bir so‘z tukumiga oid bir nechta ma’nolarni anglatishi mumkin.

1- Jadval: Bir so‘z turkumi doirasidagi omonim so‘zlar

So‘z	**So‘z turkumi**	**Ma’nosi**
O‘t	Fe’l	O‘tmoq fe’li
	O‘t	Maysa, o‘t-o‘lan
	O‘t	Inson o‘rgani
	O‘t	Olov
Oz	Ravish	Kam, miqdori nisbatan ko‘p bo‘lmagan.
	Fe'l	Oriqlamoq, etidan yo‘qotmoq.
	Fe'l	Noxush bo‘lmoq, kuchsizlanmoq, holsizlanmoq.
	Fe'l	Adashmoq, to‘g‘ri yo‘ldan chetga chiqmoq
Bo‘y	Ot	Uzunlik o‘lchovi.
	Ot	Hid, is
…	…	…

1-jadvada keltirilgan *o‘t* omonim so‘zi turli so‘z turkumlari doirasida o‘ziga xos ma’nolarni anglatadi. Shuni alohida ta’kidlash kerakki, shu so‘z turkumlarining biri doirasida bir qancha ma’nolarni anglatuvchi so‘zlar ham talaygina. Bunday omonim so‘zlarni POS tegging masalasini yechish bilan aniqlab bo‘lmaydi. Buning uchun alohida usullar, modellar zatur. Faraz qilaylik, T matn W_1, W_2, W_3, ..., W_n so‘zlar ketma -ketligidan iborat. So‘z manolarini aniqlash usullari matndagi barcha omonim so‘zlar uchun to‘g‘ri ma'noni belgilash vazifasidir. So‘z manolarini aniqlash usullarining nazariy modeli 1-rasmda ko‘rsatilgan.



1-rasm. So‘z manolarini aniqlash usullarining nazariy modeli

So‘z manosini aniqlashda ishlatiladigan ikkita asosiy yondashuv - bu *Lug‘atga asoslangan yondashuv* va *Mashinali o‘qitishga asoslangan yondashuv*dir.
Ushbu yondashuvlar ham o‘z navbatida bir qator usullarni qamrab oladi.

Lug‘atga asoslangan yondashuvda target so‘zning barcha ma'nolari lug‘atdan olinadi. Keyin bu ma’nolar kontekstda qolgan barcha so‘zlarning lug‘at ta'riflari bilan taqqoslanadi. Biz So‘z manolarini aniqlash usullari yondashuvlarini so‘z ma'nosini ajratish uchun chuqur va sayoz yondashuv sifatida tasniflashimiz mumkin. Bilimga asoslangan yondashuv kompyuter o‘qishi mumkin bo‘lgan lug‘atlar, masalan, korpus WorldNet asosida ishlaydi. So‘z manosini aniqlash uchun grammatik qoidalardan ham foydalanishi mumkin.Bilimga asoslangan yondashuvning maqsadi (lug‘atga asoslangan yondashuv) kontekstdagi so‘zlarning ma'nosini aniqlash uchun bilim resurslaridan foydalanishdir. Bilim resurslari lug‘atlar, tezauriyalar, ontologiyalar, birikmalar va boshqalardir. Yuqoridagi usullar boshqariladigan muqobil usullarga qaraganda unumdorligi pastroq, lekin ular kengroq diapazondagi afzalliklarga ega**.**

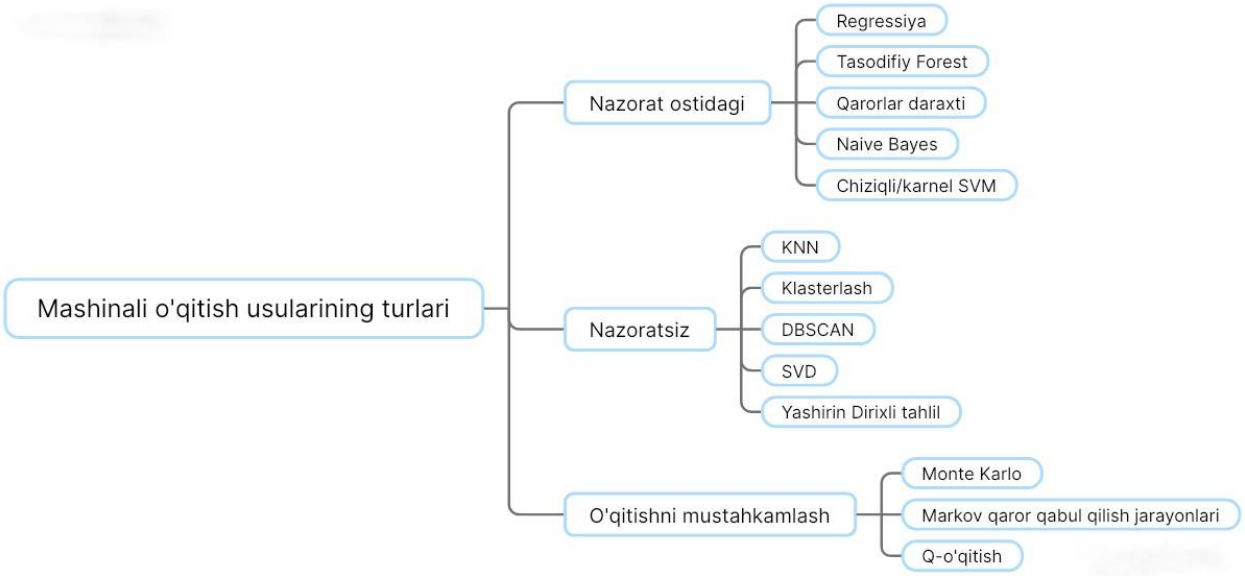
Bir-biriga o‘xshash yondashuvlar: Ushbu yondashuv mashinada o‘qiladigan lug‘at (MDR) qoidalarini talab qiladi. U kontekstdagi so‘zlarning xususiyatlari bilan bir qatorda noaniq bo‘lgan so‘zlarning ma’nolarining turli xususiyatlarini aniqlashni o‘z ichiga oladi.

Lesk algoritmi: Bir-biriga o‘xshash yondashuvga asoslangan Lesk algoritmini W so‘z ma'noni anglatuvchi so‘z, C - atrofdagi kontekstdagi so‘zlar to‘plami, S - W uchun ma’noni farqlaydi[7].

Tanlash afzalliklari: Tanlov afzalliklari so‘z turlarining mumkin bo‘lgan munosabatlari haqida ma'lumot topadi va bilim manbasidan foydalangan holda umumiy fikrni bildiradi. Masalan Son-yuz, inson a’zosi-yuz so‘zlari semantik munosabatga ega. Bunday yondashuvda noto‘g‘ri so‘z ma'nolari chiqarib tashlanadi va faqat umumiy ma'no qoidalariga mos keladigan hislar tanlanadi. Ushbu yondashuvning asosiy g‘oyasi sintaktik munosabatga ega bo‘lgan korpusda bunday so‘z juftligi necha marta sodir bo‘lishini hisoblashdir.

Mashinali o‘qitishga asoslangan yondashuvda tizimlar So‘z manolarini aniqlash usullari vazifasini bajarish uchun o‘qitiladi. Klassifikator kiruvchi gap tarkibidagi omonim so‘z(so‘zlar)ning joriy ma’nosi va xususiyatlarni aniqlash uchun foydalaniladi. Ushbu yondashuvda tizimlar *so‘z manolarini aniqlash usullari* vazifasini bajarish uchun o‘qitiladi. Klassifikatorla kiruvchi ma’lumotlarni ma’nosi va xususiyatlarini aniqlash uchun foydalaniladi. Xususiyatning qiymati so‘zning maqsadli so‘z atrofida necha marta sodir bo‘lishidir. Kontekst markazi sifatida tanlangan so‘z bilan doim birga paydo bo‘ladi. Mashinali o‘qitishga asoslangan

yondashuv usullari uch turga bo‘linadi: *nazorat qilinadigan usullar, nazoratsiz usullar va yarim nazorat qilinadigan usullar.*



2-rasm: Mashinali o‘qitish yondashuvi usullarining iyerarxiyasi

Nazorat ostidagi usullar. Bu usullarning mohiyati shundan iboratki, unda inson omili bilan semantik teglangan ma'lumotlar to‘plami asosida joriy so‘zning ma’nosini aniqlanadi.

Tasniflagichni o‘rganish uchun foydalaniladigan o‘quv majmuasi odatda ma'lum bir maqsadli so‘z va unga mos yozuvlar lug‘atining ma'no inventarizatsiyasidan olingan ma'no bilan qo‘lda teglangan misollar to‘plamini o‘z ichiga oladi. Asosan, So‘z manolarini aniqlash algoritmlari boshqa yondashuvlarga qaraganada yaxshiroq natija beradi. Quyida ushbu yondashuv usullarini ko‘rib chiqamiz.

Qaror ro‘yxatlari: ushbu usulning mohiyati shundan iboratki, test misollarini toifalarga ajratish uchun tartiblangan qoidalar to‘plamidan foydalaniladi. Buni *vaznli* [if-then-else] *qoidalari* ro‘yxati sifatida ko‘rish mumkin. O‘quv majmuasi xususiyatlar to‘plamini induktsiya qilish uchun ishlatiladi. Har qanday so‘z ko‘rib chiqilayotganda, birinchi navbatda uning paydo bo‘lishi hisoblab chiqiladi va qarorlar ro‘yxatini yaratish uchun xususiyatlar vektori nuqtai nazaridan ifodalanadi, ball hisoblanadi. Vektor uchun maksimal ball ma'noni ifodalaydi.

Qaror daraxti: Qaror daraxti ma'lumotlarini rekursiv tarzda ajratadi va daraxt tuzilishidagi tasniflash qoidalarini ifodalaydi. Ichki tomirlar xususiyatlar bo‘yicha sinovni ifodalaydi va har bir shox qaror qanday qabul qilinayotganini ko‘rsatadi va barg tomirlari natija yoki bashoratga ishora qiladi.

Naïve Bayes: Naive Bayes klassifikatori Bayes teoremasini qo‘llashga asoslangan oddiy ehtimolli klassifikatordir. U kontekstdagi fj xususiyatlarini hisobga olgan holda w so‘zning har bir Si ma'nosining shartli ehtimolini hisoblashga tayanadi. Quyidagi formulani maksimal darajaga keltiradigan S ma'nosi kontekstda eng mos ma'no sifatida tanlanadi.

$$\hat{S} = \underbrace{argmax}_{S_i \in manolar_D(w)} P(S_i|f_1, f_2, \dots, f_m) = \underbrace{argmax}_{S_i \in manolar_D(w)} \frac{P(f_1, \dots, f_m|S_i)\,P(S_i)}{P(f_1,\dots,f_m)} =$$

$$= \underbrace{argmax}_{S_i \in manolar_D(w)} = P(S_i) \neq P(f_j|S_i) \qquad (1)$$

Naïve Bayes klassifikatoridan foydalanishdan oldin target so‘zning farqlanuvchi xususiyatlari aniqlab olinadi. O‘zbek tilidagi turli so‘z turkumlari orasidagi omonimiyani aniqlashda ushbu usuldan foydalanilganiga guvoh bo‘ldik [6].
Xuddi shunday, boshqa usullarni ham tasniflash mumkin.

Nazoratsiz usul. Nazoratsiz usulning nazorat qilinadigan usuldan farqi shundaki, unda inson omili yordamida aniqlangan ma’lumotlar to‘plami kerak emas. Bu omonim so‘zlar o‘xshash kontekstlarda kelishiga asoslanadi. So‘z ma'nolari so‘zlarning voqelik klasterlarini shakllantirish orqali hosil bo‘ladi.

Kontekst klasterlash: Bu usulda birinchi kontekst vektorlari yaratiladigan, so‘ngra so‘zning ma'nosini aniqlash uchun klasterlarga guruhlanadigan klasterlash usullariga asoslanadi. Bu usuldan foydalanilganda so‘z fazosi sifatidagi vektor maydoni va uning o‘lchamlari faqat so‘zlardir. Shuningdek, bu usulda korpusda bo‘lgan so‘z vektor sifatida belgilanadi va uning kontekstda necha marta sodir bo‘lganligi hisoblanadi. Keyin birgalikda sodir bo‘lish matritsasi yaratiladi va o‘xshashlik o‘lchovlari qo‘llaniladi. Keyin farqlash har qanday klasterlash texnikasi yordamida amalga oshiriladi.

So‘z klasterlash: Ushbu usulda omonim so‘zlar bir xil klasterga biriktiriladi. So‘zlar orasidagi o‘xshashlik sintaktik bog‘liqlik orqali beriladi. Agar W w_m ga o‘xshash so‘zlardan iborat bo‘lsa, daraxt dastlab faqat bitta w_m shox bilan hosil bo‘ladi va w_1, …, $w_m$ga eng o‘xshash ma'noli so‘z ekanligi aniqlanganda w_1 shoxda w_m tugunga ega bo‘ladi. Guruh algoritmi bo‘yicha klasterlash deb ataladigan yana bir yondashuv har bir so‘zni xususiyat vektori sifatida ifodalaydi. Belgilangan so‘zlarga duch kelganda, har bir elementi w_m va w_n ikkita so‘z o‘rtasidagi o‘xshashlik bo‘lgan S_{mn}o‘xshashlik matritsasi deb ataladigan matritsa tuziladi. Ushbu algoritmning keyingi bosqichida W so‘zlar to‘plami uchun rekursiv usulda guruhlar tuziladi. Keyin klasterlash algoritmi biron bir guruhning so‘zlariga o‘xshamaydigan so‘zlarni topishga harakat qiladi. Hech qanday guruh tarkibiga kirmaydigan so‘zlar yana qo‘shimcha guruhlar tuzish uchun ishlatiladi.Yakuniy bosqichda W ga tegishli har bir belgilangan so‘z guruhning markaziy qismiga o‘xshashligiga qarab guruh a'zosi bo‘ladi.

Birgalikda sodir bo‘lgan grafiklar: Bu usulda V cho‘qqisi va E cheti bo‘lgan grafik, bu yerda V –matndagi so‘zlarni ifodalaydi va agar so‘zlar bir xil paragraf yoki matndagi sintaksisga muvofiq munosabatda birga bo‘lsa, E qo‘shiladi. Berilgan maqsadli so‘z uchun birinchi navbatda grafik tuziladi va grafik uchun qo‘shnilik matritsasi yaratiladi. Shundan so‘ng so‘zning ma'nosini topish uchun Markov klasterlash usuli qo‘llaniladi. Grafikning har bir chetiga so‘zlarning birgalikdagi chastotasi bo‘lgan og‘irlik beriladi. {m,n} chetining og‘irligi quyidagi formula bo‘yicha aniqlanadi:

$$w_{mn} = 1 - \max\{P(w_m|w_n), P(w_n|w_m)\} \qquad (2)$$

Bu yerda, $P(w_m|w_n)$–w_m va w_n so‘zlarning ketma-ket uchrash ehtimolligi. Yuqori chastotali so‘zlarga og‘irlik 0, kamdan-kam uchraydigan so‘zlarga esa 1 og‘irlik beriladi. Og‘irligi

ma'lum chegaradan oshib ketadigan chekkalar o‘tkazib yuboriladi. Keyin grafikga iterativ algoritm qo‘llaniladi va eng yuqori nisbiy darajaga ega bo‘lgan tugun markaz sifatida tanlanadi. Algoritm so‘zning o‘z markaziga chastotasi chegaradan pastroqqa yetganda tugaydi. Nihoyat, butun markaz berilgan maqsadli so‘zning ma'nosi sifatida belgilanadi. Nol og‘irlikka ega bo‘lgan maqsadli so‘zning markazlari bog‘lanadi va grafikdan minimal chegara daraxti yaratiladi. Ushbu kengaytmali daraxt maqsadli so‘zning haqiqiy ma'nosini aniqlash uchun ishlatiladi.

Yarim nazorat ostidagi usullar: Yarim nazorat ostidagi o‘qitish usullarida ma'lumotlar nazorat ostidagi kabi mavjud, ammo kamroq ma'lumot berilishi mumkin. Bu yerda aniq ma'lumot emas, faqat kritik ma'lumotlar mavjud. Misol uchun, tizim faqat maqsadli mahsulotning ma'lum bir qismi to‘g‘ri ekanligini va shunga o‘xshashligini aytishi mumkin. Yarim nazorat qilinadigan yoki minimal nazorat qilinadigan usullar kichik miqdordagi izohli ma'lumotnomalar bilan ishlash qobiliyati tufayli mashhurlikka erishmoqda va ko‘pincha katta ma'lumotlar to‘plamlarida mutlaqo nazoratsiz usullardan ustun turadi. Yordamchi ma'lumotlardan muhim xususiyatlarni o‘rganadigan va olingan ma'lumotlardan foydalangan holda ma'lumotlarni guruhlaydigan imkoniyatlar mavjud.

Xulosa.

Ushbu maqolada So‘z manolarini aniqlash usullari uchun qo‘llaniladigan turli yondashuvlarni umumlashtirildi va mavjud So‘z manolarini aniqlash usullari algoritmlarini ularning texnikasiga ko‘ra tasniflandi. Ushbu maqolada biz so‘z ma'nosini ajratib ko‘rsatishda mavjud bo‘lgan turli yondashuvlarni taqqoslash bo‘yicha so‘rovni taqdim etildi, bunda birinchi navbatda Mashinali o‘qitish yondashuvlari va lug‘atga asoslangan bilimlarga asoslangan yondashuvlarga e'tibor qaratildi. Biz nazorat ostidagi yondashuv yaxshiroq ishlaydi degan xulosaga keldik, ammo uning kamchiliklaridan biri bu katta korpusga bo‘lgan talab bo‘lib, ularsiz o‘qitish mumkin emas, uni nazoratsiz yondashuvda yengib o‘tish mumkin, chunki u noaniqlik uchun bunday keng ko‘lamli manbaga tayanmaydi. Boshqa tomondan, bilimga asoslangan yondashuv ma'lum bir kontekstda so‘zlarning ma'nosini aniqlash uchun ma'lumot manbalaridan foydalanadi.

Foydalanilgan adabiyotlar:

1. Elov B.B., Axmedova X.I. Uchta so‘z turkumi doirasidagi omonimiyani farqlovchi biznes jarayonni modellashtirish// O‘zbekiston respublikasi innovatsion rivojlanish vazirligining, Ilm-fan va innovasion rivojlanish ilmiy jurnal 2022 / 1, 150-162-b.
2. Axmedova X. I. Turli so‘z turkumlari orasidagi omonimiyani aniqlovchi matematik modellar// Science and innovation international scientific journal volume 1 issue 7 uif-2022: 8.2 | issn: 2181-3337. https://doi.org/10.5281/zenodo.7238546
3. Axmedova X.I. Chastotali usul yordamida omonimiyani aniqlash// “O‘ZBEK AMALIY FILOLOGIYASI ISTIQBOLLARI” Respublika ilmiy-amaliy konferensiyasi Toshkent: 2022.-164-170 b.
4. Elov B.B., Axmedova X.I. Determining homonymy using statistical methods. // “Hisoblash modellari va texnologiyalari (HMT 2022)” O‘zbekiston-Malayziya ikkinchi xalqaro konferensiyasi materiallari- Toshkent, 2022 16-17 sentabr,-106 b.
5. Elov B., Hamroyeva Sh., Xusainova Z., Xudayberganov N., Yodgorov U., Yuldashev A. “POS TAGGING OF UZBEK TEXTS USING HIDDEN MARKOV MODELS (HMM) AND VITERBI ALGORITHM”. “O‘zbekiston Milliy universitetining ilm-fan rivoji va

jamiyat taraqqiyotida tutgan o‘rni” mavzusidagi xalqaro ilmiy-amaliy konferensiya, 2023 yil, 12 may, 104-115-b.

6. Elov B.B., Axmedova X.I. Homonymy detection using a Naïve Bayes classifier. Journal of Computer Science Engineering and Information Technology Research (JCSEITR) ISSN(P): 2250-2416; ISSN(E): Applied Vol. 13, Issue 1, March 2023, 5–20 © TJPRC Pvt. Ltd, Hindiston
7. Elov B.B., Axmedova X.I. Primova M.H., Khudayberganov N.U. Semantic Differentiation of Uzbek Homonyms Using the Lesk Algorithm. *UBMK 2023 - Proceedings: 8th International Conference on Computer Science and Engineering*, 2023, страницы 137–140. https://doi.org/10.1109/UBMK59864.2023.10286666
8. Eneko Agirre and Philip Edmonds "Word Sense Disambiguation: Algorithms and Applications" Springer, 2007, ISBN: 978-1-84628-855-2
9. Pushpak Bhattacharyya and Josef van Genabith, "Word Sense Disambiguation in Indian Languages", 2017, Springer, ISBN: 978-981-10-4055-4
10. "La désambiguïsation lexicale et ses applications" (Lexical Disambiguation and Its Applications) by Isabelle Tellier.
11. M. Victoria Marrero Aguiar, "Desambiguación léxica y sentido en el español" (Lexical Disambiguation and Sense in Spanish) Publicacions i Edicions de la Universitat de Barcelona, 2008, ISBN: 978-84-475-3275-7
12. Huang Xiaorong, Semantic Disambiguation Techniques for Chinese Words Science Press, 2009, ISBN: 978-7-03-024173-1

ЎЗБЕК ТИЛИДА ҚАРАШЛИЛИК/ТЕГИШЛИЛИК МАЪНОСИНИНГ СЎЗ СЕМАНТИКАСИ ТАЪСИРИДА ЯШИРИН ИФОДАЛАНИШИ УСУЛИ

Исмоилова Муаттархон Илёсовна
Ўзбекистон Миллий Университети
Ўзбек тилшунослиги кафедраси мустақил тадқиқотчи.
E-mail: s42813275@gmail.ru

Аннотатция: Ўзбек тилида қарашлилик/тегишлилик маъносининг муайян сўзлар орқали уларнинг семантикаси таъсирида яширин ифодаланиши ҳам кузатилади. Бундай вазиятда ушбу сўзлар иширокида тузилган гаплардан яширин тарзда қарашлилик/тегишлилик маъноси англашилади.

Қарашлилик/тегишлилик маъносининг гапда яширин тарзда воқеланишига туртки берувчи бирликларга эга, эга бўлмоқ, эгалик қилмоқ; ихтиёрида, измида, тасарруфида, қарамоғида кабиларни киритиш мумкин.

Ўзбек тилида эга лексемаси ва у иштирокида ясалган эга бўлмоқ, эгалик қилмоқ феъл лексемалари орқали нутқда яшиин тарзда қарашлилик/тегишлиликни билдирувчи ахборот воқеланади.

Эга лексемаси кўпмаъноли сўз бўлиб, “нарса-предметга эга” маъносида қўлланганда яширин тарзда қарашлилик/тегишлилик маъносини ифодалайди. Бунда эга лексемаси “эгалланган предмет” тушунчасини англатувчи жўналиш келишигидаги исмлар билан боғланади, “предметни эгаллаган шахс” тушунчасини билдирувчи бирлик эса бош келишикдаги исмлардан таркиб топади. Масалан, Салим ноёб қўлёзмага эга гапидан ноёб қўлёзманинг Салимга тегишли эканлигини англашилади. Бунда Салим – посессор, ноёб қўлёзма – коррелят, эга лексемаси эса қарашллик/тегишлилик маъносини ифодаловчи восита. Аммо бундай конструкцияларда қарашллик/тегишлилик маъноси яширин тарзда ифодаланади. Бунда субъектнинг предметга эга бўлиши замирида ушбу предметнинг шу субъектга тегишлилигини англатувчи ахброт юзага келади. Бу ахборот гапнинг семантик структураси орқали яширин тарзда юзага чиқади.

Калит сўзлар: коррелят, эга, семантика, субъект, лексемас, сема.

Аннотация: В узбекском языке значение принадлежности/относившийся скрыто за отдельными словами под влиянием их семантики. В такой ситуации значение относившийся/принадлежности понимается из предложений, составленных с этими словами.

Иметь, владеть единицами, которые способствуют имплицитной реализации значения отношения принадлежности в предложении; можно ввести такие вещи как в распоряжение, под контроль, под опеку.

В узбекском языке образованные при его участии притяжательная лексема и глагольные лексемы владеть четко передают в речи информацию принадлежности\относившийся.

Притяжательная лексема является многозначным словом и при употреблении в значении «владеть предметом» имплицитно передает значение принадлежности. В этом случае притяжательная лексема связана с существительными в предметном договоре, что

означает понятие «принадлежащий предмет», а единица, означающая понятие «лицо, владеющее предметом», состоит из существительных в основном договоре. Например, если у Салима есть уникальная рукопись, подразумевается, что эта уникальная рукопись принадлежит Салиму. При этом Салим – обладатель, уникальная рукопись – коррелят, а лексема обладатель – средство выражения значения владения/принадлежности. Но в таких конструкциях имплицитно выражен смысл собственности/принадлежности. В этом случае возникает факт наличия у субъекта объекта, а значит, этот объект принадлежит\относиться этому субъекту.

Ключевые слова: коррелят, обладатель, семантика, субъект, лексемы, сема.

Abstract: In the Uzbek language, the meaning of belonging/belonging is hidden behind individual words under the influence of their semantics. In such a situation, the meaning of belonging/belonging is understood from the sentences made with these words.To have, to have, to own units that contribute to the implicit realization of the meaning of relation/belonging in a sentence; You can put such things as at your disposal, under control, under guardianship.In the Uzbek language, the possessive lexeme and verbal lexemes formed with his participation, posses, possess, clearly convey in speech the information of belonging/belonging. The possessive lexeme is a polysemantic word and when used in the sense of "to own an object," it implicitly conveys the meaning of belonging/belonging. In this case, the possessive lexeme is associated with the nouns in the subject contract, which means the concept "owned object," and the unit meaning the concept "person who owns the object" consists of nouns in the main contract. For example, if Salim has a unique manuscript, it is implied that this unique manuscript belongs to Salim. At the same time, Salim is the owner, the unique manuscript is the correlate, and the lexeme owner is a means of expressing the meaning of possession/belonging. But in such constructions the meaning of ownership/ belonging is implicitly expressed. In this case, the fact arises that the subject has an object, which means that this object belongs to this subject. This information is hidden behind the semantic structure of the sentence.

Keywords: correlate, possessor, semantics, subject, lexemes, seme.

Эга лексемаси қуйидаги маъносида ҳам гапда қарашлилик/ тегишлиликни билдирувчи ахборотни юзага чиқаради: "5 Нарсани "юзага келтириш, тайёрлаш ва ш.к. жиҳатидан қарашлилигини билдиради". Масалан, *Ош эгаси билан ширин* мақолида ошнинг эгасига, яъни уни тайёрлаган инсонга тегишли, қарашли эканлиги маъноси англашилади.

Шуниси характерлики, *эга* лексемасининг " Учинчи шахс ёки нарсада ўзи бирикиб келган сўз англатган нарсанинг бор, мавжуд эканини билдиради; бор, мавжуд" маъносида ҳам қарашлилик/тегишлиликка ишора кузатилади. қилади. Масалан, *Мен шу дала ҳовлига эгаман* гапидан *Менда шу далаҳовли бор* ахбороти билан бирга *Шу далаҳовли менга тегишли* ахбороти англашилади. Яъни, субъектда бирор нарса-предметнинг мавжуд эканлиги ҳақидаги ҳукм бу нарсанинг айни субъектга тегишли эканлигини англатувчи ахборотга ишора қилиб туради. Бу ахборот асосий ахборот замирида яширин тарзда юзага чиқади.

Қуйидаги гапнинг семантикасига эътибор қаратамиз:— *Саид Ҳалил мамлакатнинг диний пешвоси сифатида подшоҳ ҳокимиятини ичдан назорат қилиш ҳуқуқига* ***эга эди.*** (П.Қодиров. "Ҳумоюн ва Акбар") Ушбу гапнинг семантик структурасидан яширин тарзда

"*Саид Ҳалилга мамлакатнинг диний пешвоси сифатида подшоҳ ҳокимиятини ичдан назорат қилиш ҳуқуқи тегишли эди*"

Англашиладики, [Исм$_{б.к}$ + Исм$_{жўн.кел.}$ + эга] қолипидаги синтактик конструкцияларда доимо яширин тарзда қарашлилик/тегишлиликни билдирувчи ахборот ифодаланади.

Бундай тузилишдаги синтактик конструкцияларда бош келишикдаги исмлардан ташкил топган бирлик посессор вазифасида, жўналиш келишигидаги исмлар билан ифодаланган бирлик эса коррелятни, яъни эгалик қилинган предмет тушунчасини англатади. Масалан,

Шуниси характерлики, *эга* лексемаси гапда инкор ифодаловчи *эмас, на...* юкламалари билан бирикиб келган ҳолатда гапда қарашлилик/ тегишлиликни ифодаловчи ахорот ифодаланмайди. Қиёсланг: *Салим ноёб қўлёзмаган* ***эга эмас****. Салим* ***на*** *ноёб қўлёзмаган эга.*

Нутқда *эга* лексемаси билан ҳосил қилинган *эга бўлмоқ, эгалик қилмоқ* феъл лексемалари ҳам учрайди. Бундай ясалмалар орқали нарса-предметга эга бўлганлик ёки эгалик қилиниши ҳаракати содир этилгандагина гапда қарашлилик/тегишлиликни ифодаловчи яширин ахборот ифодаланади. Эгалик қилиш ҳаракати ҳали амалга ошмаган бўлса ёки амалга ошаётган бўлса, ҳаракатнинг инкори ифодаланса, гапда қарашлилик/тегишлилик маъноси юзага чиқмайди. Масалан:

1. *Анвар эрса аввало Раъно каби бир рафиқаға* ***эга бўлмоқчи*** *ва шу воситада бошқа масалаларга қарамоқчи эди*. (А.Қодирий. "Ўткан кунлар")

2. *Нигор ойим йигирма-ўттуз қизни ўқутиб улардан тушкан "озодлиқ" ва "панжшанбалик" пулларга ҳам* ***эга бўлолмас****, ҳар кун деярлик муҳтараб махдум поччаға ҳисобини батамом топшуриб туришқа мажбур эди.* (А.Қодирий. "Ўткан кунлар") Кўринадики, биринчи мисолда *эга бўлмоқ* феъл лексемаси орқали ҳаракатнинг келгусида режалаштирилганлиги ифодаланган, шу сабабли гапда Анварнинг ҳали рафиқага эга бўлмаганлиги англашилади.

Кейинги гапда эса *эга бўлмоқ* феъл лексемаси бўлишсиз шаклда ишлатилган. Натижасида гапнинг ушбу феъл иштирок этган қисмидан "пулга эга бўлмаганлик" маъноси ифодаланади.

Демак, *эга бўлмоқ, эгалик қилмоқ* феъл лексемалари орқали қарашлилик/тегишлилик маъносининг ифодаланиши учун ушбу феъллар англатган ҳаракат амалга ошган бўлиши шарт. Эга бўлиш, эгалик қилиш ҳаракати амалга ошгач, ҳаракат натижаси сифатида предметга эгалик, предметнинг шу шахсга тегишили эканлиги маъноси юзага чиқади.

Нутқда қарашлилик/тегишлилик маъноси *ихтиёрида, измида, тасарруфида, қарамоғида* лексемалари билан ҳам юзага чиқади. Бу бирликлар, гарчи эгалик аффикси ва ўрин-пайт келишигини қабул қилган бўлса-да, ушбу аффикслар сўз таркибида ўз вазифасидан узоқлашган. Шу сабабли *ихтиёрида, измида, тасарруфида, қарамоғида* сўзлари нутқда яхлит бирлик сифатида ишлатилади ва гапда қарашлилик/тегишлилик маъноси эгалик аффикси орқали эмас, айни шу сўзлар семантикаси таъсирида юзага чиқади.

Маълумки, эгалик аффикси сўзга қўшилиб, асосдан англашилган предметга (предмет – кенг маънода) эгалик/қарашлилик маъносини ифодалайди. Масалан, *китобим – китоб менга тегишли, укам – ука менга қарашли* каби. Аммо *ихтиёримда* бирлиги эгалик аффиксини қабул қилганлиги учун "*ихтиёр менга тегишли*" маъносини

англатмайди. Ёки *тасарруфингда* бирлиги орқали ҳам "*тасарруф сенга тегишли*" маъноси юзага чиқмайди. Худди шундай, *қарамоғимда* бирлигида ҳам "*қарамоғ менга тегишли*" маъноси англашилмайди. Бу ҳолат *ихтиёрида, тасарруфида, қарамоғида* бирликлари таркибида эгалик аффикси ҳамда ўрин-пайт келишиги маъноси ва вазифасидан узоқлашганлигини, бу бирликлар мазкур аффикслари билан яхлит ҳолда маъно касб этиши ва шу ҳолида лексик бирлик мақомига эга бўлганлигини кўрсатади.

Демак, *ихтиёрида, измида, тасарруфида, қарамоғида* гапда яхлит бирлик сифатида иштирок этади. Қарашлилик/тегишлилик маъноси ушбу бирликлар семантикаси таъсирида яширин ифодаланади.

Маълумки, *ихтиёрида* сўзининг асос қисми **"**1 Xohish-istak erki; erk, xohish. **2** Foydalanish, tasarruf yoki idora qilish huquqi; izm"[26] маъносини англатади.

Изоҳдан кўринадики, мазкур бирлик тўғридан-тўғри қарашлилик/ тегишлилик маъносини англатмайди. Аммо унинг бирор нарсадан "foydalanish, tasarruf yoki idora qilish huquqi"га эгаликни англатиши гапда *ихтиёрида* сўзи орқали яширин қарашлилик/тегишлилик маъноси ифодаланиши йўл очади. Масалан:*Ботаника боғи Фанлар академияси ихтиёрида.* Берилган гапдан "*Ботаника боғи Фанлар академиясига тегишли*" ахбороти англашилади.Ёки:*Ширин шаҳридаги чемпионлар маскани яна спортчилар ихтиёрида. ("Янги Ўзбекистон")*Ушбу гапдан "*Ширин шаҳридаги чемпионлар маскани яна спортчиларга тегишли*" ахбороти англашилади.

Ўзбек тилида *ихтиёрида, тасарруфида, қарамоғида, измида* сўзларига *бўлмоқ, қолмоқ* феълларининг бирикиб келишидан ҳосил бўлган *ихтиёрида бўлмоқ, тасарруфида бўлмоқ, қарамоғида бўлмоқ, измида қолмоқ; ихтиёрида қолмоқ, тасарруфида қолмоқ, қарамоғида қолмоқ, измида қолмоқ* ясама феъллари ҳам мавжуд.

Шуниси характерлики, феъл ясалишида мазкур сўзларнинг семантикаси буткул ўчиб кетмайди. Шу сабабли ушбу феъллар орқали ҳам уларнинг семантикаси таъсирида гапда қарашлилик/тегишлилик маъноси ифодаланади. Масалан:1. *Бу билан хорижга юборилган 100 миллиард долларлик энергия пули энди халқ ихтиёрида* ***бўлади****", – дейди Эрдўғон.* ("Даракчи")2. *Бухородa 3,3 млрд. сўмлик солиқлар тадбиркорлик субъектларининг ихтиёрида* ***қолади.*** (Zamin. uz) Биринчи гапдан яширин характердаги "*Ширин шаҳридаги чемпионлар маскани яна спортчиларга тегишли бўлади*" ахбороти воқеланади.

Иккинчи гапда *ихтиёрида қолмоқ* феъл лексемаси орқали "*Бухородa 3,3 млрд. сўмлик солиқлар тадбиркорлик субъектларига тегишли бўлади*" ахбороти яширин тарзда ифодаланади.

Нутқда *ихтиёрида* сўзи баъзан *измида* сўзи билан; *ихтиёрида бўлмоқ, ихтиёрида қолмоқ* сўзлари *измида бўлмоқ, измида қолмоқ* ясалмалари билан алмашиниб қўллана олади. Бундай вақтда *измида, измида бўлмоқ, измида қолмоқ* бирликлари "tasarruf, тасарруф қилиш, эгалик қилиш" маъносини англатиб, қарашлилик/тегишлилик маъносини яширин ифодалашга хизмат қилади. Масалан, "*Изтопар итлар истеъдодли изқуварлар* ***измида"*** гапида қўлланган *измида* сўзи "тасарруфида" маъносини билдиради. Тасарруфида бўлмоқлик эса предметга эгаликни англатади. Шу жиҳатдан ушбу гапдан

измида сўзи семантикаси таъсирида яширин тарзда **"***Изтопар итлар истеъдодли изқуварларга тегишли***"** ахбороти англашилади.

Ўзбек тилида *қарамоғида* ҳамда бу сўз иштирокида ясалган *қарамоғида бўлмоқ, қарамоғида қолмоқ* феъл лексемалари орқали ҳам гапда яширин қарашлилик/тегишлилик ифодаланади.

Маълумки, *қарамоғида* сўзи *қарамоғ* сўзи асосида ҳосил қилинган бўлса-да, бугунги кунда яхлит бирлик саналади. Аммо бу сўз қарамоғида шаклида "Ўзбек тилининг изоҳли луғати"да учрамайди. *Қарамоғ* сўзига луғатда қуйидагича изоҳ берилган:

"QARAMOG' Tarbiya, boqish, tasarruf va sh.k. jihatdan mansublik, ixtiyorida bo'lish, qaramlik. *O'n yoshimgacha onamning qaramog'ida tarbiyalandim. – Azizbek Turkiston xonligining eng zolim... beklaridan va uning o'z qaramog'ida bo'lgan Toshkent aholisiga qilgan zulmlari Farg'onaga doston.* A.Qodiriy, O'tgan kunlar. *Eshmat aka... qaramog'idagi bolalari to'g'risida spravka kelsa, uyga qaytarishlari lozimligini yozardi*. O.Yoqubov, Er boshiga ish tushsa".

Нутқда *қарамоғида* сўзи, *қарамоғида қолмоқ, қорамоғида бўлмоқ* феъл лексемалари кимсага нисбатан ишлатилганда, асосан, "тарбиясида, боқувида" маъноси реаллашади. Масалан:1. "Боқувида" маъноси ифодаланади: *Бир ўғилнинг қарамоғида* ***қолган*** *кекса отаси бор эди*. ("Даракчи") 2. "Тарбиясида" маъноси ифодаланади: *Ота-онаси хорижга кетганлиги сабабли бола бувиси* ***қорамоғида*** *эди.* (Zamin. uz) Нутқда *қарамоғида* сўзи, *қарамоғида қолмоқ, қорамоғида бўлмоқ* феъл лексемаларининг "тасарруфида, ихтиёрида бўлмоқлик" семаси юзага чиққанда "тегишлилик, қарашлилик" ифодаланади. Бироқ бу маъно яширин характерга эга бўлади.

Нутқда *қарамоғида* сўзи, *қарамоғида қолмоқ, қорамоғида бўлмоқ* феъл лексемалари нафақат кишиларга, балки *хўжалик, давлат, юрт* кабиларга нисбатан ҳам ишлатилади. Бундай вақтда ушбу бирликларнинг "таъминотида бўлмоқ, " "тасарруфида бўлмоқ" маънолари воқеланади. Қиёсланг:

1. "Таъминотида" маъноси ифодаланади*: ОТМларда таҳсил олаётган етим болалар давлат қарамоғида бўлади*. (Gazeta. uz) 2. "Тасарруфида бўлмоқ" маъноси ифодаланади*: Ўзбекистонда банк активларининг 80 фоиздан ортиғи давлат қарамоғида.* (Kun. uz) Гапда қўлланган *қарамоғида* сўзи, *қарамоғида қолмоқ, қорамоғида бўлмоқ* феъл лексемаларининг айнан "тасарруфида бўлмоқ" маъноси реаллашганда эса синтактик конструкцияда *"хўжаликка тегишли", "давлатга тегишли", "ҳокимликка тегишли"* каби турли ахборотлар ифодаланади. Масалан:1. "*Андижон" футбол клуби бундан буён Андижон вилояти ҳокимлиги қарамоғида* ***бўлади.*** (Sports. uz)

Берилган гапдан "*Андижон" футбол клуби бундан буён Андижон вилояти ҳокимлигига қарашлидир***"** ахбороти англашилади.

2. *Шунингдек, хўжалик қарамоғида 9 та туя, 4 от ва 170 та қўй мавжуд.* (Yaypan. uz) Мазкур гапдан *"9 та туя, 4 от ва 170 та қўй хўжаликка қарашли" ахбороти англашилади.*

Англашиладики, ўзбек тилида гапда қўлланган *қарамоғида* сўзи, *қарамоғида қолмоқ, қорамоғида бўлмоқ* феъл лексемалари орқали уларнинг семантикаси таъсирида яширин ахборот ифодаланади.

Шунингдек, нутқимизда *тасарруфида* ҳамда бу сўз иштирокида ясалган *тасарруфида бўлмоқ, тасарруфида қолмоқ* феъл лексемалари орқали ҳам гапда яширин тарзда қарашлилик/тегишлилик маъноси ифодаланади.

“Ўзбек тилининг изоҳли луғати”да *тасарруфида* сўзи шу шаклида учрамайди. Луғатда ексемаси изоҳи қуйидагича берилади:

“TASARRUF [a. – o‘z ixtiyoriga olish, egalik qilish; mustaqil harakat qilish; o‘zboshimchalik] **kt.** Foydalanish, egalik huquqi, ixtiyori; egalik, xo‘jayinlik. Xususiy tasarruf. *Bu yer-suvlar qadimdan boyning tasarrufida edi. – Otasi esa Bobomirzani o‘zining ko‘z qorachig‘i deb bilar, shu sababdan katta ro‘zg‘orni boshqarishni, yer-suv, mol-dunyo tasarrufini shuning qo‘liga tashlab ketish niyatida yurardi.* И.Rahim, Taqdir. [Berdiqul] *Yarim qo‘ra molning he yo‘q-be yo‘q o'gay akasi tasarrufiga o‘tib ketganidan yuragi tars yorilay dedi.* N.Norqobilov, Tog‘dagi yolg‘iz odam”.

Ўзбек тилида *тасарруф қилмоқ (этмоқ)* сўзи ҳам учрайди. Аммо бу сўз китобий услубга хосланганлиги сабабли оғзаки нутқда деярли учрамайди. Мазкур сўз гапда “эгалик қилмоқ” маъносида қўлланиб, эгалик, тегишлилик маъноси ифодалайди. Ўзбек тилининг изоҳли луғати”да *тасарруф қилмоқ (этмоқ)* сўзи қуйидагича изоҳланган:

“Tasarruf qilmoq (yoki **etmoq)** Egalik qilmoq. *Mening butun xojaligimni **taarof qilishga** undan boshqa kishining haddi yo‘q*. G‘.G‘ulom, Tirilgan murda. *Fotih Iskandar butun yer yuzini **tasarruf etib ham**, ko‘ngli zarracha shod bo‘lmagach, onasiga shunday maktub yozgan...* Gazetadan”.

Кўринадики, *тасарруф* сўзи ва ундан ҳосил қилинган *тасарруф қилмоқ (этмоқ)* сўзи “*эгалик қилмоқ”* маъносини билдиради. “Эгалик қилмоқлик” тегишлиликни англатади. Шу жиҳатдан бу сўзлар иштирок этган гаплардан яширин тегишлилик/қарашлилик маъноси англашилади. Хусусан, *Фотиҳ Искандар ер юзини **тасарруф этиб*** бирикмаси орқали “Ер юзи Фотиҳ Искандарга тегишли” ахбороти англашилади.Демак, бу сўз ҳам контекстда яширин тегишлилик/қарашлиликни ифодалаб келади.

Қарашлилик/тегишлилик маъносини яширин ифодаловчи бу бирликларнинг нутқда қўлланиши ҳамда фаоллик даражаси турлича. *Эга, ихтиёрида, измида, тасарруфида, қарамоғида* сўзлари китобий услубга хос бўлса-да, уларнинг нутқда қўлланиш даражаси бир-биридан фарқланади.

Хуллас, ўзбек тилида айрим сўзларнинг луғавий маъноси таъсирида ҳам қарашлилик/тегишлилик маъноси ифодаланади. Аммо бу сўзлар ифодалаган қарашлилик/тегишлилик маъноси яширин характерга эга бўлади.

Мазкур бирликлар тўғридан-тўғри қарашлилик/ тегишлилик маъносини англатмайди. Аммо улар гапда бирор нарсадан “foydalanish, tasarruf yoki idora qilish huquqi”га эгаликни англатганда ёки “тасарруфида, ихтиёрида бўлмоқлик”ни акс эттирганда яширин қарашлилик/тегишлилик маъноси ифодаланиши йўл очади. Ўзбек тилида *ихтиёрида, тасарруфида, қарамоғида, измида* сўзларига *бўлмоқ, қолмоқ* феълларининг бирикиб келишидан ҳосил бўлган *ихтиёрида бўлмоқ, тасарруфида бўлмоқ, қарамоғида бўлмоқ, измида қолмоқ; ихтиёрида қолмоқ, тасарруфида қолмоқ, қарамоғида қолмоқ, измида қолмоқ* феълларининг ясалишида мазкур сўзларнинг семантикаси буткул ўчиб кетмайди. Шу сабабли ушбу феъллар орқали ҳам уларнинг семантикаси таъсирида гапда қарашлилик/тегишлилик маъноси ифодаланади.

Қарашлилик/тегишлилик маъносини яширин ифодаловчи бирликларнинг нутқда қўлланиши ва фаоллик даражаси турлича. *Эга, ихтиёрида, измида, тасарруфида, қарамоғида* қаторида энг юқори частотага эга бирлик *эга* сўзидир, бу сўз корпусда 23137 марта учради. Аммо бу бирлик кўп маъноли бўлиб, унинг қарашлилик/тегишлиликни ифодалашдаги улушини "Ўзбек тили таълимий корпуси" асосида аниқлаш мушкул. Кейинги ўринда юқори частотали бирлик сифатида *тасарруфида* сўзи келади, энг паст частотали сўзлар эса *измида* бирликларидир.

Фойдаланилган адабиётлар рўйхати:

1. Абдуллина Г.Р. Категория принадлежности в башкирском языке / Научная мысль Кавказа, 2008, № 4. – С. 118-123.
2. Абдураҳмонов Ғ., Рустамов А. Қадимги туркий тил. − Тошкент, 1982.
3. Абдураҳмонов Ғ., Шукуров Ш. Ўзбек тилининг тарихий грамматикаси. − Тошкент: Ўқитувчи, 1973. − 320 б.
4. Абдураҳмонов Н. Қадимги туркий тил. − Тошкент: Ўқитувчи, 1989. − 153 б. – Б.54.
5. Абуханов Г.А. Семантика и синтаксическая функция притяжательных аффиксов в современном казахском языке: Автореф. дис. канд. филол. наук. – Алма-Ата, 1950. – 22 с.
6. Айюб Ғуломовнинг илмий мероси (4-китоб). Ўзбек тилида аниқловчилар. Ўзбек тилида келишиклар. – Тошкент, 2008. – 140 б.
7. Акбаев Ш.Х. К вопросу о происхождении посессивных аффиксов в тюркских языках // Тюркологические исследования. – М.: Наука, 1976. – С. 3-Galsworthy, J The Forsyte Saga. Book 1. The Man of Property / J. Galsworthy. – Moskow: Progress Publishers, 1974. – 383 p.

LINGVISTIK CHEKLOVLAR – KOREFERENTLIKNI AVTOMATIK HAL ETISHNING ASOSI

Abdisalomova Shahlo
ToshDO‘TAU tayanch doktoranti.

Annotatsiya: Tabiiy tilni qayta ishlash (NLP) sun’iy intellektning kichik sohasi bo‘lib, uning maqsadi kompyuterlar va odamlar o‘rtasida o‘zaro aloqalarni o‘rnatishdan iborat. Ma’lumki, tilning o‘ziga xos jihatlari matndan avtomatik ravishda ma’no chiqarish jarayonida qiyinchiliklarni yuzaga keltiradi. Inson ongi bevosita matn mazmunini tushunishga qodir. Ammo sun’iy intellekt uni to‘g‘ri talqin qilishi uchun koreferentlik yuqori aniqlikda hal etilgan bo‘lishi lozim. Bundan tashqari, mashina tarjimasi, savollarga javob berish, matnni umumlashtirish, hissiyotlarni tahlil qilish, matnni tasniflash, nutqni tanish, nomga ega obyektni aniqlash, chatbot kabi sohalar rivoji ham aynan koreferentlikning hal etilishi bilan bog‘liqdir. Ushbu maqolada NLP ning koreferentlikni hal etish vazifasi, uning ahamiyati yoritiladi, lingvistik xususiyatlarning bu jarayondagi ishtiroki masalasi o‘rganiladi.

Kalit so‘zlar: NLP, koreferentlik, antisident, anafor, lingvistik xususiyat, vektor, olmosh, matn.

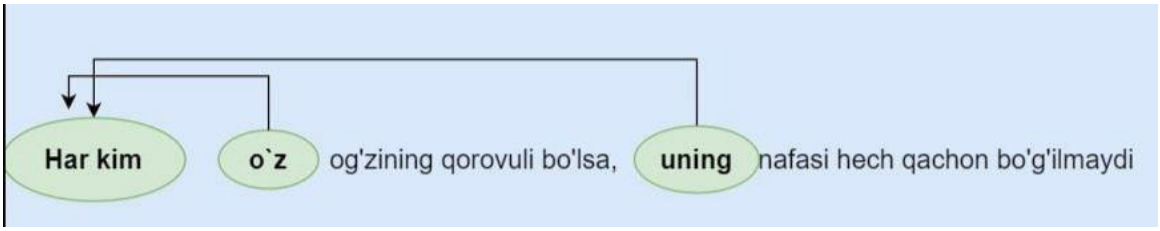
Annotation: Natural Language Processing (NLP) is a subfield of Artificial Intelligence, that aims to facilitate interactions between computers and humans. It is known, that the specific features of the language create difficulties in the process of automatically extracting meaning from the text. The human mind is able to directly understand the content of the text. But in order for the Artificial Intelligence to interpret it correctly, the coreference must be resolved with high accuracy. In addition, the development of fields such as Machine Translation, Question Answering, Text Summarization, Sentiment Analysis, Text Classification, Speech recognition, Named Entity Recognition, Chatbot is also related to the Coreference Resolution. In this article, the Coreference Resolution is a subfield of NLP, its importance is highlighted, problem of the participation of linguistic features in this process is investigated.

Keywords: NLP, coreference, antecedent, anaphor, linguistic feature, vector, pronoun, text

Аннотация: Обработка естественного языка (НЛП) — это область искусственного интеллекта, целью которой является облегчение взаимодействия между компьютерами и людьми. Известно, что особенности языка создают трудности в процессе автоматического извлечения смысла из текста. Человеческий разум способен непосредственно понимать содержание текста. Но для того, чтобы искусственный интеллект правильно интерпретировал ее, кореференция должна быть решена с высокой точностью. Кроме того, с решением кореференции связано развитие таких областей, как машинный перевод, ответы на вопросы, обобщение текста, анализ настроений, классификация текста, распознавание речи, идентификация объекта по имени, чат-бот. В данной статье подчеркивается роль НЛП в разрешении кореференции, ее значение, изучается проблем участия языковых особенностей в этом процессе.

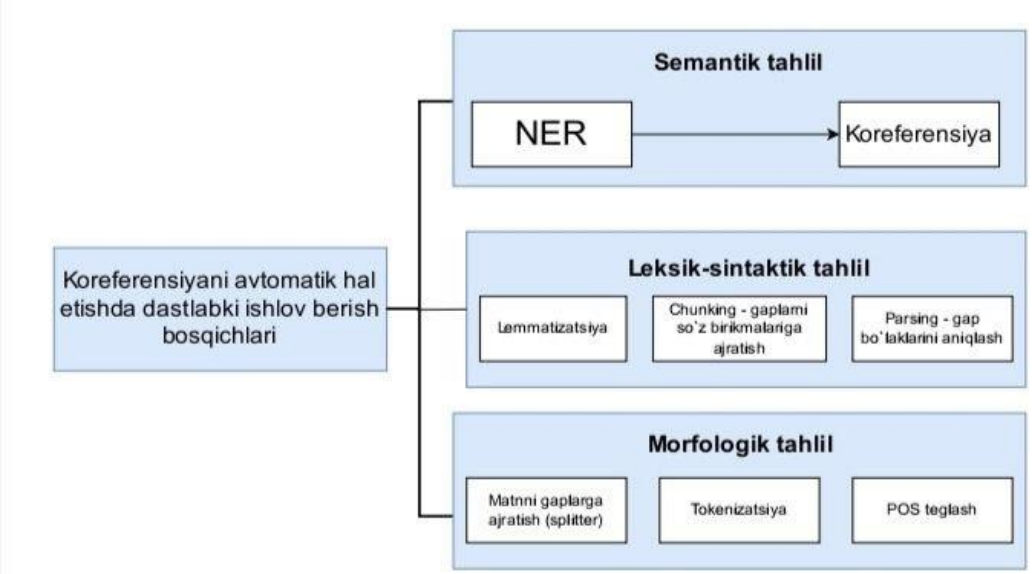
Ключевые слова: НЛП, кореференция, антисидент, анафора, лингвистический признак, вектор, местоимение, текст

Nutqdagi har qanday jumlaning mohiyati kontekstda fikr yuritilayotgan shaxs, predmet yoki voqea-hodisa yordamida anglashiladi. Inson ongi bevosita matn mazmunini tushunishga qodir. Ammo sun'iy intellekt uni to'g'ri talqin qilishi uchun matn qismlarining o'zaro sinonimiyasini, koreferentlikni hosil qiluvchi bir yadroviy zanjirga birikkan nutqiy birliklarni aniqlash muhimdir. Bu vazifa NLPda **Koreferensiyani hal etish (CR – Coreference Resolution)** tizimlariga yuklatilgan. Koreferensiyani hal etish obyekt haqidagi matn ma'lumotlarini birlashtirishning asosiy texnologiyasi [Vincent Ng, 2017; 4877], matndagi bir xil obyektga ishora qiluvchi barcha havola bo'laklarni avtomatik topish jarayoni [Elango, 2005; 1] bo'lib (1-rasm), bunda axborot uzatish mavqeyi jihatidan mustaqil referent **"antisident"** va ma'nosi antisident vositasida namoyon bo'ladigan referent **"koreferent/anafor"** deb yuritiladi.



1-rasm. Koreferensiya (koreferentlik) hodisasi

Koreferentlik dastavval qoidaga asoslangan usullar orqali hal etilgan, keyinchalik uning o'rnini mashinaviy o'rganish usullari egalladi. Koreferentlikni avtomatik hal etish imkoniyati dastlabki ishlov berish bosqichlarining sifat darajasi bilan belgilanadi. Koreferentlikni hal etishda dastlabki ishlov berish bosqichi deyarli barcha tillar uchun umumiy. Bu jarayon *morfologik ishlov berish, gapni otli birikmalarga ajratish (NP chunker), POS teglash, nomlangan obyektni aniqlash (NER), avtomatik sintaktik teglash (parsing)* ni o'z ichiga oladi [Shi Chunqi va b., 2014; 54]. Ushbu bosqichlar esa nutqiy birliklarning lingvistik tahlili asosiga quriladi (1-chizma):



1-chizma

Bugungi kunda koreferentlikni bartaraf etuvchi mashinaviy o‘rganish usullarining barchasi [Soon va b., 2001; 524] tomonidan taklif etilgan til birliklarining inobatga olish lozim bo‘lgan 12 ta xususiyatlar to‘plamidan foydalanib yaratilgan. Ushbu xususiyatlar vektori ikkita ajratib olingan ***i* (nomzod antisident) va *j* (anafor)** belgilariga asoslanadi (Ingliz tilidagi qisqartmalari bilan berildi; 1-jadval):

1. Masofa (DIST)	anafora va antisident orasida joylashgan gaplar sonini ifodalaydi. Agar 325anafora va uning antisidenti bir gap ichida qo‘llanilsa, qiymat 0 ga teng; agar ular orasida bitta gap mavjud bo‘lsa, qiymat 1 ga, ikkita gap mavjud bo‘lsa, qiymat 2 ga teng va h.k.
2. i-olmosh (i_PRONOUN)	agar antisident olmosh bo‘lsa *rost* qiymatini, aks holda, *yolg‘on* qiymatini beradi.
3. j-olmosh (j_PRONOUN)	agar 325anafora olmosh bo‘lsa *rost* qiymatini, aks holda, *yolg‘on* qiymatini beradi.
4. Matn (string) mosligi (STR_MATCH)	artikl va ko‘rsatish olmoshlari olib tashlangandan so‘ng, antisident matn anafora matniga mos kelsa, ya’ni biri ikkinchisining aynan o‘zi bo‘lsa, rost qiymatini beradi. Masalan, *dasturchi bu dasturchi* birligiga mos keladi.
5. Aniq ot birikmasi (DEF_NP)	agar antisident *“the”* artiklini olgan otli birikma bo‘lsa, *rost* qiymatini , aks holda, *yolg‘on* qiymatini beradi.
6. Ko‘rsatish olmoshiga ega ot birikmasi (DEM_NP)	agar antisident ko‘rsatish olmoshi bilan birikkan otli birikma bo‘lsa, *rost* qiymatini, aks holda, *yolg‘on* qiymatini beradi.
7. Son (birlik/ko‘plik) muvofiqligi (NUMBER)	agar antisident va anafora son bo‘yicha o‘zaro muvofiq bo‘lsa (ikkalasi ham birlikda yoki ikkalasi ham ko‘plikda bo‘lsa) *rost* qiymatini, aks holda, *yolg‘on* qiymatini beradi.
8. Semantik guruh muvofiqligi (SEMCLASS)	Har bir tizim o‘zining 325anafora guruhlarini belgilashi mumkin. Soon va b. “Erkak”, “Ayol”, “Shaxs”, “Vaqt” kabi anafora guruhlarni belgilaydi. Semantik guruhlar ISA iyerarxiyasida[27] joylashtirilgan. Anafora va uning antisidenti, agar ulardan biri ierarxiyada ikkinchisining hosila birligi bo‘lsa yoki ular o‘zaro bir xil maqomda bo‘lsa muvofiq keladi. *Ros t(1), yolg‘on (0)* yoki *noma’lum (2)* qiymatlaridan biri beriladi.
9. Jins muvofiqligi (GENDER)	agar 325anafora va uning antisidenti jins jihatidan o‘zaro muvofiq kelsa *rost (1)*, muvofiq kelmasa *yolg‘on (0)*, agar anafor yoki antisidentning jinsi noma’lum bo‘lsa, *noma’lum (2)* qiymatini beradi.
10. Atoqli ot (PROPER_NAME)	agar 325anafora ham, uning antisidenti ham atoqli ot bo‘lsa *rost*, aks holda, *yolg‘on* qiymatini beradi.

[27] ISA iyerarxiyasi narsa-predmetlar joylashuvi bo‘lib, unda asosiy va hosila birliklarning munosabati ifodalanadi. Asosiy birlik hosila birliklarning o‘zagi hisoblanadi. A ISA B tarzida keltirsak, har bir A obyekti B obyekti ham bo‘la oladi. Masalan, bizda A guruhi=shaxs, B guruhi=ayol, erkak. Agar B A bo‘lsa, B dagi har bir birlik A bo‘ladi, ya’ni har bir erkak va ayol shaxsdir.

11. Taxallus (ALIAS)	agar 326anafora antisidentning taxallusi bo‘lsa *rost*, aks holda, *yolg‘on* qiymatini beradi.
12. Izohlovchi (APPOSITIVE)	anafora antisidentning izohlovchisi bo‘lsa *rost*, aks holda, *yolg‘on* qiymatini beradi.

1-jadval. Koreferentlikni bartaraf etish uchun lingvistik cheklovlar

Keltirilgan lingvistik cheklovlarni o‘zbek tilidagi matnlarda koreferensiyani hal etish jarayoniga tatbiq etamiz. Quyida berilgan parchadagi *Qaffol Shoshiy* antisidenti va *u* anafori xususiyat vektorini o‘zaro bog‘laymiz:

Qaffol Shoshiy *Hirotdagi Nizomiya madrasasiga mudarrislik qilgan.* ***U*** *fiqh, hadis, usul, tilshunoslik fanlari va she’r san’ati bo‘yicha Movarounnahrda olimlar orasida tengsiz edi.* ***Qaffol ash-Shoshiy*** *qonunshunoslik, mantiq kabi sohalarga oid asar yozgan.*

Belgilarning xususiyat vektori (i=Qaffol Shoshiy, j=u)

Xususiyat	Qiymat	Izoh
DIST	1	i va j orasida bitta gap mavjud
i_PRONOUN	-	i olmosh emas
j_PRONOUN	+	j olmosh
STR_MATCH	-	matn mosligi mavjud emas
DEF_NP	-	j aniq ot emas
DEM_NP	-	j ko‘rsatish olmoshi emas
NUMBER	+	i va j ning ikkisi ham birlik sonda
SEMCLASS	1	i va j ning ikkisi ham shaxs
GENDER	1	i va j bir jinsga mansub
PROPER_NAME	-	Faqat i atoqli ot
ALIAS	-	j i ning taxallusi yoki laqabi emas
APPOSITIVE	-	j i ning izohlovchisi emas

Bu yerda o‘zbek tilida koreferentlikni hal etishda *u olmoshi* bilan bog‘liq noaniqliklar yuzaga keladi, ya’ni u olmoshi ko‘rsatish yoki kishilik olmoshi ekanligini aniqlash, barcha jinsni u olmoshining o‘zi ifodalashi bizdan yangi cheklov kiritishni talab etadi.

Yuqoridagi matn asosida *Qaffol Shoshiy va Qaffol ash-Shoshiy* birliklarning xususiyat vektorini ko‘rib chiqamiz:

Belgilarning xususiyat vektori (i=Qaffol Shoshiy, j=Qaffol ash-Shoshiy)

Xususiyat	Qiymat	Izoh
DIST	2	i va j orasida ikkita gap mavjud
i_PRONOUN	-	i olmosh emas
j_PRONOUN	-	j olmosh emas
STR_MATCH	+	matn mosligi mavjud
DEF_NP	+	j aniq ot
DEM_NP	-	j ko‘rsatish olmoshi emas
NUMBER	+	i va j ning ikkisi ham birlik sonda
SEMCLASS	1	i va j ning ikkisi ham shaxs
GENDER	1	i va j bir jinsga mansub
PROPER_NAME	+	i va j atoqli ot

ALIAS	-	j i ning taxallusi yoki laqabi emas
APPOSITIVE	-	j i ning izohlovchisi emas

Shu tariqa, nomzod antisident va anafor birliklar xususiyat vektorlarining qiymati aniqlangach, keyingi jarayonda mashinani o‘rganish algoritmlariga murojaat etiladi va lingvistik xususiyatlar vektori natijalariga ko‘ra antisident-anafor juftliklari klasterlanadi. O‘zbek tili matnlarida koreferentlikni bartaraf etishda [Soon va b., 2001; 524] yoki [Zhang va b., 3] tadqiqotini o‘zgarishlarsiz qo‘llash orqali yuqori natijaga erishib bo‘lmaydi. Ushbu xususiyat vektorlari ingliz tili uchun ishlab chiqilgan va ular orasida o‘zbek tili tabiatiga mos kelmaydigan cheklovlar ham mavjud. Masalan, artikllar orqali aniq va noaniq otni ajratish cheklovi nafaqat o‘zbek tili, balki barcha turkiy tillar oilasi uchun amal qilmaydi. Shu sababli bu cheklovlarni o‘zbek tilining lingvistik tabiatiga moslab o‘zgartirish, kengaytirish maqsadga muvofiqdir.

Foydalanilgan adabiyotlar:

1. Vincent Ng. Machine Learning for Entity Coreference Resolution: A Retrospective Look at Two Decades of Research. / Proceedings of the Thirty-First AAAI, Conference on Artificial Intelligence (AAAI-17), pp.4877-4884.
2. Elango P. Coreference Resolution: A Survey. – University of Wisconsin, Madison, WI, pp.1-8, 2005.
3. Shi Chunqi, Verhagen Marc, Pustejovsky James. A Conceptual Framework of Natural Language Processing Pipeline Application. / Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT, pp.53-59, Dublin, Ireland, 2014.
4. Soon, W., H. Ng, and D. Lim (2001): A Machine Learning Approach to Coreference Resolution of Noun Phrases. / Computational Linguistics 27 (4), pp. 521–544.
5. Yimeng Zhang, Yangbo Zhu. Machine Learning for Coreference Resolution: Recent Developments, pp.1-23
6. https://www.cs.cmu.edu/~yimengz/papers/Coreference_survey.pdf

MORFOANALIZATOR BOSQICHI UCHUN QORAQALPOQ TILIDAGI FE'L SO'Z TURKUMINI FORMALLASHITISH MODULI

Otemisov Aziz Zarlikbaevich
Qoraqalpoq davlat universiteti Amaliy filologiya kafedrasi mudiri, filologiya fanlari bo'yicha falsafa doktori (PhD), docent.
E-mail: utemisov.aziz@gmail.com

Sharbaev Jaras Bayrambay og'li
Qoraqalpoq davlat universiteti
Lingvistika (qoraqalpoq tili) 1-kurs magistranti.
E-mail: sharbaevj77@gmail.com

Аннотация: Каракалпакский язык принадлежит агглютинативным языкам. По этой причине степень вариации аффиксов в нем высока. В частности, глагольная группа слов лексически и грамматически очень богата по сравнению с другими группами слов. Стадия формализации глагола помогает идентифицировать корневые глаголы и другие особенности глагола в процессе анализа. Кроме того, этап формализации обеспечивает правильное добавление аффиксов к глаголам. Приведенные в статье примеры взяты из книг опытных специалистов, проводящих исследования в этой области.

Ключевые слова: морфологический анализ, формализация, глагол, категории глаголов, лемма.

Annotation: The Karakalpak language belongs to agglutinative languages. For this reason, the degree of variation of affixes in it is high. In particular, the verb group of words is lexically and grammatically very rich compared to other word groups. The verb formalization stage helps to identify root verbs and other features of the verb during the analysis process. In addition, the formalization stage ensures that affixes are added to verbs correctly. The examples given in the article are taken from books by experienced specialists conducting research in this area.

Keywords: morphological analysis, formalization, verb, verb categories, lemma.

Kirish. Morfoanalizator yaratish bir qancha bosqichlardan iborat. Biz so'z qilmoqchi bo'lgan fel so'z turkumini formallashtirish bosqichi ham shu jumladan o'z ahamiyatiga ega hisoblanadi. Hoxlagan so'z turkumuni formallashtirmasdan turib morfoanalizator yaratish imkonsiz. Foormallashtirish bosqishi morfologik tahlil jarayonida so'zlarga affikslarning tog'ri qo'shilishi, o'zak va qo'shimchalarni aniq ayirish, shaxs-son, zamon formalarini v.b. aniqlashda yordam beradi.

Asosiya bo'lim. Morfologik analizni amalga oshirish uchun tilning grammatik qoidalari formal holatta kiritiladi. Morfologiya yoki sintaksisga kompyuter orqali munosabatda bo'lish – bu tabiiyiy tilni modellashtirish, algoritm, tahlil yoki parsing orqali amalga oshiriladi [1:101].

M. Abjalovaning "Tahrir va tahlil dasturlarining lingvistik modullari" (Toshkent – 2019) monografiyasida o'zbek tilidagi fe'l so'z turkumini formallashtirishning namunasi berilgan.

Qoraqalpoq va o'zbek tillarining turkiy tillar ekanini va grammatik xususiyatlarining bir-biriga o'xshashligini hisobga olib, namunadagi kabi formallashtirishdan foydalanamiz.

Fe'l so'z turkumi boshqa turkumlarga nisbatan boy va ko'p qirrali grammatik kategoriyalarga ega. Bu izlanishda bor bo'lgan nazariy adabiyotlarga suyangan holda fe'llarno formallashtirishga etibor qaratamiz.

Tub fe'l (*lemma*) Vning kategorial shaklini yaratishda to'rt turdagi morfemalar guruhini etiborga olish kerak:

1. Nisbat kategoriyasi (x).
2. Shaxs-son kategoriyasi (y).
3. Zamon kategoriyasi (z).
4. Mayil kategoriyasi (m) [2: 16-17].

Formal grammatika nuqtaiy nazaridan kategorial shakllarni pastdagidek formallashtirish mumkin:

1Nisbat kategoriyası – x

1.1.Tub nisbat – x1,Ø morfema:teristi, juwdı, qazdı, qaradı t.b.

1.2.O'zlik nisbat – x2

x2 = -n, -ın/-in: juwındı, tarandı, saylandı, orandı h.t.b.

x2 = -l, -ıl/-il: keril, búgil, shomıl h.t.b.

x2 = -lıq/-lik: soqlıq, qıslıq, atlıq, kónlik h.t.b.

x2 = -q/-k, -ıq/-ik: jetik, tınıq, tolıq h.t.b.

1.3. Sheriklik nisbat – x3

x3= ıs/-is : salıs, kóris, tartıs h.t.b.

x3= -s : oylas, sóyles, tóbeles h.t.b.

1.4. O'zgalik nisbat- x4

x4 = -dır/-dir:aldır, kóndir, sóndir, jandır, kúldir h.t.b.

x4 = -tır/tir: tiktir, arttır, sattır h.t.b.

x4 = -t:oqıt, oynat, súyret, bosat h.t.b.

x4 = -ıt/-it: qorqıt, úrkit h.t.b.

x4 =-qız/-kiz: jatqız, satqız h.t.b.

x4 = -ǵız-giz:kitgiz, kúlgiz, oyǵız, jazǵız h.t.b.

x4-qar/-ker (-ǵar/-ger):jatqar, qutqar, ótker, pitker h.t.b.

x4-ar/-er/-ır:qaytar, shıǵar, asır, pisir, batır h.t.b.

1.5.Belgisizlik nisbat – x5

x5 = -l, ıl/-il:qazıldı, kómildi, egildi, súrildi h.t.b.

x5 = -ın/-in:salındı, julındı, tazalandı h.t.b.

2.Shaxs-son kategoriyasi *(y)*. Hozirgi qoraqalpoq tilida shaxs-son affikslari:

Shaxs-son asosi	**Shaxs**	**Birlik**		**Ko'plik**	
		Affiks	**Model**	**affiks**	**model**
	I	*-man/-men*	*y1bir 1*	-(ı/i)mız	*y1 ko'p 1*
	II	*-sań/-seń*	*y1bir2*	-sız/-siz	*y1 ko'p 2*
	III	*-dı/-di*	*y1bir3*	-dı/di	*y1 ko'p 3*

3.Zamon kategoriyası –z.

z1 – o'tgan zamon *(sintetik shakillari)*

z1 (1)=-dı/di, -tı/ti: (anıqlıq meyil forması) isledi, keldi, ketti, bardı, kórdi, isletti, oqıttı;
z1 (2)= - ǵan/-gen: (kelbetlik feyildiń forması) barǵan, kelgen, kórgen, bilgen;
z1(3)= -ıp/-ip, -p: (hal feyil formaları) barıppan, kóripseń, oqıpsań, jılaptı,
(analitik shakillar)
z1(4)= -ǵan/gen edi: barǵan edim, sóylegen edim, aytqan edim, islegen edim;
z1 (5) = –tuǵın edi: baratuǵın edi, biletuǵın edi, baratırǵan edi;
z1 (6)= -ar/-er, r edi: biler edi, sorar edi, qılar edi, ushar edi;
z1 (7)= -mas/-mes edi: tınlamas edi, islemes edi, kórmes edi;
z1 (8)= -jaq edi: barajaq edi, barajaq edim;
z1 (9)= -ıwshı/-iwshi edi: barıwshı edi, aytıwshı edi, tınlawshı edi;
z1 (10)= -maqshı/-mekshi edi:kórmekshi edi, bilmekshi edi, barmaqshı edi;
z1 (11)= -dı/di eken: baradı eken, keledi eken, biledi eken, isleydi eken.

z2 – hozirgi zamon

z2 (1)= -man/men: otırman, júrmen, kiyatırman, baratırman;
z2 (2)= -ıp/-ip, -p: júrippen, otırıptı, júripti;
z2 (3)= (ı/i) –uda/ -ude, -maqta/-mekte: gúlleniwde, rawajlanbaqta, teriwde, gúllenbekte;
z2(4) = -ıp/-ip, -p atır:kelip atır, baqlap atır, tańlap atır, uyıqlap atır;
z2 (5)= -ıp/-ip, -p tur: sóylep tur, qarap tur, aytıp tur;
z2 (6)= -ıp/-ip, -p júr: islep júr, tınlap júr, kórip júr;

z3 - kelar zamon

z3 (1) = -a, -e, -y -man/men (-ban/-ben, -pan/-pen), -sań/-seń, -dı/-di: baraman, oqıyman, keleseń, kóreseń, orınlaydı, quradı;
z3 (2) = -ar/-er, -r (bolımlı), -mas/-mes : bararman, jazarman, kelerseń, bilerseń, jazar, kelmes, kórmes, barmas, tınlamas;
z3 (3) = -a/-e, -y +jaq: barajaq, kórejaq, kelejaq, tınlayjaq, baramız, kóremiz, tınlaymız;
z3 (4) = jaq emes: barajaq emes, kórejaq emes, tınlayjaq emes;
z3 (5) = -maqshı/-mekshi: barmaqshıman, kórmekshimen, esitpekshimen;[Dáwletov A, Dáwletov M, Qudaybergenov M.,2010:159-166]

4.Mayil kategoriyasi – (m)

m1 – buyruq mayil

m1(1)= -(a) yıq/-(-e) yik: barayıq, oqıyıq, tınlayıq, bileyik;
m1(2)= -ayın/-eyin: oqıyın, tınlayın, sóyleyin, kóreyin;
m1(3)= -ǵır/-gir, -qır/-kir: qırılǵır, kórmegir, óspegir, joq bolǵır;
m1(4)= -shı/-shi: alshı, bershi, kórshi;
m1(5)= -sań/-seń, -sańa/-seńá: qoysańa, barsańa, barasań, tınlaysań, esiteseń;
m1(6)= -ma/-me: alma, kórme, barma, tınlama, ketpe;

m2 – istak mayil

m2 (1)= -ǵay/-gey, -qay/key: barǵay, kelgey, kórgey, bilgey, barǵayman, kelgeymen, kórgeymen;
m2 (2)= -ǵı/gi, -qı/-ki kel: aytqım keledi, barǵım keledi, kórgim keledi;
m2 (3) = -sa/-se edi, eken: barsa eken, kórse eken, barsa edi, kórse edi;
m2 (4) = -ar/-er, -r edim: kúter edim, oyar edim, isler edim

m3 – shart mayil

m3 (1) = -sa/-se: oqısam, oqısań, oqısa, kelsem, kelseń, kelse;

m3 (2) = -ǵan/-gen bol: kelgen bolsa, oqıǵan bolsa, kórgen bolsa;

m3 (3) = -tuǵın bol: kóretuǵın bolsa, keletuǵın bolsa, baratuǵın bolsa;

m4 – maqsad mayil

m4 (1) = -maqshı/-mekshi (-baqshı/-bekshi, -paqshı/-pekshi): barmaqshı, oqımaqshı, kórmekshi, bilmekshi;

m4 (2) = -maqshı/-mekshi (-baqshı/-bekshi, -paqshı/-pekshi) + edi, boldı: barmaqshı edi, kórmekshi boldı, ketpekshi edi;

m4 (3) = -(a/e) jaq: barajaqpan, kórejaq, bilejaq;

m5 – aniqlik mayil:

m5 (1) = -ǵan/-gen : barǵanman, kórgenmen;

m5 (2) = - ıp/-ip, -p+ atır: islep atırman, tıńlap atırsań;

m5 (3) = - ıp/-ip, -p+ júr: islep júr, tıńlap júr, kórip júr;

m5 (4) = - ıp/-ip, -p+ otır: kelip otır, esitip otır, kórip otır;

m5 (5) = -tı/-ti, -t: ketti, keldi, bardı, jetti, jolıqtı. [Dáwletov A, Dáwletov M, Qudaybergenov M.,2010:140-148]

Fe'llarni formallashtirish bosqichida etiborni qaratish kerak bo'lgan ahamiyatli vazifalardan biri bu – fe'llarning yasalish qo'idalarini bilish va unga amal qilishdir. Masalan, fe'llar yasalish bosqichida morflarning to'g'ri kelish va kelmaslik darajasiga qarab dasturni yo'naltirish lozim. Bunda asosan qoraqalpoq tilida morflarning ko'p variantliligini yoddan chiqarmaslik kerak. Yuqoridagi fikr va mulohazalardagi vazifalar to'liq va xatolarsiz amalga oshirilgan taqdirda gina ko'p vaqt tejab, morfoanalizator yaratish mumkin.

Yakuniy qism. Harqanday tilda qaysi so'z turkumini morfoanalizator bosqichi uchun modellashtirishga qaramasdan, eng birinchilardan bo'lib unda asosiy vazifani ijro etuvchi lingvist va dasturchining o'z kasbining yetuk mutaxasisi bo'lishi talab qilinadi. Tabiiy tilni modellashtirish malum darajada murakkab ishlardan biri hisoblanadi. Sababi tabiiy til malum bir formula yoki qotib qolgan qoidalar asnosida yaralmaydi. Uni insonlar yaratadi, rivojlantiradi. Mana shu rivojlanish bosqichlarida yangiliklar, yangi so'zlar hatto yangi tillar ham paydo bo'lishi mumkin. Tilni tashuvchi asosiy vosita inson bo'lgani tufayli uni cheklashning yoki qolibga solib qoyishning iloji yoq. Huddi shu sabab tufayli uning ichki va tashqi muhitini hisobga olib, teran anglagan holda modellashtirish bosqichini amalga oshirish mumkin. Shunda gina qoygan maqsadga erishisa bo'ladi.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. Kompyuter lingvistikasi. Toshkent, "Nodirabegim", 2021.
2. Abjalova M. Tahrir va tahlil dasturlarining lingvistik modullari. Toshkent, 2019.
3. Dáwletov A, Dáwletov M, Qudaybergenov M Házirgi qaraqalpaq ádebiy tili (morfemika, morfonologiya, sóz jasalıw, morfologiya). Nukus, "Bilim" – 2010.
4. Xolmanova Z. Kompyuter lingvistikasi. Toshkent, 2019.
5. Elov B., Hamroeva Sh., Elova D. Morfologik analizatorni yaratish usullari. O'zbekiston: til va madaniyat. Amaliy filologiya masalalari. Toshkent. 2022. 67-86-v. www.researchgate.net

6. Sulevmanov, D., Gatiatullin, A., Prokopyev, N., & Abdurakhmonova, N. (2020, November). Turkic morpheme web portal as a platform for turkology research. In 2020 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-5). IEEE.
7. Abdurakhmonova, N. (2016). The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta, 2(38), 12-7.

UNDOV SO‘ZLAR PRESUPPOZITSIYA SIGNALIZATORI SIFATIDA

Umurzakova Gulchehra Egamberdievna
O‘zbekiston Milliy Universiteti
Fakultetlararo ingliz tili kafedrasi kabinet mudiri.
E-mail: gulidil425@gmail.com

Annotatsiya: Undov so‘zlar matnni semantik jihatdan murakkablashtirish xususiyatiga ham ega. Bunday vaqtda matn sintaktik strukturasida birdan ortiq axborot siqiq tarzda aks etadi. Matnda ifodalangan axborotlardan biri ochiq, boshqasi undov so‘zning presuppozitsiyasi sifatida yashirin tarzda yuzaga chiqadi. Undov so‘zning presuppozitsiyaga ishorasi natijasida matn axborotlarga boy bo‘ladi.

Kalit so‘zlar: undov,axborot, sintaktik struktura, presuppozitsiya, pragmatik xususiyat.

Annotation: Exclamatory words also have the feature of semantically complicating the text. At this time, more than one information is densely reflected in the syntactic structure of the text. One of the information expressed in the text is revealed openly, and the other is hidden as a presupposition of the exclamation. As a result of the exclamation pointing to the presupposition, the text is rich in information.

Keywords: exclamation, information, syntactic structure, presupposition, pragmatic text.

Аннотация:Восклицательные слова имеют также особенность смыслового усложнения текста. В это время в синтаксической структуре текста плотно отражается более одной информации. Одна из выраженных в тексте сведений раскрывается открыто, а другая скрыта как предпосылка восклицания. В результате восклицания, указывающего на пресуппозицию, текст становится информативным.

Ключевые слова: восклицания, информация, синтаксическая структура, пресуппозиция, прагматика текста.

Undov so‘zlar matnni semantik jihatdan murakkablashtirish xususiyatiga ham ega. Bunday vaqtda matn sintaktik strukturasida birdan ortiq axborot siqiq tarzda aks etadi. Matnda ifodalangan axborotlardan biri ochiq, boshqasi undov so‘zning presuppozitsiyasi sifatida yashirin tarzda yuzaga chiqadi. Undov so‘zning presuppozitsiyaga ishorasi natijasida matn axborotlarga boy bo‘ladi. Ingliz tilidagi undov so‘zlarning pragmatik xususiyatini tekshirganS.Yu.Mamushkina ham bu so‘zlarning matnni mazmunan boyitishini qayd etadi. Uning fikricha, undov so‘zlar kogeziya vazifasini bajarar ekan, matnni hosil qilish jarayonida uni tashkillantiruvchi muhim omil hisoblanadi, undov so‘zlar nutqni emfatik jihatdan relevant hamda axborotga to‘la, boy qiladi[Мамушкина 2003.175] .Undov so‘zlar nutq egasining hissiy kechinmalarini ifodalash bilan birga, ba’zan presuppozitsiya signalizatori vazifasini bajarib, jumlani mazmunan murakkablashtiradi. Masalan:

1. *Oyim endi o‘rnidan turayotgan edi, eshik shaxt bilan ochildi. Uyga avval ammam, ketidan sovuq g‘urillab kirdi… Oyim ildam borib, uning qo‘lidagi katta tugunni oldi. Dadam ham shoshib qoldi.*

– Iye, iye! – dedi chehrasi yorishib. Ammo boshqa gapirolmadi. Og'zida nosvoy bor edi. (O'.Hoshimov. «Dunyoning ishlari»)

Birinchi matnda takror qo'llangan *iye* undov so'zi orqali nutq egasining ajablanishi bilan birga *«Kelishingizni kutmagan edik»* tarzidagi yashirin axboroti ham anglashiladi. Bu axborot undov so'zning presuppozitsiyasi sifatida matn tarkibida voqelanadi.

Undov so'zlar ba'zan gap tarkibida birdan ortiq presuppozitsiyalarni hosil qiladi. Bunday vaziyatda faqat kontekst asosida asosiy va ikkinchi darajali presuppozitsiyalarni farqlash mumkin bo'ladi. Masalan:

1. *– Kecha aytgan joyimga bordingizmi?*

– Eh! – Samad bo'shashib joyiga o'tirib qoldi.

2 . *"**Oh! oh! oh**!" A young fellow lying asleep… raised **a moan of pain...*** (J.London. «Grit of women»*1p*) (*"Вой-вой-вой-эй", – деди ёш йуловчи ётиб, симиллаб бошланаётган оғриқнинг кучайишидан қаттиқ инграб*)

Birinchi misolda qo'llangan *Eh!* undovi *"Kecha aytilgan joyga bormadim", "Aytilgan joyga borish xayolimdan ko'tarilibdi", "Aytilgan joyga bormaganimdan afsusdaman"* presuppozitsiyalarini hosil qiladi. Bunday vaziyatda so'roqqa javob sifatida anglashiluvchi *"Kecha aytilgan joyga bormadim"* presuppozitsiyasi matnda bayon etilmoqchi bo'lgan asosiy axborot sifatida muhim ahamiyatga ega. *"Aytilgan joyga borish xayolimdan ko'tarilibdi"* presuppozitsiyasi asosiy presuppozitsiyani qo'shimcha ma'lumot bilan to'ldirishga xizmat qiladi. *"Aytilgan joyga bormaganimdan afsusdaman"* presuppozitsiyasi esa nutq egasining voqelikka hissiy munosabatini ifoda etadi.

Keyingi misolda esa *"Oh! oh! oh"* undov gapi orqali nutq egasining emotsional holati bilan bog'liq quyidagi presuppozitsiyalarni anglash mumkin bo'ladi: *"Jonim og'riyapti", "Endi nima qilaman?"* kabilar.

Anglashiladiki, emotsional undovlar orqali gapda ifodalangan presuppozitsiyalarni so'zlovchi kommunikativ maqsadi asosida bir-biridan farqlash mumkin bo'ladi.

Shunisi xarakterliki, ba'zan sintaktik strukturada undov so'z ishora qilgan presuppozitsiyaga kontekstda berilgan so'zlovchi izohi yordamida oydinlik kiritiladi.

*– **Oh,** I'm sure you remember him, Mr. Haffigan.* («John Bull's other island» B.Show.36p) (*Ия, аминманки, сиз уни эслайсиз, жаноб Хаффиган*)

Ingliz tilidagi keyingi misolda esa *oh* undovi orqali *"Siz uni esingizdan chiqarishingiz mumkin emas"* presuppozitsiyasi anglashiladi. Bu axborot gapning "I'm sure you remember him, Mr. Haffigan" (Siz uni eslaysiz, janob,Xaffigan) qismi orqali ham anglashilib turadi.

Gelmxoltz (Helmholtz) va boshqalarning izlanishlariga ko'ra,og'iz bo'shlig'i va lablarning ,unli tovushning tembri bu holatga ong orqali boshqarilishi,va insonni turli holat va vaziyatlardagi psixologik jarayonga qayta javobi deb xulosa qilinadi. Hissiyot tamoyillari,og'iz va burundan chiqayotgan nafas *(un) pooh* yoki *pish* tovushi chiqishiga sabab bo'ladi. Gelmxoltz (Helmholtz) ning izlanishlariga ko'ra,insonning fiziologik tuzilishi va ongning(miya)ning hissiyotlarga javobi o'laroq tovush xosil bo'lardi. [Charles Darwin.1897].

1. *А сколько вам лет, князь?*

– Двадцать шесть.

*– **Ух!** А я думал, гораздо **меньше.***

*– Да, говорят, у меня лицо **моложавое.*** («Идиот»Федор Михайлович Достоевский.4стр)

2. *"**Wow! You've gotten so tall**!" she said, as Alex held her tight around her waist, and Carmen slowly put her arms around her, as though she were a stranger and not her child.* («The right time»Daniella Steel.27p)(*Karmen sekingina quchoqlab dedi: " Voy, rosa katta bo'libsan!" Aleks uning belidan qattiq quchoqlab olar ekan.Go'yo ular ona-bola emas begonalardek edi.*)

Ma'lum bo'ladiki, emotsional undovlarning bu kabi pragmatik vazifalari nutq egasining kommunikativ maqsadi bilan bog'liq holda voqelanadi.Ba'zi emotsional undov so'zlar bir vaqtda turli pragmatik vazifalarni ado etadi. Jumladan, gapda undov so'z orqali nutq egasining voqelikka baho munosabati hamda voqelik haqidagi qo'shimcha axboroti ifodalanadi. Masalan, *«Adham hovlisida qo'shni xotinning shang'illagan ovozini eshitdi va xo'rsinib, «Uff!» dedi»* gapida qo'llangan *obbo* undov so'zi so'zlovchining e'tirozini aks ettirish bilan birga, voqelik haqidagi qo'shimcha axborot (*Yana kelibdi*) ga ham ishora qiladi. Undov so'z orqali ifodalangan mazkur presuppozitsiya matn semantik strukturasini murakkablashtiradi.

Presuppozitsiyaning yuzaga chiqishi nutqiy vaziyat yoki kontekstni taqozo etadi. Agar mikromatn tarkibidan *«Adham hovlisida qo'shni xotinning shang'illagan ovozini eshitdi»* jumlasi tushirib qoldirilsa, *«Uff!»* undov so'zining presuppozitsiyaga ishora qilayotgani ma'lum bo'lmay qoladi. Ko'rinadiki, undov so'zlar nutqiy vaziyat yoki kontekst bilan bog'lanmasa, presuppozitsiyaga yo'l bekiladi.Emotsional undovlar orqali ifodalangan presuppozitsiya ba'zan irreal (xayoliy, tasavvurdagi) holatda bo'lishi mumkin. Ushbu matnda qo'llangan *essiz* emotsional undovi nutq egasining achinish hissini ifoda etish bilan birga, uning xayoli, tasavvuridagi voqelikka ham ishora qiladi .

Shunisi xarakterliki, emotsional undovlar dialogik nutqda matn qismlarini bog'lovchi vosita sifatida ham qo'llanadi va bu vaziyatda nutq egasining matnning oldingi qismiga nisbatan munosabatini ifoda etadi. T.V.Belousning qayd etishicha, «dialogik diskursda undov so'zlar diskurs markeri va oldingi qism bilan aloqani ta'minlovchi vosita sifatida ishtirok etib, metakommunikativ vazifa bajaradi. Ular diskurs replikalarining bog'lanishini ta'minlashda muhim rol o'ynaydi, nutq egasining suhbatdoshi replikasiga e'tibori, suhbatning umumiy kechishi haqida signal beradi, tinglovchi tomonidan diskursning qabul qilinishini yengillatish, uning yaxshi tushunilishiga erishish orqali kommunikantlar- ning nutqiy aloqasini ta'minlash asosida kommunikatsiyaning muvaffaqiyatli kechishiga erishish imkoniyatini yaratadi»[Белоус T.B.2006.187c.]

1. *"Y' all ain't seen nothing', "Karl exclaims," Till you've seen my grandmother, boy, **whowee**! Could she raise the dead with a whipping'! I remembered this one night, my grandmother wanted me to do something and I said, "No". Well, I did 'n I saw her start to come after me, so I grabbed the belt off the wall so she couldn't use it on me. In addition, do you know what she did? She went upstairs got an old pipe and come after me with it. **Whowee,** did I run. Run right out the house into the snow with no shoes on!"* (Clarence Major. «Dirty bird blues»395p)

(*– Sizlar hech nima ko'rmadinglar-a. Karl hayajonda, meni buvimni ko'rmaguningizcha, "voooy" bolalar u hattoki o'likni ham tiriltiradi-yu. Yodimda bir kuni u menga ish buyurdi, men "yuq" dedim. "haa"uni ortimdan "shoshmay tur hali men seni….", deb kelayotganini ko'rib devorda osig'lik kamarni yulib olib, menga ishlatmasligi uchun shu qadar tez qochyapmanki . Bilasizmi nima qildi? U zinadan tushib, eski quvur (truba)ni olib ortimdan yurdi. "Vuuy"yugurdimmi men, men yalang oyoq qorli ko'chaga qochdim.*

(Klerens M. «Qora qush raqsi»)

Emotsional undovlar nutq egasining shaxs xususiyatlari bilan bog‘langanda ham pragmatik xususiyatga ega bo‘ladi. Bunda emotsional undovlar yoki ular ishtirokidagi sintaktik qurilmalar orqali nutq egasining yoshi, jinsi, kasb-kori kabilarga ishora qilinishi kuzatiladi. Ba’zi hollarda emotsional undovlar orqali nutq egasining nafaqat jinsi, balki yosh xususiyatiga ham ishora qilinadi.Matnda qo‘llangan *ho* undov so‘zi orqali yosh bolaga ishora qilinayotganligini sezish mumkin. Albatta, bu o‘rinda undov so‘zning shaxs yoshiga ishora qilishi kontekstsiz reallashmaydi.

Foydalanilgan adabiyotlar:

1. Abelin A ,1999. Studies in Sound Symbolism, Doctoral Dissertation. MA Thesis, Goteborg University.
2. Anderson ER, 1998. A grammar of iconism. London: Associated University Publications.
3. Белоус Т.В. Когнитивно-прагматический анализ междометий в англоязычном дискурсе: Дисс. ... канд. филол. наук. − Санкт-Петербург, 2006. – 187 с.
4. Charles Darwin.The Expession of the emotions in man and animals. New York.D.Appleton and Company.1897
5. Мамушкина С.Ю. Семантика и прагматика междометий в современном английском языке: Дисс. ... канд. филол. наук. – Самара, 2003. – 175 с.

NLPDA MATNLARNI UMUMIYLASHTIRISH VA LEKSIK TAHLIL

Fayzullayeva Zarnigor Inatillayevna
PhD, v.b. dotent Alxorazmiy nomidagi
Toshkent axborot texnologiyalar Universiteti.
E-mail: zarnigor18z02@gmail.com

Karimov Nodirbek Nosirjon o‘g‘li
Doktarant Namangan davlat Universiteti.
E-mail: topcoder1600@gmail.com

Annotatsiya: Hozirgi kunda NLP da matnni umumlashtirish yordamida erishilgan natijalarga juda ham ko‘pligi e'tiborga molik, ushbu maqolada, biz **chuqur o'rganishdan foydalanib, matnni** yig'ish uchun zarur bo'lgan barcha tushunchalarni qamrab olgan holda, uni yaratish bo'yicha bosqichma-bosqich jarayonni ko'rib chiqdik va birinchi matnni umumlashtirish modelimizni amalga oshirdik, Umumlashtirish uchun esa xulosa chiqarish turlarini va amalga oshirish qadmlarini hamda jarayonlarini ko‘rib chiqdik.

Kalit so‘zlar: NLP, Chuqur o‘rganish, leksik tahlil, sun’iy intellekt, xulosa chiqarish, model, vector, matrissa

Abstract: It is remarkable that there is so much progress in NLP using text summarization, in this article we will use deep learning to create a step-by-step approach to text summarization, covering all the concepts needed for text summarization. We looked at the -step process and implemented our first text summarization model, and for Summarization we looked at inference types and implementation steps and processes.

Keywords: NLP, Deep learning, lexical analysis, artificial intelligence, summarization, model, vector, matrix

Аннотация: Примечательно, что в НЛП с использованием обобщения текста достигнут такой большой прогресс, что в этой статье мы будем использовать глубокое обучение для создания пошагового подхода к обобщению текста, охватывающего все концепции, необходимые для обобщения текста. Пошаговый процесс и реализовали нашу первую модель суммирования текста, а для суммирования мы рассмотрели типы вывода, а также этапы и процессы реализации.

Ключевые слова: НЛП, глубокое обучение, лексический анализ, искусственный интеллект, вывод, модель, вектор матрица.

Tabiiy tilni qayta ishlash (NLP) - bu mashinalarga inson tilini o'qish, shifrlash, tushunish va qayta ishlash imkonini beruvchi eng mashhur AI texnologiyasi hamda inson tili ma'lumotlarni yozilishi, og‘zaki va tartibga solinishi bilan boshqara oladigan mashinalarni qurishdir. Matnni bashorat qilish va his-tuyg'ularni tahlil qilishdan nutqni aniqlashgacha, NLP mashinalarga inson aqli va qobiliyatlarini ta'sirchan tarzda taqlid qilish imkonini beradi. NLP ilovalarini yaratish uchun til modellari kalit hisoblanadi. Biroq, noldan murakkab NLP tili modellarini yaratish ancha murakkabdir, shuning uchun AI va ML ishlab chiquvchilari va tadqiqotchilari oldindan

tayyorlangan til modellari bilan birgalikda ishlashadi. Tilni modellari nutqni tanib olish, belgilarni optik aniqlash, qo‘l yozuvini aniqlash, mashina tarjimasi va imloni tuzatish kabi NLPning turli masalalarni yechish uchun qo‘llanadi. Ushbu modellar matnni mashina tiliga o‘qitish uchun teglash texnikasidan foydalanadi, bunda model topshiriqni bajarish uchun bitta ma’lumotlar to‘plamida o‘qitiladi hamda oldingi so‘zning xususiyatidan kelib chiqqan holda, gapdagi keyingi so‘zning qanday bo‘lishini taxmin qilish taklifnii beriladi. Bazi modellar yangi ma’lumotlar to‘plamida turli NLP funksiyalarini bajarish uchun o‘zgartiriladi. Ushbu vazifaning maqsadi ma’lum bir tilda so‘zlar ketma-ketligining paydo bo‘lish ehtimolini o‘rganishdir. Oldindan o‘rgatilgan NLP til modellarning bir nechta turlari majud bo‘lib, ularning ishlash funksiyalarini ko‘rib chiqamiz.

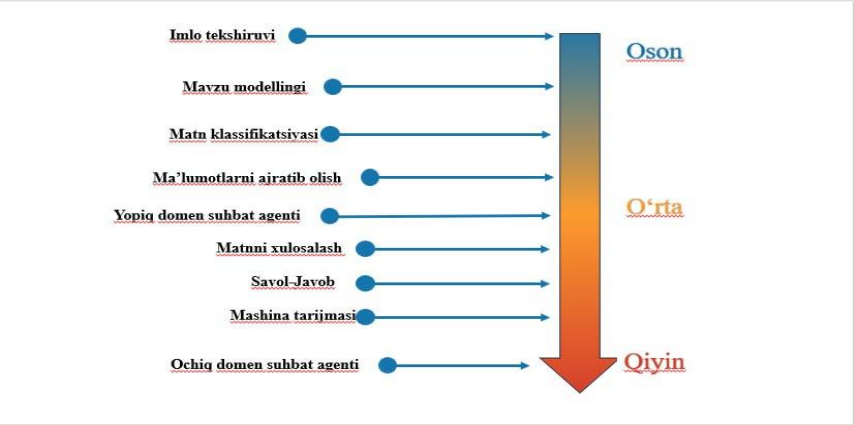
GPT-4 (generativ oldindan o'rgatilgan transformator) - OpenAI tomonidan ishlab chiqilgan katta til modeli (LLM). Bu GPT til modellari seriyasining to‘rtinchi avlodi bo‘lib, 2023-yil 14-martda birinchi marta tadbiq qilingan. GPT-4 multimodal model bo‘lib, u matn hamda tasvirlarni qabul qiladi. Bu faqat matn kiritish sifatida qabul qilinishi mumkin bo‘lgan oldingi GPT modellariga qaraganda ko‘p qirralidir. Ushbu model ChatGPT Plus orqali ommaga ochiq tarzda nomoyon qilingan, uning tijorat API-siga esa kutish ro‘yxati orqali kirish mumkin. Ushbu model rivojlanish davomida keyingi tarkibni oldindan bilishga harakat qiladi. Bu uning insoniy qadriyatlarga mos kelishini va kerakli siyosatlarga mos kelishini ta’minlash uchun qilingan.

U yanada ijodiy va hamkorlikka asoslangan, ijodiy va texnik yozish vazifalarida foydalanuvchilar bilan yaratish, tahrirlash va takrorlash imkoniyatiga ega. U hali ham ishlab chiqilmoqda, lekin u turli xil ilovalar, jumladan, kontent yaratish, tarjima, yozishda yordam, ta’lim, mijozlarga xizmat ko‘rsatish va tadqiqotlar uchun kuchli vosita bo‘lish potentsialiga ega.

BERT (Transformatorlardan ikki tomonlama kodlovchi tasvirlar)- Google tomonidan ishlab chiqilgan NLP-dan oldingi trening texnikasi. U tilni tushunish uchun o‘z-o‘ziga e’tibor mexanizmiga asoslangan yangi neyron tarmoq arxitekturasi transformerdan foydalanadi. U ketma-ketlik transduktsiyasi yoki neyron mashina tarjimasi muammosini hal qilish uchun ishlab chiqilgan. Bu shuni anglatadiki, u kirish ketma-ketligini chiqish ketma-ketligiga aylantiradigan har qanday vazifa uchun mos keladi, masalan nutqni aniqlash, matnni nutqqa o‘zgartirish va hokazo.

Ushbu model ikkita alohida mexanizmni o‘z ichiga oladi: kodlovchi (matn kiritishni o‘qiydi) va dekoder (vazifa uchun bashorat ishlab chiqaradi). BERT mexanizmining maqsadi til modelini yaratishdir. BERT algoritmi 11 ta NLP vazifasini samaralibajaradi.

Matnni mashina tiliga o‘girish uchun ML (machine learning) foydalanamiz bu jarayon ko‘p sonli misollar asosida vazifalarni avtomatik tarzda bajarishni ko‘rishimiz mumkin. Matnni umumiyl ashtirish uchun Chuqur o‘rganish (deep learning, DL) sun’iy neyron tarmoqlari arxitekturalariga asoslangan mashinali o‘rganib chiqamiz. ML, DL va NLP sun’iy intellekt sohadagi barcha kichik maydonlar bo‘lib, ular orasidagi munosabatlar 1-rasmda tasvirlangan.



1-rasm. ML, DL va NLP orasidagi munosabatlar.

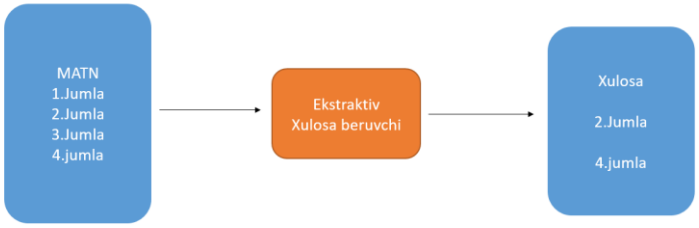
Ma'lumki matnni umumlashtirish bizning hayotimizga katta ta'sir ko'rsatishi mumkin bo'lgan tabiiy tilni qayta ishlash (NLP) ilovalaridan biridir. Raqamli ommaviy axborot vositalari va tobora o'sib borayotgan nashriyotlar bilan - kim foydali yoki yo'qligini aniqlash uchun butun maqolalar / hujjatlar / kitoblarni ko'rib chiqishga hozirgi kunda ko'pgina insonlar vaqtini qizg'onadi shuning uchun ham bu texnologiya allaqachon mavjud lekin o'zbek tili uchun ushbu model hali ishlab chiqilmagan.

Matnni umumlashtirish - bu hujjatning muhim nuqtalarini ajratib ko'rsatish orqali erishilgan asl matnning asosiy ma'lumotlari va asosiy g'oyalarini qamrab oladigan qisqa va yaxshi yozilgan xulosaga uzun matnli hujjatni ixchamlashtirish texnikasini anglatadi. Matnni umumlashtirish uchun ikki xil yondashuv qo'llaniladi:

- Xulosa chiqarish
- Abstraktiv xulosa

Xulosa chiqarish bu butun matn bo'yicha ko'p takrorlanadigan so'zlarni hamda ularning sinonimini izlaydi va **asl matndan muhim jumlalarni yoki iboralarni aniqlaymiz va faqat matndan ajratib olib,** chiqarilgan jumlalar xulosa sifatida taqdim qiladi. Xulosa chiqarishni ikki toifaga bo'lish mumkin – **ekstraktiv xulosa** va **mavhum xulosa**.

Ekstrativ xulosa chiqarish: Bu usullar matndan iboralar va jumlalar kabi bir nechta qismlarni ajratib olishga va xulosa yaratish uchun ularni bir joyga to'plashga tayanadi. Shuning uchun xulosa qilish uchun to'g'ri jumlalarni aniqlash ekstraktiv usulda eng muhim ahamiyatga ega.



2-rasm. Ekstrativ xulosa.

Mavhum xulosa: Bu usullar mutlaqo yangi xulosani yaratish uchun ilg'or NLP usullaridan foydalanadi. Ushbu xulosaning ba'zi qismlari hatto asl matnda ham ko'rinmasligi mumkin.

Abstraktiv xulosa Bu juda qiziq yondashuv. Bu erda biz asl matndan yangi jumlalar yaratamiz. Bu biz ilgari ko'rgan ekstraktiv yondashuvdan farqli o'laroq, biz faqat mavjud

jumlalarni ishlatganmiz. Mavhum umumlashtirish orqali tuzilgan jumlalar asl matnda mavjud bo'lmasligi mumkin va bu jarayon rasmda tasvirlangan:

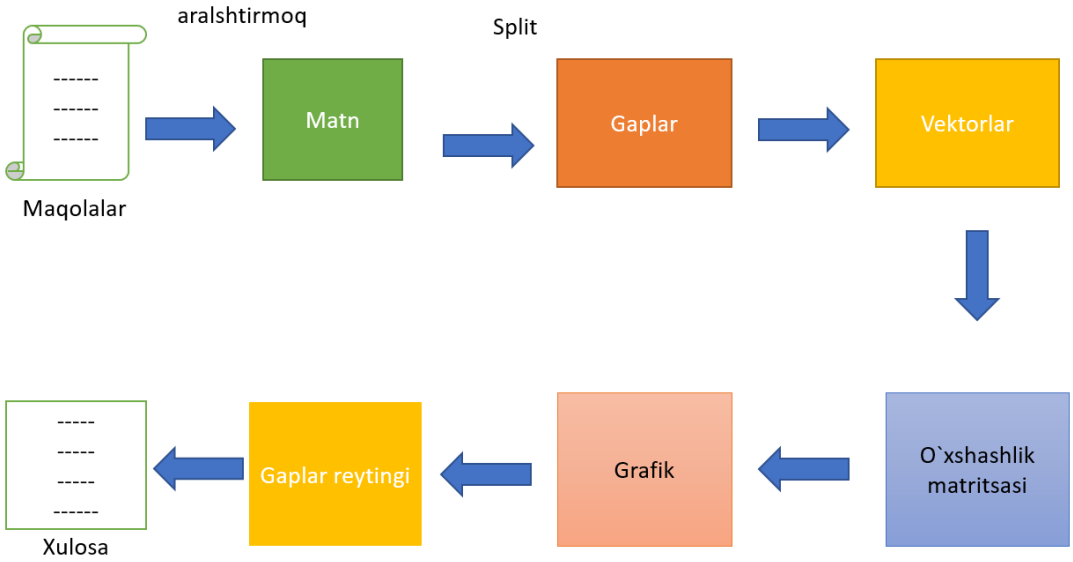


3-rasm. Abstraktiv xulosa.

Biz matnni umumiylashtirish uchun TextRank algoritmidan foydalandik.TextRank algoritminini amalaga oshirish uchun PageRank foydalandik. Ushbu ikki algoritm o‘rtasidagi o'xshashliklarni quyidagicha:

- jumlalarni ishlatamiz
- har qanday ikkita jumla o‘rtasidagi o‘xshashlik web-sahifaga o‘tish ehtimoliga ekvivalent sifatida ishlatiladi
- o‘xshashlik ballari PageRank uchun ishlatiladigan M matritsasiga o‘xshash kvadrat matritsada saqlanadi.

TextRank - bu ekstraktiv va nazoratsiz matnni umumlashtirish usuli. Keling, biz amal qiladigan TextRank algoritmi oqimini ko'rib chiqaylik:



4-rasm. TextRank algoritmini ishlash jarayoni.

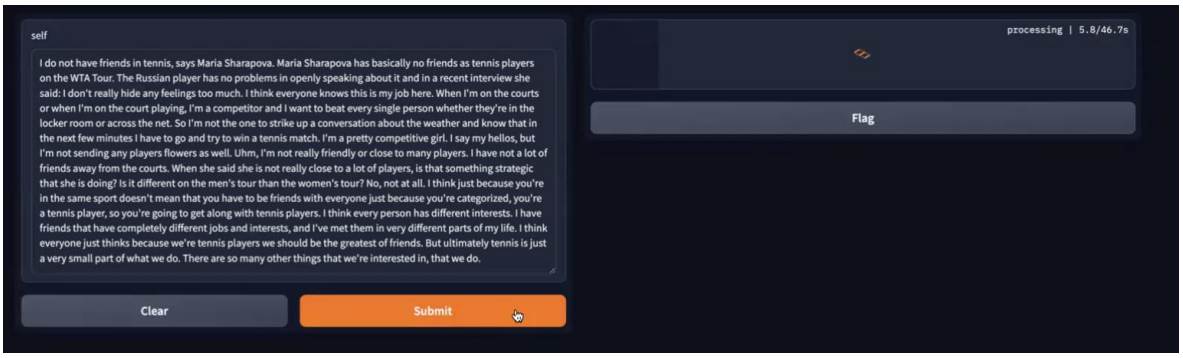
- birinchi qadam maqolalardagi barcha matnlarni birlashtiriladi
- matnni alohida jumlalarga ajratiladi
- keyingi bosqichda har bir jumla uchun vektor tasviri (so‘zlarni joylashtirish) topiladi
- jumla vektorlari orasidagi o‘xshashliklar hisoblab chiqiladi va matritsada saqlanadi
- keyin o‘xshashlik matritsasi jumlalar darajasini hisoblash uchun cho‘qqilar sifatida jumlalar va qirralar sifatida o‘xshashlik ballari bilan grafikga aylantiriladi.

- nihoyat, ma’lum miqdordagi yuqori o‘rinli jumlalar orqali yakuniy xulosa tashkil qilinadi

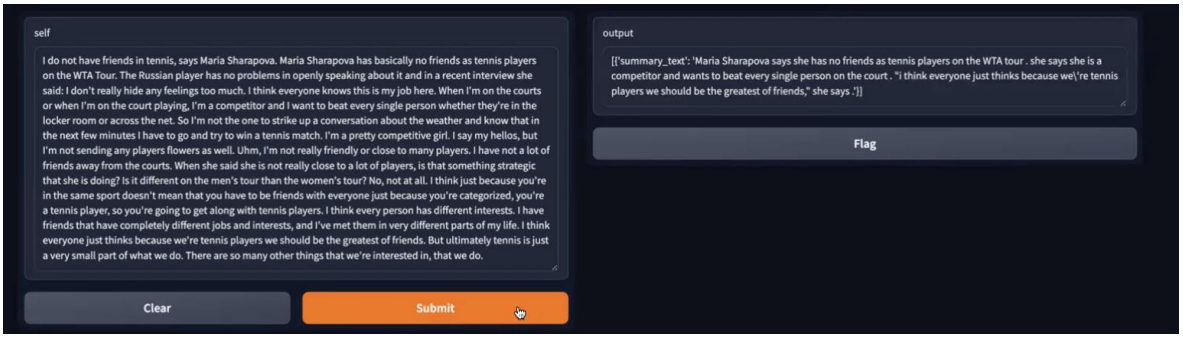
Matnni Pyhton dasturlash tilida umumiylashtirish jarayonnini quyidagicha



5-rasm. Matnni umumiylashtirih jarayoni interfeysi.



6-rasm. Matnni umumiylashtirih jarayoni.



7-rasm. Matnni umumiylashtirih jarayoni natijasi.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. Mashina tarjimasining lingvistik ta’minoti(monografiya).-Toshkent:Muharrir, 2018. 66-86 b.
2. Abduraxmonova N. O‘zbek tili elektron korpusining kompyuter modellari(monografiya).-Toshkent:Muharrir, 2021.165 b.
3. Напр., Krovetz R. (2000). Viewing morphology as an inference process. In *Artificial Intelligence*, 118, 277–294.
4. Напр., Sheremetyeva S., Nirenburg S., Nirenburg I. (1996). Generating Patent Claims from Interactive Input. In Proceedings of the 8th International Workshop on Natural Language Generation (Herstmonceux, Sussex, June 1996), 61–70.
5. Goldsmith, J. (2001). Unsupervised Learning of the Morphology of a Natural Language. In *Computational Linguistics*, 27(2), 153–198.
6. Ford A., & Singh R. (1991). Propedeutique Morphologique. *Folia Linguistica*, 25 (3–4), 549–575; Neuvel Sylvain (2002). Whole Word Morphologizer: Expanding the Word-Based Lexicon: A Nonstochastic Computational Approach. In *Brain and Language* 81, 454–463.

7. Neural Machine Translation (seq2seq) Tutorial – Режим доступа: https://www.tensorflow.org/tutorials/seq2seq.
8. Word2vec – Режим доступа: https://ru.wikipedia.org/wiki/Word2vec.
9. Klymenko, N. F. et al. (2014). Morfemno-slovotvirnyj fond ukrajins'koji movy jak doslidnyc'ka ta informacijno-dovidkova systema. In *Klymenko N. F. Vybrani praci*, pages 545–558, Kyiv.
10. Korpus ukrajins'koji movy. Accessible at: http://www.mova.info/corpus.aspx, retrieved 2017-03-15.

O‘ZBEK TILI KORPUSI MATNLARI ASOSIDA TIL MODELLARINI YARATISH

Elov Botir
Texnika fanlari falsafa doktori, dotsent.
Alisher Navoiy nomidagi
Toshkent davlat o‘zbek tili va adabiyoti universiteti.
E-mail: elov@navoiy-uni.uz
ORCID: 0000-0001-5032-6648

Abdullayev Abdulla
O‘qituvchi. “Urganch innovatsion university”
NTM ta’limni kredit tizimini boshqarish bo‘lim boshlig‘i.
E-mail: abdulla_abdullayev9270@mail.ru

Xudayberganov Nizomaddin
O‘qituvchi. Alisher Navoiy nomidagi
Toshkent davlat o‘zbek tili va adabiyoti universiteti.
E-mail: nizomaddin@navoiy-uni.uz
ORCID: 0000-0002-6213-3015

Annotatsiya: Nutqni tanib olish, mashina tarjimasi, POS teglash, matnni intellektual tahlil qilish, yangi matnni yaratish va so‘zlarni bashorat qilish kabi NLP vazifalarini hal qilish uchun til modellaridan foydalaniladi. Til modellari har qanday NLP vazifasining muhim komponent hisoblanadi. Ushbu maqolada til modellarini ishlab chiqish bosqichlari va o‘zbek tili korpusi matnlari asosida til modellarini yaratishning n-gramm metodi keltiriladi.

Kalit so‘zlar: til modellari, NLP, til korpusi, UzbNlp paketi, unigram, bigram, trigram, n-gramm, so'zni bashorat qilish.

Kirish

Tabiiy tilni qayta ishlash (Natural Language Processing, NLP) dagi til modeli – oldingi so‘zlar asosida gapda berilgan so‘zlar ketma-ketligining hosil bo‘lish ehtimolini aniqlaydigan statistik model. Ushbu model asosida matn konteksti asosida mos so‘zni bashorat qilish imkonini beradi. Til modellaridan nutqni aniqlash, mashina tarjimasi, imloni tuzatish va boshqa NLP vazifalarida keng foydalaniladi. Til modelini ishlab chiqish odatda katta hajmdagi namunaviy gaplarini mashinali o‘qitish (ML) qadamidan boshlanadi. O‘qitish jarayonida model so‘zlar ketma-ketligi bo‘yicha ehtimollik taqsimotini hisoblaydi. Kontekstdagi so‘zni qo‘yilgan NLP vazifasiga muvofiq bashorat qilish uchun *unigram, bigram, trigram* yoki *n-gram* usullaridan foydalanish mumkin.

Til modellari

Til modellari tabiiy tilni qayta ishlashning asosini tashkil qiladi. Til modeli matn haqidagi ma’lumotni mashinalar tushunadigan miqdoriy ma’lumotga aylantiradi [1,2]. Bugungi kunda til modellari asosida moliya, sog‘liqni saqlash, harbiy va shu kabi turli sohalardagi NLP ilovalari ishlab chiqilmoqda [3]. Kundalik hayotda mobil telefonda bashoratli matn kiritish yoki Google

qidiruv tizimida til modellaridan foydalaniladi. Demak, til modellari har qanday tabiiy tilni qayta ishlash dasturining ajralmas qismini tashkil qiladi. Ushbu maqolada til korpusidagi strukturlanmagan matn asosida unigram, bigram va trigram til modellarini yaratish va ular yordamida keyingi so‘zlarni bashorat qilish usullari keltiriladi [1,2,4].

Til korpusidan matnni o‘qish

Til modelini ishlab chiqish uchun o‘zbek tili milliy korpusidagi 2024 yil mart oyida kun.uz va qalampir.uz onlayn nashrlaridagi post/maqolalarni yuklab olamiz. Strukturlanmagan matnlar yuklab olinganidan so‘ng, matn faylidagi belgilarning umumiy sonini aniqlash lozim.

```
file = open("UzbCorpus.txt", "r")
rawReadCorpus = file.read()
print ("O'qitiladigan ma'lumotlar to'plamidagi belgilar soni: {}".format(len(rawReadCorpus)))
```

Toshkent davlat o‘zbek tili va adabiyoti universiteti kompyuter lingvistikasi va raqamli texnologiyalar kafedrasi ilmiy tadqiqotchilari tomonidan ishlab chiqilgan UzbNlp paketida o‘zbek tilidagi matnlarni tahlil qiluvchi bir qator funksiya va metodlar mavjud [5]. Ushbu paketni import qilish lozim:

```
import UzbNlp
from UzbNlp.tokenize import uzb_word_tokenize, uzb_sent_tokenize
```

Strukturlanmagan matnni boshlang‘ich qayta ishlash

Birinchi navbatda, matnlardagi barcha ortiqcha qatorlarni va maxsus belgilarni olib tashlash kerak [6]. Buni quyidagi kod bilan amalga oshiramiz:

```
import string
string.punctuation = string.punctuation +'“'+'”'+'-'+'’'+'‘'+'—'
string.punctuation = string.punctuation.replace('.', '')
file = open('UzbCorpus.txt').read()
# yangi qatorlar va maxsus belgilarni olib tashlash uchun ma'lumotlarni boshlang`ich qayta ishlash
file_new = ""
for line in file:
   line_new = line.replace("n", " ")
   file_new += line_new
preprocessedCorpus = "".join([char for char in file_new if char not in string.punctuation])
```

Yangi qatorlar va maxsus belgilarni olib tashlaganimizdan so'ng, **UzbNlp.tokenize** dan **uzb_sent_tokenize** va **uzb_word_tokenize** metodlari yordamida so‘zlar va gaplarni ajratib olish uchun korpus matnlarini parchalash mumkin [7,8]. Quyidagi dastur kodida korpusdan olingan dastlabki 5 ta gaplani va birinchi 5 ta so‘z chop qilingan:

```
sentences = sent_tokenize(preprocessedCorpus)
print("Korpusdagi birinchi 5 ta gap: ")
print(sentences[0:5])
words = word_tokenize(preprocessedCorpus)
print("Qayta ishlangan korpusdagi birinchi 5 ta so`z/token: ")
print(words[0:5])
```

[Bilgil va ogoh bo‘lg‘ilki dunyoda ilm o‘qimoq va ilm istamoqdin g‘araz-u maqsud ikki dunyoning saodatiga yetmoqdur.]
[Lekin qancha mulla va qancha olim bo‘lsang ham dilingdan axloqi zamima (buzuq xulqlar) va yomon fe’llarni chiqarib ul dilingni axloqi hamida va axloqi pisandidalar (yaxshi va maqtovga sazovor xulqlar) bilan ziynat-u oro bermasang hanuz ikki dunyoni saodatiga yeta olmassen, nechuk?]
[Zeroki ko‘ngul bir podshoh va o‘zga a’zolar fuqaro manzilindadur.]
[Agar ko‘ngul durust bo‘lsa, boshqa a’zolar ham durust bo‘lur.]
[Va agar ko‘ngul buzuq bo‘lsa, o‘zga a’zolar ham buzuq bo‘lur.]

[Bilgil], [va], [ogoh], [bo‘lg‘ilki], [dunyoda]

Keyingi qadamda, korpusdagi **nomuhim so'zlar (stopwords)**ni olib tashlash kerak. Stopwords – bu gapga hech qanday maxsus ma’no yoki ahamiyatga ega bo‘lmagan **"va", "biroq", "ammo"** kabi keng qo`llaniladigan so‘zlar. O‘zbek tili uchun nomuhim so‘zlar ro‘yxatiga namuna quyida keltirilgan:

```
UzbNlp.download('uzb_stopwords')
from UzbNlp.corpus import stopwords
stop_words = set(stopwords.words('uzbek'))
filtered_tokens = [w for w in words if not w.lower() in stop_words]
```

Unigram, bigram va trigram til modellarini yaratish

O‘zbek tilidagi matnlarga mos n-gramm qiymatlarni UzbNlp paketidagi metodlar yordamida yaratish mumkin [5]. N-grammalar korpusda uchraydigan n ta ketma-ket so‘zdan iborat ketma-ketlikdir. Masalan, **"Men kitob o‘qishni yaxshi ko‘raman"** gapida **"men", "kitob", "o‘qishni", "yaxshi"** va **"ko‘raman"**, so‘zlari unigramlar, **"men kitob", "kitob o‘qishni", "o‘qishni yaxshi"** va **"yaxshi ko‘raman"** esa bigrammalardir. Quyidagi kod yordamida korpusdagi unigramlar, bigramlar va trigramlar aniqlanadi:

```
from collections import Counter
from UzbNlp.util import ngrams
unigrams=[]
bigrams=[]
trigrams=[]
for content in (sentences):
    content = content.lower()
    content = word_tokenize(content)
    for word in content:
        if (word =='.'):
            content.remove(word)
        else:
            unigrams.append(word)
    bigrams.extend(ngrams(content,2))
    trigrams.extend(ngrams(content,3))
```

```
print ("n-gramlar:n" + "-------------------------")
print ("--> UNIGRAMLAR: n" + str(unigrams[:5]) + " ...n")
print ("--> BIGRAMLAR: n" + str(bigrams[:5]) + " ...n")
print ("--> TRIGRAMLAR: n" + str(trigrams[:5]) + " ...n")
```

_____

UNIGRAMLAR:

[‘Bilgil’, ‘va’, ‘ogoh’, ‘bo‘lg‘ilki’, ‘dunyoda’, ‘ilm’, ‘o‘qimoq’, ‘va’, ‘ilm’, ‘istamoqdin’, ‘g‘araz-u’, ‘maqsud’, ‘ikki’, ‘dunyoning’, ‘saodatiga’, ‘yetmoqdur’]

BIGRAMLAR:

[(‘Bilgil’, ‘va’), (‘va’, ‘ogoh’), (‘ogoh’, ‘bo‘lg‘ilki’), (‘bo‘lg‘ilki’, ‘dunyoda’), (‘dunyoda’, ‘ilm’), (‘ilm’, ‘o‘qimoq’), (‘o‘qimoq’, ‘va’), (‘va’, ‘ilm’), (‘ilm’, ‘istamoqdin’), (‘istamoqdin’, ‘g‘araz-u’), (‘g‘araz-u’ , ‘maqsud’), (‘maqsud’, ‘ikki’), (‘ikki’, ‘dunyoning’), (‘dunyoning’, ‘saodatiga’), (‘saodatiga’, ‘yetmoqdur’)]

TRIGRAMLAR:

[(‘Bilgil’, ‘va’, ‘ogoh’), (‘va’, ‘ogoh’, ‘bo‘lg‘ilki’), … (‘dunyoning’, ‘saodatiga’, ‘yetmoqdur’)]

Keyingi qadamda, korpusdan **ko‘makchi, bog‘lovchi** yoki **olmosh** kabi nomuhim so‘zlarga ega bo‘lmagan unigramma, bigramma va trigrammalarni shakllantirish lozim. N-gramlardan nomuhim so'zlarni olib tashlash uchun quyidagi koddan foydalanamiz.

```
def stopwords_removal(n, a):
    b = []
    if n == 1:
        for word in a:
            count = 0
            if word in stop_words:
                count = 0
            else:
                count = 1
            if (count==1):
                b.append(word)
        return(b)
    else:
        for pair in a:
            count = 0
            for word in pair:
                if word in stop_words:
                    count = count or 0
                else:
                    count = count or 1
            if (count==1):
                b.append(pair)
        return(b)

unigrams_Processed = stopwords_removal(1,unigrams)
```

```
bigrams_Processed = stopwords_removal(2,bigrams)
trigrams_Processed = stopwords_removal(3,trigrams)
print ("Korpusni boshlang`ich qayta ishlangandan keyingi n-gramlar ro`yxati:")
print ("--> UNIGRAMLAR: n" + str(unigrams_Processed[:5]) + " ...n")
print ("--> BIGRAMLAR: n" + str(bigrams_Processed[:5]) + " ...n")
print ("--> TRIGRAMLAR: n" + str(trigrams_Processed[:5]) + " ...n")
```

_____

UNIGRAMLAR:

['Bilgil', 'ogoh', 'bo'lg'ilki', 'dunyoda', 'ilm', 'o'qimoq', 'ilm', 'istamoqdin', 'g'araz-u', 'maqsud', 'ikki', 'dunyoning', 'saodatiga', 'yetmoqdur']

BIGRAMLAR:

[('Bilgil', 'ogoh'), ('ogoh', 'bo'lg'ilki'), ('bo'lg'ilki', 'dunyoda'), ('dunyoda', 'ilm'), ('ilm', 'o'qimoq'), ('o'qimoq', 'ilm'), ('ilm', 'istamoqdin'), ('istamoqdin', 'g'araz-u'), ('g'araz-u', 'maqsud'), ('maqsud', 'ikki'), ('ikki', 'dunyoning'), ('dunyoning', 'saodatiga'), ('saodatiga', 'yetmoqdur')]

TRIGRAMLAR:

[('Bilgil', 'ogoh', 'bo'lg'ilki'), ... ('dunyoning', 'saodatiga', 'yetmoqdur')]

Yuqoridagi natijalar asosida korpusda hosil bo'ladigan har bir n-grammning sonini yoki chastotasini aniqlash mumkin. Ushbu natija n-grammlar asosida kontekstdagi keyingi mumkin bo'lgan so'zning ehtimolini hisoblash uchun kerak bo'ladi. Quyida keltirilgan **get_ngrams_freq_dist** funksiyasi orqali har bir n-gramga mos keladigan chastota aniqlanadi [9]. Ushbu funksiya orqali korpusdagi barcha unigramlar, bigramlar va trigramlarning chastotalari hisoblanadi.

```
def get_ngrams_freq_dist(n, ngramList):
    ngram_freq_dict = {}
    for ngram in ngramList:
        if ngram in ngram_freq_dict:
            ngram_freq_dict[ngram] += 1
        else:
            ngram_freq_dict[ngram] = 1
    return ngram_freq_dict
unigrams_freqDist = get_ngrams_freqDist(1, unigrams)
unigrams_Processed_freqDist = get_ngrams_freqDist(1, unigrams_Processed)
bigrams_freqDist = get_ngrams_freqDist(2, bigrams)
bigrams_Processed_freqDist = get_ngrams_freqDist(2, bigrams_Processed)
trigrams_freqDist = get_ngrams_freqDist(3, trigrams)
trigrams_Processed_freqDist = get_ngrams_freqDist(3, trigrams_Processed)
```

Bigram va trigram til modellari yordamida keyingi uchta so'zni bashorat qilish

Til modelida gapning ehtimolini hisoblash uchun zanjir qoidasi qo'llaniladi. $w_1w_2\ldots w_n$ gap, undagi w_1, w_2, w_n alohida so'zlar/tokenlar bo'lsin. Ushbu gapning hosil bo'lish ehtimoli quyidagi formula bilan aniqlanadi:

$$P(w_1, w_2 \dots w_n) = \prod_i P(w_i | w_1, w_2 \dots w_{i-1})$$

Masalan, **"Men kitob o`qishni yaxshi ko'raman"** gapining ehtimoli quyidagicha aniqlanadi:

$P("Men\ kitob\ o`qishni\ yaxshi\ ko'raman"=$
$P("men")P("kitob"|"men")P("o`qishni"|"men\ kitob")$
$P("yaxshi"|"men\ kitob\ o`qishni")$
$P("ko'raman"|"men\ kitob\ o`qishni\ yaxshi")$

Yuqoridagi formuladagi individual ehtimolliklar quyidagi usul orqali aniqlanadi:

$P(men) = Count('men') \ / \ umumiy\ so`zlar\ soni$
$P("kitob"|"men") = Count('men\ kitob') \ / \ Count('men')$
$P("o`qishni"|"men\ kitob") = Count('men\ kitob\ o`qishni') \ / \ Count('men\ kitob')$
$P("yaxshi"|"men\ kitob\ o`qishni")$
$= Count('men\ kitob\ o`qishni\ yaxshi') \ / \ Count('men\ kitob\ o`qishni')$
$P("ko'raman"|"men\ kitob\ o`qishni\ yaxshi")$
$= Count('men\ kitob\ o`qishni\ yaxshi\ ko'raman')$
$/ \ Count('men\ kitob\ o`qishni'\ yaxshi')$

Bunda,
$Count('men')$,$Count('men\ kitob')$,
$Count('men\ kitob\ o`qishni')$,
$Count('men\ kitob\ o`qishni\ yaxshi')$ va $Count('men\ kitob\ o`qishni\ yaxshi\ ko'raman')$ qiymatlar yuqoridagi get_ngrams_freq_dist(n, ngramList) funksiyasidan foydalangan holda hisoblagan tegishli unigram, bigram va trigramma chastotalaridir.

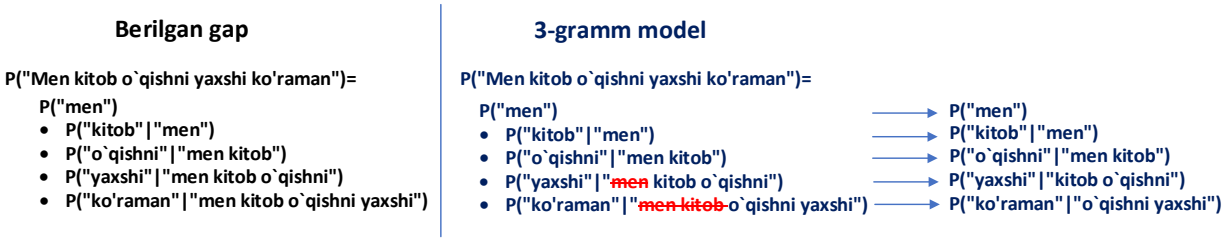
Yuqordagi formulalrdagi ehtimolliklarni hisoblash uchun bigram modelidan foydalanganda, har bir yangi so'zning ehtimolligi faqat oldingi so'zga bog'liqligini qayd etish mumkin. Ya'ni, oldingi misol uchun, gaplaning ehtimoli quyidagicha hisoblanadi:

$P("Men\ kitob\ o`qishni\ yaxshi\ ko'raman"=$
$P("men")P("kitob"|"men")P("o`qishni"|"kitob")$
$P("yaxshi"|"o`qishni")P("ko'raman"|"yaxshi")$

Xuddi shunday, trigramm modeli uchun ehtimollik quyidagicha hisoblanadi:

$P("Men\ kitob\ o`qishni\ yaxshi\ ko'raman"=$
$P("men")P("kitob"|"men")P("o`qishni"|"men\ kitob")$
$P("yaxshi"|"kitob\ o`qishni")$
$P("ko'raman"|"o`qishni\ yaxshi")$

Trigramm modellashtirishni quyidagi diagramma bilan yaxshiroq tushuntirish mumkin:



Biroq, bunday modellashtirishning o‘ziga xos jihati bor. O‘quv namunalarida (dataset) mavjud bo‘lmagan, ammo joriy matn/korpusda mavjud qandaydir bigramma bor. Yuqoridagi

hisob-kitoblar asosida ushbu bigrammaga **0** ehtimol belgilanadi va gapdagi umumiy ehtimollik qiymati **0** ga teng bo‘ladi. Ushbu muammoni bartaraf etish uchun **“silliqlash” (smoothing)** amalga oshiriladi. Ba‘zi ehtimollik qiymatlarini hisobga olimagan (ko‘rib chiqilmagan) hodisalarga qayta belgilash uchun parametrlar tartibga solinadi. Bunday yumshatish usullaridan biri **Add-one** yoki **Laplace smoothing** bo‘lib, yuqoridagi til modelini ishlab chiqishga joriy qilamiz ushbu maqolada foydalanamiz. Buning uchun barcha bigramm qiymatlariga **1** va barcha unigramma qiymatlarga **V** (korpusdagi unikal so‘zlar soni)ni qo‘shish lozim.

Avvalgi baho:

$$P_{MLE}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i)}{c(w_{i-1})}$$

Add-one baho:

$$P_{Add-1}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i) + 1}{c(w_{i-1}) + V}$$

Silliqlangan bigram va trigram modellari haqida ma’lumot hosil bo‘lganidan so‘ng, modelda mos o‘zgarishlarni amalga oshirish kerak. Bashorat qilish uchun qayta ishlanmagan bigramm va trigram elementlaridan (nomuhim so`zlarga ega matn) foydalanamiz.

```
smoothed_bigrams_probDist = {}
V = len(unigrams_freqDist)
for i in bigrams_freqDist:
   smoothed_bigrams_probDist[i] = (bigrams_freqDist[i] + 1)/(unigrams_freqDist[i[0]]+V)
smoothed_trigrams_probDist = {}
for i in trigrams_freqDist:
   smoothed_trigrams_probDist[i] = (trigrams_freqDist[i] + 1)/(bigrams_freqDist[i[0:2]]+V)
```

Keyingi qadamda, hisoblangan bigram va trigram tili modellari yordamida keyingi uchta so‘z bashorat qilinadi.

```
testSent1 = "Qolaversa, durustgina "
testSent2 = "Har bir odam uchun"
testSent3 = "Aytmat bobom qo‘li gul usta edi"
```

Berilgan namunaviy gaplarni tokenlarga ajratilib, ulardagi oxirgi unigramlar va bigramlar aniqlanadi.

```
token_1 = word_tokenize(testSent1)
token_2 = word_tokenize(testSent2)
token_3 = word_tokenize(testSent3)
ngram_1 = {1:[], 2:[]}
ngram_2 = {1:[], 2:[]}
ngram_3 = {1:[], 2:[]}
for i in range(2):
   ngram_1[i+1] = list(ngrams(token_1, i+1))[-1]
   ngram_2[i+1] = list(ngrams(token_2, i+1))[-1]
   ngram_3[i+1] = list(ngrams(token_3, i+1))[-1]
print("1-gap: ", ngram_1,"2-gap: ",ngram_2,"3-gap: ",ngram_3)
```

So‘ngra, yangilangan bigram modelidan foydalanib, berilgan namunaviy gaplarga mos keyingi 3 ta so‘zni bashorat qilish funksiyalarini shakllantiramiz.

```
def predict_next_word(last_word,probDist):
    next_word = {}
    for k in probDist:
        if k[0] == last_word[0]:
            next_word[k[1]] = probDist[k]
    k = Counter(next_word)
    high = k.most_common(1)
    return high[0]
def predict_next_3_words(token,probDist):
    pred1 = []
    pred2 = []
    next_word = {}
    for i in probDist:
        if i[0] == token:
            next_word[i[1]] = probDist[i]
    k = Counter(next_word)
    high = k.most_common(2)
    w1a = high[0]
    w1b = high[1]
    w2a = predict_next_word(w1a,probDist)
    w3a = predict_next_word(w2a,probDist)
    w2b = predict_next_word(w1b,probDist)
    w3b = predict_next_word(w2b,probDist)
    pred1.append(w1a)
    pred1.append(w2a)
    pred1.append(w3a)
    pred2.append(w1b)
    pred2.append(w2b)
    pred2.append(w3b)
    return pred1,pred2
print("Silliqlashtirilgan bigram modeli yordamida keyingi 3 ta mumkin bo‘lgan so‘zlar ketma-ketligini bashorat qilish:")
pred1,pred2 = predict_next_3_words(ngram_1[1][0],smoothed_bigrams_probDist)
print("1a)" +testSent1 +" "+ '33[1m' + pred1[0][0]+" "+pred1[1][0]+" "+pred1[2][0] + '33[0m')
print("1b)" +testSent1 +" "+ '33[1m' + pred2[0][0]+" "+pred2[1][0]+" "+pred2[2][0] + '33[0m')
pred1,pred2 = predict_next_3_words(ngram_2[1][0],smoothed_bigrams_probDist)
print("2a)" +testSent2 +" "+ '33[1m' + pred1[0][0]+" "+pred1[1][0]+" "+pred1[2][0] + '33[0m')
print("2b)" +testSent2 +" "+ '33[1m' + pred2[0][0]+" "+pred2[1][0]+" "+pred2[2][0] + '33[0m')
pred1,pred2 = predict_next_3_words(ngram_3[1][0],smoothed_bigrams_probDist)
print("3a)" +testSent3 +" "+ '33[1m' + pred1[0][0]+" "+pred1[1][0]+" "+pred1[2][0] + '33[0m')
print("3b)" +testSent3 +" "+ '33[1m' + pred2[0][0]+" "+pred2[1][0]+" "+pred2[2][0] + '33[0m')
```

Yangilangan bigram modelining bashoratlari:

1a) Qolaversa, durustgina baxshi ham edi
1b) Qolaversa, durustgina rassom ham edi
2a) Har bir odam uchun dunyo ostonadan boshlanadi
2b) Har bir odam uchun hayot juda go`zal
3a) Ahmad bobom qo‘li gul usta edi
3b) Ahmad bobom qo‘li juda chiroyli edi

Xuddi shunday uslubda yangilangan trigram modelidan bashorat olamiz.

```
def predict_next_word(last_word,probDist):
    next_word = {}
    for k in probDist:
        if k[0:2] == last_word:
            next_word[k[2]] = probDist[k]
    k = Counter(next_word)
    high = k.most_common(1)
    return high[0]
def predict_next_3_words(token,probDist):
    pred = []
    next_word = {}
    for i in probDist:
        if i[0:2] == token:
            next_word[i[2]] = probDist[i]
    k = Counter(next_word)
    high = k.most_common(2)
    w1a = high[0]
    tup = (token[1],w1a[0])
    w2a = predict_next_word(tup,probDist)
    tup = (w1a[0],w2a[0])
    w3a = predict_next_word(tup,probDist)
    pred.append(w1a)
    pred.append(w2a)
    pred.append(w3a)
    return pred
print("Silliqlashtirilgan trigram modeli yordamida keyingi 3 ta mumkin bo‘lgan so‘zlar ketma-ketligini bashorat qilish:")
pred = predict_next_3_words(ngram_1[2],smoothed_trigrams_probDist)
print("1)" +testSent1 +" "+ '33[1m' + pred[0][0]+" "+pred[1][0]+" "+pred[2][0] + '33[0m')
pred = predict_next_3_words(ngram_2[2],smoothed_trigrams_probDist)
print("2)" +testSent2 +" "+ '33[1m' + pred[0][0]+" "+pred[1][0]+" "+pred[2][0] + '33[0m')
pred = predict_next_3_words(ngram_3[2],smoothed_trigrams_probDist)
print("3)" +testSent3 +" "+ '33[1m' + pred[0][0]+" "+pred[1][0]+" "+pred[2][0] + '33[0m')
```

---

1) Qolaversa, durustgina omadli biznesmenga aylanishdi
2) Har bir odam uchun eng yaxshi kun

3) Ahmad bobom qo‘li yarador holda qaytdi

Xulosa.

Tilni modellashtirish (Language modeling, LM) kontekstda so‘zlarni bashorat qilish uchun korpus matnlarini tahlil qiladi va tabiiy til modelini shakllantiradi. Ushbu modellar gapda ma’lum bir so‘z ketma-ketligining yuzaga kelish ehtimolini aniqlash uchun statistik va ehtimollik usullaridan foydalanadi. NLP kontseptsiyalariga asoslangan ilovalar va dasturlar audiodan matnga o‘tkazish, hissiyotlarni tahlil qilish, nutqni aniqlash va imlo tuzatish kabi vazifalarda til modellaridan foydalaniladi. Til modellari matn ma’lumotlarining tahlil qilingan qismidagi so‘z ehtimolini aniqlash orqali ishlaydi. Ma’lumotlar ushbu tabiiy tilda kontekst qoidalarini qidiradigan mashinali o‘rganish algoritmiga berilgandan so‘ng qo‘llaniladi. Ushbu maqolada o‘zbek tili kospusi matnlarini n-gramm usuliga asoslangan til modeli ishlab chiqildi va namunaviy gaplarga qo‘llash usullari keltirildi.

Foydalanilgan adabiyotlar:

1. Jurafsky, D., & Martin, J. H. (2019). Chapter 3: N-Gram Language Models N-Gram Language Models. *Speech and Language Processing*.
2. Chen, M., Suresh, A. T., Mathews, R., Wong, A., Allauzen, C., Beaufays, F., & Riley, M. (2019). Federated learning of N-gram language models. *CoNLL 2019 - 23rd Conference on Computational Natural Language Learning, Proceedings of the Conference*. https://doi.org/10.18653/v1/k19-1012
3. Republic, C., & Mikolov, T. (2012). Statistical Language Models Based on Neural Networks. *Wall Street Journal*, *April*. https://doi.org/10.1016/j.csl.2015.07.001
4. P.~Brown, V.~Della Pietra, de Souza, P., J.~Lai, & R.~Mercer. (1992). Class-based n-gram models of natural language. *Computational Linguistics*, *18*.
5. Boltayevich, E. B., Mirdjonovna, H. S., & Ilxomovna, A. X. (2023). Methods for Creating a Morphological Analyzer. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13741 LNCS*. https://doi.org/10.1007/978-3-031-27199-1_4
6. Bommasani, R., Liang, P., & Lee, T. (2023). Holistic Evaluation of Language Models. *Annals of the New York Academy of Sciences*, *1525*(1). https://doi.org/10.1111/nyas.15007
7. Konstantopoulos, S. (2010). Learning language identification models: A comparative analysis of the distinctive features of names and common words. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*.
8. Wallace, E., Gardner, M., & Singh, S. (2020). *Interpreting Predictions of NLP Models*. https://doi.org/10.18653/v1/2020.emnlp-tutorials.3
9. Roh, J., Park, S., Kim, B. K., Oh, S. H., & Lee, S. Y. (2021). Unsupervised multi-sense language models for natural language processing tasks. *Neural Networks*, *142*. https://doi.org/10.1016/j.neunet.2021.05.023

АЛГОРИТМИЧЕСКИЙ ПОДХОД К АНАЛИЗУ ЭМОЦИОНАЛЬНОЙ ОКРАСКИ И ПОЭТИЧНОСТИ УЗБЕКСКИХ ТЕКСТОВ

Барахнин Владимир Борисович
д.т.н., проф., Новосибирский государственный университет

Менглиев Давлатёр Бахтиярович
Ургенчский филиал ТУИТ
E-mail: shogunuz@gmail.com

Abstract: This article presents the development of an algorithm for recognizing poetic texts in the Uzbek language using a lexical approach. The research focuses on creating an extensive dictionary containing more than 10 thousand poetic words and expressions collected from contemporary literary works. The algorithm analyzes texts, comparing each word with the dictionary database, determining its emotional connotation and presence in the dictionary. A feature of the developed dictionary is the inclusion of full forms of words, which allows you to accurately preserve their meaning in the text. The results of testing the algorithm on a sample of texts showed its high.

Keywords: Uzbek poetry, recognition algorithm, vocabulary base, sentiment analysis, algorithm accuracy, artificial intelligence, lexicons, emotional richness, morphological structure, literary research.

Аннотация: Эта статья представляет разработку алгоритма для распознавания поэтических текстов на узбекском языке с использованием лексического подхода. Исследование фокусируется на создании обширного словаря, содержащего более 10 тысяч поэтических слов и выражений, собранных из современных литературных произведений. Алгоритм анализирует тексты, сравнивая каждое слово с базой данных словаря, определяя его эмоциональную окраску и наличие в словаре. Особенностью разработанного словаря является включение полных форм слов, что позволяет точно сохранять их значимость в тексте. Результаты тестирования алгоритма на выборке текстов показали его высокую эффективность, подтверждая его потенциал для глубокого анализа поэтических текстов.

Ключевые слова: Узбекская поэзия, алгоритм распознавания, словарная база, сентимент-анализ, точность алгоритма, искусственный интеллект, лексиконы, эмоциональная насыщенность, морфологическая структура, литературные исследования.

Введение

В области компьютерной лингвистики задача анализа поэтических текстов, особенно в условиях богатого многообразия узбекского языка, представляет собой сложную задачу. Поэзия отличается своими стилистическими и структурными элементами, такими как ритм, метр и образное использование языка, что требует тонкого анализа текста.

Предложенный подход основан на создании и использовании специализированных лексиконов, отражающих наиболее важные элементы поэзии, включая поэтическую лексику.

Статья начинается с введения и предоставления информации о узбекском языке, в частности, о его особенностях в поэтическом контексте. Далее в статье представлен обзор существующих исследований по данной тематике или близких к ней работ. В последующих разделах авторы излагают свое решение, а также результаты тестирования предложенного алгоритма. В заключительном разделе содержится обобщение результатов исследования, а также возможные направления для дальнейшего развития алгоритма.

Морфологические особенности

Агглютинативный характер узбекского языка подразумевает, что слова часто формируются путем последовательного добавления морфем (суффиксов)[1-2], каждый из которых выполняет определенную грамматическую или семантическую функцию[3-4]. Например, слово «kitoblaringizdan» (из ваших книг) демонстрирует, как существительное «kitob» (книга) может быть дополнено несколькими суффиксами, указывающими на притяжательность («laringiz» для вашего), множественное число («lar») и аблатив («dan» означает «из»).

Аналогичные работы

Несмотря на большое количество научных работ в области узбекской компьютерной лингвистики, авторам статьи не удалось найти научные исследования, связанные с распознаванием поэтических текстов на узбекском языке. К примеру, статья [5] посвящена разработке информационных систем для анализа поэтических текстов. Авторы рассматривают сложность анализа поэзии и используют в своих исследованиях как традиционные методы и алгоритмы классической математики, так и современные подходы, связанные с машинным обучением. Статья [6] посвящена анализу зависимости семантического содержания поэтических текстов от их метроритмических характеристик и строфической структуры. Авторы ставят целью сформулировать определение «текстуры» как структурной модели поэтического текста, которая уникально определяет метроритмический узор и строфическую структуру текста, а также разработку алгоритмов для автоматического определения текстуры.

Предлагаемое решение

Для обеспечения эффективного функционирования алгоритма была разработана словарная база, охватывающая более 10 тысяч поэтических слов и выражений. Источниками для составления словаря послужили литературные произведения современных авторов XX и XXI веков, тогда как произведения более ранних периодов в словарь не вошли в связи с их оригинальным созданием на персидском или арабском языке. Структурированный словарь представлен в виде табличного файла (.xlsx) с четырьмя столбцами:

1) идентификатор слова или фразы;
2) само слово;
3) часть речи слова;
4) уровень настроения (присущ только прилагательным, варьируется от -2 до 2, что указывает на спектр от очень отрицательной до очень положительной эмоциональной окраски).

Важно отметить, что слова в словаре представлены в полной форме с учетом всех аффиксов и приставок, чтобы избежать потери смысла при морфологическом анализе. Основная причина этого в том, что при образовании литературных слов образующиеся корни слов могут иметь совершенно разное значение, благодаря чему мы можем получить совершенно разный результат. Все элементы словаря были созданы вручную

Алгоритм работает следующим образом:

1. Входной текст разбивается на предложения, а затем на слова.
2. Каждое слово сравнивается со словарем, чтобы определить, существует оно там или нет.
3. Определяется эмоциональный оттенок слова, если он есть.
4. Подсчитывается общее количество обнаруженных стихотворных слов и словосочетаний в анализируемом тексте.
5. Численные значения тональных слов суммируются для оценки общего уровня тональностей в тексте.
6. Результаты представлены по количеству выявленных поэтических элементов и общему уровню тональности текста.

Тестирование и результаты алгоритма

Главным критерием отбора была литературная значимость и доступность текстов, в результате было отобрано 500 предложений из разных произведений. Этот образец охватывает широкий спектр тем и стилей, представляющих современную узбекскую поэзию, что позволяет оценить алгоритм в различных контекстах использования.

По результатам тестирования алгоритм демонстрирует высокую эффективность выявления поэтических элементов в тексте, достигая точности 99%. Это подтверждает его способность точно распознавать и классифицировать поэтические слова и фразы в узбекской поэзии. Кроме того, алгоритм также показал высокую точность оценки настроений, достигнув 89%. Эти результаты подтверждают эффективность разработанного алгоритма и его пригодность для анализа поэтических текстов на узбекском языке.

Наш словарь, хотя и включает более 10 тысяч поэтических слов и словосочетаний, не является исчерпывающим. В узбекской поэзии существует огромное разнообразие стилей и выражений, и некоторых уникальных слов или новых явлений может не быть в словаре. Это объясняет случаи, когда алгоритм не мог идентифицировать определенные поэтические элементы.

Кроме того, несмотря на высокую точность алгоритма определения настроений, достигающую 89%, он имеет небольшие недостатки. Основная причина – субъективность настроения, то есть эмоциональный оттенок слов может меняться в зависимости от контекста употребления, что делает его определение сложной задачей. В некоторых случаях тональность слова может интерпретироваться по-разному в зависимости от литературного контекста исходного текста.

Решением подобных проблем является изменение технологии расчета уровня настроенности предложений. Например, вместо словарного подхода использовать технологии искусственного интеллекта, хотя для этого потребуется совершенно другой тип и объем данных, значительно превышающий текущий объем размеченного словаря.

Подробности результатов показаны в Таблице I и Таблице II.

Таблица I.

Type of analyzing	Correctly detected words and phrases / all poetic words	Precision
Poetic words detection	990 / 1001	99%
Sentiment words detection	389 / 437	89%

Таблица II.

Type of sentiment words in text	Correctly detected words / all words	Precision
Very negative	23 / 35	66%
Negative	89 / 94	95%
Neutral	115 / 120	96%
Positive	92 / 99	93%
Very positive	70 / 89	79%
Overall	**389 / 437**	**89%**

Заключение

Алгоритм для распознавания поэтических текстов на узбекском языке, основанный на методе использования словарной базы, был разработан и протестирован. Алгоритм показал высокую точность в определении поэтических элементов и анализе сентиментов, достигая соответственно 99% и 89%. Эти результаты подтверждают значительный потенциал подхода, основанного на специализированных лексиконах для анализа текстов в стиле узбекской поэзии и их эмоциональной насыщенности.

Тем не менее, достигнутые результаты являются важным шагом для задач анализа поэтических текстов, а также их эмоциональной окраски. Дальнейшее развитие алгоритма, включая интеграцию с современными технологиями искусственного интеллекта, может открыть новые горизонты для исследований в области литературы.

Использованная литература:

1. D. Mengliev, E. Akhmedov, V. Barakhnin, Z. Hakimov, O. Alloyorov, "Utilizing Lexicographic Resources for Sentiment Classification in Uzbek Language," 2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE), Novosibirsk, Russian Federation, pp. 1720-1724, 10-12 November 2023.
2. D. Mengliev, V. Barakhnin, M. Eshkulov, B. Palvanov, N. Abdurakhmonova, S. Khamraeva, "Dictionary-Based Medical Text Analysis in Uzbek: Overcoming the Low-Resource Challenge," 2023 IEEE Ural-Siberian Conference on Computational Technologies in Cognitive Science, Genomics and Biomedicine (CSGB), Novosibirsk, Russian Federation, pp. 85-89, 28-30 September 2023.
3. M. Sharipov, J. Mattiev, J. Sobirov, R. Baltayev, “Creating a morphological and syntactic tagged corpus for the Uzbek language”, The International Conference and Workshop on Agglutinative Language Technologies as a challenge of Natural Language Processing (ALTNLP), June 7-8, 2022.

4. G. Oripova, “Uzbek poetry and the world literature in the years of independence”, Periodyk naukowy akademii polonijnej,, vol. 32, pp. 116-130, 2019.
5. O. Kozhemyakina, “Information systems for the analysis of poetic texts: history, methods and algorithms”, Journal of Computational Technologies, pp. 136-166, 2023.
6. V. Barakhnin, O. Kozhemyakina, I. Kuznetsova, V. Karpova, “The model of facture of Russian poetic texts”, Journal of Computational Technologies, pp. 107-117, 2021.

SIYOSIY MATNLARNI O‘RGANISHDA NUTQIY YONDASHUV

Yaxshiboyev Mirjalol
O‘zbekiston Milliy universiteti o‘qituvchisi.

Annotation: this article examined the discursive approach to the study of political texts - an important rule of modern political linguistics, as well as the relationship of each specific text with other texts within the current political situation.

Keywords: text, political text, speech, political communication, political linguistics.

Annotatsiya: v dannoy state bыl izuchen diskursivnыy podxod k izucheniyu politicheskix tekstov - vajnoye pravilo sovremennoy politicheskoy lingvistiki, a takje sootnosheniye kajdogo konkretnogo teksta s drugimi tekstami v ramkax sozdavsheysya politicheskoy situatsii.

Ключевые слова: текст, полититческий текст, речь, политическая коммуникация, политическая лингвистика.

Matn – nutq ko‘rinishi bo‘lib, vazifasi jihatidan tugal nutqiy butunlikdir. Har bir matn murakkab tuzilish va mazmun mundarijasiga ega bo‘lib, u og‘zaki hamda yozma ijod namunasi hisoblanadi. Tilshunoslikda matn tilning alohida birligi va matn tilshunosligi deb atalayotgan sohaning asosiy obyekti sifatida talqin qilinadi. Matn so‘z birikmasi va gapdan farqlanadi, chunki matnning ham o‘z kategoriyasi va qonuniyatlari mavjud. Matn so‘zining lug‘aviy ma’nosida birikish, bog‘lanish tushunchalarining borligi va matn tarkibi o‘zaro qaysidir bog‘lovchilar yordamida birikishni o‘rganish “Matn tilshunosligi” sohasining asosiy muammolaridan biridir.

Tilshunos E.Qilichev “Matnning lingvistik tahlili” nomli kitobida “Matn – hamma elementlari o‘zaro zich aloqada bo‘lgan va muallif nuqtayi nazaridan ma’lum bir maqsadga yo‘naltirilgan nominativ-estetik axborotni ifodalovchi murakab tuzilma» [Qilichev E., 2000: 45] degan ta’rifni beradi. I.Rasulov esa matnni quyidagicha ta’riflaydi: “Gapdan katta birlik murakkab sintaktik butunlik bo‘lib, u fikran va sintaktik jihatdan o‘zaro bog‘liq bo‘lgan gaplar birlashmasidan iborat. Unda fikr gapga nisbatan ancha to‘liq bo‘ladi” [Rasulov I., 2021: 45].

So‘nggi o‘n yilliklar mobaynida birmuncha istiqbolli ilmiy yo‘nalishlar aksariyat hollarda turli bilim sohalarining tutashuvi doirasida yuzaga kelgan. Shu kabi fanlardan biri, biz uchun yangi fan hisoblangan siyosiy lingvistika fani bo‘lib, etnologiya, ijtimoiy psixologiya, sotsiologiya va boshqa gumanitar fanlarga erishishni nazarda tutuvchi lingvistikaning siyosatshunoslik bilan kesishmasida vujudga kelgan. Yangi ilmiy yo‘nalishning vujudga kelish va rivojlanish zarurati jamiyatning siyosiy kommunikatsiyalar mexanizmlari va shartlariga qiziqishning ortib borishi bilan belgilanadi.

Siyosiy lingvistika til va jamiyatning o‘zaro bog‘liqligi muammolari bilan shug‘ullanuvchi sotsiolingvistika, matn lingvistikasi va kognitiv lingvistika, mumtoz va zamonaviy ritorika, funksional stilistika va ayniqsa publitsistik uslubni tadqiq qilish kabi qator lingvistik yo‘nalishlar bilan chambarchas bog‘liq.

Zamonaviy tilshunoslikning ko‘p tarmoqlilik (turli fanlar metodologiyasini qo‘llash), antropotsentrizm (lisoniy holatlarni tadqiq qilish uchun hisoblash nuqtasiga aylangan lisoniy shaxs, inson), ekspansionizm (lingvistik izlanishlar sohasini kengaytirish tendensiyalari), funksionalizm (o‘z funksiyalarini amalga oshirishda tilning nutqdagi harakatlarini o‘rganish) va

eksplanatorlik (dallilarni shunchaki yoritishga urinish emas, balki ularga izoh berish) kabi jihatlari to‘laqonli ravishda siyosiy lingvistikaga xosdir.

Siyosiy lingvistika individual, sotsiumga doir va milliy tafakkurni o‘rganuvchi ijtimoiy psixologiya, madaniyatshunoslik, sotsiologiya, siyosatshunoslik va etnologiya kabi fanlar bilan chambarchas bog‘liq.

Siyosiy lingvistikaning siyosiy kommunikatsiya, ya’ni nutqiy faoliyat, jamiyatdagi fikrlar rang-barangligi sharoitida ijtimoiy-siyosiy qarorlar qabul qilish va ularni asoslash, mamlakat fuqarolariga hissiy ta’sir o‘tkazish orqali ijtimoiy hamjihatlikka erishish uchun ularni siyosiy harakatlarga undash kabi muayyan g‘oyalarni targ‘ib qilishga yo‘naltirilgan nutqiy faoliyatdir. Har bir inson hech bo‘lmasa gohida gazeta o‘qiydi, radio tinglaydi yoki televizor ko‘radi, va shu tariqa siyosiy kommunikatsiya ishtirokchisiga aylanadi. Qachonki ushbu inson saylovga borsa, u siyosiy hayotda ishtirok etadi va u buni siyosiy kommunikatsiyalar subyektlari ta’sirida amalga oshiradi. Va shunda siyosiy lingvistika manfaatlari sohasi sirasiga nafaqat siyosiy ma’lumotlarni uzatishni, balki kommunikativ faoliyat jarayonidagi siyosiy voqe’likni idrok etish va uni baholash bilan bog‘liq bo‘lgan barcha narsalarni kiritish mumkin.

Siyosiy kommunikatsiyaning asosiy vazifasi - kommunikativ faoliyatni qo‘llash asosida siyosiy hokimiyat uchun kurashishdan iborat: siyosiy kommunikatsiya hokimiyatni (saylovlar, tayinlashlar, ijtimoiy fikrni yaratish va h.k. yo‘llar orqali) taqsimlashga hamda uni qo‘llash (qonunlar qabul qilish, farmonlar va farmoyishlar chiqarish)ga bevosita va bilvosita ta’sir ko‘rsatishga undaydi. Siyosiy kommunikatsiya mavjud siyosiy voqe’likni aks ettirgan holda, u bilan birga o‘zgaradi va unda kechayotgan islohotlarda ishtirok etadi.

Siyosiy kommunikatsiyaning asosiy maqsadi – siyosiy faoliyatning maqbul strategiyasi va taktikasini ishlab chiqish uchun sharoit yaratuvchi jamiyatning siyosiy holati va siyosiy faoliyati subyektlari, kommunikatsiya, mushohada va til o‘rtasidagi xilma-xil o‘zaro munosabatlarni tadqiq qilishdan iborat.

Siyosiy kommunikatsiya siyosiy qarorlar qabul qiluvchi insonlar ongiga ta’sir o‘tkazuvchi vosita sifatida xizmat qilishi sharofati tufayli hokimiyatdan foydalanish va uni taqsimlashga ta’sir ko‘rsatadi. Siyosiy kommunikatsiya nafaqat axborotni uzatadi, balki qabul qiluvchiga hissiy ta’sir ko‘rsatgan holda, inson ongida mavjud bo‘lgan olamning siyosiy manzarasini o‘zgartiradi [Chudinov A., 2001: 17].

Zamonaviy siyosiy lingvistika siyosiy kommunikatsiyaning umumiy muammolari bilan faol shug‘ullanadi (uning boshqa sohalardagi kommunikatsiyalar bilan farqini tahlil qiladi), siyosiy nutq (shior, dastur, gazeta maqolasi, mitinglarda so‘zga chiqish, parlamentdagi munozaralar) janrlari muammolarini hamda siyosiy matnlar funksiyalari xususiyatlarini o‘rganadi. Bizning ilmu-fanimiz ayrim siyosatchilar, siyosiy partiyalar va yo‘nalishlarning idiostillari muammolariga faol murojaat qiladi, siyosiy kommunikatsiyalar taktikalari, strategiyalari va usullarini o‘rganib chiqadi, siyosiy matnlar kompozitsiyasi, leksikasi va frazeologiyasini hamda ularda keng turli obrazli vositalarning qo‘llanishini o‘rganadi. Shuningdek, jamiyat rivojlanishining turli bosqichlarida turli mamlakatlardagi siyosiy kommunikatsiyaning qiyosiy tadqiqoti boshqa davlatlar fuqarolari tomonidan muayyan davlatlarning siyosiy voqe’liklarini anglab yetish muammolariga murojaati, tegishli til va milliy madaniyat doirasidagi alohida siyosiy konseplarni o‘rganish siyosiy lingvistikaning muhim yo‘nalishlari sirasiga kiradi.

Siyosiy matnlarni o‘rganishda nutqiy yondashuv - zamonaviy siyosiy lingvistikaning muhim qoidasidir. Bu esa ushbu matnning turli insonlar tomonidan idrok etilishi spetsifikasi, muallifning

siyosiy qarashlari va shaxsiy fazilatlarini hisobga olgan holda, u yaratilgan siyosiy vaziyat doirasida har bir aniq matnning boshqa matnlar bilan nisbati o‘rganib chiqilishini anglatadi. Ushbu matning siyosiy matnlar tizimidagi roli va – keyinroq - mamlakat siyosiy hayotida o‘ynashi mumkin bo‘lgan muhim roli albatta hisobga olinadi. Jumladan, bir xil g‘oyalar va bir xil mulohazalar gazeta maqolasidagi jurnalist matnida va prezidentlar rasmiy bayonotida mutlaqo turlicha idrok etiladi. Birgina siyosatchining ular muhim davlat vazifasiga belgilanishida saylovoldi kurashida so‘zlagan nutqi va mulohazasi mutlaqo turlicha salmoqqa ega bo‘lishi mumkin [Chudinov A., 2003: 37].

An’anaga ko‘ra hokimiyat yuksak qadriyatlar sirasiga kiradi, binobarin hokimiyat uchun kurashda barcha insoniy imkoniyatlar va fan yutuqlaridan foydalaniladi. Siyosiy matnlar mualliflari zaruriy natijaga erishish uchun ulkan kuch sarflaydi. Shu sababli siyosiy kommunikatsiyani tahlil qilishda mualliflarning yuksak mahorati va ijodiy no‘noqligi, ularning kommunikativ muvaffaqiyatlari va muvaffaqiyatsizliklari ayniqsa seziladi.

Sotsiolingvistika manfaatlari sohasiga mansub bo‘lgan davlatning til siyosati sohasi tadqiqotlarida siyosiy kommunikatsiyani o‘rganishga yo‘naltirilgan siyosiy lingvistikani farqlay bilish zarur. Muammoni birinchi soddalashtirgan holda, aytish joizki, siyosiy lingvistika bo‘yicha mutaxassislarni siyosatchilarning bu borada qanday fikr bildirayotganliklari qiziqtirsa, til siyosati bo‘yicha mutaxassislarni esa tildan samarali foydalanish uchun siyosatchilarning nima bilan shug‘ullanayotgani qiziqtiradi. Shu sababli mamlakatdagi (davlat tilining rasman tan olinishining zarurligi to‘g‘risidagi masaladan boshlab) davlat tili va boshqa tillarning xizmat qilishi muammolari tahlili kabi ishlarning g‘oyat muhim yo‘nalishini, millatlararo tillar muammolari va xalqaro muloqot tillari, tillarni o‘zlashtirish borasidagi masalalar siyosiy lingvistika manfaatlari sohasiga taalluqli emas.

Demokratik jamiyatda barcha fuqarolar nutqni ishlab chiqish sohasidagi kabi, uni qabul qilish sohasida ham siyosiy kommunikatsiyalarning barqaror ko‘nikmalariga ega bo‘lmog‘i zarur. Zamonaviy siyosiy kommunikatsiyalarning tanqidiy tahlili yangi avlod siyosiy yetakchilari va jurnalistlarini kommunikativ amaliyotini yanada uyg‘unlashtirishiga yordam qiladi. Jahon siyosiy kommunikatsiyalarini o‘rganishda, jahon andozalarini qo‘llash xorijiy jaomatchilik tasavvurida ijobiy nufuzini shakllantirishga ko‘maklashadi [Chomsky N., 2017: 38].

Siyosiy lingvistika yo‘nalishi keyingi o‘n yilliklari davomida birmuncha muvaffaqiyatga erishdi. Ushbu sohaning zamonaviy holatiga xos bo‘lgan jihatlaridan biri, uning xorijiy tadqiqotlar bilan yaqinlashuvi va tadqiqot sohasining sezilarli ravishda kengayishidir.

Siyosiy lingvistika – bu ulkan tatbiqiy ahamiyat kasb etuvchi ilmiy yo‘nalishdir. Siyosiy sohadagi kommunikativ faoliyat tajribasini o‘rganish tilshunoslar, siyosatshunoslar, jurnalistlar va jamoatchilik bilan aloqalar bo‘yicha mutaxassislar uchun aniq tavsiyalarni ishlab chiqishga mezon yaratadi.

Foydalanilgan adabiyotlar:

1. Chomsky N. Language and Politics. Montreal and New York, 2017. – B. 200.
2. Qilichev E. Matnning lingvistik tahlili. Buxoro, 2000. – B. 200.
3. Rasulov I. Eksperimental psixologiya va praktikum. Toshkent, 2021. – B. 147.
4. Chudinov A.P. Metaforicheskaya mozaika v sovremennoy politicheskoy kommunikatsii. Yekaterinburg, 2003. – B. 200.
5. Chudinov A.P. Rossiya v metaforicheskom zerkale: kognitivnoye issledovaniye politicheskoy metaforы (1991—2000). Yekaterinburg, 2001. – B. 270.

METHOD OF DEVELOPING ARTIFICIAL PATTERNS TO ANALYZING SEMANTIC SIMILARITY OF DOCUMENTS

Tuliev Ulugbek
PhD, Senior Lecturer
National University of Uzbekistan.

Abdurakhmonova Nilufar
Doctor of science, professor
National University of Uzbekistan.

Annotation: A technique for creating artificial patterns using data from textual documents is considered. There is a need for patterns to reduce the dimensionality of the feature space and analyse the semantic similarity of thematic documents. The composition of patterns is formed according to the rules of hierarchical agglomerative grouping. The formation process is based on the selection of a latent feature space using the technology for calculating generalized assessments of objects.

Keywords: semantic similarity of documents, latent feature.

Аннотация: Рассматривается методика создания искусственных паттернов по данным из текстовых документов. Существует потребность в паттернах для снижения размерности признакового пространства и анализа семантической близости тематических документов. Состав паттернов формируется по правилам иерархической агломеративной группировки. Процесс формирования основывается на выборе латентного признакового пространства по технологии вычисления обобщённых оценок объектов.

Ключевые слова: семантической близости документов, латентный признак.

Studying the structure of object relationships is one of the sources of obtaining new knowledge in intelligent systems. In this work, an artificial pattern is understood as a stable set of entities (words, concepts) of natural language (NL) for distinguishing text documents from subject areas. It is assumed that for the stability (compactness) of sets, numerical indicators are defined, the calculation of which is based on strict mathematical principles.

One of the problems when analyzing the structure of document descriptions in problems of mathematical linguistics is the problem of the curse of dimensionality. It is believed that the number of patterns is determined algorithmically based on the rules of hierarchical agglomerative grouping.

Reducing the dimension of the feature space is carried out in order to solve the problem of blurred relationships between objects. The blurriness assessment is based on quantitative indicators of the structure of relations between class objects. As a rule, when measuring indicators, distance measures between objects are used. Due to the large number of features used to describe topical documents, this inevitably leads to the well-known phenomenon of the curse of dimensionality in data analysis. The negative effect of this phenomenon is a significant limitation of the possibilities for applying classical methods of data mining. Using methods of linear or

nonlinear mapping of object descriptions into spaces of lower dimensions also does not answer the questions:

- What dimension of space should be after transformation?
- How to measure display quality for decision making?
- By what properties can we distinguish between document descriptions in different languages?

The selection of noisy features is considered as a constrained optimization problem. A separate problem is the choice of restrictions for the implementation of numerical solution methods, taking into account the availability of classification of thematic documents. Criteria [Ignatyev 2018, 590; Ignatev, Tuliyev 2022, 1185;]given below is used to find optimal boundary of the values of features

$$\left(\frac{\sum_{p=1}^{2}\sum_{i=1}^{2}u_i^p\left(m-|K_i|-\sum_{j=1}^{2}u_j^p+u_i^p\right)}{2|K_1||K_2|}\right)\left(\frac{\sum_{p=1}^{2}\sum_{i=1}^{2}u_i^p\left(u_i^p-1\right)}{|K_1|\left(|K_1|-1\right)+|K_2|\left(|K_2|-1\right)}\right)\rightarrow\max_{c_1<c_2<c_3}\ (1)$$

where u_i^j-number of objects which belong to K_i class in j-interval, m-number of objects in data set, $|K_i|$-number of objects which belong to K_i class.

The problems of reducing the dimension of the feature space, semantic relatedness of documents and taking into account polysemy in their various manifestations are solved [Tuliyev 2019, 102]. With the help of patterns, it becomes possible to automate the process of generating dictionaries for presenting documents in tabular form. It is relatively easy to interpret polysemy in scientific texts through glossaries and thesauri. The problems are aggravated in works of art in which the authors, through polysemy, seek to show the richness of the language.

The stability of the feature $x_c \in X(n)\backslash Z$ is calculated as []

$$\varphi(c)=\frac{1}{m}\sum_{r=1}^{m}\begin{cases}b_{rc}, b_{rc}>0.5,\\ 1-b_{rc}, b_{rc}<0.5.\end{cases}\quad (2)$$

The set of admissible values (2) belongs to (0.5;1]. Stability $\varphi(c)$=1 if, according to the boundary [Ignatev, Tuliyev 2022, 1185], objects are divided into classes K_1 and K_2 without errors. The ordering of words according to (2) in general does not depend on dimensions of space for describing a collection of documents.

The point of calculating the coefficient of content authenticity is to find the optimal number of clusters that determine the division of documents into topics in the latent feature space [Ignatyev 2018, 590; Tuliyev 2021, 2365]. To form a latent space and a common dictionary based on it, it is proposed:

- order the features with indices from D in non-decreasing order of their stability values (2) as
$$\varphi(\varepsilon_1),\dots,\varphi(\varepsilon_j),\dots,\varphi(\varepsilon_{dim}),\varepsilon_j\in D, dim=|D|;\quad (3)$$
- generate a set of latent features $Y(t)$= (y_1, . . ., yt) according to (3);
- select $Y(\sigma)\subset Y(t)$, $\sigma\leq t$ and the set of raw features related by D to $Y(\sigma)$ as a common dictionary.

The number of words for the general vocabulary from (3) is, in general, a free parameter. This parameter can be set for heuristic reasons, or based on the results of hierarchical agglomerative grouping when forming a set of latent features.

Let us denote by TYPLAM the set of indices of raw features included in the group by the hierarchical grouping algorithm, lugat the limit on the number of words in the general dictionary, guruh the number of latent features. The implementation of the algorithm in steps will be as follows [Ignatev, Tuliyev 2022, 1185].

Step 1. j=0. *guruh*=0.

Step 2. Computing $j=j+1$. *crit*=10. $u=\varepsilon_j$. *TYPLAM* = $\{u\}$. *guruh*=*guruh*+1.

Loop for $t\in\{1,...,m\}$ $R(S_t)=\eta_u(a_{tu})$. End of loop;

Step 3. $u=\varepsilon_{j+1}$. Loop for $t\in\{1, ..., m\}$ $b_t = R(S_t) + \eta_u(a_{tu})$. End of loop;

$M_1 = \sum_{S_t\in K_1} b_t$. $M_2 = \sum_{S_t\in K_2} b_t$. $M_1=M_1/|K_1|$. $M_2=M_2/|K_2|$. θ=0. γ=0. Loop for $t\in\{1, ..., m\}$ If $S_t\in K_1$, то $\theta=\theta+|b_t-M_1|$, $\gamma=\gamma+|b_t-M_2|$. Else $\theta=\theta+|b_t-M_2|$, $\gamma=\gamma+|b_t-M_1|$. End of loop;

Step 4. If θ/γ < *crit*, then *crit*= θ/γ , *TYPLAM* =*TYPLAM* $\cup\{u\}$, $j=j+1$, go to 3.

Step 5. Print $\{R(S_t)\}_{t\in\{1,...,m\}}$, *TYPLAM*.

Step 6. If $j<lugat$, then go to 2; Else print *guruh*.

Step 7. End.

To demonstrate the hierarchical algorithm described above in Table. 1 we present the results of its implementation on text documents presented by abstracts of scientific dissertations. Class K_1 is represented by documents in physics and mathematics, K_2 - from 10 different subject areas.

Table 1. Latent features and their composition from raw features

№	Number of features	Features	Value of criteria (1)
1	267	179, 55, 171, … , 414, 297, 299	0.6360
2	87	423, 66, 450, … , 186, 36, 265	0.5061
3	66	70, 67, 192, …. , 290, 15, 21	0.3917
4	10	393, 61, 75, 253, 163, 405, 53, 24, 44, 490	0.2793
5	23	322, 408, 295, … , 311, 364, 316	0.3493
6	1	89	0.2442
7	16	139, 189, 166, … , 305, 199, 304	0.3492
8	8	50, 158, 223, 195, 427, 302, 433, 369	0.2372
9	2	98, 197	0.2127
10	1	177	0.1852
11	2	57, 279	0.2344
12	2	120, 135	0.2460

References:

1. Ignatyev N. A. Structure Choice for Relations between Objects in Metric Classification Algorithms // Pattern Recognition and Image Analysis. 2018. V. 28. № 4. P. 590–597.
2. N. A. Ignatev, U. Y. Tuliyev, “Semantic structuring of text documents based on patterns of natural language entities”, Computer Research and Modeling, 14:5 (2022), 1185–1197 http://doi.org/10.20537/2076-7633-2022-14-5-1185-1197
3. N. Abdurakhmonova, U. Tuliyev and A. Gatiatullin, "Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus.uz," 2021 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2021, pp. 1-4, http://doi.org/10.1109/ICISCT52966.2021.9670043

4. Tuliyev U. (2021). Space formation for the description of thematic documents. AIP Conference Proceedings. 2365. 070007. http://doi.org/10.1063/5.0056963
5. Тулиев У. Ю. Кластерный анализ текстовых документов по отношению их связности // Проблемы вычислительной и прикладной математики. — 2019, No 6(24). — С. 102–109.
6. N.A. Ignatyev, Sh.F.Madrakhimov, D.Y.Saidov. Stability of object classes and selection of the latent features // International journal of engineering technology and sciences, 2017, Malaysia, Vol. 7, pp. 1-10.
7. Игнатьев Н.А., Саидов Д.Ю. Анализ данных и принятие решений с помощью логических закономерностей в форме полуплоскостей // Известия СамНЦ, 2017, Том 19, № 4(2), С. 294-300.

IV SHO‘BA
MASHINA TARJIMASINING LINGVISTIK VA DASTURIY TA’MINOTI

COUNTVECTORIZER YORDAMIDA SO‘ZLAR STATISTIKASINI ANIQLASH

Alayev Ruhillo
O‘zbekiston Milliy universiteti dotsenti, t.f.f.d.

Maxmudjonova Gulshaxnoz
ToshDO‘TAU Kompyuter lingvistikasi magistranti.

Annotatsiya: Ushbu maqola tabiiy tilni qayta ishlash va matnga asoslangan mashinali o‘rganishda muhim vosita bo‘lgan CountVectorizer haqida umumiy ma’lumot beradi. CountVectorizer metodologiyasini tushuntiradi hamda so‘z chastotalarini hisoblash va hujjat atamasi matritsasini yaratishni o‘z ichiga oladi. CountVectorizer so‘z statistikasini aniqlash va matnni tahlil qilish vazifalariga yordam beradi.
Kalit so‘zlar: so‘zlar statistikasi, matnni qayta ishlash, chastota, matn, tokenizatsiya,

Annotation: This article provides an overview of CountVectorizer, an important tool in natural language processing and effective machine learning for text. Explains the CountVectorizer methodology and retrieves word frequencies and a document term matrix. CountVectorizer helps with word statistics and text analysis.
Keywords: word statistics, text processing, frequency, text, tokenization

Аннотация: В этой статье представлен обзор CountVectorizer, важного инструмента для обработки естественного языка и эффективного машинного обучения текста. Объясняет методологию CountVectorizer и извлекает частоты слов и матрицу терминов документа. CountVectorizer помогает со статистикой слов и анализом текста.
Ключевые слова: статистика слов, обработка текста, частота, текст, токенизация.

CountVectorizer – bu matn ma’lumotlarini vektorlashtirish uchun ishlatiladigan vosita bo‘lib, u matnni mashinani o‘rganish algoritmlarida ishlatilishi mumkin bo‘lgan raqamli CountVectorizer usuli hujjatda so‘zning paydo bo‘lish chastotasini hisoblab chiqishga asoslangan. Ushbu usul orqali korpusdagi bir necha gaplar asosida so‘zlar matritsasi aniqlanadi va u gapdagi har bi so‘zning chastotasi bilan to‘ldiriladi ma’lumotlarga aylantiradi [Elov B, Hamroyeva Sh, Xusainova Z, Xudayberganov N,. 2023: 81]. CountVectorizer dasturining asosiy g‘oyasi hujjat atamasi matritsasini yaratishdan iborat bo‘lib, unda har bir satr matn korpusidagi hujjatni, har bir ustun esa korpusda topilgan unikal so‘z yoki “termin”ni ifodalaydi. Matritsadagi qiymatlar har bir hujjatdagi har bir atamaning paydo bo‘lish chastotasini ko‘rsatadi.

Matnni oldindan qayta ishlash deganda shovqinli, tartiblanmagan matnni aniq maqsadga tayanib tartiblash tushuniladi. Bunda jumlaning kontekstual ma’nosiga hissa qo‘shmaydigan so‘zlar olib tashlanadi. Tabiiy tilni qayta ishlashning muhim qismi bo‘lgan matnni oldindan qayta ishlash turli usullar bilan amalga oshirilishi mumkin, chunki bu usullar muammoning talabiga ko‘ra tanlanadi, ammo mashinali o‘rganish modellarining ishlashini yaxshilash uchun har bir

muammoda qo‘llanilishi kerak bo‘lgan bir nechta usullar mavjud. Mumkin bo‘lgan eng yaxshi natijalarni ta’minlash uchun usullar ma’lum bir tartibda qo‘llanilishi kerak [Aachal J., Harshada J., 2022: 878]

Matnlarini qayta ishlash jarayoni garchi iterativ bo‘lsa ham, mashinali o‘rganish modeli/algoritmi uchun muhim rol o‘ynaydi. Matnli ko‘rinishlarini ikki qismga ajratish mumkin:

1. Matnni diskret ko‘rinishlari (Discrete text representations);
2. Taqsimlangan/uzluksiz matn ko‘rinishlari (Distributed/Continuous text representations)

Korpus matnlarini diskret ko‘rinishda tasvirlashda korpusdagi so‘zlar bir-biridan mustaqil tarzda ifodalanadi. Ushbu yondashuvda so‘zlar korpus(lar) lug’atidagi o‘rniga mos keladigan indekslar bilan ifodalanadi. Ushbu toifaga CountVectorizer metodi ham kiradi [Elov, B., Hamroyeva, Sh., Alaev, R., Xusainova, Z., Yodgorov, U., 2023:117-129]. CountVectorizer imkoniyatlarini ochib berishda ixtiyoriy tanlangan to‘rtta badiiy adabiyotlarning so‘zlar statistikasi va asarda eng ko‘p ishlatilgan 10 ta so‘zlar aniqlanadi. Avvalo, badiiy asardagi matnni qayta ishlash uchun matnga qo‘yilgan maqsaddan kelib chiqib quyidagilar tanlandi:



1-rasm. Matnga qayta ishlov berish

Alifbodan tashqari barcha belgilar olib tashlandi. Alifbodan tashqari belgilarni olib tashlash matnni soddalashtiradi, tahlil qilish va qayta ishlashni osonlashtiradi. Bu, ayniqsa, matnning lingvistik mazmunini tushunishga qaratilgan NLP) vazifalarida foydali bo‘lishi mumkin.

Qisqa so‘zlarni olib tashlash. Qisqa so‘zlar odatda uzunroq so‘zlarga qaraganda kamroq semantik ma’noga ega. Ularni olib tashlash orqali matn ma’lumotlaridagi shovqinni kamaytirish mumkin, bu esa yanada mazmunli tarkibga e’tibor qaratishni osonlashtiradi. Biroq, qisqa so‘zlarni olib tashlash matnni qayta ishlash vazifasining aniq talablari va maqsadlarini hisobga olgan holda amalga oshirilishi kerak. Ba’zan, ma’lum kontekstlarda qisqa so‘zlar muhim ma’lumotlarni o‘z ichiga olishi mumkin. Shu sababli, shovqinni kamaytirish va tegishli kontentni saqlashda muvozanat juda muhimdir.

Matnni kichik harflarga o‘tkazish. Bu qadam matnni oldindan qayta ishlashning eng oddiy va samarali bosqichidir. Matn odatda qisqartma so‘zlardan yoki barcha so‘zlari bosh harflardan iborat bo‘lishi mumkin. Misol uchun matn boshida kelgan “Ona” so‘zi bosh harfda yozilib, matn ichida esa kichik harf bilan yozilgan bo‘lsa bu so‘zlar kompyuter tomonidan ikki xil so‘z sifatida qabul qilinadi va so‘zlarni joylashtirishning keyingi bosqichlarida ikki xil so‘z vektorlari hosil bo‘ladi. Shunday qilib, barcha so‘zlarni kichik harflar bilan yozish matnni qayta ishlashda eng yaxshi amaliyot bo‘ladi.[Alayev, R., Maxmudjonova G., 2023: 79]

Tokenizatsiya. Tokenizatsiya oddiy jarayo bo‘lib, qayta ishlanmagan ma’lumotlarni oladi va uni foydali ma’lumotlarga aylantiradi. Tokenizatsiya tabiiy tilni qayta ishlashda paragraflar va gaplar ma’nosini osonroq belgilash mumkin bo‘lgan kichikroq birliklarga bo‘lish uchun ishlatilinadi [Xusainova, Z., Elov, B., Yodgorov, U., 2023: 27]

Nomuhim so‘zlarni olib tashlash. Nomuhim so‘zlar hujjatning katta qismini tashkil etuvchi matn tarkibidagi so‘zlar bo‘lib, ularning umumiy xususiyati shundaki, hujjatda muhim ma’lumotga ega emas; faqat grammatika tufayli ishlatiladi. Nomuhim so‘zlarni olib tashlash ma’lum bir ma’noga ega bo‘lgan so‘zlarni necha marotaba kelganini aniqlashga yordam beradi [Maxmudjonova G., 2023: 204-211]. Phyton dasturlash tili yordamida bu jarayon amalga oshirildi. Quyida qayta ishlanmagan va qayta ishlangan matn keltirilgan (1-rasm).

	document	clean_doc
0	Quyoshni bahorning boshi va kuzning adog‘ida h...	quyoshni bahorning boshi kuzning adog‘ida qach...
1	XX asrning buyuk yozuvchisi Chingiz Aytmatov b...	asrning buyuk yozuvchisi chingiz aytmatov tani...
2	BUYUK QALB EGASI\nChingiz Aytmatov 1928-yilnin...	buyuk qalb egasi chingiz aytmatov yilning deka...
3	Chol qayiqda yolg‘iz o‘zi Golfstrimda baliq ov...	chol qayiqda yolg‘iz golfstrimda baliq ovlardi...

2-rasm. Matn qayta ishlangandan oldin va keyin

Ixtiyoriy tanlangan badiiy asarlardan olingan natijalar:

Badiiy asarlar	So‘zlar statistikasi
Ahmad Lutfiy Qozonchi “O‘gay ona”	10638
Chingiz Aytmatov, Muxtor Shoxonov “Cho‘qqida qolgan ovchining ohi-zori”	24462
Ernest Xeminguey “Chol va dengiz”	6752
Chingiz Aytmatov “O‘tar qush nolasi”	9509

1-jadval. Badiiy asarlardagi so‘zlar statistikasi

Ahmad Lutfiy Qozonchining “O‘gay ona” asarida eng ko‘p foydalanilgan 10 ta so‘zlar:

Ahmad Lutfiy Qozonchi “O‘gay ona”		Asarda eng ko‘p uchragan so‘zlar	Pythonda Count Vectorizerdan olingan natija
1	bir	539	Top 10 words for 0: 0 bir 539 husayn 176 ona 169 fotimaxonim 148 ikki 134 bor 132 odilbek 101 kun 98 bo‘lgan 85 bo‘ldi 84
2	husayn	176	
3	ona	169	
4	fotimaxonim	148	
5	ikki	134	
6	bor	132	
7	odilbek	101	
8	kun	98	
9	bo‘lgan	85	
10	bo‘ldi	84	

2-jadval. Eng ko‘p ishlatilgan so‘zlar statistikasi

Chingiz Aytmatov, Muxtor Shoxonovning “Cho‘qqida qolgan ovchining ohi-zori” asarida eng ko‘p foydalanilgan 10 ta so‘zlar:

	Chingiz Aytmatov, Muxtor Shoxonov “Cho‘qqida qolgan ovchining ohi-zori”	Asarda eng ko‘p uchragan so‘zlar	Pythonda Count Vectorizerdan olingan natija
1	bir	196	Top 10 words for 2: 0 bir 196 suv 106 bo‘lib 105 qoratoy 105 bo‘lsa 89 sen 88 bo‘ldi 70 isabekov 66 ketdi 64 qiz 62
2	suv	106	
3	bo‘lib	105	
4	qoratoy	105	
5	bo‘lsa	89	
6	sen	88	
7	bo‘ldi	70	
8	isabekov	66	
9	ketdi	64	
10	qiz	62	

3-jadval. Eng ko‘p ishlatilgan so‘zlar statistikasi

Ernest Xemingueyning “Chol va dengiz” asarida eng ko‘p foydalanilgan 10 ta so‘zlar:

	Ernest Xeminguey “Chol va dengiz”	Asarda eng ko‘p uchragan so‘zlar	Pythonda Count Vectorizerdan olingan natija
1	chol	352	Top 10 words for 3: 0 chol 352 baliq 197 bir 178 dedi 148 bo‘lsa 114 bo‘lib 96 o‘yladi 73 bola 73 qayiq 67 olib 62
2	baliq	197	
3	bir	178	
4	dedi	148	
5	bo‘lsa	114	
6	bo‘lib	96	
7	o‘yladi	73	
8	bola	73	
9	qayiq	67	
10	olib	62	

4-jadval. Eng ko‘p ishlatilgan so‘zlar statistikasi

Chingiz Aytmatovning “O‘tar qush nolasi” asarida eng ko‘p foydalanilgan 10 ta so‘zlar:

	Chingiz Aytmatov “O‘tar qush nolasi”	Asarda eng ko‘p uchragan so‘zlar	Pythonda Count Vectorizerdan olingan natija
1	bir	808	
2	bo‘lgan	379	
3	bo‘lib	335	
4	bo‘ladi	264	
5	bor	248	

6	xalq	215	Top 10 words for 1: 0 bir 808 bo‘lgan 379 bo‘lib 335 bo‘ladi 264 bor 248 xalq 215 qilib 198 bitta 183 ikki 157 bo‘lsa 157
7	qilib	198	
8	bitta	183	
9	ikki	157	
10	bo‘lsa	157	

5-jadval. Eng ko‘p ishlatilgan so‘zlar statistikasi

CountVectorizerni har tomonlama o‘rganish uning tabiiy tilni qayta ishlash va matnga asoslangan mashinali o‘rganishda muhim roli bor. So‘z chastotalarini hisoblash va hujjat muddatli matritsalarini yaratishni o‘z ichiga olib, matnni tahlil qilish va tushunishni yaxshilashga yordam beradi. CountVectorizer so‘z statistik tahlillarini osonlashtirish orqali matnli ma’lumotlar haqida chuqurroq ma’lumotga ega bo‘lish imkonini beradi va shu bilan lingvistik ma’lumotlarni qayta ishlashda mashinali o‘rganish algoritmlarining imkoniyatlarini oshiradi.

Foydalanilgan adabiyotlar:

1. Achal J, Harshada J, Bavik J, Charmi Ch “Text Pre-Processing Techniques in Natural Language Processing: A Review” International Research Journal of Engineering and Technology (IRJET), 2022. –B. 878
2. Alayev R, Maxmudjonova G, “O‘zbek tilidagi matnli hujjatlarda izlashni amalga oshirishni takomillashtirish”, Toshkent: O‘zbekistan: til va madaniyat 2023. –B. 79,
3. Elov B, Hamroyeva Sh, Xusainova Z, Xudayberganov N (2023). “O‘zbek tili korpusi matnlarini qayta ishlashda CountVectorizer, TF-IDF hamda Co-occurrence matrix usullarining ahamiyati” Elektron lug’atlar yaratishning nazariy va amaliy asoslari mavzusidagi xalqaro ilmiy-amaliy anjuman materiallari ., Andijon-2023. – B. 81
4. Elov B., Hamroyeva Sh., Alaev R., Xusainova Z., Yodgorov U., “O‘zbek tili korpusi matnlarini qayta ishlash usullari” Raqamli Transformatsiya va Sun’iy Intellekt ilmiy jurnali, 2023. –B. 117-129.
5. Maxmudjonova G., “Nomuhim so‘zlar tushunchasi va uning ahamiyati”. Kompyuter lingvistikasi: muammolar, yechim, istiqbollar Xalqaro ilmiy-amaliy konferensiya materiallari, 2023. 204-211.
6. Xusainova, Z., Elov, B., Yodgorov, U., O‘zbek tili matnlari uchun tokenizayorni ishlab chiqish.., MUHAMMAD AL-XORAZMIY AVLODLARI ilmiy-amaliy va axborat-tahliliy jurnal, 2023. –B. 27

MASHINA TARJIMASIDA SINONIMLAR BILAN BOG‘LIQ MUAMMOLAR

Sotvoldiyeva Nilufar Sirojiddin qizi
Mirzo Ulug‘bek nomidagi O‘zbekiston
Milliy universiteti 2-bosqich magistranti.
E-mail: sotvoldiyevanilufar0@gmail.com

Annotatsiya: ushbu maqolada turk va o‘zbek tillaridagi sinonimlarning mashina tarjimasi jarayonida kelib chiqadigan muammolari o‘rganilib chiqildi. Ma’lumki bugungi kunda turk adabiyoti durdonalariga qiziqish katta. Shu sababli ko‘plab turk asarlari o‘zbek tiliga tarjima qilinmoqda. Bunda, albatta, mohir tarjimonlar bizga ko‘mak bo‘lmoqda. Biroq mashina orqali ham qiziquvchilar tomonidan tarjimalar amalga oshirilmoqda. Bu, albatta, vaqtni hamda mablag‘ni tejashi mumkin, biroq bunday yo‘l bilan qilingan tarjimlar o‘zini qanchalik oqlamoqda? Bilamizki, bizning tilimiz boy va imkoniyatlari keng til hisoblanadi. Bu imkoniyatlarni hosil qiluvchi tilimizda omonim, antonim, ko‘p ma’noli so‘zlar hamda sinonim kabi leksik birliklar mavjud. Xo‘sh, mashina qilayotgan tarjimalarda tilimizning asosi bo‘lgan bunday birliklar nechog‘lik o‘zining asl ma’nosini saqlab qolmoqda? Konteks bilan bog‘liq semantik muammolar yuzaga kelmayaptimi? Biz ushbu maqolada mana shu muammolarni ko‘rib chiqdik.

Kalit so‘zlar: Mashina tarjimasi, CAT, HAMT, FAMT, turk va o‘zbek tilidagi sinonimlar, sinonimlarning semantik va lingvistik xususiyati.

Abstract: this article examines the problems that arise in the process of machine translation of synonyms in Turkish and Uzbek languages. It is known that today there is great interest in the masterpieces of Turkish literature. For this reason, many Turkish works are translated into Uzbek. Of course, experienced translators help us with this. However, those who wish can make transfers using a machine. Of course, this can save time and money, but how worthwhile are translations done this way? We know that our language is a rich and multifaceted language. In our language there are lexical units such as homonyms, antonyms, polysemantic words and synonyms that create these possibilities. So how many of these units, which form the basis of our language, retain their original meaning in machine translations? Are there semantic issues related to context? We examined these problems in this article.

Keywords: Machine translation, CAT, HAMT, FAMT, Turkish and Uzbek synonyms, semantic and linguistic characteristics of synonyms.

Аннотация: В данной статье изучены проблемы, возникающие в процессе машинного перевода синонимов турецкого и узбекского языков. Известно, что сегодня существует большой интерес к шедеврам турецкой литературы. По этой причине многие турецкие произведения переводятся на узбекский язык. Конечно, в этом нам помогают опытные переводчики. Однако переводы желающие осуществляют с помощью машины. Конечно, это может сэкономить время и деньги, но насколько оправдывают себя переводы, выполненные таким способом? Мы знаем, что наш язык – богатый и многогранный язык. В нашем языке существуют такие лексические единицы, как омонимы, антонимы, многозначные слова и синонимы, которые создают эти возможности. Итак, сколько таких единиц, составляющих основу нашего языка, сохраняют свое первоначальное значение в

машинных переводах? Есть ли семантические проблемы, связанные с контекстом? Эти проблемы мы рассмотрели в этой статье.

Ключевые слова: Машинный перевод, CAT, HAMT, FAMT, турецкие и узбекские синонимы, семантическая и лингвистическая характеристика синонимов.

Kirish

Jamiyatimizda hozirgi davrda avtomatlashgan informatsion texnologiyalar va kompyuter muhim ahamiyat kasb etmoqda. Bu esa yangi yo‘nalishlarning yuzaga kelishiga sabab bo‘lmoqda. Mashina tarjimasi kompyuter lingvistikasining muhim yo‘nalishi hisoblanib, fanga oid barcha tadqiqotlar dasturiy ta`minot va lug‘atshunoslik bilan uzviy bog‘liqdir. Tabiiy tilning milliy leksik boyligi, imkoniyatlari lug‘atlarda namoyon bo‘ladi. Shu bois lug‘atlar mashina o‘qiy oladigan formatda bo‘lishi, qidiruv tizimiga egaligi, indekslanganligi, kategoriyalashtirilganligi, klasterlashtirilganligi va foydalanuvchiga undan foydalanish imkoniyatining qulayligi va kengligi bilan xarakterlanadi. Dasturiy ta’minot esa ana shunday mukammal lug‘atlar bilan lingvistik ta`minot ba’zasini ta’minlab beradi. Mashina tarjimasi haqida so‘z borar ekan, tarjimon turlari haqida alohida to‘xtalib o‘tish maqsadga muvofiq bo‘ladi.

♦ mashina yordamidagi inson tarjimasi (machineassisted human translation – MAHT) – inson tomonidan matnni tarjima qilishda kompyuter elektron lug‘atlaridan foydalanish va unga kiritilgan ko‘rsatmalardan foydalanish;

♦ kompyuter yordamidagi tarjima (computer assistedtranslation – CAT) – matnni kompyuter vositasida tarjima qilish va unda anglashilgan asosiy tushunchalarni ajratishga xizmat qiladi;

♦ inson yordamidagi mashina tarjimasi (human-assisted machine translation – HAMT) – kompyuter dasturiy ta’minoti orqali amalga oshiriladigan tarjimaga inson resursi tomonidan tahrirlash vazifalarini yuklash;

♦ to‘liq avtomatik mashina tarjimasi (fully automatic machine translation – FAMT) – tarjimon dasturi orqali matnni ham tarjima, ham tahrirlovchi vazifalarini amalga oshirish.

Qisqacha shuni ta’kidlash o‘rinliki, kompyuter lingvistikasining yo‘nalishi sifatida o‘rganilayotgan mashina tarjimasi har bir rivojlanayotgan jamiyat uchun zaruriy sohalardan biridir. Bildirilgan ijobiy va salbiy fikrlar uning yanada takomillashuviga xizmat qiladi [Абрурахмонова Н., 2012: 16]

Sinonimlar – ma’nolari o‘xshash bo‘lgan so‘zlar bo‘lib, bunga oid soha o‘xshash so‘zlar orasidagi ma’noning nuanslari va soyalarini tushunishni o‘z ichiga oladi. Tarjimada ushbu nuanslarni o‘quvchiga samarali yetkazish uchun kontekstga eng mos bo‘lgan sinonimni tanlash muhim ahamiyat kasb etadi. Ma’noning nuanslari va soyalari: sinonimlar umumiy ma’noga ega bo‘lishi mumkin, lekin ular ko‘pincha konnotatsiya, ohang yoki foydalanishda nozik farqlarga ega. Masalan, “kulgi” va “qahqaha” sinonimlari ko‘rib chiqiladigan bo‘lsa, ikkalasi ham ijobiy his-tuyg‘ularni bildirsa-da, “qahqaha” yanada kuchli yoki jo‘shqin tuyg‘uni ko‘rsatadi, “kulgi” esa, albatta, hayajonlanmasdan qoniqish hissini anglatadi[Abdurakhmonova, N.2020,2022,2023].

Sinonim so‘zlarning kontekstga mosligini aniqlash va to‘g‘ri sinonimlarni tanlab olish ko‘p jihatdan ularni ishlatadigan kontekstga bog‘liq. Ba’zi sinonimlar ma’lum vaziyatlarda bir-birini almashtirsa, boshqalarda biri boshqasidan ko‘ra ko‘proq mos keladi. Misol uchun, ba’zi so‘zlar bir-birining o‘rnini bosadigan bo‘lib tuyulishi mumkin, ammo ulardan foydalanish muayyan kontekst yoki mintaqalar bilan cheklanishi mumkin [Eric Stanley. "Polysemy and Synonymy and how these Concepts were Understood from the Eigtenth Century onwards in Treatises, and

Applied in Dictionaries of English" in Historical Dictionaries and Historical Dictionary Research, papers from the International Conference on Historical Lexicography and Lexicology. - University of Leicester, 2002. – P 157-1]

Biz bu masalani turk va o‘zbek tilida ko‘rib chiqamiz. Masalan turk tilidagi “Baş-Kafa” sinonimlari bizdagi bosh va kalla sinonimlariga to‘g‘ri keladi. Biroq bu so‘zlar ba’zan jumladagi ma’nosiga ko‘ra o‘zgarishi mumkin.

Çocuğun kafasına taş geldiği için kafası kanadı.

Bolaning boshiga tosh tekkani uchun boshi qonadi.

Sende de hiç kafa yokmuş be arkadaş !

Senda ham hech aql yo‘q ekan-da, do‘stim!

Bu yerda kafa so‘zi qo‘llangani bilan konteksdan kelib chiqib bosh so‘zi aql so‘ziga to‘g‘ri kelyapti va shu so‘z bilan sinonimlik hosil qilyapti.

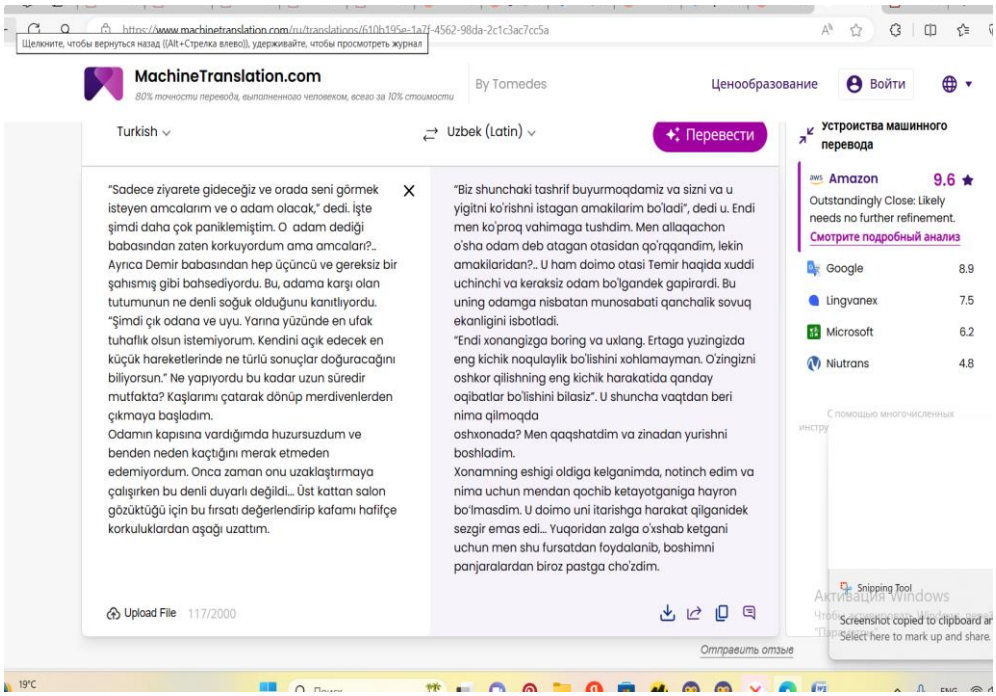
Siyah = Kara

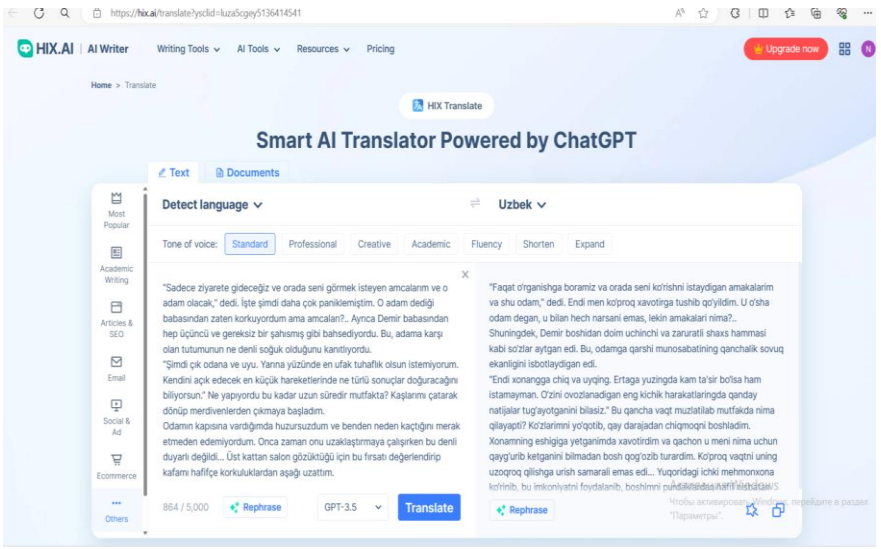
Evinin üzerine kara bulutlar çöküverdi

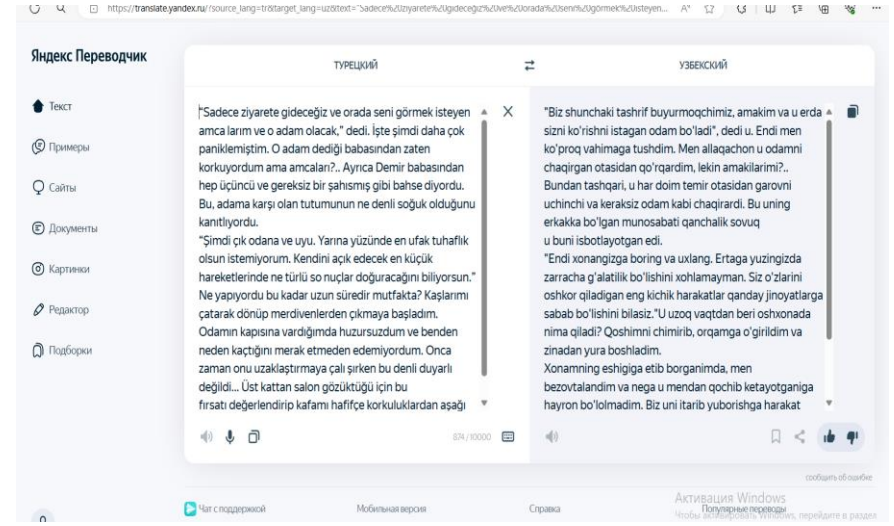
Kara günler bir gün sona erecektir

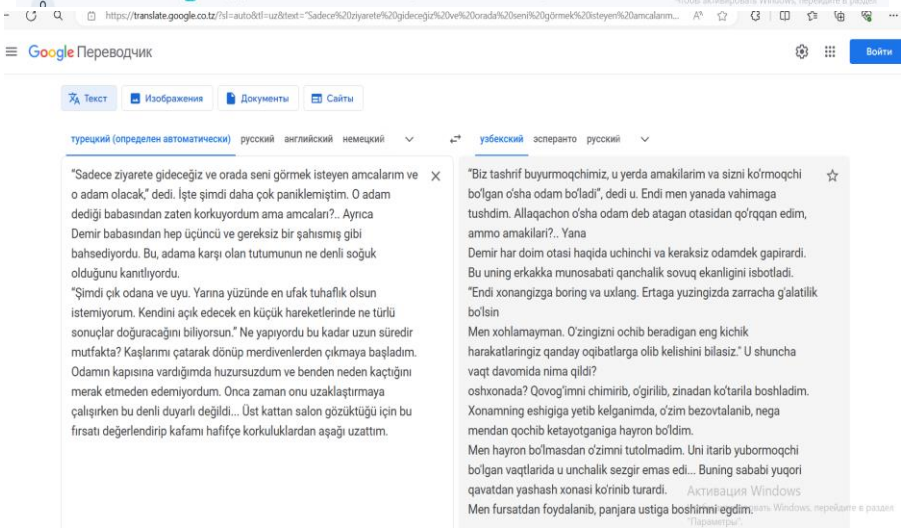
Odatda qora so‘zining sinonimi kara so‘zidir. Biroq “Kara günler bir gün sona erecektir”, (Qora kunlar bir kun tugaydi)gapida “kara günler”, kötü günler (yomon kunlar)ma’nosida qo‘llanilgan. Shu sababli “kara günler” o‘rniga “siyah günler” qo‘llay olmaganimiz uchun bu jumladagi “kara” so‘zining sinonimi “kötü” so‘zidir.

Quyidagi bir xil turkcha matnni bir nechta mashina tarjimasi yordamida tarjima qilib ko‘ramiz.









Yuqoridagi rasmlardan ko‘rinib turibdiki, tarjimon dasturlar matnni deyarli bir xil xatoliklar bilan o‘zbek tiliga o‘girmoqda. “Şahıs” (shaxs) va adam (erkak) so‘zlarining ikkalasi ham odam deb tarjima qilinyapti. Bizda shaxs, kishi, odam, inson so‘zlari o‘zaro sinonimlik hosil qiladi[Abdurakhmonova, N.2020,2022,2023].

Xulosa o‘rnida shularni ta’kidlash lozimki mashina tarjimasida sinonimiya hodisasi alohida xarakterlanadi va bu masala maxsus tadqiqotlar olib borishni talab qiladi. Biz mashina tarjimasini yanada mukammallashtirish va rivojlantirish uchun har bir sinonim bilan alohida ishlangan lingvistik ta’minot yaratishimiz zarur.

Foydalanilgan adabiyotlar ro‘yxati:

1. Eric Stanley. “Polysemy and Synonymy and how these Concepts were Understood from the Eigtenth Century onwards in Treatises, and Applied in Dictionaries of English” in Historical Dictionaries and Historical Dictionary Research, papers from the International Conference on Historical Lexicography and Lexicology. - University of Leicester, 2002. – P 157-1
2. Н.Абдурахмонова. Машина таржимасининг лингвистик асослари, Тошкент: Akademnashr, 2012. 16.
3. Ismailov, A. S., Shamsiyeva, G., & Abdurakhmonova, N. (2021). Statistical machine translation proposal for Uzbek to English. *Science and Education*, *2*(12), 212-219.
4. Абдурахмонова, Н., & Бойсариева, С. (2023). TABIIY TILNI QAYTA ISHLASHDA (NLP) OKKAZIONALIZMLARNING MORFEM TAHLILI. *МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА*, *6*(3).
5. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In *Компьютерная обработка тюркских языков. TURKLANG 2022* (pp. 231-240).
6. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
7. Khusainov, A., Suleymanov, D., Gilmullin, R., Minsafina, A., Kubedinova, L., & Abdurakhmonova, N. (2020, November). First Results of the TurkLang-7 Project: Creating Russian-Turkic Parallel Corpora and MT Systems. In *CMCL* (pp. 90-101).
8. https://translate.yandex.ru
9. https://translate.google.com
10. https://www.machinetranslation.com
11. https://hix.ai.translate.com

INGLIZCHA-O‘ZBEKCHA MASHINA TARJIMASI UCHUN FRAZEOLOGIK BIRLIKLAR BAZASINI YARATISH TAVSIFI

Abduraxmonov Umbar
O‘zbekiston Milliy universiteti magistranti.
E-mail: umbarabdurahmonov@gmail.com

Annotatsiya: Mazkur maqolada inglizcha-o‘zbekcha mashina tarjimasi uchun frazeologik birliklar bazasini yaratish, lug‘atlarda inglizcha frazeologik birliklarning ishlash tamoyillari, frazeologizmlarning turlari, frazeologik birliklarning o‘rganish tarixi, birliklarni o‘zbek tiliga tarjima qilish usullari hamda leksik vositalarining semantik-strukturaviy xususiyatlariga alohida e’tibor qaratildi.

Kalit so‘zlar: frazeologik birliklar, mashina tarjimasi, ma'lumotlar bazasi, emotsional-ekspressiv, nutq sintezatori.

Annotation: In this article, the creation of a database of phraseological units for English-Uzbek machine translation, the working principles of English phraseological units in dictionaries, types of phraseological units, the history of the study of phraseological units, methods of translating units into Uzbek and the semantic-structural analysis of lexical tools special attention was paid to its features.

Keywords: phraseological units, machine translation, database, emotional-expressive, speech synthesizer.

Аннотация: В данной статье описано создание базы фразеологизмов для англо-узбекского машинного перевода, принципы работы английских фразеологизмов в словарях, виды фразеологизмов, история изучения фразеологизмов, методы перевода единиц на Узбекский язык и при семантико-структурном анализе лексических средств особое внимание было уделено его особенностям.

Ключевые слова: фразеологизмы, машинный перевод, база данных, эмоционально-выразительные средства, синтезатор речи.

Zamonaviy lingvistik qarashlar hozirgi tilshunoslikning muhim yo‘nalishlaridan biri bo‘lgan til birliklarining nutq faoliyati vositasi sifatida qo‘llanish qonuniyatlari va tamoyillarini aniqlash, shuningdek, til birliklarining o‘ziga xos xususiyatlarini o‘rganishga qaratilgan. Bir tomondan, til tizimi va uning asosiy birligi bo‘lgan so‘zning semantik-tarkibiy barqaror birliklarini tahlil qilishga alohida e’tibor beradi. Frazeologiya til leksikasining bir qismi bo‘lgani uchun tilshunoslikning leksikologiya sohasiga mansub bo‘limi sifatida qaralib, ikki yoki undan ortiq so‘zlarning turg‘un birikmasi sanaladi. Frazeologizmlar katta so`zni o`z ichiga olishiga qaramay, ular bir umumiy ma`no bilan birikadi va emotsional-ekspressiv ma`no ifodalaydi. Frazeologizmlar badiiy adabiyotda obrazli va ta’sirli vosita sifatida keng qo‘llanilib, odatda, tildagi erkin so‘z birikmalari asosida hosil bo‘lsa-da, biroq, ular ko‘chma ma’noga ega bo‘ladi. Turli tillardagi frazeologizmlar o'sha xalqning ba’zi odatlarini aks ettirib, ularni bir tildan boshqa tilga tarjima qilish, mazmuniga mos keladigan iboralarsiz ishlatiladi. Frazeologizmlar semantik jihatdan

umumlashgan ko‘chma ma’no bilan izohlansa, grammatik jihatdan ular yaxlitlik komponentlarining o‘zaro bog‘lanishi va gapda bir xil tartibni saqlab turishi bilan izohlanadi.

Frazeologik mashina tarjimasi- avtomatik tarjimaning sifatini oshirish uchun Rossiyada professor G.G.Belonogov tomonidan ishlab chiqilgan konsepsiya hisoblanadi. Unga ko‘ra, matndagi tushunchalar ko‘pincha alohida so‘zlar yordamida emas, balki so‘z birikmalari yordamida ifodalanadi. O‘z-o‘zidan kelib chiqadiki, matnni bir tildan boshqasiga tarjima qilish jarayonida ham mazmunni yuzaga chiqaruvchi asosiy til birliklari alohida so‘zlar emas, balki tipik vaziyatlarni ifodalovchi frazeologik birikmalar (bu yerda frazeologik birlik keng ma’noda tushuniladi) hisoblanadi[Abdurakhmonova, N.2020,2022,2023]. Mazkur yondashuvga muvofiq matnda tez-tez uchraydigan so‘zlar, iboralar, so‘z birikmalari, nutqiy etiket namunalarining tarjima ekvivalentlari bi-lan boyitilgan ma’lumotlar bazasi frazeologik mashina tarjimasi tizimining asosini tashkil etadi. Bunda so‘zlarning distributsiyasiga eng ko‘p e’tibor beriladi. Bu metod polisemantik so‘zlarning tarjimada mos keluvchi to‘g‘ri ekvivalentini tan- lashga, omonim so‘zlarning kontekstdagi holatini aniqlashga, iboralarning ham idiomatik xususiyatini belgilashga imkon beradi.

Taniqli rus tilshunosi A. Shaxmatovning “Sintaksis” asari bu borada kata ishlarni amalga oshirdi. Bu ishni keyinchalik akademik V.V.Vinogradov davom ettirdi. Ingliz frazeologiyasi sohasidagi ulkan ishlar professor Kunin, I. Made by Arnold va boshqalar tomonidan nashr etilgan. Frazeologik birliklar, odatda, hech qanday asosga ega bo‘lmagan, nutqda erkin shakllana olmaydigan, lekin tayyor birlik sifatida ishlatiladigan iboralar deb tan olinadi [Abdurakhmonova, N.2020,2022,2023]. Frazeologik birliklarning asosiy xususiyatlaridan biri leksik komponentlar va grammatik tuzilishning barqarorligidir. Muloqot jarayoni natijasida o‘zgarib turadigan erkin so‘z birikmalaridan farq qiluvchi frazeologik birikmalar o‘zgarmas butun frazeologik birikmalar sifatida ishlatiladi.

Shuningdek, akademik V.V. Vinogradov uchta asosiy turni ajratib ko‘rsatadi:

Frazeologik bitimlar- bular ma’nosi jihatidan bir-biriga mos kelmaydigan frazeologik birliklar, ularning butun ma’nosi tarkibiy so‘zlarning ma’nolari bilan mutlaqoasoslanmagan bo‘ladi.

Frazeologik birlashish - bu frazeologik birliklar, ularning yaxlit ma’nosi tarkibiy qismlarning ma’nosidan iborat va shu bilan birga tarkibiy qismlardan biri birlashma deb ataladi.

Frazeologik birlashma- tuzilishining bo‘linishiga qaramay, ular umumlashtirilgan yaxlit ma’noga ega bo‘ladi.

Shu bilan birga, tarjimon tomonidan amalga oshiriladigan tarjima o‘zgarishlarini matnlar o‘rtasidagi sintaktik, morfologik, leksik va semantik jihatdan kuzatish mumkin. Parallel matnda shunday holatlarni erkin kuzatish, konteksdan kelib chiqib, so‘zlarning o‘zaro muqobillarini aniqlash imkoniyati tug'iladi. Kalit so‘zlar esa so‘zning nechog'lik kontektda qo‘llanganlik chastotasiga qarab belgilanadi. Parallel korpus nafaqat mashina tarjimasi uchun lingvistik resurs, balki bilingval lug‘atshunoslik, qiyosiy tarjima, uchun muhim manba bo‘lib hizmat tarjimasi uchun parallel Endilikda neyro yoki statistik mashina ma'lumotlar bazasini yaratish va uning tarjima xotirasiga yog'ishtirma tilshunoslik kabi sohalar lingvistik jihatlarini chuqur asosiy obyekti sanaladi. Nechog‘lik tarjima matnlarning muqobil juftliklari aniqlanib, boyitib borilsa, kelajakda turli uslubga tegishli avtomatik tarjima qiluvchi tarjimon dasturlarning sifatini yaxshilashda qo‘l keladi. Shuningdek, parallel ravishda yaratilgan og'zaki (Parallel WaveNet) va yozma matnlar

nutq sintezatoriga doir tadqiqotlar olib borishda mashina fondi vazifasini o‘tamoqda.[N. Abduraxmonova.2018]

Shu o'rinda, CLARIN infrastrukturasida 86 turdagi bilingval korpuslarning bazasi kirilganligini alohida qayd etish o‘rinli. [Abdurakhmonova, N.2020,2022,2023]. Undagi mavjud parallel korpuslar Yevropa tillariga mansub va boshqa til oilalariga til juftliklarining hajmi, lingvistik razmetkasi, foydalanish imkoniyati, muqobillik konkordansi, matnlarning uslubiy ko‘lami va undagi janrlar xilma-xilligiga ko'ra farqlanadi. Ikki tildagi parallel korpuslarni V.Zaxarov shartli ravishda ikki turga ajratadi:

1) muayyan tildagi asliyat matnning ikkkinchi tildagi aynan tarjimasi;

2) biror sohaga oid matnlarning mavzulashtirilgan guruhlar bo'yicha bir tasnifga kiritilishi.

Parallel korpuslar yordamida so‘z va so‘z birikmalarining o‘zaro valentliklarini belgilash muayyan guruhga kiruvchi leksik-semantik guruhdagi so‘zlarning semantik maydonini aniqlash, qolaversa, mashina tarjimasi uchun tarjima sifatini oshirishga xizmat qiladi .

Xulosa o‘rnida shuni aytish mumkinki, mashina tarjimasi bazasi ta’limoti uchun o‘zbek-ingliz tillari frazeologik birliklarining mukammal bazasini yaratish tarjimon dasturlarning rivojiga asos bo‘lib xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Kuganova, Sh. Asiljonova. Frazeologik birliklar va tarjima. Central Asian Research Journal for Interdisciplinary Studies, 2022.
2. B.Abdumuminov. Lingvistik dasturlar uchun o‘zbek tilidagi frazemalar bazasini yaratish tamoyillari. Kompyuter lingvistikasi:muammolar, yechim, istiqbollar respublika I ilmiy-texnikaviy konfrensiyasi, 2021.
3. Абдурахмонова, Н., & Бойсариева, С. (2023). TABIIY TILNI QAYTA ISHLASHDA (NLP) OKKAZIONALIZMLARNING MORFEM TAHLILI. *МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА*, *6*(3).
4. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In *Компьютерная обработка тюркских языков. TURKLANG 2022* (pp. 231-240).
5. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
6. Khusainov, A., Suleymanov, D., Gilmullin, R., Minsafina, A., Kubedinova, L., & Abdurakhmonova, N. (2020, November). First Results of the TurkLang-7 Project: Creating Russian-Turkic Parallel Corpora and MT Systems. In *CMCL* (pp. 90-101).
7. F. Yusupjanova. Ingliz tilida frazeologik birliklarni o‘qitishdagi muammolar. International journal of innovative achivements in science, 2023.
8. N. Abduraxmonova. Mashina tarjimasining lingvistik ta’minoti(monografiya). - Toshkent:Muharrir, 2018. 165b.
9. N. Abduraxmonova. Kompyuter lingvistikasi (darslik) - Toshkent, 2021. 400b.

MASHINALI TARJIMA TIZIMLARI UCHUN O‘ZBEK TILI SO‘ZLARINI LEMMALASH VA GAP STRUKTURALARINI TAHLIL QILISH

Sharipov Maqsud Siddiqovich
Texnika fanlari nomzodi, dotsent
UrDU, Kompyuter ilmlari kafedrasi dotsenti.
E-mail: m.sharipov@urdu.uz

Sobirov Og‘abek Ollayor o‘g‘li
UrDU, Kompyuter ilmlari kafedrasi tayanch doktoranti.
E-mail: sobirov.o@urdu.uz

Sharipova Gavharjon Maqsud qizi
UrDU, Kompyuter ilmlari kafedrasi talabasi.
E-mail: gavharjonsharipova1@gmail.com

Annotatsiya: Tarjima butun dunyodagi odamlar o‘rtasida samarali muloqot qilish imkonini beradi. Mashina tarjimasi - bu tabiiy tilni bir tildan boshqasiga avtomatik ravishda, ma’nosini saqlab qolgan holda o‘girishdir. Ushbu maqolada mashinali tarjima tizimlari uchun o‘zbek tili gap strukturasi va o‘zbek tili so‘zlarini lemmalash, lemmalashning tarjima uchun muhimligi haqida ma’lumot berilgan. Mashina tarjimasi turlari bir nechta hisoblanadi, lemmalash aynan qaysi turlar uchun ishlatilishi ham aytib o‘tilgan va misollar keltirilgan.
Kalit so‘zlar. Mashinali tarjima, so‘z, lemma, lemmalash, gap, gap bo‘laklari, gap turlari.

Annotation: Translation enables effective communication between people all over the world. Machine translation is the automatic translation of natural language from one language to another, preserving the meaning. This article provides information on Uzbek sentence structure and Uzbek word lemmatization for machine translation systems, and the importance of lemmatization for translation. There are several types of machine translation, the use of lemmatization specifically for those types is mentioned and examples are given.
Keywords: Machine translation, word, lemma, lemmatization, sentence, sentence fragments, sentence types.

Аннотация: Перевод обеспечивает эффективное общение между людьми по всему миру. Машинный перевод — это автоматический перевод естественного языка с одного языка на другой с сохранением смысла. В этой статье представлена информация о структуре предложений на узбекском языке и лемматизации узбекских слов для систем машинного перевода, а также о важности лемматизации для перевода. Существует несколько видов машинного перевода, упоминается использование лемматизации именно для этих типов и приводятся примеры.
Ключевые слова: Машинный перевод, слово, лемма, лемматизация, предложение, фрагменты предложений, типы предложений.

Mashina tarjimasi (MT) - bu tabiiy tilni bir tildan boshqasiga avtomatik ravishda, ma'nosini saqlab qolgan holda o'girishdir. Tarjima butun dunyodagi odamlar o'rtasida samarali muloqot qilish imkonini beradi. Ta'lim, iqtisod, meditsina, turizm va san'at sohalarida MT tizmlaridan keng ko'lamli va samarali foydalanib kelinmoqda. Bundan tashqari, MT katta hajmdagi matnlarni tezda tarjima qilish uchun ishlatilishi mumkin, bu esa an'anaviy tarjima usullari yordamida imkonsiz bo'lib qoladi .

MTning asosiy turlari:

- Qoidalarga asoslangan mashina tarjimasi (Rule-based machine translation)
- Lug'atga asoslangan mashina tarjimasi (Dictionary-based machine translation)
- Statistik mashina tarjimasi (Statistical machine translation)
- Misol asosidagi mashina tarjimasi (Example-based machine translation)
- Neyron mashina tarjimasi (Neural machine translation)
- Gibrid mashina tarjimasi (Hybrid machine translation)

Har bir narsaning ustun va kamchilik tomonlari bo'lganidek, MT tizimlarini qurishda ham o'ziga yarasha kamchilik tomonlari bor. MT tizimlarini qurishda uchraydigan kamchiliklar sifatida quyidagilarni aytish mumkin, *bir nechta ma'noli so'zlar, bir nechta grammatik tuzilishga ega bo'lgan jumlalar* va *grammatikaning boshqa muammolari*. Ushbu maqolada, o'zbek tili uchun MT tizimlari qurishda yuqoridagi muammolardan, aynan *murakkab grammatik tuzilishga ega jumlalar (gaplar)* haqida to'liq ma'lumot beriladi. Bundan tashqari maqolada mashinali tarjima tizimlari uchun o'zbek tili so'zlarini lemmalash, lemmalashning tarjima uchun muhimligi haqida ma'lumot beriladi. Mashina tarjimasi turlari bir nechta hisoblanadi, lemmalash aynan qaysi turlar uchun ishlatilishi ham aytib o'tilgan va misollar keltirilgan.

Lemmalash. Lemma so'zning asosiy yoki lug'at shaklidir. Tilshunoslikda, ayniqsa morfologiya va leksikografiyada lemma so'zning kanonik shaklini ifodalaydi, undan uning barcha flektiv shakllari hosil qilish mumkin. Lemmalash - bu so'zlarni asosiy yoki lug'at shakliga, ya'ni lemmasiga qisqartirish jarayonidir. Lemmalash o'zgaruvchanlikni kamaytirish uchun turli xil tabiiy tillarni qayta ishlash [inglizcha: Natural Language Processing (NLP)] vazifalarida, jumladan, mashina tarjimasi, so'z turkumlarini belgilash va hissiyotlarni tahlil qilishda qo'llaniladi. Misol uchun, *'matn', 'matnni', 'matnga', 'matnlar'* kabi so'z shakllarining lemmasi *'matn'* hisoblanadi.

Gap strukturasini o'rganish orqali MTning bir nechta turlari, xususan, misol asosidagi mashina tarjimasi (example-based machine translation) orqali quriladigan MT tizimlarini yaxshilash mumkin. Misol asosidagi mashina tarjimasi - bu MTning turi bo'lib, u ko'pincha oniy vaqtda parallel matnlardan iborat ikki tilli korpusdan foydalanish bilan tavsiflanadi. Quyidagi misolda MAMT tizimini qurishda gap struktursi va lemmalashning ahamiyatini ko'rib chiqamiz:

I. Juft jumlalarni o'z ichiga olgan ikki parallel korpus olinadi.
*Ana u **ko'k soyobon**ning narxi qancha? - How much is that **blue umbrella**?*
*Ana u **katta kamera**ning narxi qancha? - How much is that **big camera**?*

II. Bu juftliklar o'xshash va farq qiluvchi qismlarga ajratiladi, bunda farqlanuvchi va o'xshash qismlar so'zlari lemmalash orqali topiladi. Keyin o'xshash va farqli so'zlar o'zaro moslanadi, shu tariqa matnlar tarjima qilinadi.
*Ana u **X**ning narxi qancha? – How much is that **X**?*
ko'k soyobon – blue umbrella

katta kamera – big camera

O‘zbek tili ilmiy faoliyatida lemmalash va mashinali tarjima jarayonlari uchun muhim ishlar amalga oshirilib natijalari e’lon qilingan va davom etmoqda. Ushbu maqolani yozishda shunday ilmiy ishlar bilan tanishib qilgan. Ushbu maqolada qoidaga asoslangan lemmalash jarayonining algoritmini qurish haqida aytilgan, bunda so‘zlar va qo‘shimchalar bazasi yig‘ilgan va qoidalar asosida qo‘shimchalarni olib tashlab so‘zlarning lemmasini topish chekli holatlar mashinasi orqali amalga oshirilgan [1]. Ushbu maqolada o‘zbek tilida qo‘llaniladigan uchta yozuv shakli: kirill, hozirda rasmiy lotin va yangi e’lon qilingan yangi lotin alifbolari o‘rtasida mashina transliteratsiyasi (harflarni almashtirish) vositasini taqdim etishgan bo‘lib, bu jarayon qoidalarga asoslangan va sozlash usullari kombinatsiyasidan foydalangan holda yaratilgan [2]. Bu maqolada, avtomatik morfologik tahlil qiluvchi dasturning jarayoni qisqacha ko‘rsatib o‘tilgan bo‘lib. ushbu dastur mashina tarjimasi va o‘zbek tilidagi so‘zlar birliklarining tartibi va qoidalari morfologik jihatdan o‘rganilgan [3]. Bu maqolada agglyutinativ tillar lug‘atida mavjud bo‘lmagan so‘zlar uchun teglash masalasi, lemmalash va o‘zakni topish jarayonlari bilan hamohangligida tushuntirilib misollar bilan keltirib o‘tilgan [4]. Ushbu maqolada teglash metodi bo‘yicha ya’na bir muhim ma’lumotlar keltirib o‘tilgan bo‘lib, bu jarayonning yechimlari turli xil chizmalar va jadvallar asosida keltirilgan, aynan *‘uzun’* so‘zining ko‘p manoligi 1-jadvalda keltirib o‘tilgan [5].

Mashinali tarjima uchun lemmalashning muhimligini xorijiy ilmiy ishlarda ham ko‘rish mumkin. Ushbu maqolada mualliflar ingliz-malay parallel korpusini olish orqali ingliz-malay statistik mashina tarjimasini amalga oshirishgan, bunda, ular lug‘at va inglizcha lemmalash jarayoni yordamida ingliz-malay statistik mashina tarjimasini yaxshilashga harakat qilishgan va tadqiqotilari shuni ko‘rsatadiki, ikki tilli lug‘at va inglizcha lemmatizatsiyadan foydalanish yondashuvining kombinatsiyasi ingliz tilidan malay tiliga tarjima uchun BLEU ballini 12,90 dan 15,41 gacha oshirdi [6]. Bu maqolada xitoy tilidan ingliz tiliga statistik mashina tarjimasini yaxshilash uchun ingliz lemmatizatsiyasidan foydalanishni taklif qilishgan. Bunda mualliflar CLAW-5 teglar to‘plami uchun ingliz tilidagi so‘z turkumlarini teglash vositasini joriy qilishgan va ingliz tilini lemmatizatsiya qilish uchun ingliz morfologik analizatoridan foydalanishgan, ularning yondashuvi oz miqdordagi o‘quv ma’lumotlari bilan statistik mashina tarjimasini uchun juda samarali ekanligi isbotlangan va ular keng qamrovli tajribalar o‘tkazishgan, xulosalar chiqarish uchun eng zamonaviy baholash ko‘rsatkichlaridan (BLEU, METEOR va TER) foydalanishgan [7]. Ushbu maqolada ma’lumotlarning siyrakligi statistik mashina tarjimasini yomonlashtiradigan omillardan biri ekani aytib o‘tilgan, ularning tegishli ishlar shuni ko’rsatdiki, morfo-sintaktik ma’lumotlardan foydalanish ma’lumotlarning siyrakligini samarali hal qiladi, biroq, inglizcha morfo-sintaktik tahlildan foydalangan holda xitoy-ingliz tilidagi SMT uchun kamroq harakatlar qilingan, tajribalarda ingliz lemmalaridan foydalanish so‘zlarni siyrakligi sifatini sezilarli darajada yaxshilashi mumkinligini aniqlashgan, bu esa tarjima samaradorligini oshirishga olib keladi [8].

Sintaksis. Gaplarning qurilishi hamda ifoda xususiyatlarini o‘rganuvchi tilshunoslik bo‘limi sintaksis deyiladi. Sintaksis so‘zlarning bir-biriga bog‘lanishi hamda gap tuzilishi qonun-qoidalarini o‘rganishiga ko‘ra ikkiga bo‘linadi: *so‘z birikmasi sintaksisi* va *gap sintaksisi*. So‘z birikmasi mustaqil so‘zlarning bog‘lanishidan hosil bo‘ladi, masalan: *do‘stimning kitobi, maktab bog‘i*. Ohang va fikr tugalligiga ega bo‘lib, kesimlik shartlri orqali ifodalanuvchi so‘z va so‘zlar qo‘shilmasiga gap deyiladi, masalan: *Bugun xavo ochiq. Yurtimiz qanday chiroyli!* Ilmiy va rasmiy

uslublarda gapda so‘zlarning odatdagi tartibi, asosan, qat’iy saqlanadi. Lekin so‘zlashuv, badiiy va publistik uslublarda bu tartib o‘zgarishi, ya’ni inversion tartib qo‘llanishi mumkin.

Gap bo‘laklari. Gap tarkibida ma’lum so‘roqqa javob bo‘luvchi mustaqil so‘z yoki so‘zlar birikmasi gap bo‘laklari hisoblanadi. Gap bo‘laklari 1) *bosh bo‘laklar* va 2) *ikkinchi darajali gap bo‘laklaridan* iborat bo‘ladi. Bundan tashqari 3) *uyushuq bo‘laklar*, 4) *ajratilgan bo‘laklar* va gap bo‘laklari bilan aloqaga kirishmaydigan bo‘laklar: 5) *undalma* va 6) *kiritmalarga* ham bo‘linadi.

1) Bosh bo‘laklar

- Ega *(kim? nima? qayer?)* - *Bulbul sayradi.*
- Kesim *(nima qildi? nima bo‘ldi?)* - *Qushlar sayraydi (fe’l). Osmonimiz musaffo (ot).*

2) Ikkinchi darajali bo‘laklar

- Aniqlovchi *(qanday? qancha?)* - *Aytar (sifatlovchi) so‘zni ayt. Mehnatning (qaratqich) ziynati ko‘p.*
- To‘ldiruvchi *(kimga?, nimaga?)* - *Inson bilimga (vositali) intiladi. Aqlni (vositasiz) beaqldan o‘rgan.*
- Hol *(qachon? nima uchun?)* - *Aqlni o‘stirmoq uchun hadeb o‘qiyvermasdan ko‘proq fikrlash kerak.*

3) Uyushiq bo‘laklar. Gapda bir xil sintaktik vazifa bajarib, bir xil so‘roqqa javob bo‘luvchi, o‘zaro teng aloqada, boshqa bo‘laklar bilan esa tobe aloqada bo‘lgan bo‘laklar uyushiq bo‘laklar deyiladi. Misol: *Yo‘lda ba’zan katta-katta maydonni qoplab olgan oq, pushti, sariq va qizil gular uchraydi.*

4) Ajratilgan bo‘laklar. Gapning ajratilgan bo‘laklari o‘zaro aloqada bo‘lgan bo‘laklarning ma’nosini izohlab, bo‘rttirib, aniqlashtirib yoki ta’kidlab keladi. Ular boshqa bo‘laklardan maxsus to‘xtam bilan ajralib, gap urg‘usini (mantiqiy urg‘u) oladi. Misol: *Sizga, o‘rta bo‘yli qizga, atlas juda yarashibdi.*

Gap bo‘laklari bilan aloqaga kirishmaydigan bo‘laklar.

5) Undalmalar. Undalma - deb so‘zlovchining nutqi (fikri) qaratilgan shaxs (obyekt) yoki narsalarni bildiruvchi so‘z va so‘zlar birikmasiga aytiladi. Misol: *Bahor, ketma bizning bog‘lardan.*

6) Kiritmalar (Kirish so‘zlar). Kiritmalar so‘zlovchining o‘zi bayon etayotgan fikriga munosabati (*ishonchi, gumoni, tasdig‘i, inkori va h.k.*), fikrning birovga nisbatlanishi (*mansubligi, aloqadorligi*), o‘z fikri tarkibiy qismlarining ahamiyatliligi darajasi (*birinchidan, ikkinchidan va h.k.*), fikri va uning tarkibiy qismlariga doir bayon etadigan qo‘shimcha axboroti yoki izohi kabi rang-barang ma’nolarni ifodalaydi. Misol: *Xayriyat, vaqtida yetib keldik.*

Gapning ifoda maqsadiga ko‘ra turlari. Har bir gapda ma’lum bir maqsad, niyat yoki his-hayajon ifodalanadi. So‘zlovchi biror narsa, voqea-hodisa haqida xabar qiladi, yoki o‘ziga noma’lum bo‘lgan narsa va voqea-hodisalar to‘g‘risida ma’lumot olishni istaydi, tinglovchiga biror ishni bajarish bo‘yicha buyruq-xitobini bildiradi. Gaplar ana shu maqsadlarga ko‘ra turlicha quriladi, o‘ziga xos grammatik xususiyatlarga ega bo‘lib, alohida ohang bilan talaffuz etiladi. Ifoda maqsadiga ko‘ra gaplar: 1) darak gaplar, 2) so‘roq gaplar, 3) buyruq gaplar, 4) istak gaplarga bo‘linadi.

1) Darak gap. Darak, xabar ma’nosini bildirib, kesimi aniqlik maylidagi fe’l orqali ifodalanadi. Misol: *Filiologiya fakulteti shahar markazidan xiyla chetda joylshgan.*

2) So‘roq gap. Narsalar, voqea-hodisalar haqida so‘rash ma’nosini ifodalaydigan gaplar. Misollar: *1. Bugungi darsga kim kelmadi?(So‘roq olmoshlari yordamida: kim? nima? qayer?*

qachon? ...). 2. Kyingi darsga kelasanmi? (So'roq yuklamalari yordamida: -mi, -chi, -a, -ya). 3. Soat ikki bo'ldi. Ikki bo'ldi?

3) Buyruq gap. Davat qilish ma'nosini bildirib, kesimi buyruq shaklidagi fe'l orqali ifodalanadi. Misol: *Avval o'yla, keyin so'yla.*

4) Istak gap. Istak ma'nosini bildirib, kesimi shart (*-sa*) shaklidagi fe'l orqali ifodalanadi. Misol: *Millatning dardiga darmon bo'lsangiz.*

Gapning tuzilishiga ko'ra turlari. Grammatik asoslarning miqdoriga ko'ra gaplar ikkiga bo'linadi: 1) soda gaplar va 2) qo'shma gaplar.

1) Sodda gap (kesim=1). Grammatik asosi bitta bo'lib, ma'lum bir fikr ifodalovchi gaplar sodda gaplar deyiladi.

2) Qo'shma gap (kesim≥2). Ikki va undan ortiq sodda gaplarning birikuvidan hosil bo'lgan gap qo'shma gap hisoblanadi.

Foydalanilgan adabiyotlar:

1. M. Sharipov and O. Sobirov, 'Development of a Rule-Based Lemmatization Algorithm Through Finite State Machine for Uzbek Language', in *CEUR Workshop Proceedings*, V. J. and K. B., Eds., CEUR-WS, 2022, pp. 154 – 159. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85146112590&partnerID=40&md5=e1080c39d101c0e351cfed1a8228d391
2. U. Salaev, E. Kuriyozov, and C. Gómez-Rodríguez, 'A Machine Transliteration Tool Between Uzbek Alphabets', *CEUR Workshop Proc*, vol. 3315, pp. 42 – 50, 2022, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85146119140&partnerID=40&md5=be670d829670d883b2f8326559ce954a
3. N. Abdurakhmonova and U. Tuliyev, 'Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language', *Literature, Education*, vol. 3, p. 68, 2018.
4. E. B. Boltayevich, E. Adali, K. S. Mirdjonovna, A. O. Xolmo'minovna, X. Z. Yuldashevna, and X. Nizomaddin Uktamboy O'G'li, 'The Problem of Pos Tagging and Stemming for Agglutinative Languages (Turkish, Uyghur, Uzbek Languages)', in *UBMK 2023 - Proceedings: 8th International Conference on Computer Science and Engineering*, 2023, pp. 57 – 62. http://doi.org/10.1109/UBMK59864.2023.10286792 .
5. A. M. Abdurashetona and I. O. Ismailovich, 'Methods of Tagging Part of Speech of Uzbek Language', in *Proceedings - 6th International Conference on Computer Science and Engineering, UBMK 2021*, 2021, pp. 82 – 85. https://10.1109/UBMK52708.2021.9558900
6. Y.-L. Yeong, T.-P. Tan, and S. K. Mohammad, 'Using Dictionary and Lemmatizer to Improve Low Resource English-Malay Statistical Machine Translation System', *Procedia Comput Sci*, vol. 81, pp. 243–249, 2016, doi: https://doi.org/10.1016/j.procs.2016.04.056
7. R. Zhang, H. Yamamoto, and E. Sumita, 'On the Use of Lemmatization for Statistical Machine Translation'.
8. R. Zhang and E. Sumita, 'Boosting statistical machine translation by lemmatization and linear interpolation', in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, 2007, pp. 181–184.

MASHINA TARJIMASIDA SINTAKTIK QOLIPLARNING QIYOSIY TAHLILI (QORAQALPOQ TILI MISOLIDA)

Xudoyberganova Munisa
O'zbekiston Milliy universiteti magistranti.
E-mail: munisaxudayberganova13@gmail.com

Annotatsiya: Mazkur maqolada qoraqalpoq tilini mashina orqali oson o'rganish, modellashtirish, bevosita boshqa tillar bilan solishtirish imkoniyatlari, o'zbek tillaridagi sintaktik qoliplar haqida so'z boradi. Qoraqalpoq va o'zbek tili so'z birikmalari taqqoslanadi.
Kalit so'zlar: Modellashtirish, sintaktik qolip, so'z birikmasi, so'z birikmasi qolipi.

Abstract: This article talks about the possibilities of easy machine learning of the Karakalpak language, modeling, direct comparison with other languages, syntactic patterns in Uzbek languages. Karakalpak and Uzbek word combinations are compared.
Keywords: Modeling, syntactic pattern, word formation, word formation pattern.

Аннотация: В данной статье говорится о возможностях легкого машинного обучения каракалпакского языка, моделировании, прямом сравнении с другими языками, синтаксических закономерностях в узбекских языках. Сравниваются каракалпакские и узбекские словосочетания.
Ключевые слова: Моделирование, синтаксическая закономерность, словообразование, схема словообразования.

XXI boshiga kelib axborot texnologiyalari rivojlanishi tufayli globallashuv jarayoni yangi bosqichga ko'tarildi. Texnika taraqqiyoti mahsuli bo'lgan kompyuter tizimi deyarli barcha sohalarda qulayliklar yaratishga ulgurdi, shu jumladan, axborotning tezkor yetkazib berilishi, matnni avtomatik tarjima qilish, tahrir jarayonlarining qisqa muddatda amalga oshirilishi, turli millat vakillari o'rtasida muloqotni amalga oshirishdagi vosita sifatida sun'iy til shakllanishini ta'minlab bermoqda. Tabiiy tilni qayta ishlash (NLP), mashinali ta'lim (Machine Learning) hamda sun'iy intellekt (Artifificial Intelligence) kabi yangi sohalar kesimida o'zbek tili kompyuter lingvistikasi ham rivojlanib bormoqda.

Ma'lumki, muayyan so'z yoki matnlarning bir tildan boshqa tilga kompyuter yordamida qilingan tarjimasi avtomatik yoki mashina tarjimasini shakllantiradi. Mashina tarjimasi (avtomatik tarjima) kompyuter lingvistikasining asosiy yo'nalishlaridan biri bo'lib, avtomatik tarjima tizimi lingvistik ta'minot yaratish, mavjud tizimlar va lug'at yaratish tarixi, til haqidagi bilim va ko'nikmalarni shakllantirish, mashina tarjimasi tizimining lingvistik protsessorini shakllantirishda lingvistik tahlilni algoritmlash tizimini amalga oshiradi[Abdurakhmonova, N.2020,2022,2023]. Adabiy nutqning rivojlanishi uchun uning grammatik xususiyatlarini o'rganish nazariy hamda amaliy jihatdan muhim ahamiyatga ega bo'lib, uning keng imkoniyatlarga ega til sifatida bir qancha yo'nalish va usullar orqali o'rganilmoqda. Bugungi kunda o'zbek tilida ma'lum sintaktik qoliplar yaratilgan, qoraqalpoq tilida esa bunday harakatlar endi boshlanmoqda. Ushbu sintaktik qoliplar orqali bir tilni ikkinchi tilga qiyoslab o'rganish muhim ahamiyat kasb etadi.

Har bir inson o‘z ona tili grammatikasini bolalik paytidanoq o‘zlashtirib boradi, intuitiv ravishda his qiladi va undan g‘ayrishuuriy tarzda foydalanadi. Ushbu jihatlar esa har qanday tilning grammatikasini formallashtirish masalasi juda murakkab ekanligini tasdiqlaydi. Mashhur tilshunos olim K.Fossler grammatikani ideal darajada tavsiflovchi qoidalar sistemasining mavjud emasligi haqida to‘xtalib, quyidagi fikrlarni yozadi: “Grammatikani fonetika, morfologiya va sintaksisga bo‘lish hech kimga sir emas. Tilni uning shakllanish jarayoni ichida emas, balki uning o‘z holati ichida o‘rganish zarur. Uning ustida anotomik operatsiya o‘tkaziladi. Jonli nutq gaplarga, gap bo‘laklariga, so‘zlarga, bo‘g‘inlarga va tovushlarga ajratilib chiqiladi. Ushbu metod to‘liq o‘zini oqlaydi va ahamiyatga molik natijalarga olib borishi mumkin, biroq shu bilan birga xatolarning manbai hisoblanishi ham mumkin. Xatolar ajratilgan bo‘linishning nutq organizmi uchun asos bo‘lishini da’vo qilgan paytdan boshlanadi, aslida bu bo‘linish insonning ixtiyoriga bog‘liq ravishda mexanik va majburiy tarzda parchalash edi. Mazkur holat anatomiyadagi holatni eslatadi: anatom o‘zining yorish, kesish ishlarini albatta o‘zboshimchalik bilan asossiz amalga oshirmaydi, o‘ziga birmuncha qulaylik tug‘diruvchi jarayonlarni tanlab oladi[Abdurakhmonova, N.2020,2022,2023]. Xuddi shuningdek, grammatist ham nutqni tovush, bo‘g‘in, asos, qo‘shimcha kabilarga ajratishda tabiiy bo‘lmagan, lekin o‘ziga qulay yo‘ldan boradi”. Shu vaqtgacha olimlar tomonidan ishlab chiqilgan barcha grammatik modellar o‘ta shartli va nisbiydir. Murakkab tizim hisoblangan tilning barcha qirralarini formallashtirishda muayyan cheklovlar mavjud bo‘lib, bu xususda nemis olimi 1931-yilda K.Gedelning fikricha, yetarli darajada nazariya formallashtirilsa, bu nazariya formal tizimda to‘la aks ettirish mumkin emas. Bunda aniqlanmagan, formallashtirilmagan qoldiq qoladi”.

Ma’lumki, model bevosita kuzatish imkoni bo‘lmagan hodisani o‘rganish uchun yaratilgan sun’iy mexanizmdir. “Model (qolip) (fran. modele<ital. modello- namuna, qolip)- biron til hodisasi tarkibi yoki qismlarining joylashish holatini ko‘rsatuvi sxema yoki namuna”[Abduraxmonova.N.Z., 2012.]. Lingvistik model termini E.Sepir, L.Blumfild, R.Yakobson, X.Chomiskiy, Z.Xarris, Ch.Xokket kabi tilshunoslari tomonidan fanga kiritilgan. Uning rivojlanishi XX asrning 60-70-yillariga to‘g‘ri keladi.

Qoliplashtirish, ya’ni har bir fan obyektini soddalashtiruvchi metoddir. Lingvistik birliklarni qoliplashtirish bu so‘z birikamsi, gaplarning barqaror munosabatiga asoslanadi. Tilshunoslik sohasida qoliplashtirish strukturalizm yo‘nalishi ta’sirida faol tadbiq qilina boshlandi. XX asrning 50-yillarida amerikalik tilshunos Charlz Friz tomonidan gap tuzilishini qoliplashtirish g‘oyasi olg‘a surildi. Olimning 1952-yildagi “The Structure of English” nomli tadqiqot ishida, o‘z konsepsiyasini distributiv model deb nomlangan. Unga ko‘ra, gap muayyan so‘z turkumlariga oid bo‘lgan so‘zlar zanjiri hisoblanadi va tahlilda morfologiya bazasiga tayanadi. Masalan, “The young man painted the door yesterday” jumlasi distributiv model asosida quyidagicha tahlil qilinadi: D 3 I^a 2-d D I^b 4. Bu yerda D- otning aniqlovchisi (inglizcha determiner), 3-sifat, I^a-birlikdagi ot, 2-d o‘tgan zamon shaklidagi fe’l, 4 -ravishni anglatadi. Mazkur qolipda turli so‘z turkumlariga mansub so‘z shakllarining nutqdagi distributsiyasi gap strukturasini qoliplashtirishning asosiy mezoni sanaladi[Abdurakhmonova, N.2020,2022,2023].

Prof. Nilufar Abduraxmonovaning fikricha, tilni matn orqali kuzatish mumkin, lekin insonning nutqiy faoliyat jarayonida yuzaga chiquvchi tabiiy til hodisalarini to‘liq tushunish uchun yetarlimasligi adabiyotlarda ko‘p bora ta’kidlangan. Shuning uchun tilshunoslikda obyektni tushunishning asosiy vositalaridan biri sifatida modellardan foydalaniladi.

Sintaktik qolip so‘z birikmasi, gapning qolipi bo‘lib, bir qolipning imkoniyatidan foydalanib, voqelikka nutq vaziyatidan kelib chiqib, bir necha so‘z birikmasi yoxud gap hosil qilish mumkin, bunday lisoniy sintaktik qoliplar bir tilni o‘rganish, o‘zlashtirish boshqa tillarga uning grammatikasini solishtirish uchun tilning ixchamlashtirilgan, muqobil vositasidir. So‘z birikmalari qolipi cheklangan, 18 ta, ammo yuqorida ta’kidlanganidek, ularning har biridan cheksiz so‘z birikmasi hosil qilinadi. Ammo bunday so‘z birikmasi, gapning sintaktik qoliplar yaratish harakati qoraqalpoq tili grammatikasida endi boshlanmoqda.

Ma’lumki, o‘zbek tilida so‘z birikmalari asosan uch usulda birikadi, bular boshqaruv, moslashuv va birikuv. Qoraqalpoq tilida esa bunday birikish usullari boshqaruv (basqarıw), moslashuv (izafet), bitishuv (jupкerlesiw) va kelishuv (кelisiw) bo‘lib to‘rtga bo‘linadi.

1) Boshqaruv (basqarıw) usulida tobe so‘z ma’lum grammatik vosita kelishik formalari yoki ko‘makchilar orqali boshqa so‘zga bog‘lanadi. Masalan, mallardı suwg’arıw (mollarni sug‘ormoq), atızda islew (dalada ishlamoq), кo’pir arqalı o’tiw (ko‘prik orqali o‘tmoq).
2) Moslashuv (izafet) ikki otning qaratqich kelishigi yordamida birikuvidan hosil bo‘lgan. Masalan, ten’izdin’ hawası (dengizning havosi), boyannın’ tamırı (bo’yanning tomiri), tereкtin’ japırag’ı (daraxtning yaprog’i), senin’ da’pterin’(sening daftaring), onın’ кitabı (uning kitobi) va boshqalar.
3) Bitishuv (jupкerlesiw) so‘zlar grammatik vositalar yordamisiz o‘zaro aloqaga kirib, so‘z birikmasi hosil qiladi. Bunday so‘z birikmalari mazmun va intonatsiya yordami bilan o‘zaro birikadi. Masalan,tınıq suw (tiniq suv), ashıq hawa (ochiq havo), salqın samal (salqin shamol), bes da’pter (beshta daftar), tez ju’riw (tez yurmoq), jaqsı oqıydı (yaxshi o‘qiydi), tez-tez soylew (tez-tez gapirmoq) va h.

Qoraqalpoq tilida so‘z birikmalari birikishining yana bir usuli bo‘lib, kelishuv (кelisiw) deb ataladi, asosan, ega va kesim munosabatidagi so‘zlar orasidagi bog‘lanishni ifodalaydi.

Masalan, men meкtepte isleymen (men maktabda ishlayman), sen qurılısta injenersen’(sen qurilishda muhandissan), ol awılda isleydi (u qishloqda ishlaydi). Bunday bog’lanish turi o‘zbek tilida yo‘qligi uchun bunday munosabat gap sintaksisi tarkibiga kiradi. Qoraqalpoq tili agglyutinativ tillar oilasiga mansubligi uchun o‘zbek tilida yaratilgan so‘z birikmasi qoliplari bilan qiyoslab o‘rganildi va ushbu qoliplarga mos tushganini quyidagi misollar oraqali ko‘rishimiz mumkin:

[ot + tushum kelishigi +fe’l] – gulni uzmoq, daraxtni ekmoq, onani e’zozlamoq, she’rni yodlamoq.

[atliq+ tabıs sepligi+feyil] – gu’ldi uziw, terekti egiw, anani qa’dirlew, qosiqti yadlaw.

Ushbu qolip o‘zbek tilidagi so‘z birikmalari uchun umumiy bo‘lsa-da, qoraqalpoq tilidagi so‘z birikmalariga birdek tushadi. Xuddi shunday har bir qolipdan hosil bo‘lgan so’z birikmasining hosilasi qoraqalpoq tili uchun quyidagicha keltirsak bo‘ladi:

[ism + qaratqich kelishigi +ism egalik qo’shimchasi] – daftarning varag‘i (da’pterdin’ beti), daraxtning bargi (terektin’ ja’pirag’i);

[ot + jo‘nalish kelishi + fe’l] – maktabga bormoq (mektepke bariw), uyga qaytmoq (u’yge qaytiw), daftarga yozmoq (da’pterge ja’ziw);

[ot + chiqish kelishigi +fe’l] – shahardan chiqmoq (qaladan ketiw), ishdan bo‘shamoq (ju’misdan bosaw);

Yuqorida keltirilib o‘tilgan so‘z birikmalarining ayrimlari o‘zbek, qoraqalpoq tillarida turli xil sintaktik model ko‘rinishiga ega va quyidagicha:

Kitobni o‘qimoq – kitapti oqiw - [to V + the N]
She’rni yodlamoq – qosiqti yadlaw - [to V + the N]
Daraxt ekmoq – terekti egiw- [to V + the N] ushbu so‘z birikmasi qolipida;
Maktabga bormoq –mektepke bariw- [to V + to N]
Uyga qaytmoq- uyge qaytiw- [to V + to N]
Qishloqqa jo‘namoq – awilg’a bariw- [to V + to N] ushbu so‘z birikmalari qolipida,
Uzukning ko‘zi – ju’zu’ktin’ kozi - [N’s + N]
Daraxtning bargi – terektin’ ja’pirag’i - [N’s + N] yuqoridagi so‘z birliklari qolipidan farqli ravishda hosilaning cheksizligi so‘z birikmasi – [N+N] so‘z birikmasi qolipida beriladi.
Shahardan chiqmoq –qaladan shig’iw- [to V + N]
Yoddan aytmoq – yatdan aytiw- [to V + N]
Ishdan bo‘shamoq – ju’mistan bosaw – V

Demak, o‘zbek tilidagi bitta so‘z birikmasi qolipida tuziladigan bir nechta so‘z birikmalari ham qoraqalpoq tili ham ingliz tilida turli modellarga mos tushmoqda.

Xulosa o‘rnida shuni aytish mumkinki, mashina tarjimasini amalga oshirish modellar bilan bog‘liq bo‘lib, ular matnning avtomatik tarjimasini amalga oshiradilar, ammo bu jarayonni inson tarjimon uni tahrirlashdan keyin amalga oshiradi. Mashina tarjimasining yana bir afzalligi shundaki, u juda tez ishlaydi, bir necha soniya ichida millionlab so‘zlarni bir tildan ikkinchi tilga tarjima qila oladi. Qoraqalpoqcha matnlarni o‘zbek va boshqa tillarga avtomatik tarjima qilishda, avvalo, qoraqalpoq tilining xususiyatlaridan kelib chiqqan holda uning modellarini aniqlashtirish, qoraqalpoq va o‘zbek tillarining o‘zaro tahlilida ularning o‘xshash va farqli tomonlarini ishlab chiqib, tarjimon dasturi uchun yaratiladigan modellarga aniqlik kiritishga yordam beradi.

Foydalanilgan adabiyotlar ro‘yxati:

1. Abduraxmonova.N.Z. Mashina tarjimasining lingvistik asoslari. Toshkent: –2012.
2. Bekbergenov.K. Qaraqalpaq tilinin’ izzertlewshileri. Qaraqalpaqstan: –2000.
3. Абдурахмонова, Н., & Бойсариева, С. (2023). TABIIY TILNI QAYTA ISHLASHDA (NLP) OKKAZIONALIZMLARNING MORFEM TAHLILI. *МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА*, *6*(3).
4. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In *Компьютерная обработка тюркских языков. TURKLANG 2022* (pp. 231-240).
5. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
6. Khusainov, A., Suleymanov, D., Gilmullin, R., Minsafina, A., Kubedinova, L., & Abdurakhmonova, N. (2020, November). First Results of the TurkLang-7 Project: Creating Russian-Turkic Parallel Corpora and MT Systems. In *CMCL* (pp. 90-101).
7. Da’wletov.A. Ha’zirgi qaraqalpaq a’debiy tili. Sintaksis. Nukus: – 2009.
8. Rahmatulloyeva.M. O‘zbek tilida sintaktik qoliplar. Academic Research in Academic Sciences, 2023.
9. Sayfullayeva.R, Kurbonova.M. Hozirgi o‘zbek adabiy tili. T.: –2009.

THE ROLE OF LINGUISTICS IN MACHINE TRANSLATION SOFTWARE DEVELOPMENT

Otemisov Aziz Zarlikbaevich
head of the Department of
Practical Philology of Karakalpak State University,
doctor of philosophy (PhD), associate professor.
E-mail: utemisov.aziz@mail.ru

Abdullaeva Ayzada
KSU Linguistics:
Karakalpak language 2nd year master's student.

Abstract: This paper explores the pivotal role of linguistics in the development of machine translation (MT) software. It emphasizes how linguistic principles and theories inform the design, implementation, and improvement of MT systems. By drawing on insights from various linguistic subfields such as syntax, semantics, morphology, and pragmatics, developers can address the inherent complexities of natural language and enhance the accuracy and fluency of translation outputs. The paper highlights the interdisciplinary nature of MT development and underscores the importance of integrating linguistic expertise into the software engineering process.

Keywords: Machine Translation**,** linguistics, software development, computational linguistics, syntax, semantics, morphology, pragmatics

Аннотация: В данной статье исследуется ключевая роль лингвистики в разработке программного обеспечения для машинного перевода (МП). В нем подчеркивается, как лингвистические принципы и теории влияют на проектирование, внедрение и улучшение систем МП. Опираясь на знания из различных лингвистических областей, таких как синтаксис, семантика, морфология и прагматика, разработчики могут решать проблемы, присущие естественному языку, и повышать точность и беглость результатов перевода. В документе подчеркивается междисциплинарный характер разработки МП и подчеркивается важность интеграции лингвистических знаний в процесс разработки программного обеспечения.

Ключевые слова: машинный перевод, лингвистика, разработка программного обеспечения, компьютерная лингвистика, синтаксис, семантика, морфология, прагматика.

Introduction

Machine translation has witnessed significant advancements in recent years, driven by advances in computational linguistics, machine learning, and artificial intelligence. While technological innovations play a crucial role in improving MT systems, linguistic insights remain indispensable for achieving high-quality translations. This paper aims to elucidate the multifaceted role of linguistics in MT software development, shedding light on how linguistic theories and methodologies contribute to the optimization of translation algorithms and models.

Linguistics, as the scientific study of language, plays a pivotal role in the development of machine translation (MT) systems. MT aims to automate the process of translating text or speech

from one language to another, a task that involves understanding the intricate structures and nuances of human language. In this introduction, we will explore the intersection of linguistics and MT, highlighting how linguistic principles inform the design, implementation, and improvement of MT systems. [Abdurahmanova N.Z, 2011, 74]

Linguistics provides the theoretical foundation for understanding language structure, meaning, and communication patterns, which are essential for effective translation [Abdurakhmonova, N.2021,2022,2023]. By leveraging insights from various linguistic subfields such as syntax, semantics, morphology, and pragmatics, developers can address the complexities of natural language and enhance the accuracy and fluency of translation outputs.

In the following sections, we will delve into specific aspects of linguistics and their role in MT software development, including syntactic analysis, semantic representation, morphological processing, and pragmatic considerations [Abdurakhmonova, N.2021,2022,2023]. Through interdisciplinary collaboration between linguists, computer scientists, and engineers, we can harness the power of linguistic expertise to create more robust and reliable MT systems.

Ultimately, this exploration aims to underscore the indispensable role of linguistics in advancing MT technology and facilitating cross-linguistic communication and understanding in an increasingly interconnected world.

The Role of Linguistics in MT Software Development:

1. Syntax and Sentence Structure: Linguistic knowledge of syntax informs MT systems' ability to analyze and generate grammatically correct sentences in the target language. Syntactic parsing techniques and grammar rules play a crucial role in ensuring syntactic coherence and fidelity in translation outputs.

2. Semantics and Meaning Representation: Semantic analysis enables MT systems to capture the meaning of source texts accurately and convey it effectively in the target language. Linguistic theories of semantics, including lexical semantics and compositional semantics, guide the development of algorithms for semantic representation and alignment in MT.

3. Morphology and Word Formation: Morphological analysis is essential for handling inflectional and derivational morphology in different languages. Linguistic knowledge of morphological structures and processes facilitates the generation of morphologically accurate translations and improves lexical selection and alignment.

4. Pragmatics and Contextual Understanding: Pragmatic considerations are crucial for producing translations that are contextually appropriate and communicatively effective. Linguistic insights into discourse structure, speech acts, and implicature inform MT systems' ability to capture and convey pragmatic nuances and sociocultural conventions in translation. [Mukhamedova S., 2007, 89]

In the field of linguistics and machine translation, analysis refers to the process of examining linguistic structures, patterns, and features to inform the development and improvement of machine translation systems [Abdurakhmonova, N.2021,2022,2023]. This analysis involves various levels of linguistic analysis, including:

1. Syntactic Analysis: This involves parsing the structure of sentences and phrases in source and target languages to ensure grammatical correctness and syntactic coherence in translation outputs. Syntactic analysis helps identify sentence boundaries, word order, and grammatical relationships between words, which are crucial for generating accurate translations.

2. Semantic Analysis: Semantic analysis focuses on capturing the meaning of source text and representing it in a way that preserves its intended semantics in the target language. This involves identifying semantic roles, relationships, and entailments encoded in the source text and ensuring they are accurately conveyed in the translation.

3. Morphological Analysis: Morphological analysis deals with the study of word structure and formation, including inflectional and derivational morphology. In machine translation, morphological analysis is essential for handling linguistic variations such as verb conjugation, noun declension, and word derivation across different languages.

4. Pragmatic Analysis: Pragmatic analysis considers contextual factors, communicative intentions, and sociocultural conventions that influence language use and interpretation. This analysis helps ensure that translations are contextually appropriate, culturally sensitive, and effectively convey the intended message in the target language.

5. Error Analysis: Error analysis involves identifying and analyzing errors or inaccuracies in translation outputs generated by machine translation systems. This helps developers understand the limitations and challenges of current MT technologies and guide efforts to improve translation quality through error correction and optimization. [Piotrovsky R.G. 1988, 35]

Overall, linguistic analysis plays a crucial role in shaping the design, implementation, and evaluation of machine translation systems, providing valuable insights into the linguistic structures and phenomena that underlie human language use and communication. By conducting thorough linguistic analysis, developers can enhance the accuracy, fluency, and usability of machine translation software, ultimately facilitating cross-linguistic communication and understanding in diverse linguistic contexts. In conclusion, the integration of linguistic expertise is indispensable for the development of robust and reliable machine translation software. By leveraging insights from various linguistic subfields, developers can address the linguistic complexities inherent in natural language and enhance the accuracy, fluency, and usability of MT systems. Moving forward, interdisciplinary collaboration between linguists, computer scientists, and software engineers will be key to advancing the state-of-the-art in machine translation and realizing its full potential in facilitating cross-linguistic communication and understanding.

Literatures:

1. Abdurahmonova N. Z. Problems of processing texts on a computer: Graduation qualification work. - Tashkent: UzMU 2009;
2. Abdurahmanova N.Z. Translation of a simple English sentence into Uzbek English support of the program. Master's thesis. -Tashkent- UzMU, 2011
3. Mukhamedova S. Computer based on Uzbek action verbs creating linguistic support for programs. - Tashkent, 2006. - 143 B.
4. Mukhamedova S. Computer linguistics (methodical use). - T., 2007;
5. Mukhamedov S.A. Statistical analysis of lexicomorphological structures of Uzbek texts: author's abstract. diss. cand., philol. Sci. – Tashkent, – 1980
6. Абдурахмонова, Н., & Абдувахобов, Г. (2021). О ‘QUV LUG ‘ATINI TUZISHNING NAZARIY METODOLOGIK ASOSLARI. *МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА*, *4*(6).

7. Abdurakhmonova, N., Shakirovich, I. A., & O‘G‘Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. *Science and Education*, *3*(9), 146-156.
8. Ismailov, A. S., Shamsiyeva, G., & Abdurakhmonova, N. (2021). Statistical machine translation proposal for Uzbek to English. *Science and Education*, *2*(12), 212-219.
9. Абдурахмонова, Н., & Бойсариева, С. (2023). TABIIY TILNI QAYTA ISHLASHDA (NLP) OKKAZIONALIZMLARNING MORFEM TAHLILI. *МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА*, *6*(3).
10. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In *Компьютерная обработка тюркских языков. TURKLANG 2022* (pp. 231-240).
11. Piotrovsky R.G. Computerization of language teaching (textbook according to a special course). - L.: 1988.
12. Pulatov A.K., Djuraeva H. Razrabotka formal model grammar Uzbek Language // Uzbek Mathematical Journal. — Tashkent, 2002.
13. Pulatov A. Problems of linguistics in the era of globalization// Current issues of linguistics. T., 2004
14. Pulatov A. Computational linguistics (textbook). Tashkent: Akademnashr, 2011.
15. Rahmatullaev Sh. Phraseological dictionary of the Uzbek language. – Tashkent: Teacher, 2001.

PARALLEL KORPUS MASHINA TARJIMASINING O‘RGANISH OBYEKTI SIFATIDA

Jo‘raqulova Mushtariy Abdumalikovna
O‘zbekiston Milliy universiteti 2-kurs magistranti.
E-mail: mushtariy00712@gmail.com

Annotatsiya: Ushbu maqola parallel korpuslarning turlari uning ahamiyati, va uni tuzishdagi muammolari hamda mashina tarjimasini rivojlantirish masalalariga bag‘ishlangan.

Kalit so’zlar: korpus tilshunosligi, parallel korpus matnlari (PKM), mashina tarjimasi, Ingliz va Norvegiya Parallel Korpusi, CzEng parallel korpusi, Scielo parallel korpusi, FAPESP parallel korpusi.

Abstract: This article is devoted to the types of parallel corpora, their importance, and the problems of their creation, as well as the development of machine translation.

Keywords: corpus linguistics, parallel corpus texts (PKM), machine translation, English and Norwegian Parallel Corpus, CzEng parallel corpus, Scielo parallel corpus, FAPESP parallel corpus.

Аннотация: Данная статья посвящена важности типов параллельных корпусов, их проблемам и совершенствованию машинного перевода.

Ключевые слова: корпусная лингвистика, параллельный корпус текстов (PKM), машинный перевод, параллельный корпус английского и норвежского языков, параллельный корпус CzEng, параллельный корпус Scielo, параллельный корпус FAPESP.

Zamonaviylashib borayotgan jamiyatimizda avtomatik informatsion texnologiyalar va kompyuter muhim ahamiyat kasb etmoqda. Kompyuter hozirgi kunda barcha sohalariga kirib borayotgani hammamizga birday ma’lumdir. Buning natijasida ko‘plab yangi sohalar rivojiga o‘z hissasini qo‘shmoqda.Shu jumladan, kompyuter texnologiyalarining tilshunoslik fani bilan hamkorligini boshlash natijasida kompyuter lingvistikasi fani vujudga keldi.

Korpus lingvistikasining ilk bo‘g‘ini Naom Chomskiy nomi nilan bog‘liq bo‘lib, u generativ tilshunoslikka oid qarashlarni ilgari surgan bo‘lsa ham, uning matnlarni amaliy jihatdan o‘qiy oladigan shaklda to‘plab, uni kompyuter tahlil qiladigan darajaga ko‘targan bosqich korpus lingvistikasi uchun dastlabki qadam bo‘ldi.

Kompyuterda yaratilgan birinchi matnlar korpusi Braun korpusi (БК , inglizcha Brown Corpus, BC) hisoblanadi, u 1961 -yilda Braun universitetida yaratilgan, har biri 2000 so‘zli 500 ta matn fragmentini o‘z ichiga oladi. 1970-yillarda 1 mln so‘zni o‘z ichiga olgan matnlar korpusi asosida rus tilining chastotali lug‘ati yaratildi. 1980-yillarda Shvetsiyaning Upsala universitetida ham rus tilida matnlar korpusi yaratildi[Abdurakhmonova, N.2021,2022,2023].

Parallellik mezoniga ko‘ra korpuslar bir tili, ikki tili va ko‘p tiliga bo‘linadi. Bir tili korpuslarda dialektlar va til variantlariga qarama –qarshi bo‘ladi. Masalan, ingliz tili ona tili va ingliz tili chet tili kabi yangi texnologiyalar paydo bo‘lgunga qadar ilmiy qiziqish doirasidan tashqarida qoldi, bu esa qarama-qarshi tildagi nutq asarlarini ancha ko‘p jalb qilishga imkon berdi.

Ingliz va Norvegiya Parallel Korpusi (ENPC) asl badiiy va badiiy bo'lmagan ingliz va norvegcha matnlar va ularning tarjimalaridan olingan ikki million tokendan iborat. Asl matnlar ularning tarjimalari bilan jumla darajasida moslashtirildi. Ushbu korpus haqiqiy matnlar va ularning tarjimalari, har ikki tildagi asl nusxalar, ikkala tildagi tarjimalar va asl nusxalar hamda bir tildagi tarjimalarni taqqoslashga qaratilgan edi. Keyinchalik ENPCga nemis-norvegiya, fransuz-norvegiya, rus va norvegiya parallel korpuslari qo‘shildi va bu Oslo ko‘p tilli korpusini tashkil qildi[H. Hasselgard, 2010; 98-101].

ENPC dan keyin ingliz tilini o‘z ichiga olgan yana bir qancha parallel korpuslar yaratildi. Ular quyidagilarni tashkil qiladi: ingliz-shved parallel korpusi, ingliz-fransuz korpusi, ingliz-german korpusi va ingliz-ispan korpuslaridir. Bunga misol qilib, JRC-ACQUIS Multilingual Parallel Corpusini ham ko‘rsatib o‘ta olamiz. U 22 tildan bir milliarddan ortiq tokendan iborat[G. R. Yepes,2011;65-80]. Qo‘shimcha qilib,Ochiq Parallel Korpus (OPUS) Yevropa va Osiyo tillarining parallel korpusi, shu jumladan arab tillaridagi 352 millionga yaqin tokenlardan iborat.

Arabcha parallel korpus 1999 yild Job Xobkins universitetida mashina tarjimasini osonlashtirish maqsadida ingliz-arab parallel Misr korpusi ishlab chiqarilgandan so‘ng vujudga keldi. Bu korpus ingliz va arab tilidagi Qur’ondan iborat edi. Qur’on arabcha Corpus [http://corpus.quran.com/] Scielo korpusidir morfologik izohli klassik arab tilini tasvirlash uchun juda yaxshi manba hisoblandi.

Keyinchalik esa Al-Ajmiy tomonidan 2004-yilda ingliz-arab parallel korpusi yaratildi va unda tarix, iqtisod san’at, fan va adabiyotga oid 3 million so‘zdan iborat tarjima kitoblar turlaridan bor edi. Birlashgan Millatlar Tashkilotining Ingliz-Arab Parallel Korpusi (EAPCOUNT) 2013 yilda Hammuda Salhi tomonidan tuzilgan.

Keyingi CzEng korpusi bo‘lib, chex-ingliz parallel korpusi notijorat tadqiqot va ta’lim maqsadlarida erkin foydalanishlari mumkin. Bu korpusning aksariyat qismi kitoblar,film subtitrlari hamda Yevropa Ittifoqi qonunchilik manbalaridan iborat. Bu korpus tokeni har bir til uchun taxminan 200 millionni tashkil qiladi. Bu korpusning oxirgi nashriga ko‘ra morfologik teglar, sirt sintaktik va avtomatik havola xususiyatlari kiritilgan[O. Bojar ., 2016; 231-238].

Yana bir parallel korpus uchun na’munasi Scielo korpusidir. Biotibbiyot sohasi (biologiya fanlari va sog'liqni saqlash fanlari) uchun bepul mavjud bo'lgan parallel ilmiy nashrlar korpusidir. Bu korpus Scielo ma’lumotlar bazasidan olingan. Korpus uchta til juftligi uchun tuzilgan: portugal-ingliz (jami 86 000 ga yaqin hujjat), ispan-ingliz (taxminan 95 000 ta hujjat) va fransuz-ingliz (2000 ga yaqin hujjat). [M.Neves., 2016].

FAPESP korpusi ilmiy yangiliklar matnlari asosida ikki til juftlikda, portugal-ingliz va portugal-ispan tillarida tuzilgan. Ushbu matnlar onlayn ko‘p tilli Braziliya jurnalidan (Pesquisa FAPESP) avtomatik ravishda o‘rganildi. Korpus hujjat va jumla qilib tizimlashtiriladi. Unda 2700 ga yaqin parallel hujjatlar mavjud bo‘lib, ular har bir tilda 150 000 dan ortiq tizimlangan jumlalarni o‘z ichiga oladi.[28]

Bu ikki va ko‘p tilli parallel korpuslar qurilishi jihatdan qiyin bo‘lganligi sababli yuqoridagilargina bepuldir.

Parallel korpuslar mashina tarjimasi (MT) tizimlarini rivojlantirishda muhim rol o'ynaydi. Mashina tarjimasida uchta yondashuv mavjud bo‘lib, ular quyidagilarni tashkil etadi: lingvistik,

statistik va kompyuter yordamida amalga oshiriladigan bilimlardir[Abdurakhmonova, N.2021,2022,2023].

Birinchi yondashuv qoidaga asoslangan bo‘lib, manba va maqsad tillarning morfologik, sintaktik, semantik va idiomatik bilimlari kabi lingvistik bilimlarga bog‘liq. Bunday yondashuvda bosh gap tahlilchi yordamida maqsadli gapga tarjima qilinadi. Ushbu tahlilchi manba jumlani NP, VP, AdvP va PP kabi tarkibiy qismlariga tahlil qiladi va keyin ularni TL ekvivalentlari bilan almashtiradi.

Ikkinchi yondashuv - statistik mashina tarjimasi. U parallel korpusni tahlil qiladi, eng tez-tez mos keladigan SL-TL naqshlarini tanlaydi va ularni tarjimada ishlatadi. Masalan, Google Translate kabi statistik tizimlarda parallel korpus, monolingual korpus va ularning ma'lumotlarining statistik modellari matnlarni bir tildan boshqa tilga avtomatik ravishda ko‘rsatish uchun keng qo‘llaniladi[F.J.Och, 2005; Thailand].

Uchinchi yondashuv - kompyuter yordamida tarjima bo‘lib, u mashina va tarjimon o‘rtasidagi interaktiv jarayonni o‘z ichiga oladi. Tarjima jarayonini osonlashtirish va avtomatlashtirish uchun kompyuter yordamida tarjima dasturlarining ko‘p turlari (masalan, elektron lug‘atlar va tarjima xotiralari) ishlab chiqilgan. Ikki tilli va ko‘p tilli elektron lug'atlar nutqning bir qismi, talaffuz va birikmalar kabi SL va TL so'zlari haqidagi ma'lumotlarni o‘z ichiga oladi. Ushbu lug‘atlarni turli shakllarda topish mumkin: maxsus qurilmalar (masalan, Atlas zamonaviy lug'ati inglizcha-arabcha), kompyuter dasturlari (masalan, Parallel korpusni ishlab chiqish oddiy ish emas. Bu qurilishning ko'p bosqichlarida duch keladigan texnik va lingvistik qiyinchiliklar bilan bog'liq: matnni tanlash, o'zgartirish, segmentatsiya qilish, stamplash, tekislash va izohlash[Abdurakhmonova, N.2021,2022,2023].

Parallel korpusni yaratishda ko‘plab qiyinchililklar mavjud. Ular texnik va lingvistik qiyinchiliklar bilan bog‘liq: matnni tanlash, o'zgartirish, segmentatsiya qilish, shtamplash, tekislash va izohlash kabi bir qancha muammolari mavjud. Bunga sabab qilib esa, tilning avtomatik moslashtirilishiga yetarlicha parallel matnlarga manba yo‘qligi va texnalogik jihatdan kam rivojlanganlik hisoblanadi. Parallel korpusni loyihalash uchun universal retsept yo‘q. Parallel korpus uchun ma’lumotlarni tanlash jarayoni tarjima qilingan ma’lumotlarning mavjudligi, tarjima qilingan materialning sifati, maqsadi tillar va tarjima yo‘nalishlari uchun tarjima qilingan materialning proportsional mavjudligi kabi bir qator cheklovlar bilan tavsiflanadi.

Foydalanilgan adabiyotlar ro‘yxati:

1. N. Abduraxmonova. O‘zbek tili elektron korpusining kompyuter modellari (Monografiya). GlobeEdit, 2021.
2. H. Hasselgard. Contrastive Analysis / Contrastive Linguistics. The Routledge Linguistics Encyclopedia, K. Malmkjær, Ed., Third Edition, London, NY: Routledge, pp. 98-101, 2010.
3. G. R. Yepes. Parallel Corpora in Translator Education. Redit: Revista Electrónica de Didáctica de la Traducción y la Interpretación, pp. 65-80, 2011.
4. H. Al-Ajmi. A New English–Arabic Parallel Text Corpus for Lexicographic Applications. Lexikos, vol. 14, pp. 326-330, 2004
5. Абдурахмонова, Н., & Абдувахобов, Г. (2021). О ‘QUV LUG ‘ATINI TUZISHNING NAZARIY METODOLOGIK ASOSLARI. *МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА*, *4*(6).

6. Abdurakhmonova, N., Shakirovich, I. A., & O‘G‘Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. *Science and Education*, *3*(9), 146-156.
7. Ismailov, A. S., Shamsiyeva, G., & Abdurakhmonova, N. (2021). Statistical machine translation proposal for Uzbek to English. *Science and Education*, *2*(12), 212-219.
8. Абдурахмонова, Н., & Бойсариева, С. (2023). TABIIY TILNI QAYTA ISHLASHDA (NLP) OKKAZIONALIZMLARNING MORFEM TAHLILI. *МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА*, *6*(3).
9. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In *Компьютерная обработка тюркских языков. TURKLANG 2022* (pp. 231-240).
10. O. Bojar, O. Dusek and others. Enlarged czech-english parallel corpus with processing tools dockered, in: Int. Conf. Text Speech Dialogue, Springer, 2016, pp. 231–238.
11. M. Neves, A.J. Yepes, A. Neveol. The scielo corpus: a parallel corpus of scientific publications for biomedicine, Proc. Tenth Int. Conf. Lang. Resour. Eval. LREC 2016 Paris Fr. Eur. Lang. Resour. Assoc., ELRA, 2016.
12. F. J. Och. Statistical Machine Translation: Foundations and Recent Advances, presented at the Tutorial at MT Summit, Phuket, Thailand, 2005.
13. L. Bowker and J. Pearson. Working with Specialized Language: A Practical Guide to Using Corpora, London, NY: Routledge, 2002.
14. Abduraxmonova N. Mashina tarjimasining lingvistik asoslari. GlobeEdit, 2012.

A LINGUISTIC STUDY OF COLLOCATION IN ENGLISH

Yuldasheva Rukhshona
Master's degree student
of National University of Uzbekistan.

Abstract: In the field of ELT, a growing awareness has been paid to the use of correct and appropriate word combinations. Of methods and techniques existed in the literature, the issue of lexical collocation gets a great deal of attention. However, one of the main obstacles, particularly for non-native writers (NNW), is indeterminate knowledge of word combinations. Through the acquisition of collocation, it may be possible for NNW to increase their lexical competence. The present study attempted to investigate the use of English lexical collocations in the texts written by native writers of English (NW) and non-native writers of English (NNW), and to examine whether there are any statistically significant differences between NW and NNW in terms of employing collocations in their written productions.

Keywords: collocation, grammatics, translation, language, structure.

Аннотация: В области ELT все большее внимание уделяется использованию правильных и подходящих словосочетаний. Из существующих в литературе методов и приемов вопросу лексической словосочетания уделяется большое внимание. Однако одним из главных препятствий, особенно для писателей-неносителей языка, является неопределенное знание словосочетаний. Благодаря приобретению словосочетаний NNW может повысить свою лексическую компетентность. В настоящем исследовании предпринята попытка изучить использование английских лексических словосочетаний в текстах, написанных носителями английского языка и авторами, не являющимися носителями английского языка, а также проверить, существуют ли какие-либо статистически значимые различия между NW и NNW с точки зрения использования словосочетаний. в их письменные произведения.

Ключевые слова: словосочетание, грамматика, перевод, язык, структура.

Introduction

Writing proficiency is an essential component of language competency as well as an aspiration of NNW. In addition to the fact that writing proficiency has an important place in every sphere of life, it is a leading ability particularly in scholarly writing. On the other hand, linguistic literature shows that there are some basic requirements in order to have writing fluency, one of which is the knowledge of collocations. There are some definite evidences that establish a link between native-like writing fluency and collocation knowledge of the author. One reason of foreignness in the scientific texts of NNW is shown as the lack of collocation knowledge, which NW could easily and intuitively acquire.

Materials and methods

Collocations are crucial in a variety of ways, one of which is the lexical contribution that it provides to the authors. To be more precise, there exist some studies that create a strong establish a link between lexical diversity and collocation knowledge, which further enables and make an

opportunity for authors to have more academic writings strived for (for example, Hyland). That important link between lexical diversity and collocation has led to many studies to investigate collocations which are used in academic texts by authors consciously or unconsciously [Abdurakhmonova, N.2021,2022,2023]. The common results indicated that collocations have been widely employed from the most basic English courses to the extreme English proficiency levels. There is nearly no way of using a language without referring to collocations because they are intricately interwoven with the language itself. That is why; you absolutely have a high chance of encountering collocations whether you are a reader, speaker, or writer.

Results and discussion

The frequency of occurrence of collocations does not make it an apparent issue for the language producers. In contrast, it is a convoluted issue requiring a huge amount of attention. The very first reason that makes collocations so indispensible for academic scholars lies behind the absolute power of collocation that allows a potential native-like written production. Because native speakers of English use "ready-made chunks" [Robins, 1967, p. 21], it is sure that a large many of non-native writers aspire to have native-like written productions, which would be impossible or at least too arduous with insufficient collocational knowledge in the process of composing a text. In order for that, not to create a scientific paper crammed with odd expressions, the author must be aware of collocations that native writers of English use in their texts. But for that, there may be no point of mentioning an academic paper which is free of inappropriate lexical bundles. The authentic detrimental effect of that would be language inappropriateness, stylistic infelicities, and foreignness to native readers. That highly possible case for non-native writers is the key problem as Fox reported. Further, Fox indicated that the fundamental trouble regarding an academic writing with full of odd expressions is of collocations but nothing else. What is more, it is a rigour issue for even very proficient non-native writers.

The literature provides insight about that errors concerning inappropriate use of collocations, among all error types, have been experienced as the most commonly by non-native speakers. Undoubtedly, these troublesome errors disrupt the transmission of knowledge to the reader, which is a situation that mars the persuasion or convincing power of the author [Abdurakhmonova, N.2021,2022,2023]. On account of that key fact, they must be studied if writing fluency is demanded. Nation, pertaining to low proficiency learners, stated that they have a tendency to "encode words in memory on the basis of sound and spelling rather than by association meaning" (p.3). The situation is not different for non- native writers. A paucity of collocational proficiency is associated with lexical proficiency of the writer. Similarly, collocations tie in with lexical development, says Ellis. For this, the reason why non-native writers are not able to use collocations as proficient as native writers may be because non-native writers fail to correlate words on the basis of true and specific word partnership [Sung, 2003]. Despite the obvious significance of collocations for scholarly writing and covetable need from the aspects of non-native scholars of English, it was not studied in a way that would show the differences and similarities between native and non-native academics, hence would provide valuable suggestions on how to use collocations in order for more native-like written products[Abdurakhmonova, N.2021,2022,2023].

Collocations are a single title but instead they are divided into four different subcategories which are as follows:

1) Lexical collocations: As indicated by Bahns lexical collocations do not contain clauses, infinitives, or prepositions; instead, various combinations of nouns, verbs, adverbs, and adjectives. Again, if you compare lexical collocations to closed class structure of grammatical collocations, it can be easily seen that lexical collocations include no subordinate element, and are composed of two equal open-class lexical items [Barhudarov, 2015, 2].
2) Grammatical collocations: Grammatical collocations, different from lexical ones, include a verb, a noun, and an adjective, plus a preposition, an infinitive, or a clause. The patterns of a phrasal grammatical collocations form from a lexical unit and a pattern that specifies the sub-categorization property of the head (Bentivogli & Pianta).
3) Solid lexical collocations: When they are compared to lexical collocations, the occurrence possibility of two words is much more rigid in solid lexical collocations while constructing a lexical word combination. In here, the lexical items have a strong bound to one another[Abdurakhmonova, N.2021,2022,2023]. There is such a powerful interconnection between lexical items that the native speaker hardly regards them as free combinations or separate items. *Acute pain, high winds, light drizzle* can be considered as common examples of solid lexical collocations.

It can be clearly said that there are robust differences between native and non-native writers in terms of using lexical collocations. In all categories, considerable differences existed in favour of native writers who used much more collocations than Turkish authors. However, the situation was different in the category of *noun + verb* in which Turkish authors used more collocations than their counterparts. In other words, Turkish authors used *noun + verb* collocations like “the table shows...; the study reveals...; the data indicate...etc.”, more than native authors[Abdurakhmonova, N.2021,2022,2023]. Meanwhile, the results put forth that native writers have a tendency of using *adjective + noun* collocations like “prefer to use *booster* (assertive words) adjectives for nouns and adverbs for verbs like “effective ways, key research, intense criticism”. As seen from the examples, native authors preferred to increase their force of statements by using boosters such as “effective, key, and intense” [Walter LaFeber, 2011, 4].

The other thing that the present study revealed is the close relation between nativity of the authors and the number of collocation which were used. It is highly apparent that native authors used much more collocations than Turkish authors, which may be refer to a sign of strong correlation between being a native speaker and non-native speaker of English [Abdurakhmonova, N.2021,2022,2023]. Prodromou, who saw collocations as a potential difficulty that non-native writers encounter, claims that a close relationship existed between collocations and native fluency. Some other researchers made experimental and/or theoretical investigations so that the relationship might be proven. One of them belongs to Martynska who carried out a study aiming to reveal collocational competence of non-native English speakers and the role of it in the process of L2 learning.

Conclusion

As last, the present study provided valuable insight for non-native writers who desire to have native-like academic writing. However, a study with larger data could provide more accurate results. Furthermore, the present study only investigated ELT in terms of detecting lexical collocations, however; because the use of word combinations may be unique to the discipline itself,

each discipline needs to be studied in terms of collocations usage in order to give detailed insight about the use of collocations.

References:

1. America and Russian and the Cold War, Walter LaFeber, 6th Edition, Cornell University 2011.
2. Абдурахмонова, Н., & Абдувахобов, Г. (2021). O ‘QUV LUG ‘ATINI TUZISHNING NAZARIY METODOLOGIK ASOSLARI. МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА, 4(6).
3. Abdurakhmonova, N., Shakirovich, I. A., & O‘G‘Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. Science and Education, 3(9), 146-156
4. Ismailov, A. S., Shamsiyeva, G., & Abdurakhmonova, N. (2021). Statistical machine translation proposal for Uzbek to English. Science and Education, 2(12), 212-219.
5. Абдурахмонова, Н., & Бойсариева, С. (2023). TABIIY TILNI QAYTA ISHLASHDA (NLP) OKKAZIONALIZMLARNING MORFEM TAHLILI. МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА, 6(3).
6. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
7. English Grammar, L.S. Barhudarov & D.A. Schteling, Moscow 2015
8. Translation features, Basnett-McGuire, New York Publishing house 2010.
9. Translation difficulties, T.R. Levitskaya & A.M. Fitterman, “International Relations” Publishing house, Moscow 201
10. U. Khoshimov I., Yokubov I. Methods of teaching English. - Tashkent, 2013. 117p.

THE ESSENCE OF THE CHARMING POETRY OF ZAHIRUDDIN MUHAMMAD BABUR

Kholbekova Begoyim
Associated professor
Uzbekistan State World Language University.
E-mail: begoyimholbekova@mail.ru

Annotation: This article talks about the unique content of the beautiful poetry of the great poet and statesman Zahiruddin Muhammad Babur. Also, translations of Babur's best ghazals into English are presented to the readers with parallel texts in the original language.

Keywords: Babur, charming poetry, ghazals, love, nostalgia, devotion, loyalty;

Аннотация: В данной статье говорится об уникальном содержании прекрасной поэзии великого поэта и государственного деятеля Захируддина Мухаммада Бабура. Также читателям представлены переводы лучших газелей Бабура на английский язык с параллельными текстами на языке оригинала.

Introduction

Zahiruddin Muhammad Babur, who knew well the past of literature and history, music and art, was devoted to religious teaching, was always in the circle of scientists and virtues, in particular, he showed sincere respect for creative people, patronized people of professions and supported them financially and spiritually. This loving attitude towards people of creativity and art was not in vain. Babur was creative by nature. From his youth until the end of his life, he was engaged in effective creative activity, did not stop his creativity in any conditions and situations, and as a result left a rich scientific and literary heritage.

Discussion and analysis

The main theme of Zahiruddin Muhammad Babur's poetry is love and nostalgia. In his ghazals Babur raises such problems of human relationships as love, friendship, and the desire for beauty. With great sincerity, he expresses his readiness to sacrifice himself for the sake of love.

Love for Babur is loyalty, devotion, nobility and humanity. In his wonderful ghazals Babur creates the image of a beautiful beloved, endowing her with an incomparable beautiful appearance and spiritual perfection. At the same time, he skillfully uses original artistic means:

Qaddi shoxi guldek nigorim qani?
Labi g'unchadek gul'uzorim qani?
Ne tong, yuz nigor o'lsa qon yosh bila -
Ki, ul yuzi gul yosh nigorim qani?
Qaro kechalarda ko'zim yoritur,
O'shal sham'i shabhoi torim qani?

Zahiruddin Muhammad Babur

My beloved with body like a cypress is where?
My sweetheart with lips like a flower bud is where?
There are hundred beauties with bloody tears, but
She who with flower face is where?

She who lights my eyes at dark nights,
That sun shine of my soul is where?

Translated by Begoyim Kholbekova

The theme of the homeland occupies a special place in Babur's lyrics. In his poems, especially in the quatrains, longing for his homeland and boundless love for it are expressed with great impressive way:

Ey yel, borib ahbobga nomimni degil,
Har kim meni bilsa, kalomimni degil.
Mendin demagil gar unutulg'on bo'lsam,
Har kimki meni so'rsa, salomimni degil.

Zahiruddin Muhammad Babur

Hey breeze, to people my best regards, give,
To those who know me, my words, give.
If I was forgotten don't say it is from me,
Each who remember me, my regards give.

(*The translation is ours. B.Kh.*)

From the following verses it is not difficult to notice the poet's strong longing for his native country, his grassy laments:

Ko'pdin berikim, yoru diyorim yo'qtur,
Bir lahzayu bir nafas qarorim yo'qtur.
Keldim bu sori o'z ixtiyorim birla,
Lekin borurimda ixtiyorim yo'qtur.

Zahiruddin Muhammad Babur

For long neither beloved nor homeland I've had,
Neither time nor a moment for decision I had.
I came here of my own free will,
But I cannot go back of my own accord.

(*The translation is ours.* ***B.Kh.***)

Speaking about the artistic language of Babur's works, it is necessary to note its simplicity, accessibility and clarity. The poet does not like loud phrases and complex expressions. Babur's poetic mastery is expressed in his artistic and literary style and skillful use of the most expressive means of his native language and in the creative recreation of sources of folk art. The following ghazal, written about a beautiful beloved, can charm anyone, because it sounds as natural as the pleas and confessions of lovers:

Ul ahd ila paymon qani, ey yor, ne bo'ldi?
Ul lutf ila ehson qani, ey yor, ne bo'ldi?
Ketdim men hayron eshigingdan, demading hech
Ul telbai hayron qani, ey yor, ne bo'ldi?

Zahiruddin Muhammad Babur

Where is your oath, promise, hey beloved, what happened?
Where's your caresses, signs of love, hey beloved, what happened?
I left your door, I was surprised but you didn't say
Where that mad man, crazy person, what happened?

(*The translation is ours.* ***B.Kh***.)

Results

While studying the poetry of Zahiruddin Muhammad Babur, his following charming ghazals also inspired us to translate them into English.We hope they will bring unique joy to your hearts.

***

Mening ko‘nglumki, gulning g'unchasidek tah-batah qondur,
Agar yuz ming bahor o’lsa, ochilmog'i ne imkondur.

Agar ul qoshi yosiz bog' gashtin orzu qilsam,
Ko‘zumga o‘qdurur sarv-u ko‘ngulga g'uncha paykondur.

Bahor-u bog‘ sayrin ne qilaykim, dilistonimning,
Yuzi gul, zulfi sunbul, qomati sarvi xiromondur.

Visoli lazzatidin zavq topmog‘liq erur dushvor,
Firoqi shiddatinda yo‘qsa jon bermaklik osondur.

Boshidin evrulur armoni birla o‘ldum, ey Bobur,
Mening na’shimni bori ul pari ko‘yidin aylondur.

***

Like flower buds, my heart is full of blood,
Thousands of springs come, but sorrows in my heart!

If I dream of walking in the garden without her,
An arrow pierces my eye and cypress pierces the heart!

What ambling through spring flowers to me?
Her rose face, her cypress figure – all are in my heart!

It’s difficult to taste the spirit of passion, o,
It’s easier to die if you don’t have it, hey bard!

I died with a desire in my heart, o Babur.
When my corpse is carried pass her – the rose sobs so hard.

***

Agarchi sensizin sabr aylamak, ey yor, mushkildur,
Sening birla chiqishmog‘liq dog‘i bisyor mushkildur.
Mizojing noziku sen — tundu, men — bir beadab telba,
Sanga xolimni qilmoq, ey pari, izhor mushkildur.
Ne osig‘ nolayu faryod xobolud baxtimdin,
Bu unlar birla chun qilmoq ani bedor mushkildur.
Manga osondurur bo‘lsa agar yuz ming tuman dushman,
Vale bo‘lmoq jahonda, ey ko‘ngul, beyor mushkildur.

Visolinkim tilarsen, nozini xush tortg'il, Bobur -
Ki, olam bog'ida topmoq guli bexor mushkildur.

***

Spending a day apart from you is difficult for me.
But also getting along with you, is difficult for me.

Your taste is delicate, you are quiet, I am a silly fool,
To express my feelings is difficult for me.

What is my cry, what is my groan, if my sweetheart in deep sleep?
Keeping her awake with these pleadings is difficult for me.

It is easy to defeat a hundred thousand enemies, but
Dwelling in the world without lover is difficult for me.

If you want a true love, be careful, Babur -
To find a flower without a thorn is difficult for me.
Translated into English by Begoyim Kholbekova

Conclusion

Babur's rich literary heritage is still widely studied and researched. In a word, the beautiful poetry of the king and great poet Zahiruddin Muhammad Babur has captivated hearts for more than five and a half centuries. We believe that the unique poetry of the great poet will attract the attention of poetry lovers and researchers for many years to come.

Reference:

1. Bobur.Yoz fasli yor fasli. Gafur Gulom nashriyoti.Toshkent 2016.
2. Бабур. Избранное. Лирика Востока. Издательство Sharq 2006.
3. А.Ҳожиаҳмедов. Мумтоз Бадиият Луғати. Т. 2020.

MASHINA TARJIMASI UCHUN GRAMMATIK SATHDA MODELLASHTIRISH BOSQICHLARI

Shamsiyeva Gulshoda
E-mail: gulshodashamsiyeva5@gmail.com

Annotatsiya. Ushbu maqolada grammatik modellar va ularning avtomatik tarjimadagi o‘rni haqida so‘z boradi. Modellashtirish jarayoni, model va uning turlari muhokama qilingan. Shuningdek, o‘zbek va ingliz tilidagi sodda gap modellari tahlil qilingan bo‘lib, ushbu tillarda ega va kesimning moslashuvini modellashtirishga e’tibor qaratilgan.

Kalit so‘zlar: lingvistik model, orginal model, funksional model, struktor model, gap shakllarni modellashtirish.

Annotation: This article talks about grammatical models and their role in automatic translation. The modeling process, model, and its types are discussed, and simple sentence models in Uzbek and English are analyzed. Emphasis is placed on modeling the adaptation of possessive and participle in these languages.

Keywords: linguistic model, original model, functional model, structural model, modeling of speech forms.

Аннотация: В данной статье рассказывается о грамматических моделях и их роли в автоматическом переводе. Обсуждается процесс моделирования, модель и ее виды, анализируются модели простых предложений на узбекском и английском языках. Особое внимание уделяется моделированию адаптации притяжательного падежа и причастия в этих языках.

Ключевые слова: лингвистическая модель, оригинальная модель, функциональная модель, структурная модель, моделирование речевых форм.

Lisoniy birliklarni modellashtirish masalasi jahon tilshunosligida bir qator tadqiqotchilar tomonidan tahlil qilingan.

Lingvistik model tushunchasi haqida qarashlar struktur tilshunoslikning E.Sepir, L.Blumfild, R.Yakobson, X.Chomskiy, Z.Xarris, Ch.Hokket kabi namoyondalari asarlarida aks ettirilgan bo‘lib, XX asrning 60-70-yillari (matematik va kibernetik lingvistika rivojana boshlagan davr) da taraqqiy etadi. [Abduraxmonova N., 2021; 225].

Modellashtirishga oid bir qancha ilmiy tadqiqotlar amalga oshirilgan. Modellashtirish masalasiga system tilshunoslik namoyandalaridan bir qator tilshunoslar e’tibor qaratganlar. Jumladan, rus tilshunoslaridan Y. F. Kirov, L.N. Murzin, I.I. Revzin, o‘zbek tilshunosligida morfemik, morfologik, sintaktik sathda modellashtirish masalalari T. Mirzaqulov kabilarning ilmiy izlanishlarida yoritib o‘tilgan.

Boduen asarlarida fonema goh so‘zlovchi ongida tovushlarni umumlashtirish natijasida hosil bo‘lgan ideal mohiyat sifatida, goh ilmiy mavhumlashtirish (abstraksiya), lingvistik modellashtirish natijasi sifatida izohlanadi.

A.Gardinerning fikricha, har qanday lug‘at va har qanday sintaktik modellar yig‘indisi tildir. Qisqa muddatli, tarixan betakror so‘zdan foydalanish faoliyati nutqdir. [Nurmonov A., 2009; 160].

L.N.Murzin obyekti bevosita kuzatuvga berilmagan har qanday fan obyektni modellashtirishga majbur deb hisoblaydi.

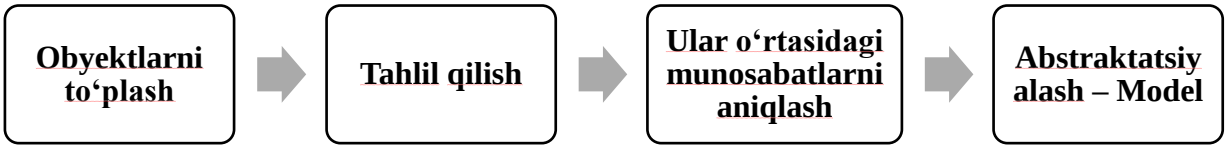
Modellashtirish dinamik va kommunikativ lingvistika, psixolingvistika va struktur tilshunoslikning obyekti sifatida namoyon bo‘ladi.

Lingvistik modellashtirish berilgan semiotik elementlar o‘rtasidagi munosabatlar tizimi sifatidagi til bilan yoki verbal kommunikatsiya jarayoni sifatidagi til bilan ish ko‘radi.

Til modeli haqida fikr yuritar ekan, I.I.Revzin shunday yozadi: “Model quyidagicha quriladi. Muayyan fan tomonidan to‘plangan xilma-xil tushunchalar yig‘indisidan birlamchi deb qabul qilinuvchi ayrimlari tanlab olinadi. Birlamchilar o‘rtasidagi munosabatlar aniqlanadi va ular pastulat sifatida qabul qilinadi. Qolgan barcha da’volar qat’iy deduktiv ravishda terminlarda namoyon bo‘ladi. Ular birlamchi tushunchalar orqali aniqlanadi. Shu ma’noda model sistema sifatidagi tilning qismi emas, balki qandaydir konstrukt, gipotetik ilmiy qurilma sifatida qabul qilinadi”. [Ревзин И., 1962; 9].

Yadro gap modeli va uning turli hosilalari ro‘yxatining olinishi u yoki bu tilda muayyan axborotni turlicha ifodalash imkoniyatlarini ochib beradi. [Nurmonov A., 2009; 145].

Manbalarga ko‘ra, modellashtirish to‘rt bosqichda amalga oshiriladi:



Bu holatda model obyektlarni intellektual, fikriy abstraktlashtirish natijasidir.

Kompyuter lingvistikasiga “lingvistik model” tushunchasi struktur lingvistikadan kirib kelganligi uchun unga struktur lingvistika va kompyuter lingvistikasini bog‘lab turuvchi mustahkam ko‘prik deb ta’rif beradi.

Model bevosita kuzatish imkoni bo‘lmagan hodisani o‘rganish uchun yaratilgan sun’iy mexanizm. Ammo matn inson nutq faoliyati asosida yotuvchi til mexanizmini to‘liq tushunish uchun yetarli emas. Shuning uchun tilshunoslikda obyektni tushunishning asosiy vositasi sifatida modellardan foydalaniladi.

Lingvistik strukturalar model bo‘lishi mumkin. Biroq sintaktik jihatdan qaralganda modellashtirish lingvistik o‘ziga xos jihatlarni, ya’ni sintaktik omonimiya masalasini qamrab ololmaydi [Abdurakhmonova, N.2021,2023,2024].

Tilni matn orqali kuzatish mumkin, ammo matnning inson nutqiy faoliyati jarayonida yuzaga chiquvchi tabiiy til hodisalarini to‘liq tushunish uchun yetarli emasligi adabiyotlarda ko‘p bora ta’kidlangan. Shuning uchun tilshunoslikda obyektni tushunishning asosiy vositalaridan biri sifatida modellardan foydalaniladi. Matn nutqiy vaziyatning yozma shakli bo‘lganligi uchun u tilshunoslikning barcha sathlari uchun asos vazifasini bajaradi. [Abduraxmonova N., 2021; 319].

Avtomatik tarjima uchun matn tushunchasi bir-biri bilan grammatik va mantiqiy jihatdan uzviy bog‘langan gaplar tuzilmasini anglatadi, shu sababli ham tarjima tayyor matnlar asosida amalga oshiriladi. Bunda matnni yaratishga emas, balki mavjud bir tilda kiritilgan matnni ikkinchi tilga grammatik va mantiqiy shakllangan matn holatida qanday qayta tiklashga e’tibor qaratiladi.

Buning uchun tayyor matn holatidagi matn kompyuter yordamida bazaga kiritiladi. Matematik ta'rifga ko'ra matn – harflar, raqamlar, shartli belgilar, turli xil ishoralar kabi maxsus belgilarning ketma-ketligidan iborat to'plamdir. Shunga ko'ra kodlashtirish amalga oshiriladi. Buning uchun maxsus belgilar tartiblanib, chekli tartiblangan to'plam holatiga o'tkaziladi va har bir belgini tartib bo'yicha raqamlab, sonlarning ikkilik sanoq sistemasidagi ko'rinishi bilan nisbiy holda kodlashtirib olinadi. Tarjima jarayoni uchun til juftliklarining har bir fragmenti kodlashtiriladi va dasturiy ta'minotga beriladigan buyruq shu kodlashtirish asosida amalga oshiriladi[Abdurakhmonova, N.2021,2023,2024].

Modellashtirishda quyidagi jihatlar e'tiborga olinadi: [Abduraxmonova N., 2021; 315-316].

I. Aniqlashtirish: obyekt butun va qism sifatida o'rganiladi:

1) umumiy g'oya ayrim vositalari orqali tekshirish;

2) obyektni qayta nomlash;

3) u yoki bu muhitdan obyektni material holatda ajratish mumkin.

II. Strukturani soddalashtirish: obyekt mantiqiy ichki tuzilma sifatida tekshiriladi, bunda turli qismlardagi farqlar o'rganiladi.

III. O'zaro mustaqillik, funksiyallik.

IV. Originallik.

V. Barqarorlik.

Modellashtirish metodida tadqiqotchi obyektning o'zini emas, balki uning modelini o'rganadi. Original obyekt bilan model o'rtasida o'zaro bog'liqlik va mutanosiblik mavjud bo'ladi.

N.Abduraxmonova tilshunoslikda modelning 3ta turi haqida ma'lumot beradi:

1. Original modellar.
2. Funksional modellar.
3. Struktur modellar.

Original modellar – obyektning tuzilishini, funksional modellar – original modellarning qanday ishlashini, struktur modellar – har ikkalasini o'rganadi.

Orginal modellar quyidagi talablarga javob berishi kerak:

1) model tabiiy obyektning aynan nusxasi bo'lishi;

2) original model o'zida elementlarning murakkab tuzilishini namoyon eta olishi;

3) tabiiy obyektning hamma xususiyatlari original modelga to'g'ri kelishi;

4) model evristik funksiyaga ega bo'lishi, ya'ni u yangi g'oyalar bera olishi va uni amaliyotda sinab ko'rish imkoniyati bo'lishi zarur.

Sosnina qarashlariga asoslanadigan bo'lsak, model lingvistik obyektning formal tavsifini berishda qulay vosita bo'lib, uning quyidagi turlari mavjud: [Соснина Е., 2012; 36].

1) Model – chizmalar tilga oid obyektlarning mazmuni va strukturasini ifodalashda qo'llaniladi;

2) Model – tur (language pattern) muayyan barqaror birliklar uchun (so'z birikmalari, sintaktik strukturalar);

3) Model nazariya – formallashgan ilmiy nazariya, elementlar strukturasi va metatilga moslashtirilgan qoidalar asosidagi tavsif.

Kompyuter texnologiyalarga asoslangan inson bilimi, tili va tafakkurini modellashtirishda sun'iy intellekt texnologiyasi shug'ullanadi.

Avtomatik tarjima muhiti lingvistik ta’minotida modellashtirishni ingliz va o‘zbek tillarida gapshakllarini ko‘rib chiqishdan boshlaymiz. Ingliz tilida sodda gapshakli quyidagicha:

1) [S+VP] – S – Subject (Ega) VP – fe’lshakl – I read - fe’l kesimli gapshakl;

2) [S + NP] – S – Subject (Ega) NP – otshakl – I am a pupil ot kesimli gapshakl.

Gap qoliplaridan ma’lumki, demak, kesim egalik, zamon va modal ma’no anglatuvchi grammatik ko‘rsatkichlar bilan birga qo‘llanadi. O‘zbek tilida esa sodda gap modeli [WPm] – W – atov birlik, P – predikat, m – marker (Bor-di-m. O‘quvchi-san) holida qabul qilingan[Abdurakhmonova, N.2021,2023,2024].

Tarjima uchun sodda gapshakllarini modellashtirishda gapning umumiy tahlili emas, balki gap bo‘laklari tahlili asos qilib olinadi. Ingliz tilida sodda gapshakllarini modellashtirishda ularning kesimiga tayaniladi va fe’l-kesimli va ot-kesimli tarzida ajratiladi. Fe’l kesimli gapshakllarini modellashtirish jarayonida kesimning tranzitiv (o‘timli) va intranzitiv (o‘timsiz) ligi inobatga olinadi.

Kesimi o‘timli fe’ldan tarkib topgan gap modellarini ko‘rib chiqamiz[29]:

1. N – V – DO ⟹ *The **girl kicked** the dog.* (***Qiz** itni **tepdi***)

2. Adj – N – V – DO ⟹ *The **young** girl kicked the dog. (**Yosh** qiz itni tepdi)*

3. Adj – N – Adv – V – DO ⟹ *The young girl **viciously** kicked the dog. (Yosh qiz itni **shafqatsizlarcha** tepdi)*

4. Adj – N – Adv – DO – PP ⟹ *The young girl viciously kicked the dog **down the stairs**. (Yosh qiz itni **zinadan** shafqatsizlarcha tepdi)*

5. Adj – N – V – Conj – Adv – V – PP ⟹ *The young girl **and her brother** viciously kicked the dog down the stairs. (Yosh qiz **va uning akasi** itni zinadan shafqatsizlarcha tepishdi)*

6. *The young girl and her brother viciously kicked the dog down the stairs **and laughed**. (Yosh qiz va uning akasi itni zinadan shafqatsizlarcha tepib, **kulib yuborishdi**)*

7. *The young girl and her brother viciously kicked the dog down the stairs anf laughed **nervously**. (Yosh qiz va uning akasi itni zinadan shafqatsizlarcha tepib, **asabiy** kulishdi)*

8. *The young girl **felt threatened** and viciously kicked the dog down the stairs. (Yosh qiz **xavf ostida qolganini sezdi** va itni zinadan shafqatsizlarcha tepdi)*

9. *The young girl**, feeling threatened**, viciously kicked the dog down the stairs. (Qiz o‘zini **tahdid ostida his qilib**, zinadan itni shafqatsizlarcha tepdi)*

10. ***Feeling threatened**, the young girl viciously kicked the dog down the stairs. (**O‘zini tahdid ostida his qilgan** yosh qiz itni shafqatsizlarcha zinadan tepdi)*

11. *Jhon broke his leg when his car hit the tree. (Mashina daraxtga urilganda Jonning oyog‘i singan)*

12. *Mary bought her dress after she fought with her mother. (Meri ko‘ylagini onasi bilan urushganidan keyin sotib oldi)*

13. *Mary brought his mother befor he made dinner. (Maryam kechki ovqat tayyorlashdan oldin onasini olib keldi)*

Quyida kesimi to be (am is, are, was, were), will be, woud be kabi yordamchi fe’llar bilan yasaladigan asosiy gap modellari keltirilgan[30]:

1. N/P – V – Adj – Sarah was brilliant.

[29] Qarang: https://en.wikiversity.org/wiki/TESOL/English_sentence_structure/Exercises
[30] Qarang: https://en.wikiversity.org/wiki/TESOL/English_sentence_structure/Exercises

2. N/P – V – N – Sarah was a student.
3. N/P – V – Adj – N – Sarah was a brilliant student.
4. N/P – V – N – N – Sarah was a brilliant student and a kind friend.
5. N/P – N/P – N – V – Adj – N – Her sister and she were brilliant students.
6. N/P – N/P – V – Adj – Sarah was a brilliant student and expected to go to college.
7. N/P – V – V – VP free modidiers – Sarah was a brilliant student but did not excpect to go to college.
8. A brilliant student, Sarah expected to go to college.
9. Mary, a brilliant student, expected to go to college.
10. Mary was a brilliant student who expected to go to college.
11. Mary and I remained hungry.
12. Mary and I remained hungry and tired.
13. Mary and I remained hungry and tired, but we worked hard.
14. Mary and I remained hungry and tired, but we still worked hard.

Avtomatik tarjimada gap modellari borasida N. Abduraxmonova tadqiqot olib borgan bo‘lib, mashina tarjimasining lingvistik ta’minotiga bag‘ishlangan monografiyada jami hisobda ingliz tili gap modellarining 12ta darak gap tipi, so‘roq gaplarning 10 ta turi, undov gaplarning 12 turi modellashtirilgan. [Abduraxmonova N., 2018]

Mashina tarjimasi uchun ma’lumotlar bazasi va lingvistik ta’minotning to‘liq va to‘g‘ri shakllanganligi asosiy o‘rin egallaydi. Chunki matndagi fragmentlarni ajratish uchun dasturga avtomatik buruq berilganida, natija kompyuter xotirasiga joylangan lingvistik ta’minot orqali olinadi. Shu sababli ham database uchun yaratilgan gap modellarining to‘g‘riligi va lingvistik ta’minotning mukammalligi jarayonning samaradorligini ta’minlaydi[Abdurakhmonova, N.2021,2023,2024].

Matnni sintaktik analiz qilishda ingliz tilidagi bosh bo‘laklar – ega va kesimni aniqlab ajratib olish tarjima jarayoning muhim bosqichi sanaladi. N. Abduraxmonovaga ko‘ra modellashtirishda matematik shartli belgilardan foydalaniladi:

{ } – ixtiyoriy tanlov
[ ] – model belgisi
| - bo‘lishi mumkin
=> kelib chiqish

Tarjima jarayonida, avvalo, ega va kesim ajratib olinadi va morfologik lug‘atda keltirilgan guruhlar bo‘yicha joylanadi.

Ingliz-o‘zbek avtomatik tarjima muhitini yaratishda ingliz tilidagi gapshakllarni o‘zbek tiliga o‘girishda ega va kesimning moslashuvi ancha murakkab jarayon hisoblanadi. Bunga matnni gap bo‘laklari bo‘yicha sintaktik tahlil qilish paytida ikki tilning struktura va mazmun tomondan farqli paradigmalarga egaligining namoyon bo‘lishi sabab bo‘ladi. Ega va kesim moslashuvini modellashtirishda quyidagi shartli grammatik qisqartmalardan foydalaniladi: [Abduraxmonova N., 2018]

PS-Predicate (Singular)- KB-kesim birlik	**Sub.Thing-Epred.**-(narsa-buyum oti)	**Art.**-artıcle-artikl
PP-Predicate (Plural)- KK)- KK-kesim ko‘plik	**Sub.P(Person) -Otsh.**-shaxs oti	**OtA**- atoqli ot

SubS.-subject-EB-(ega birlik)	**Pneg**.-negatıve pronoun	**Ol**-olmosh
SubP.-Subject-E**K**-(ega ko‘plik)	**Puniv**. –universal pronoun	**Obj**.-object-**O**-to‘ldituvchi
Conj.-(Conjunction)-B.(bog lovchi)	**Sub.**ⁿ-uyushiq ega	**Att.**-atributive-**A**-aniqlovchi
VAux- Auxilary verb - yordamchi fe’l	**OKB**-Object birlik yoki ko‘plikda	**ZM-zamon**
S-sıngular-birlik	**P**-plural-ko‘plik	

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. Kompyuter lingvistikasi (darslik), Toshkent: Nodirabegim, 2021. 400-b
2. Abduraxmonova N. Mashina tarjimasining lingvistik ta’minoti. (Monografiya) Toshkent .2018.
3. Abduraxmonova, N. Z. “Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref.” (2018).
4. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016;2 (38):12-7.
5. Abdurakhmonov N. Modeling Analytic Forms of Verb in Uzbek as Stage of Morphological Analysis in Machine Translation. Journal of Social Sciences and Humanities Research. 2017;5(03):89-100.
6. Alessandro Agostini, Timur Usmanov, Ulugbek Khamdamov, Nilufar Abdurakhmonova, and Mukhammadsaid Mamasaidov. 2021. UZWORDNET: A Lexical-Semantic Database for the Uzbek Language. In Proceedings of the 11th Global Wordnet Conference, pages 8–19, University of South Africa (UNISA). Global Wordnet Association.
7. D. B. Mengliev, N. Abdurakhmonova, D. Hayitbayeva and V. B. Barakhnin, “Automating the Transition from Dialectal to Literary Forms in Uzbek Language Texts: An Algorithmic Perspective,” 2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE), Novosibirsk, Russian Federation, 2023, pp. 1440-1443, http://doi.org/10.1109/APEIE59731.2023.10347617
8. Gatiatullin, A., Suleymanov, D., Prokopyev, N., Abdurakhmonova, N. (2024). “Turkic Morpheme”: From the Portal to the Linguistic Platform. In: Aliev, R.A., et al. 12th World Conference “Intelligent System for Industrial Automation” (WCIS-2022). WCIS 2022. Lecture Notes in Networks and Systems, vol 912. Springer, Cham. https://doi.org/10.1007/978-3-031-53488-1_22
9. Mengliev Davlatyor, Vladimir Barakhnin, and Nilufar Abdurakhmonova. 2021. “Development of Intellectual Web System for Morph Analyzing of Uzbek Words” Applied Sciences 11, no. 19: 9117. https://doi.org/10.3390/app11199117
10. Мирзакулов Т. Ўзбек тили морфемикасининг асосий аспектлари ва уларнинг муносабатлари. -Хўжанд.1992.
11. Мурзин Л.Н. Лингвистическое моделирование и деривация речеврй деятельности. – Деривация в речевой деятельности. Пермь, 1988.

12. Nurmonov A. Struktur tilshunoslik: ildizlari va yo‘nalishlari. – Toshkent: Ta’lim, 2009.
13. Ревзин И.И. Модели языка. – М., 1962. – С.9.
14. Соснина Е. П. Введение в прикладную лингвистику Учебное пособие Ульяновск УлГТУ. 2012.
15. Qodirova Z.G. Semantik texnologiyalar yaratishda ontologiyalardan foydalanish. Filolog tadqiqotlar: Til, Adabiyot, Ta'lim 132, 74/79, 2023.
16. Qodirova Z.G. Bilimlar bazasini ontologik modellarda ifodalash. International Scientific Conferences with Higher Educational Institutions. 1.., 2023.
17. Qodirova Z.G. Terminologik ma’lumotlar bazasini loyihalashda ontologik modellashtirish. RESEARCH AND EDUCATION 1 (2), 154-160, 2022.
18. Qodirova Z.G. Tibbiyot sohasi terminlarining ontologik modelini yaratishda sinf hamda munosabatlar tavsifi. Academic research in educational sciences 3(5), 1134-1140, 2022.
19. Кодирова З.Г. Tabiiy tillar jarayoni uchun semantik munosabatlar tahlilida ontologiyaning o‘rni. Международный журнал искусство слова 5 (2), 2022.

BREAKING DOWN LANGUAGE BARRIERS: THE POWER AND PROGRESS OF MACHINE TRANSLATION

Shamieva Nargiza
PhD student, English Teacher
National University of Uzbekistan.

Abstract: In our increasingly interconnected world, effective communication across language barriers is paramount. This is where machine translation comes into play, revolutionizing the way we bridge linguistic gaps. As learning and development specialists, we will appreciate how technology is reshaping industries, including the trucking business, by making communication more accessible and efficient. Neural machine translation, on the other hand, has brought about a paradigm shift. Inspired by the human brain, NMT uses deep learning models called neural networks to process entire sentences and even paragraphs. This enables it to consider context, idiomatic expressions, and linguistic nuances, resulting in more accurate translations. While machine translation has made significant strides, it's not without challenges. It can still struggle with idiomatic expressions, cultural nuances, and low-resource languages. However, ongoing research and development promise continued improvement in these areas.

Keywords: machine translation, linguistic gaps, communication, neural networks, trucking Industry, compliance standards

Annotatsiya: Bizning tobora o'zaro bog'liq dunyomizda til to'siqlari orqali samarali aloqa juda muhimdir. Aynan shu erda til tanaffuslarini bartaraf etish usullarini inqilobiy ravishda o'zgartiradigan mashina tarjimasi paydo bo'ladi. Ta'lim va rivojlanish bo'yicha mutaxassislar sifatida biz texnologiya tarmoqlarni, shu jumladan yuk tashishni qanday o'zgartirayotganini baholaymiz, bu esa muloqotni yanada qulay va samarali qiladi. Boshqa tomondan, neyromashina tarjimasi paradigma o'zgarishiga olib keldi. Inson miyasidan ilhomlangan NMP butun jumlalar va hatto paragraflarni qayta ishlash uchun neyron tarmoqlar deb ataladigan chuqur o'rganish modellaridan foydalanadi. Bu unga kontekstni, idiomatik iboralarni va lingvistik nuanslarni hisobga olishga imkon beradi, bu esa aniqroq tarjimaga olib keladi. Mashina tarjimasi katta yutuqlarga erishgan bo'lsa-da, u qiyinchiliklardan xoli emas. U hali ham idiomatik iboralar, madaniy nuanslar va kam resursli tillar bilan shug'ullana olmaydi. Biroq, doimiy tadqiqotlar va ishlanmalar yanada takomillashtirishni va'da qilmoqda

Kalit so'zlar: mashina tarjimasi, lingvistik bo'shliqlar, aloqa, neyron tarmoqlar, yuk tashish, muvofiqlik standartlari

Аннотация. В нашем все более взаимосвязанном мире эффективная коммуникация через языковые барьеры имеет первостепенное значение. Именно здесь на помощь приходит машинный перевод, революционно меняющий способы преодоления языковых разрывов. Как специалисты по обучению и развитию, мы оценим, как технологии меняют отрасли, включая грузоперевозки, делая общение более доступным и эффективным. Нейромашинный перевод, с другой стороны, привел к смене парадигмы. Вдохновленный человеческим мозгом, НМП использует модели глубокого обучения, называемые нейронными сетями, для обработки целых предложений и даже абзацев. Это позволяет ему

учитывать контекст, идиоматические выражения и языковые нюансы, что приводит к более точному переводу. Хотя машинный перевод добился значительных успехов, он не лишен трудностей. Он все еще не справляется с идиоматическими выражениями, культурными нюансами и языками с низким содержанием ресурсов. Однако постоянные исследования и разработки обещают дальнейшее совершенствование в этих областях.

Ключевые слова: машинный перевод, лингвистические пробелы, коммуникация, нейронные сети, грузоперевозки, стандарты соответствия

A Brief History

Machine translation dates back to the mid-20th century, with early rule-based systems. These systems struggled with nuances and context. However, the field has advanced with the rise of statistical and neural machine translation, driven by machine learning approaches [Grace Hui-chin Lin, 2009, 133].

Machine translation is the automated process of translating text or speech from one language to another using computer algorithms. It eliminates the need for human translators in many scenarios, making it faster and more cost-effective. It's a topic that might not seem directly related to your role in learning and development, but understanding its potential impact on the trucking industry is crucial.

The Evolution of Machine Translation

Machine translation has come a long way since its inception. Early attempts at translation relied on rule-based systems, which had limitations due to the complexity of human languages. However, recent advancements in artificial intelligence and machine learning have transformed the landscape.

Recent developments in neural machine translation have led to more context-aware and accurate translations. However, challenges remain, such as handling idiomatic expressions and low-resource languages. Researchers are continually working to improve these systems [Haifeng Wang, 2022, 1].

Statistical Machine Translation (SMT): This approach used statistical models to predict the best translation based on large bilingual corpora. While it improved translation quality, it still struggled with context and idiomatic expressions.

Neural Machine Translation (NMT): NMT revolutionized the field by using deep neural networks to translate entire sentences at once, capturing context and nuances much better than previous methods. This has become the dominant approach in recent years [Abdurakhmonova, N.2021,2022,2023].

Statistical machine translation relies on statistical models and algorithms to decipher languages. It learns from vast amounts of parallel text data, mapping words and phrases from one language to another based on probability. While effective for its time, SMT had limitations in capturing context and nuance.

Neural machine translation, on the other hand, has brought about a paradigm shift. Inspired by the human brain, NMT uses deep learning models called neural networks to process entire sentences and even paragraphs. This enables it to consider context, idiomatic expressions, and linguistic nuances, resulting in more accurate translations.

Numerous studies have compared the performance of SMT and NMT systems. They often highlight NMT's superiority in terms of fluency, coherence, and context understanding. NMT has significantly reduced the infamous "lost in translation" moments [8].

Both SMT and NMT have practical applications across industries. From translating websites and user manuals to facilitating cross-cultural business negotiations, these technologies are bridging language divides.

The Role of Machine Translation in the Trucking Industry

Machine translation can be a game-changer for trucking companies operating internationally. Here's how:

Multilingual Communication: Truck drivers often cross borders and interact with people who speak different languages. Machine translation can help them communicate with customers, colleagues, and authorities more effectively, reducing misunderstandings and improving efficiency [Abdurakhmonova, N.2021,2022,2023].

Documentation and Compliance: In an industry with strict regulations, accurate translation of documents is vital. Machine translation can ensure that legal and safety documents are available in multiple languages, helping trucking companies meet compliance standards.

Training and Development: As learning and development specialists, we can leverage machine translation to create multilingual training materials. This facilitates the onboarding process for employees from diverse linguistic backgrounds and enhances their understanding of safety protocols and procedures.

Machine translation has found applications in academia, aiding in the translation of research papers and facilitating global knowledge exchange. It can improve accessibility and understanding of scientific literature [6][7].

Dictionary-based translation is a method that enhances the humanization of translated content by providing accurate word meanings. In addition to internet tools like the Google and Yahoo toolbars, dictionary entries, such as Dr. Eye, contribute to making machine-translated words sound more natural and human-like. High-priced translation software, designed by linguistic and translation experts, often produces translations that are perceived as more truthful and readable compared to translations generated by Google or Yahoo toolbars.

To improve the comprehensibility of machine-translated text, the translation process requires human intervention. Re-encoding through human understanding of the target language and careful examination of the machine-translated text are two crucial steps in achieving more humanized translations. The act of translation involves complex cognitive operations, where the translator must decode the source text's original ideas into a logically coherent whole. This process demands a deep understanding of grammar, semantics, syntax, idioms, and linguistic nuances in both the source and target languages [Abdurakhmonova, N.2021,2022,2023]. Additionally, a professional translator should possess knowledge of the traditions, cultures, and linguistic preferences of the intended audience to ensure the translation's acceptability [6].

Bilingual or Interlingual machine translation represents a rule-based approach to software-driven translation that is readily usable by human translators. This modern method allows English text to be translated, albeit potentially with some ambiguity, into Uzbek or Russian. Translators typically refine these ambiguous translations using techniques like repetition and elaboration to make them more understandable.

Based on the experiences of researchers who translated a course at MIT, it was observed that Google and Yahoo toolbar's translation functions make it relatively easy to translate and comprehend English. According to the researchers, the machine translation tool accurately generated around eighty percent of the translations, leaving only a small amount of content

requiring additional clarification and interpretation. This suggests that English majors in Taiwanese universities could easily adopt and translate courses, requiring minimal effort and time after their formal education [6].

Challenges and Future Outlook

Machine translation is an interdisciplinary field, bringing together linguists, lexicologists, and computer scientists. Collaboration across these domains is crucial for advancing the technology further [Mojca Brglez and Vintar Špela Vintar, 2022, 5].

While machine translation has made significant strides, it's not without challenges. It can still struggle with idiomatic expressions, cultural nuances, and low-resource languages. However, ongoing research and development promise continued improvement in these areas.

Conclusion

In a world where global communication is essential, machine translation is a valuable tool that extends beyond its origins in language pairs and into various industries, including the trucking sector. We have the opportunity to explore how machine translation can enhance training and communication within your company, contributing to a more efficient and inclusive work environment.

References:

1. Abduraxmonova N. Kompyuter lingvistikasi (darslik), Toshkent: Nodirabegim, 2021. - B. 400
2. François Yvon, Machine Translation, France, 2022, pp 20-22.
3. Grace Hui-chin Lin, Machine Translation for Academic Purposes, Proceedings of the International Conference on TESOL and Translation, 2009, pp.133-148.
4. Абдурахмонова, Н., & Абдувахобов, Г. (2021). O ‘QUV LUG ‘ATINI TUZISHNING NAZARIY METODOLOGIK ASOSLARI. МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА, 4(6).
5. Abdurakhmonova, N., Shakirovich, I. A., & O‘G‘Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. Science and Education, 3(9), 146-156.
6. Ismailov, A. S., Shamsiyeva, G., & Abdurakhmonova, N. (2021). Statistical machine translation proposal for Uzbek to English. Science and Education, 2(12), 212-219.
7. Абдурахмонова, Н., & Бойсариева, С. (2023). TABIIY TILNI QAYTA ISHLASHDA (NLP) OKKAZIONALIZMLARNING MORFEM TAHLILI. МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА, 6(3).
8. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
9. Haifeng Wang, Progress in Machine Translation, Birmingham, 2022, pp. 1-2.
10. Mojca Brglez and Vintar Špela Vintar, Lexical Diversity in Statistical and Neural Machine Translation, 1000 Ljubljana, Slovenia, 2022, pp.5-10
11. https://www.researchgate.net/journal/Machine-Translation-1573-0573

TERMINOLOGY AS A LEARNING OBJECT IN MACHINE TRANSLATION

Zokirova Hulkar
Senior teacher, the department of
Foreign languages and literature of Angren University.

Annotation: Machine translation (MT) has seen significant advancements in recent years, yet challenges persist in accurately translating specialized terminology across various domains. This paper explores the role of terminology as a learning object in machine translation. It examines the significance of terminology in ensuring translation accuracy, consistency, and contextual understanding. Through a comprehensive review of literature and empirical evidence, this article highlights the importance of incorporating terminology into MT models as a learning object to enhance translation quality.

Keywords: Terminology, Machine Translation, Learning Object, Translation Quality, Contextual Understanding

Аннотация: В последние годы в машинном переводе (МП) наблюдался значительный прогресс, однако сохраняются проблемы с точным переводом специализированной терминологии в различных областях. В данной статье исследуется роль терминологии как объекта обучения в машинном переводе. В нем рассматривается значение терминологии в обеспечении точности, последовательности и контекстуального понимания перевода. Благодаря всестороннему обзору литературы и эмпирических данных в этой статье подчеркивается важность включения терминологии в модели машинного перевода в качестве объекта обучения для повышения качества перевода.

Ключевые слова: терминология, машинный перевод, объект обучения, качество перевода, понимание контекста.

Machine translation (MT) has revolutionized the way we communicate across languages, enabling rapid translation of texts with varying degrees of complexity. However, the accurate translation of specialized terminology remains a challenge for MT systems. Terminology, comprising domain-specific vocabulary and jargon, plays a crucial role in ensuring translation accuracy, coherence, and contextual understanding. In this paper, we delve into the significance of terminology as a learning object in machine translation, exploring its implications for translation quality, domain adaptation, customization, and quality control.

Terminology encompasses specialized vocabulary, concepts, and terminological conventions specific to various domains, such as medicine, law, engineering, and finance. In the context of machine translation, terminology refers to the key terms and expressions that require accurate translation within specific domains. Machine translation systems encounter several challenges in accurately translating terminology, including polysemy, synonymy, context dependency, and domain-specific nuances. These challenges necessitate specialized approaches and techniques for handling terminology within MT systems. "Terminology plays a critical role in domain adaptation, allowing machine translation models to specialize in translating texts from specific domains." (Chu, C., Wang, Y., Wu, W., & Ma, W., 2018).

Consistent translation of terminology is essential for maintaining coherence and readability in translated texts. Inconsistent or erroneous translation of key terms can lead to confusion and misinterpretation, undermining the overall quality of translation. Accurate translation of specialized terminology is crucial for preserving the intended meaning and conveying domain-specific concepts accurately. Machine translation systems must learn and understand the contextual nuances of terminology to produce precise translations. "The accurate translation of specialized terminology is essential for preserving the meaning and coherence of translated texts." (Koehn, P., 2017)

The integration of domain-specific terminology into training data and corpora is essential for enhancing the performance of machine translation systems. "Terminology serves as a valuable metric for evaluating the performance and quality of machine translation systems." (Chatterjee, S., & Gupta, K., 2018) By exposing MT models to diverse terminology contexts, they can learn to accurately translate specialized terms within specific domains. "The integration of external terminology resources into machine translation systems enhances translation accuracy and consistency across various domains."(Qi, F., & Davison, B. D., 2019) Terminology serves as a key component in domain adaptation, allowing machine translation systems to specialize in translating texts from specific domains. Domain-specific terminology models and dictionaries facilitate more accurate and contextually relevant translations. "Incorporating domain-specific terminology as a learning object enhances the performance of machine translation systems by improving contextual understanding."(Bentivogli, L., Bisazza, A., & Cettolo, M., 2016). Some organizations may have their own terminology preferences or conventions that they wish to adhere to in translated documents. Customizing machine translation systems to incorporate organizational terminology standards enhances translation consistency and alignment with organizational preferences. Effective terminology management is essential for quality control and evaluation in machine translation. Analyzing terminology usage and consistency helps identify areas for improvement and refine translation models to enhance overall translation quality. "Terminology management is crucial in machine translation to ensure consistency and accuracy across diverse domains."(Pinto, M., Nisi, V., & Palmonari, M., 2018). Several studies have evaluated the impact of terminology management on the performance of machine translation systems. Empirical evidence suggests that incorporating domain-specific terminology improves translation accuracy and consistency across various domains. User feedback and satisfaction with machine translation outputs are influenced by the accuracy and consistency of translated terminology. Studies have demonstrated that users perceive translations to be of higher quality when key terms are accurately translated within their respective contexts. "User satisfaction with machine translation outputs is influenced by the accurate translation of key terms within their respective contexts."(Specia, L., 2016)

The advancement of neural machine translation (NMT) holds promise for improving the handling of specialized terminology. Future research should focus on developing NMT models capable of learning and adapting to domain-specific terminology nuances more effectively. Integrating external terminology resources, such as domain-specific glossaries and ontologies, into machine translation systems can further enhance translation accuracy and consistency. Research efforts should explore methods for seamlessly incorporating external terminology resources into MT models. "Advancements in neural machine translation hold promise for improving the handling of specialized terminology and domain adaptation." (Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I., 2017)

Terminology plays a critical role in machine translation, influencing translation quality, accuracy, and contextual understanding. By treating terminology as a learning object, machine translation systems can improve their handling of specialized terminology across various domains. The integration of terminology into MT models enhances translation consistency, coherence, and accuracy, ultimately improving the usability and performance of machine translation systems.

References:

1. Abdurakhmonov N. Modeling Analytic Forms of Verb in Uzbek as Stage of Morphological Analysis in Machine Translation. Journal of Social Sciences and Humanities Research. 2017;5(03):89-100.
2. Abdurakhmonova N, Tuliyev U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
3. Abdurakhmonova N, Urdishev K. Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL). 2019;6(1-2019):131-7.
4. Abduraxmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref." (2018).
5. Aripov, M., Sharipbay, A., Abdurakhmonova, N., Razakhova B.: Ontology of grammar rules as example of noun of Uzbek and Kazakh languages. In: Abstract of the VI International Conference “Modern Problems of Applied Mathematics and Information Technology - Al-Khorezmiy 2018”, pp. 37–38, Tashkent, Uzbekistan (2018)
6. Bentivogli, L., Bisazza, A., & Cettolo, M. (2016). Neural versus phrase-based machine translation quality: a case study. arXiv preprint arXiv:1608.04631.
7. Cao, J., Chen, K., Jin, Y., & Li, X. (2020). A Review of Machine Translation Technology Based on Deep Learning. In Advances in Intelligent Systems and Computing (Vol. 1131, pp. 587-595). Springer.
8. Chatterjee, S., & Gupta, K. (2018). Sentiment Analysis of Amazon Product Reviews Using Machine Learning Algorithms. Procedia Computer Science, 132, 1078-1085.
9. Chu, C., Wang, Y., Wu, W., & Ma, W. (2018). Domain Adaptation for Neural Machine Translation with Domain-Invariant Feature Embeddings. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 1743-1755).
10. Koehn, P. (2017). Neural machine translation. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 9-12).
11. Pinto, M., Nisi, V., & Palmonari, M. (2018). A methodology for the construction of multilingual domain-specific terminologies: the case of environment. Data Science Journal, 17(0), 13.
12. Popović, M., & Ney, H. (2018). Towards language adaptive machine translation. Machine Translation, 32(3), 225-253.
13. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30, 5998-6008.

MATHEMATICAL MODELING OF MACHINE LEARNING ALGORITHMS

Shikhnazarova Guzal Alisherovna
PhD student, Tashkent University
of Information Technologies named
after Muhammad al- Khwarizmi.

Gulyamova Dilfuza Rakhmatullaevna
Senior Lecturer, Tashkent University
of Information Technologies named
after Muhammad al- Khwarizmi.

Samandarova Nilufar Rakhimovna
Lecturer, Chilanzar 1st vocational school.

Abstract: The article provides an overview of the main mathematical models in machine learning methods, such as linear regression, logistic regression, decision trees, support vector machine (SVM) and neural networks. The study highlights the importance of mathematical modeling for the development and application of machine learning algorithms and offers new perspectives in the field.

Keywords: machine learning, linear regression, logistic regression, decision trees, support vector machine (SVM), neural networks.

Machine learning algorithms play an important role in the development of artificial intelligence and are widely used in many fields. In this article we will look at the main methods of machine learning and the application of machine learning models in new areas [Abdurakhmonova, N.2021,2022].

Linear regression is one of the basic machine learning methods for regression analysis when you want to predict a continuous dependent variable based on one or more independent variables. The mathematical model of linear regression is a linear equation that determines the relationship between independent and dependent variables. In general, the linear regression equation can be written as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_r x_r + \varepsilon \qquad (1)$$

where y is the dependent variable, x_1, x_2, ..., x_r are independent variables, β_0, β_1, β_2, . .., β_r - model coefficients, ε - random error. The purpose of linear regression is to find the optimal coefficients β_0, β_1, β_2, . .., β_r, which minimize the sum of squared errors (the sum of squared differences between the predicted and actual values). To find coefficients, the least squares method (OLS) or its variations are often used.

Logistic regression is a classification algorithm used to solve binary classification problems (where data needs to be divided into two classes). It can also be generalized to a multi-class classification. The logistic regression mathematical model uses the logistic function (also known as the sigmoid) to predict the probability of membership in a particular class. For binary classification, the logistic regression equation can be written as:

$$P(y = 1|x) = 1/(1 + \exp(-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r))) \quad (2)$$

where P(y=1|x) is the probability of belonging to class 1 for given values of the independent variables, exp is the exponential function. The goal of logistic regression is to find the optimal coefficients $\beta_0, \beta_1, \beta_2, \ldots, \beta_r$, which maximize the likelihood of the data or minimize a loss function, such as a logarithmic loss function.

Support Vector Machines (SVM) Support Vector Machines (SVM) is a powerful machine learning algorithm used for classification and regression problems. It is based on finding an optimal hyperplane that maximally separates data from different classes in the feature space [Abdurakhmonova, N.2021,2022]. The SVM mathematical model defines a separating hyperplane as a linear combination of support vectors. For linearly separable data, the SVM mathematical model can be written as:

$$y(x) = w * x + b \quad (3)$$

where y(x) is the predicted class, w is the vector of weights, x is the vector of features, b is the bias (free term). In the case of nonlinear data, SVM can use kernel functions to project the data into higher dimensions where it becomes linearly separable.

Development of new methods - mathematical modeling stimulates the development of new methods and approaches in the field of machine learning. It allows us to explore new mathematical concepts, develop more efficient algorithms, and explore the principles behind deep learning, recurrent neural networks, generative models, and other advanced techniques. In general, mathematical modeling plays a central role in the development and application of machine learning algorithms. It allows us to understand, optimize, generalize and manage the complexity of algorithms, and stimulates the development of new methods. Thanks to mathematical models, we can extend the power of machine learning and apply it to various fields, from data science to medicine, finance, robotics and many more.

References:

1. Mathematical analysis: textbook / compiled by E. P. Yartseva. - Stavropol: NCFU, 2017. - 256 p. — Text: electronic // Lan: electronic library system. — URL: https://e.lanbook.com/book/155295 (access date: 06/30/2023). — Access mode: for authorization. users.
2. Abdullayeva, Z. S., & Shihnazarova , G. A. (2023, April). INTELLEKTUAL TAHLIL ASOSIDA CHET TILI BILIMINI TEKSHIRISHNI MODELLASHTIRISH. In INTERNATIONAL SCIENTIFIC CONFERENCES WITH HIGHER EDUCATIONAL INSTITUTIONS (Vol. 1, No. 14.04, pp. 150-152).
3. Shixnazarova , G. A., & Gulyamova , D. R. (2016). Modeling of the process of examination data mining techniques in the study of English language. Computer science: problems, methodology, technologies, 900.
4. Abdullayeva, Z. S., & Shikhnazarova , G. A. (2023). Priority factors of the nutritional model that influence the prenatal development of the fetus. In E3S Web of Conferences (Vol. 420, p. 05011). EDP Sciences.
5. Абдурахмонова, Н., & Абдувахобов, Г. (2021). О ‘QUV LUG ‘ATINI TUZISHNING NAZARIY METODOLOGIK ASOSLARI. МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА, 4(6).

6. Abdurakhmonova, N., Shakirovich, I. A., & O‘G‘Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. Science and Education, 3(9), 146-156.
7. Ismailov, A. S., Shamsiyeva, G., & Abdurakhmonova, N. (2021). Statistical machine translation proposal for Uzbek to English. Science and Education, 2(12), 212-219.
8. Абдурахмонова, Н., & Бойсариева, С. (2023). TABIIY TILNI QAYTA ISHLASHDA (NLP) OKKAZIONALIZMLARNING MORFEM TAHLILI. МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА, 6(3).
9. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
10. Shamshaddinovna, A.Z., Alisherovna, S.G., Karimullaevna, K.G., & Baxtiyorovich, A.S. (2023). RATIONAL NUTRITION OF PREGNANT WOMEN IN THE PERIOD OF PRENATAL DEVELOPMENT OF THE FETUS. Web of Medicine: Journal of Medicine, Practice and Nursing, 1(3), 1-6.
11. Abdufattokhov, S., Ibragimova, K., & Gulyamova, D. (2021, November). The applicability of machine learning algorithms in predictive modeling for sustainable energy management. In International Conference on Forthcoming Networks and Sustainability in the IoT Era (pp. 379-391). Cham: Springer International Publishing.
12. Gulyamova, D. (2022, April). The Applicability of Machine Learning Algorithms in Predictive Modeling for Sustainable Energy Management. In Forthcoming Networks and Sustainability in the IoT Era: Second International Conference, FoNeS-IoT 2021, Volume 1 (Vol. 129, p. 379). Springer Nature.

V SHO‘BA
KOMPYUTER LEKSIKOGRAFYASI: ONTOLOGIYA, TEZAURUS

KORPUSNING LINGVISTIK VA TA’LIMIY AHAMIYATI

Abduraxmonova Nilufar
O‘zbekiston Milliy universiteti
professori, filologiya fanlari doktori.
E-mail: n.abduraxmonova@nuu.uz

Askarova Umida
O‘zbekiston Milliy universiteti o‘qituvchisi.
E-mail: samina090822@gmail.com

Yulbarsov Ochilbek
O‘zbekiston Milliy universiteti o‘qituvchisi.

Annatatsiya: Ushbu maqolada xalqaro ta’limiy korpuslarning lingvistik instrumentariylari, lingvistik jihatdan aniqlangan xatolarni teglash tizimi, ta’limiy korpus loyihasi shakllantirilishning asosiy kategoriyalari, ta’limiy korpuslar toksonomiyasi kabi masalalar tahlilga tortilgan.
Kalit so‘zlar: TAGSELECT, ICEMARK, ICECUP, razmetkalash, kollokatsiyalar, subkorpus, transliteratsiya, token, korpuslar toksonomiyasi.

Annotation: In this article, the issue of linguistic instruments of international educational corpora, system of tags of linguistically identified errors, main categories in the formation of educational corpus project, taxonomy of educational corpora is highlighted.
Keywords: TAGSELECT, ICEMARK, ICECUP, sorting, collocations, subcorpus, transliteration, token, taxonomy of corpora.

Аннотация: В данной статье освещен вопрос лингвистического инструментария международных образовательных корпусов, системы меток лингвистически определяемых ошибок, основных категорий при формировании проекта учебного корпуса, а также вопрос таксономии учебных корпусов.
Ключевые слова: TAGSELECT, ICEMARK, ICECUP, сортировка, словосочетания, субкорпус, транслитерация, токен, таксономия корпусов.

Til va tilshunoslikni o‘rganilishida korpus manba sifatida foydalanish bo‘yicha omma qiziqishining ortib borishi ICAME (the international computer archive of modern and medieval English) a’zolari o‘rtasida bo‘lib o‘tgan muhokama natijasida 1994- yil TaLC yaratilishiga sabab bo‘ladi. Dastlabki ta’limiy korpusga oid konferensiya 1994- yilda Lankasterda bo‘lib, tadqiqotda kompyuter matn korpuslaridan foydalanish yaxshi yo‘lga qo‘yilgani va ular o‘qitish maqsadlarida keng foydalanib kelinyotganligini, korpus ma’lumotlaridan o‘quv materiallarini tayyorlashda

hamda xorijiy tillarni va lingvistik ta'limda talabalar tomonidan bevosita korpuslarning yaratilishiga doir muhokmalar kun tartibiga qo'yiladi.

Dastlabki ta'limiy korpuslarni yaratilishi Sylviane Granger va uning ilmiy jamoasi tomonidan "Ingliz tili o'rganuvchilari uchun xalqaro korpus" (ICLE – International corpus of learner English) yaratilgan. Ushbu korpusning maqsadi til o'rganuvchilarining umumiy yuzaga keluvchi xatolarini tasniflash orqali ularni tuzatish metodlarini yaratishga qaratilgan [D. Stewart. 2004:3].

Ushbu ta'limiy korpusni yaratishda ingliz tili o'rganuvchilarning quyidagi belgilari hisobga olingan [Sylviane Granger.1994:24]: ta'lim muhiti, yosh, ona tilisi, ta'lim bosqichi, vazifalarning tabiati.

Ushbu korpus uchun quyidagi kriteriyalari doirasida til o'rganuvchilar tanlab olingan:

1) o'rganuvchilar turi;
2) yoshi: yoshi kattalar (adults)
3) ta'lim bosqichi: yuqori daraja (advanced);
4) vazifasi: esselar

Ingliz tilining ta'limiy xalqaro korpusi ingliz tilining ikkinchi til sifatida franzus, nemis, xitoy, yapon, chex kabi millatga tegishli o'rganuchilar asos qilib olingan.

Ingliz tili xalqaro ta'limiy korpusi bir nechta lingvistik instrumentariylardan tarkib topgan.

- The ICE Markup Assistant – ushbu instrumentariy razmetkalash uchun standart belgilar to'plami asosida jarayonni soddalashtirilgan avtomat tizimi sanaladi;
- TAGSELECT – the Ice tag Selection system (teglash tizimi) korpusdagi har bir so'z uchun TOSCA teggeri yordamida generatsiya qilingan so'z guruhlari teglarini avtomatik saralash tizimi sanaladi;
- ICEMARK – TOSCA parseridan oldin teglangan matnni sintaktik razmetkalash uchun qo'llanadi. Shuningdek, boshqa funksiyalaridan biri so'z guruhlari teglarini o'zgartirish, matnda uchragan so'zlarni xatolari tahrirlash, teglanmagan gaplarni ko'rish kabi imkoniyatga ham egadir.
- ICECUP – korpus Utility dasturi ushbu instrumentariy korpusni tahlil qilish, konkordanslash va qidiruv kabu funksional jihatdan ahamiyatga ega.

XIX asrning oxiri va XX asrning boshlarida so'zning klaasik nazariyasi iki yirik leksik tadqiqotning yaratilishida katta rol o'ynadi. Roget tezaurus (1852) va Oksford ingliz lug'ati (Murray 1884-1928). Matnlarga qarab so'zlarning giponimi, sinonimi, antonimi va ularning ma'nolari leksik munosabatlarga ko'ra lug'atda beriladi. Har bir so'z bundan tashqari bir yoki bir nechta grammatik kategoriyalarga tegishli bo'ladi va ularning talaffuzi, etimologiyasi va tarixi beriladi. Ushbu nazariyaga ko'ra leksik birlik boshqa tizimlarga ham aloqador bo'lib, grammatika va fonologiya o'rtasida o'zaro munosabat yotadi. Sistem-funksional lingvistikaning markazida ham leksis turadi, zero bu modeldagi tizimlar leksik tanlov eng oxirgisi bo'lganidek ba'zan uni shu tarzda hosil qiladi [D. Stewart. 2004: 22]. XIX asrda yaratilgan Tezaurus lug'ati ta'sirida ko'pgina nazariy tadqiqotlar jonlana boshlaydi. XX asrda korpusga asoslangan o'quv lug'atlari Collins Cobuild, Oksford, Longman, Makmillan kabi lug'atlar yartailganligini guvohi bo'lamiz.

Ta'limiy korpuslar 1990-yillarda ommalashgan bo'lsa-da, biroq 1960-70-yillarda tillarni o'rganishda "xatoni tahlil qilish" (error analysis) tarzda metodologik yondashuv sifatida o'rganilgan.

Ma’lumki, so‘zlar muayyan darajada kombinatsiya hosil qilish uchun ma’lum turdagi matnlarda qo‘llanadi. Bu an’anaviy tilshunoslida so‘zlarning leksik ma’no qurshivini ko‘rsatishda matn, gap yoki kontekst jumlalari bilan qo‘llanganini ko‘ramiz. Endilikda so‘zni bir vaqtda turli maqsadlardan kelib chiqib tahlil qilish imkoniyatini yaratib beruvchi lingvistik instrumentariy sifatida korpusni ta’kidlash o‘rinli bo‘lar edi. So‘zning kollokatsiyalar bilan (so‘z birikma yoki turg‘un birikma) birgalikda uning tarkibida qo‘llanganda so‘zga doir u yoki bu ma’no anglashiladi. Demak ayrim so‘zlar uchun uning grammatik kategoriyaga xoslanishi gap tarkibida kelganda aniqlashadi: *o‘qish* (ot, harakat nomi yoki fe’lning buyruq shakli), *qil* (ot, fe’l), *yosh* (sifat, ot) kabilar.

Qachonki til foydalanuvchisi uchun biror so‘z semantik to‘plam yoki guruh bilan birgalikda kelganda semantik assotsiatsiyani aniqlash mumkin bo‘ladi [D. Stewart. 2004: 25].

Ta’limiy korpusga asoslangan til o‘qitish yoki o‘rganish jarayoni statistik yoki chastotaga asoslangaligi bilan ahamiyatlidir. Ayniqsa, u yoki bu so‘zning kontekstda birikma tarzda so‘z qurshovida kela olish imkoniyati so‘zlarning ko‘p ma’noli yoki omonim bo‘lgan holatda juda muhim ahamiyat kasb etadi.

Ta’limiy korpus til o‘rganuvchilar tomonidan tayyorlangan matnlarning tizimli kompyuterlashtirilgan majmuasi sifatida ta’rif keltiriladi [N.Nesselhauf. 2005:40]. Tizimli deganda, muayyan til o‘rganuvchilar guruhi sifatida tanlab olingan til bilish darajasiga ko‘ra jamlangan matnlardan tuzilgan korpus nazarda tutiladi. Ona tili korpusiga ixtisoslashgan korpusda matnning tabiiy yaratilganligi va tayyor olingan matnlar resurs bo‘lib hizmat qilishi bilan ajralib turadi. Ta’limiy korpusda xorijiy til yoki ikkinchi til sifatida o‘rganuvchilar uchun mo‘ljallangan bo‘lib, nazoratdagi til o‘rganuvchilar tomonidan yaratilgan chekli matnlardan tashkil topadi. Uar janr jihatdan esselar, og‘zaki intervyular bo‘lishi mumkin [D. Stewart. 2004: 40].

Rus tili ta’limiy korpusi rus tilida gapiruvchi turli millat vakillari tomonidan tayyorlangan materiallar asosida yaratilgan.

Korpusdagi matn va muallifga tegishli metama’lumotlar quyidagi ikki maydonga ajratilgan[7]:

I. Asosiy maydon:
- Og‘zaki / yozma matn
- Xorijiy / ona tili sifatida rus tili
- Muallifning dominant tili
- Rus tilini bilish darajasi

II. Ixtiyoriy maydon:
- Muallif jinsi
- Matn yaratilgan sanasi
- Janri

Ushbu korpus iqtisodiyot oliy maktab korpus texnologiyalari lingvistik laboratoriyasida xalqaro loyiha doirasida yaratilgan. Korpus 730000 tokendan iborat bo‘lib, matnlarning 56 foizi rus tili ikkinchi til sifatida so‘zlovchilar nutqidan, 44 foizi mahalliy (universitet va kollej talabalari) yuqori va o‘rta til bilish darajasiga ega rusiyzabon vakillar tomonidan olingan materiallarga asoslanilgan. Korpusning dastlabki versiyasi faqat Amerika ingliz tilida ona tilisi bo‘lgan loyihada cheklangan ishtirokchilar tomonidan yozib olingan. RULEC deb nomlangan subkorpus tarkibida akademik yozishmalardan iborat. Barcha ishtirokchilar ononim shaklda qatnashgan

[Abdurakhmonova, N.2019,2022]. Boshqa respondentlar ID kod bilan belgilangan [E.Rakhilina. 2016:66].

Lingvistik jihatdan aniqlangan xatolarni teglar tizimida ko‘rsatilgan:

Til sathi	TEG	Izohi
Imloviy xato	Graph	Lotin alifbosidan foydalanish
	Hyphen	Chiziqchani qo‘llash bilan xatolik
	Space	So‘zlar o‘rtasidagi bo‘shliqning tushib qolishi
	Translit	Atoqli otlarni transliteratsiya qilishdagi xatolik
	Ortho	Xato yozilgan harf
	Misspel	Bir tokendagi bir nechta o‘rinda yo‘l qo‘yilgan imloviy xato
Morfologik xato	Infl	Grammatik qo‘shimchlar xato qo‘llangan so‘z
	Deriv	Yasalgan so‘z
	Altern	So‘z o‘zagini qo‘llashdagi xato
	Num	So‘zdagi son kategoriyasini qo‘llash bilan bog‘liq xato
	Gender	Jins bilan bog‘liq xato
	Morhp	Boshqa morfologik xato
Sintaktik xato	AgrCase	Kelishik bilan bog‘liq xato
	AgrGender	Jinsning o‘zaro moslashish bilan bog‘liq xato
	AgrNum	Son bilan moslashishga doir xato
	AgrPers	Shaxs-son bilan bog‘liq xato
	Asp	Fe’l mayli bog‘liq xato
	Passive	Majhul nisbatni qo‘llash bilan bog‘liq xato
	Tense	Fe’l zamoni bilan bog‘liq xato
	Mode	Fe’lni qo‘llash bilan bog‘liq xato
	Refl	O‘zlik nisbatdagi fe’lni qo‘llash bilan bog‘liq xato
	Gov	Xato kelishik
	WO	So‘z tartibidagi xato
	Ref	O‘zlik olmoshi bilan bog‘liq xato
	Conj	Bog‘lovchini noto‘g‘ri qo‘llanishi
	Neg	Bo‘lishsizlikning xato qo‘llanishi
	Aux	Ko‘makchilarni xato qo‘llash
	Brev	Sifatlarning qisqa formasini qo‘llashdagi xato
	Syntax	Boshqa sintaktik xato
Tuzilish	Constr	Gap tuzilishfdagi xato
Leksik xato	Lex	Leksik xatolik
	CS	Kodlash bilan bog‘liq xato
	Par	Paronim bilan bog‘liq xato
	Idiom	Idiom bilan bog‘liq xato
Qo‘shimcha teglar	Del	So‘z yoki biror morfemada belgining tushib qolishi
	Insert	So‘z yoki biror morfemada belgining qo‘shilib qolishi
	Subst	So‘z yoki biror morfemada belgining almashib qolishi kelishi
	Transp	So‘z yoki biror morfemada belgining noo‘rin kelishi
	Transfer	Tildagi kelishikni transfer qilish
	Not-clear	Boshqa noma’lum fragmentlar

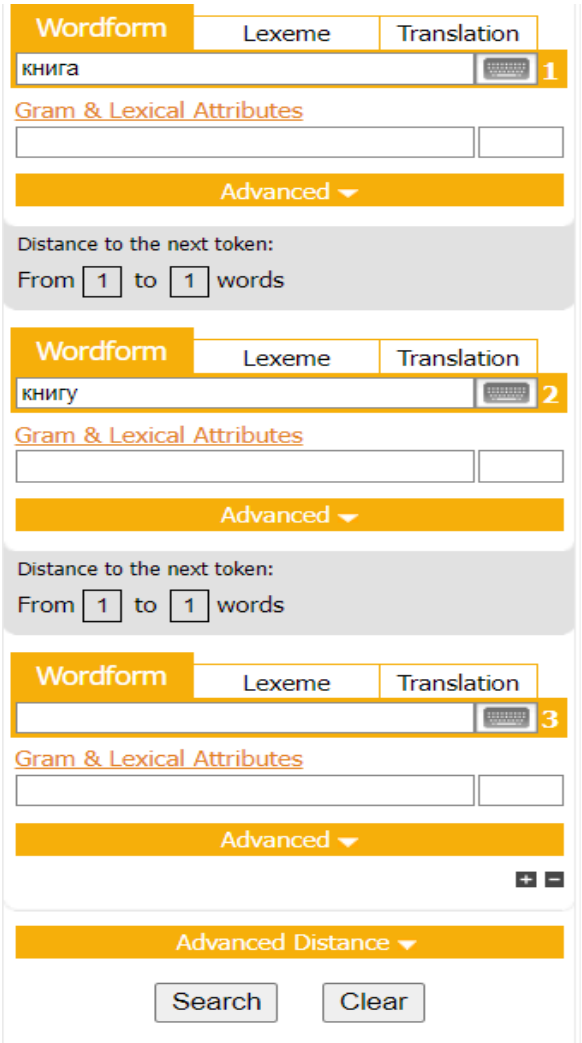
Rus tili ta’limiy korpusi 5 tilda gaplashuvchi (Amerika, franzus, koreys, qozoq va nemis tili) sodda, o‘rta va yuqori til bilish ko‘nikmasiga ega talabalardan olingan materiallardan tashkil topgan[Abdurakhmonova, N.2019,2022].

Y.Tononing fikricha, ta'limiy korpus loyihasi uch asosiy kategoriya asosida shakllantiriladi [Y.Tono.2003:800]: a) til bilan bog'liq kriteriya (janr, mavzu); b) mavzu bilan bog'liq kriteria (davomiy yoki yo'nalishlar kesimida; beixtiyoriy yoki tayyorlangan); c) o'rganuvchiga bog'liq kriteriya (ingliz tili xorijiy yoki ikkinchi til ekanligi (EFL/ESL), yosh, gender, ona tili, xorijiy tajriba).

Rus tili ta'limiy korpusi yuqoridagi me'yordan kelib chiqib, 8turdagi metama'lumotdan tarkib topgan 2 kategoriyaga biriktirilgan: muallifga tegishli hamda matnga tegishli. Muallifga tegishli belgilarga muallif IDsi, gender, til bilish darajasi, qaysi til vakili ekanligi. Muataxassislik darajasi CEFR (Common European Framework of Reference for Languages) va ACTFL (American Council on the Teaching of Foreign Languages) bo'yicha tanlanadi.

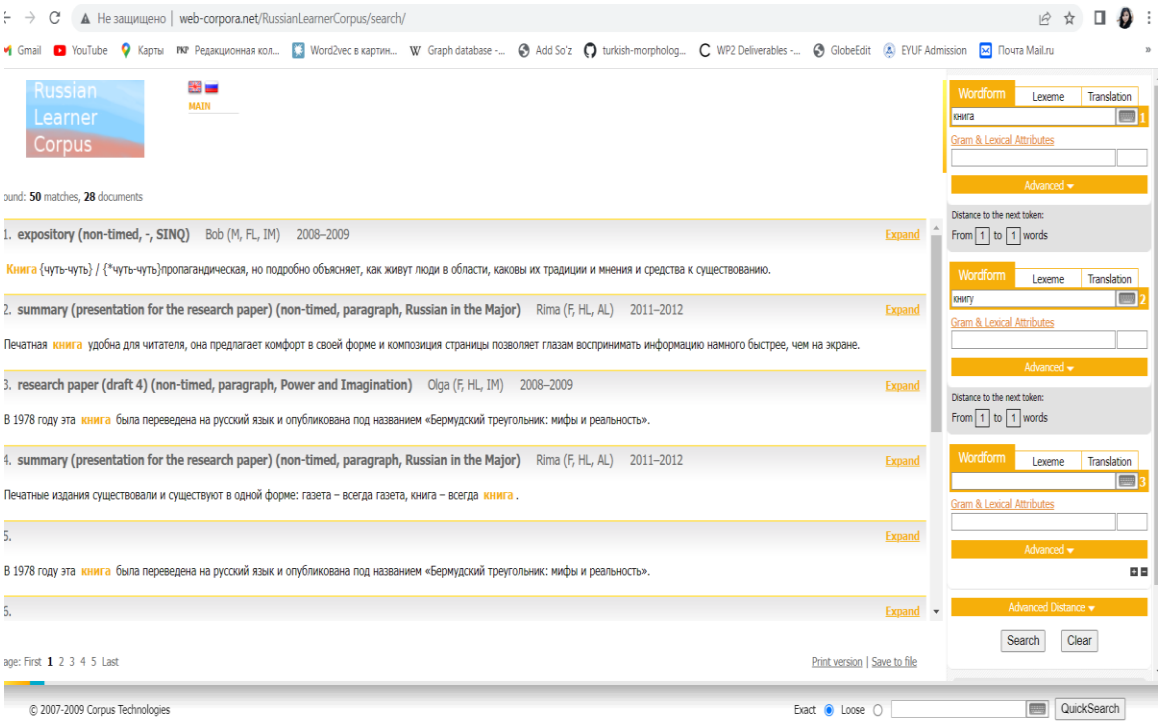
Y.Tononing fikriga ko'ra kamida ikki aspekt xatolarni belgilash annotatsiyalashdan iborat bo'lishi lozim. Birinchi guruh teglar lingvistik belgisiga ko'ra (qo'shimchalarni qo'shish, sintaktik moslashish) aniqlanaydi.

Rus tili ta'limiy korpusi annotatsiya instrumentariysi sifatida ochiq resursli JavaScript library Annotator.js ga asoslangan. Annotatsiya uch jihatga ega: 1-jihat asliyatdagi gapni ifodalash (annotator xatolarni belgilash), 2-jihat morfologik xatolar va imlolarni tuzatish orqali asliyatdagi matnni ko'rsatish, 3- ganing to'g'rilangan shaklini barcha sintaktik va leksik o'zgarishlar bilan borgalikda ko'rsatish.



2-rasm

Mazkur korpusning qidiruv tizimi quyidagi funksional imkoniyatlardan iborat: lemma, so‘z turkumlari va boshqa grammatik kategoriyalar (gender, son, nisbat, zamon).



3-rasm

Y.Tono ta’limiy korpuslar toksonomiyasi sifatida quyidagilarni keltiradi [Y.Tono. 2003:809]**:**

Proyekt	**Mavzusi/topshiriqlar** **Hajmi**	**Annotatsiya**
Yevropa:		
Xalqaro ingliz til ta’limiy korpusi **(ICLE** – international corpus of learner English)	- ingliz tili xorijiy til sifatida o‘qiydigan 3-4 kurs universitet talabalari -15 ta millat - Yozma esselar - 3 mln	-xato bo‘yicha teglash -POS (so‘z turkumi bo‘yicha teglash)
LINDSEI (og‘zaki ingliz tili Luven xalqaro ma’lumotlar bazasi)	-50 intervyu -ingliz tili xorijiy til sifatida o‘qiydigan 3-4 kurs universitet talabalari 100000so‘z	-orfografik
Longman ta’limiy korpusi **(LLC –** Longman Learners’ corpus)	-Barcha til darajalari -yozma esselar -10 mln	-POS- so‘z turkum bo‘yicha teglash -tijoriy maqsadlarga yo‘naltirilgan
Polyak-ingliz tili korpusi tadqiqi va ilovalari **PELCRA (**Polish-English Language Corpus Research and Applications**)**	-Barcha til darajalari -og‘zaki/yozma esselar Polyak -o‘rganuvchilar	POS bo‘yicha teglangan Mavjud emas
UAM korpusi	Talaba va o‘qituvchilarning tayyorlagan korpusi	

ISLE korpusi (ingliz tili xorijiy til bo‘lganlar uchun)	-20 minutlik nutq -o‘rta darajada ingliz tili darajasiga ega nemis va intalyan o‘rganuvchilar	-orfografik -urg‘u -ELRA orqali mavjud
JPU (Janus Pannonius University) korpusi	-universitet til o‘rganuvchilar Yozma materiallar	-oddiy matn
Kembridj ta’limiy korpusi **CLC (**Cambridge Learners Corpus**)**	-Barcha til daraja -10 mln.	-POS bo‘yicha teglangan -Xatolar bo‘yicha teglangan
IBLC (Indianapolis Business Learner Corpus**)**	-Biznes yozishmalar -oddiy matn	-oddiy matn -mavjud emas
OSIYO:		
JEFLL korpusi (Yaponiya)	-barcha darajalar uchun -og‘zaki va yozma nutq -1 mln.	-POS bo‘yicha teglangan -Xatolar bo‘yicha teglangan
Yaponiya o‘rganuvchilar uchun ingliz tili korpusi	-barcha darajalar uchun - yozma nutq -1 mln. Tarjimat	-oddiy matn -web orqali foydalanish mumkin
TAO /SST (Standart speaking test)	-barcha darajalar uchun - yozma nutq -1 mln. -15 min. intervyu	-Xatolar bo‘yicha teglangan (qisman)
TELEC talabalar korpusi	-Hong Kong o‘rganuvchilari -Universitet imtihonlaridagi yozuvlar -3 mln.	Oddiy matn
Poly U korpus	-universitet bitiruvchilari -Tezislar -282.000	Oddiy matn
NTOU korpus	-Ingliz tili xorijiy til sifatida o‘rganuvchilar -53.000	Oddiy matn
Ingliz tilini o‘rganuvchi yapon zabon vakillarining parallel korpusi	-ingliz va yapon tilidagi tarjimalar	Ma’lumotlar bazasi formatida
HKUST korpusi	Ingliz tili o‘rganuvchi xitoylik talabalar 10 mln. Yozma esselar va imtihon topshiriqlari	POS bo‘yicha teglangan

Foydalanilgan adabiyotlar:

1. D. Stewart, S. Bernardini, G. Aston Introduction: Ten years of TaLC / Corpora and language learners, Vol 17, John Benjamins Publishing Company Amsterdam/Philadelphia, 2004, - P. 3.
2. Sylviane Granger The learner corpus: a revolution in applied linguistics / English Today 39, Vol.10, No.3 (July 1994), Cambridge university press. – P. 26.
3. D. Stewart, S. Bernardini, G. Aston Introduction: Ten years of TaLC / Corpora and language learners, Vol 17, John Benjamins Publishing Company Amsterdam/Philadelphia, 2004, - P. 22.
4. Nadja Nesselhauf Collocations in a lerner corpus John Benjamins Publishing company, Amsterdam, 2005. – P. 40.

5. Abdurakhmonova, N., & Urdishev, K. (2019). Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL), 6(1-2019), 131-7.
6. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
7. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
8. Abduraxmonova, N. Z. Q., & Urazaliyeva, M. Y. (2022). O ‘ZBEK TILI ELEKTRON KORPUSIDA (http://uzbekcorpus. uz/) OG ‘ZAKI MATNLAR KORPUSINI YARATISHNING NAZARIY VA AMALIY MASALALARI. Academic research in educational sciences, 3(3), 644-650.
9. Ekaterina Rakhilina, Anastasia Vyrenkova, Elmira Mustakimova, Alina Ladygina, and Ivan Smirnov. 2016. Building a learner corpus for Russian. In Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition, pages 66–75, Umeå, Sweden. LiU Electronic Press.
10. Tono Y. Learner corpora: design, development and applications. In Proceedings of the 2003 Corpus Linguistics Conference, 2003. - P. 800–809.
11. http://web-corpora.net/RLC

O‘ZBEK TILI TEZAURUSI UCHUN SIFAT SO‘Z TURKUMI LINGVISTIK BAZASINI YARATISH

Po‘latova Gulhayo
Namangan davlat universiteti mustaqil tadqiqotchisi.

Annotatsiya: Jahonda inson faoliyatining bilim talab qiladigan zamonaviy sohalari rivojlanishi kompyuter texnologiyalari rolining oshishi bilan belgilanadi. Bugungi kunda axborot oqimi sezilarli darajada ko‘paymoqda, endi ularni saqlash, taqdim etish, formallashtirish va tartibga solish, shuningdek, avtomatik qayta ishlashning yangi usullarini izlash zarurati yuzaga kelmoqda. Shu bois turli amaliy maqsadlarda qo‘llanilishi mumkin bo‘lgan keng qamrovli bilim bazalariga qiziqish ortib bormoqda. Ayniqsa, inson omilisiz matndan har qanday ma’lumotni chiqarib oluvchi neyroto‘rlarga asoslangan tizimlarga ehtiyoj katta. Ushbu maqolada tezauruslar, sifat so‘z turkumi lingvistik bazasini yaratish bo‘yicha taklif va tavsiyalar berilgan.
Kalit so‘zlar: tezaurus, lingvistik baza, korpus, lug‘at, sifat so‘z turkumi, “oq” leksemasi.

Abstract: The development of modern fields of human activity that require knowledge in the world is determined by the increasing role of computer technologies. Today, the flow of information is increasing significantly, now there is a need to search for new methods of storing, presenting, formalizing and organizing them, as well as automatic processing. Therefore, there is a growing interest in comprehensive knowledge bases that can be used for various practical purposes. In particular, there is a great need for systems based on neural networks that extract any information from the text without the human factor. This article provides suggestions and recommendations for the creation of thesauruses and the linguistic base of adjective vocabulary.
Keywords: thesaurus, linguistic base, corpus, dictionary, adjective phrase, "oq" lexeme.

Аннотация: Развитие современных областей человеческой деятельности, требующих знаний, в мире определяется возрастающей ролью компьютерных технологий. Сегодня потоки информации значительно увеличиваются, теперь возникает необходимость поиска новых методов их хранения, представления, формализации и организации, а также автоматической обработки. Поэтому растет интерес к всеобъемлющим базам знаний, которые можно использовать для различных практических целей. В частности, существует большая потребность в системах на основе нейронных сетей, извлекающих из текста любую информацию без человеческого фактора. В данной статье даны предложения и рекомендации по созданию тезаурусов и лингвистической базы прилагательной лексики.
Ключевые слова: тезаурус, лингвистическая база, корпус, словарь, прилагательное словосочетание, «oq» лексема.

Kirish

Texnologiya asrida taraqqiyot sur’atining tezlashgani, intellektual salohiyat, texnikaning yuqori darajada yuksalishi kishilik jamiyati oldiga qator vazifalarni qo‘ymoqda. Mintaqamizdagi globallashuv jarayoni barcha sohalarda tezkorlikni, jadal taraqqiyotni talab etmoqda. Kompyuter tizimi qulayliklar, imkoniyatlar majmuyiga va tarkibiy qismiga aylandi [Пулатов А.К., Мухамедова С., 2008: 54].

Kishilik taraqqiyotining har bir davrida fan-texnika yangiliklari muhim ahamiyat kasb etib kelgan. Insoniyat uchinchi ming yillikka, axborot asriga qadam qo‘ydi. Inson uchun ma’lumot olish zaruriyatga aylandi. Axborot – ma’lumot oltindan ham qimmatroq deb topildi. Axborotni topish, saqlash, qayta ishlash va boshqalarga yetkazishning qulay usullariga bo‘lgan ehtiyoj kun sayin ortib bormoqda. Bu esa XX asrning buyuk kashfiyoti bo‘lgan kompyuter va kompyuter texnologiyalari sohasi uchun muhim vazifalarni belgilab berdi. Kompyuter texnologiyalari xalq xo‘jaligi, sport, san’at, tibbiyot, umuman, ijtimoiy hayotning barcha tarmoqlariga kirib keldi. Fan yo‘nalishlarini kompyuterlashtirish ilmning asosiy shartlaridan bo‘lgan obyektivlik va aniqlikka amal qilish asosida tez rivojlana boshladi va barcha sohalarda yutuqlarga erishildi [Xolmanova Z., 2020: 9]. Xususan, kompyuter lingvistikasi sohasida ham qator ilmiy izlanishlar olib borildi.

Asosiy qism

O‘zbek tilshunoslari tomonidan o‘zbek tili milliy korpusining yaratilganligi, til ixlosmandlari uchun yangi imkoniyatlar eshigini ochdi.

Korpus (korpus) lotincha “tana” degan ma’noni bildiradi. “Korpus so‘z, so‘z birikmasi, grammatik shakllarni, so‘z ma’nosini muayyan qidiruv tizimi orqali topishni anglatuvchi elektron ko‘rinishdagi matnlar jamlanmasidir” [http://rusorpora.ru]. Korpus tushunchasi bilan yonma-yon “matnlar korpusi” atamasi ishlatilmoqda. Matnlar korpusi elektron holda saqlanadigan fonema, grafema, morfemalar, leksema, gap va matnlardan tashkil topishi mumkin bo‘lgan yaxlit butunlikdir. Korpuslar aslida ma’lumotlar bazasi sifatida shakllantiriladigan, tilshunoslik masalalarini hal etish maqsadida va turli yo‘nalishdagi tadqiqotlarni amalga oshirish uchun material sifatida xizmat qiladigan jamlanmadir.

Jahon tilshunosligida korpusga doir ilk ma’lumotlar XX asrning 40-yillarida qayd etilgan. Korpuslar tarixi haqida so‘z borar ekan, birinchi navbatda, 1961-1964 yillarda yaratilgan Braun korpusi tilga olinadi. Bu korpus Braun universitetida yaratilgan, har biri 2000 so‘zli 500 ta matn fragmentini o‘z ichiga oladi. O‘zbek tilshunosligida A.Po‘latov kompyuter lingvistikasi, tabiiy tilni qayta ishlash, statistik tahlil masalalariga doir izlanishlarda korpus lingvistikasiga ham to‘xtalib o‘tgan. Kompyuter lingvistikasi yo‘nalishlari monografik tadqiqot obyekti sifatida o‘rganila boshlandi. Olimalarimiz Nilufar Abdurahmonova, Manzura Abjalova va boshqalar bu sohada tadqiqotlar olib bormoqdalar. Kompyuter lingvistikasining yo‘nalishi sifatida shakllangan, hozirda o‘z taraqqiyot yo‘nalishiga ega bo‘lgan alohida soha hisoblanuvchi korpus lingvistikasi masalalarini atroflicha o‘rgangan o‘zbek tilshunosi Shaxlo Hamroyeva monografik planda tadqiq etdi. Tezauruslarning ma’lumotlar bazasi sifatidagi qiymatlari yoritilgan tadqiqotlar ham korpuslar haqida muayyan tasavvur berishga xizmat qiladi. Tezauruslarning tuzilishi, ishlash tamoyillari, kompyuter bazasi sifatidagi imkoniyatlari, WordNet tezaurus bazasi haqidagi ma’lumotlar ham ilmiy-amaliy ahamiyatga ega.

Katta hajmdagi ma’lumotlarni izlash va qayta ishlashni nazarda tutuvchi o‘quv – axborot materiallarini yaratish asosi o‘qitishning zamonaviy axborot usullarini rivojlantirish va takomillashtirish bilan bog‘liq. Ushbu usullarni amalga oshirishning muhim va zarur sharti innovatsion loyihalar va ta’lim axborot qidiruv dasturlarini tayyorlash va yaratishdir. Ushbu ilmiy yo‘nalishdagi o‘qitish uslublarining keng miqyosda rivojlanishi innovatsion jarayonlarni o‘rganishga va yangi avtomatlashtirilgan axborot-qidiruv tizimlarini yaratishga olib keladi.

Har qanday tilning ma’lumotlar bazasini yaratishda mazkur tilning to‘liq lug‘atini yaratish muhim rol o‘ynaydi. Hozirgi zamon talablariga ko‘ra, samaradorligi katta bo‘lgan lug‘at

kompyuter lug‘atidir. Kompyuter lug‘ati, an’anaviy lug‘atdan farqli o‘laroq, quyidagi qismlardan iborat bo‘ladi:

1.An’anaviy lug‘atlarga kiritiladigan leksik ma’noga ega, o‘zbek tilida keng iste’molda bo‘lgan so‘z va frazeologik iboralar. Lug‘atga barcha tub so‘zlar (o‘tsimon, kompyuter, yaxshi va h.k.), yasama so‘zlar (badavlat, chiroyli, ishchi va h.k.), qo‘shma so‘zlar (gultojixo‘roz, oshqozon, kungabotar va h.k.), murakkab qisqartma so‘zlar (aeroport, mikroskop va h.k.) kiritiladi.

2.Barcha dialekt (sheva)ga xos so‘zlar.

3.Ayrim sohalarda ishlatiladigan atamalar.

4.Vulgar so‘z va iboralar.

5.Kishi nomlari.

6.Geografik nomlar.

7.Qisqartma so‘zlar (abbreviaturalar) [Po‘latov A., 2011: 213].

Lingvistik bilimlar bazasi tabiiy tildagi matlarni analiz va sintez qiluvchi dasturiy majmuadir. Uning uch asosiy komponenti mavjud:

– lingvistik (tabiiy tilning formal modeli, lug‘at, grammatika, lingvistik jadvallar, qoidalar);

– matematik-algoritmik (formal til translyatori, matnni qayta ishlovchi algoritmlar);

– dasturiy ta’minot.

Lingvistik bilimlar bazasi mashina uchun matnlari qayta ishlashda zarur bo‘lgan barcha ma'lumotlarni o‘z ichiga qamrab oladi. Tabiiy tilining lingvistik sathlari quyidagilardan iborat: leksik, morfologik, sintaktik, semantik. Bularning har biri o‘zining struktur ma’lumotlar bazasiga ega. Har bir lingvistik birlik muayyan umumiy belgilariga ko‘ra paradigmalarga birlashadi. Lingvistik ma’lumotlar bazasi (ta’minoti) deyilganda, tilga oid barcha ma’lumotlar tushuniladi. Til tuzilishi hamda matnni analiz va sintez qilishning aniq vazifalariga, asosan, lingvistik ta’minotida quyidagilarning mavjud bo‘lishi talab qilinadi: matn – leksik asos – leksemashakl – so‘z birikmasi – sintaktik gap strukturalari[Abdurakhmonova, N.2020,2021,2022].

Tarjimadan avvalgi tahlil jarayonida uch asosiy ierarxik bosqich amalga oshiriladi: matn – jumla – leksema shakl. Bunda leksema shakl keng ma’noda matndan olingan istalgan segment birlik sifatida qaraladi. Gap va so‘z birikmalarining barcha strukturaviy hamda boshga xususiyatlari leksema shakl tushunchasida ifodalanadi.

Lingvistik bilimlar bazasiga quyidagilar kiradi:

Leksik resurslar:

– kompyuter lug‘atlari;

– tezaurus, ontologiya.

Matniy resurslar:

– matnlarning sohaviy majmuasi;

– matn korpusi [Abduraxmonova, 2021: 33].

Quyida lingvistik bilimlar bazasiga kiruvchi tezaurus haqida kichik bir tadqiqot o‘tkazamiz. Tezaurus (yun. “xazina”) muayyan so‘zning leksik-semantik, kontekstual ma’nolarini qamragan lug‘at hisoblanadi, umuman olganda, maxsus terminologiyadir. [Лукашевич, 2010: 89]. Zamonaviy ilmiy paradigmada tezaurusni bilimlar tizimi sifatida tushunishda bir nechta yondashuvlar mavjud. Jahon tilshunosligida tezaurus termini ostida ideografik lug‘at turi tushuniladi, dunyo haqidagi bilimlarning kognitiv tizimi (bu ma’noda tezaurus olam tasviri va mental leksika terminlariga muvofiq) hisoblanadi, keyingi yillarda esa xorij kompyuter

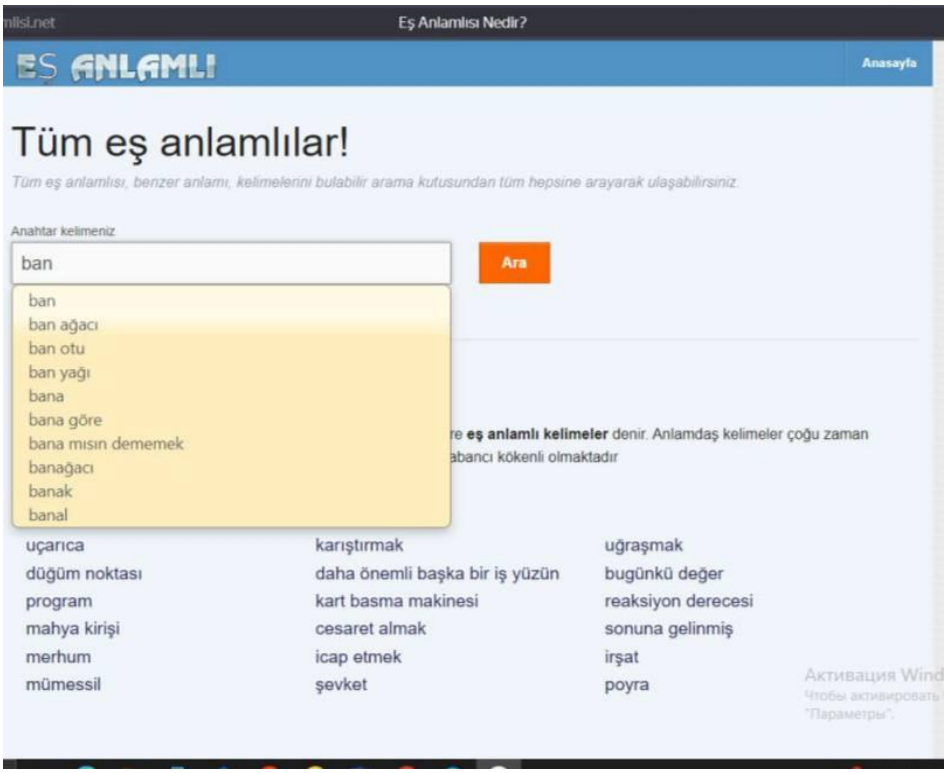
lingvistikasida axborot qidiruv tezauruslari va lingvistik ontologiyalar rivojlanib bormoqda, shuningdek, sun'iy intellekt tizimlarini ishlab chiqishda faol qo'llanilmoqda [Abjalova M., 2022: 117].

Tezaurus, WordNet va boshqa ingliz tili tezauruslari mualliflari ta'kidlaganlaridek, tabiiy til leksikasi yordamida ifodalanadigan bilimlar tizimi, tildagi so'zlarning muayyan ma'nosi bilan cheklanib qolmagan mantiqiy kategoriyalar yordamida boyitilgan ontologiya hisoblanishidan qat'iy nazar tezaurus axborot va ma'lumotlarni taqdim etuvchi semantik (kognitiv) tizim sifatida o'rganiladi. Bunda semantik tizim elementlaridan konsept, tushuncha va "madaniy konstantalar"ga e'tibor qaratiladi. Ontologiyada ushbu elementlardan tashqari semantik munosabatlar birlamchi o'rinda turadi[Abdurakhmonova, N.2020,2021,2022].

Tezauruslar matnlar to'plami yoxud til korpuslaridagi tushunchalarni, maxsus bilim sohasi yoki faoliyat sohasining tushunchalari, ta'riflari va terminlarini qamrab oluvchi leksikografik manbalar hisoblanadi.

Tezaurus – yaqin ma'noga ega so'zlar va so'z birikmalari tushunchalar, konseptlar yoki deskreptor (tavsiflovchi)lar deb ataladigan birliklarga birlashtirilgan hamda ushbu tushunchalar o'rtasida ularning iyerarxik shaklidagi semantik munosabatlari ko'rsatilgan lug'at hisoblanadi [Abjalova M., 2022: 195].

O'zbek tilshunosi Manzura Abjalova "O'zbek tezaurus lug'ati uchun sifat turkumi bazasini yaratish" bo'yicha kichik tadqiqot davomida https://live.babelnet.org, https://www.dictionary.com, https://www.thesaurus.com saytlarini farqli va o'xshash tomonlarini o'rgangan. Ushbu saytlarga qo'shimcha sifatida turk tilshunoslari tomonidan yaratilgan https://esanlamlisi.net saytini keltirishimiz mumkin. Bu saytning afzalligi turk tilidagi ayrim so'zlarning omonimlari kiritilgan bo'lib, lug'atdan har qanday inson onlayn foydalanish imkoniyatiga ega.



O‘zbek tili lingvistik bazasini yaratishda yuqorida tanishib chiqqan bir nechta bazalarni umumlashtirib, ularning har biridan foydalangan holda ish olib kerak. Sifat so‘z turkumi misolida “oq” leksemasining tezaurusi uchun baza tuzib olishimiz kerak.

Masalan:

“Oq” sifatining tezaurus lug‘atini yaratib ko‘ramiz:

Turkumi: Sifat;

Sinonimik qatori: bo‘z, oqish, tiniq

Antonimi: qora;

Omonimi: oqmoq – fe’l;

Paronimi: yo‘q;

Yasalishiga ko‘ra: sodda sifat;

Turiga ko‘ra: Asliy sifat;

Bog‘lanadigan so‘zlari: Ot so‘z turkumidagi barcha so‘zlar;

Qaysi tildan kirib kelgan: qadimgi turkiy tlda ham shunday ma’noni anglatgan bu sifat **a:q** tarzida talaffuz qilingan.

Misollar: Oq it, qora it – baribir it. *(Maqol)*. Xotin o‘tirmoqchi bo‘lgan edi, paranji ichidan uning atlas ko‘ylaklari va nafis oq qo‘llari ko‘rinib ketdi. *(Abdulla Qodiriy “O‘tgan kunlar”)*. Chapay askarlari keldilar bosib, Bosib keldi ular oqlar ustiga. *(A.Oripov “Yillar armoni”).*

Shu shaklda o‘zbek tilida mavjuda bo‘lgan sifatlarning har biriga to‘xtalib o‘tamiz. Ilmiy ishda asosiy manba sifatida “Izohli lug‘at” va https://uzbekcorpus.uz/ dan foydalansak, qo‘shimcha manbalar sifatida badiiy va ilmiy asarlardan foylaniladi.

Xulosa

O‘zbek tilida 50 000 dan ortiq leksemalar mavjud bo‘lib, korpus va lingvistik kompyuter dasturlari bazasida har bir leksema turkumining aniqlab berilishi o‘ta muhim masala hisoblanadi [Abjalova M., 2022: 5]. Korpus lingvistikasida so‘z turkumlarini teglash, grammatik kategoriyalarni teglash va so‘zlarni toifalashda noaniqliklarni bartaraf etish uchun so‘zni faqat uning lug‘atdagi shakliga asoslanib emas, balki matn (jumla)dagi ifodasi bo‘yicha uning turkumlik tegi va jumla (xatboshi, ibora)da boshqa so‘zlar bilan birikish imkoniyatini hisobga olish muhim sanaladi.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. Kompyuter lingvistikasi. Darslik Toshkent –2021. – B.33.
2. Abjalova M. O‘zbek tili ontologiyasi: yaratish texnologiyasi va konsepsiyasi Monografiya (qayta nashr) “Nodirabegim” nashriyoti Toshkent – 2022. – B.203.
3. Abjalova M., Sharipov E. O‘zbek tezaurus lug‘ati uchun sifat turkumi bazasini yaratish masalasi. “Kompyuter lingvistikasi: muammolar, yechim, istiqbollar” Respublika ilmiy-texnikaviy konferensiya. Vol. 1 №. 01 (2021). –B.189.
4. Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – М., 2010. – С.89.
5. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
6. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language

Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.

7. Abdurakhmonova, N., & Ismailov, A. S. (2022). APPLYING WEB CRAWLER TECHNOLOGIES FOR COMPILING PARALLEL CORPORA AS ONE STAGE OF NATURAL LANGUAGE PROCESSING. In СОВРЕМЕННАЯ ФИЛОЛОГИЯ. СОЦИАЛЬНАЯ И НАЦИОНАЛЬНАЯ ВАРИАТИВНОСТЬ ЯЗЫКА И ЛИТЕРАТУРЫ (pp. 22-27).
8. Abduraxmonova, N., & Abduvaxobov, G. I. (2021). O ‘quv lug ‘atini tuzishning nazariy metodologik asoslari. СЎЗ САНЪАТИ ХАЛҚАРО ЖУРНАЛИ, 103.
9. Sulevmanov, D., Gatiatullin, A., Prokopyev, N., & Abdurakhmonova, N. (2020, November). Turkic morpheme web portal as a platform for turkology research. In 2020 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-5). IEEE.
10. Po‘latov A. Kompyuter lingvistikasi – T.: Akademnashr, 2011-yil. – B.213.
11. Пулатов А.К., Мухамедова С. Компьютер лингвистикаси (укув кулланма). – Т., 2008. – B.56.
12. Xolmanova Z. Kompyuter lingvistikasi. Toshkent – 2020 “Asian Book House” – B.9.

TEZAURUS TUSHUNCHASI VA ILMIY YONDASHUV

Ubaydullayeva Ziroatxon
Mirzo Ulug‘bek nomidagi O‘zbekiston
Milliy universiteti magistranti.
E-mail: ubaydullayevaziroatxon9@gmail.com

Annotatsiya: Ushbu maqolada “tezaurus” tushunchasi, uning kelib chiqishi va o‘rganishga bo‘lgan turlicha qarashlar ilmiy yondashuvlar asosida ochib berilgan. Tezaurusning hayotning har bir jabhasida, har bir sohada qanchalik muhim ahamiyat kasb etishi to‘g‘risida fikrlar bayon etilgan.

Kalit so‘zlar: tezaurus, indekslash, iyerarxik, adekvat, assotsiativ, sinonimik, anaforistik, elliptik, konvolyutsiya.

Abstract: In this article, the concept of “thesaurus”, its origin and different views on learning are revealed based on scientific approaches. Opinions are expressed about how important the thesaurus is in every aspect of life, in every field.

Keywords: thesaurus, indexing, hierarchical, adequate, associative, synonymous, anaphoristic, elliptical, convolution.

Аннотация: В данной статье раскрываются понятие «тезаурус», его происхождение и различные взгляды на обучение, основанные на научных подходах. Высказываются мнения о том, насколько важен тезаурус во всех аспектах жизни, в каждой области.

Ключевые слова: тезаурус, индексация, иерархическая, адекватная, ассоциативная, синонимическая, анафористическая, эллиптическая, свертка.

Tezaurus tarixi haqidagi dastlabki ma’lumot Alan Gilchristning 1971-yilda chop etilgan “Tezaurus in Retrieval” asarida uchraydi hamda muallif bu so‘zni yunon tilidan olinganligini ta’kidlaydi. Shorter Oksford Dictionary lug‘atiga ko‘ra, ingliz tilida birinchi marta 1736-yilda paydo bo‘lgan va “lug‘at sifatidagi bilimlar xazinasi yoki ombori” ma’nosini bildirgan. Eng mashhur tezaurus 1852-yilda Peter Mark Roget tomonidan ishlab chiqilib, “Thesaurus of English Words and Phrases” deb nomlangan. Bu alifbo tartibida emas, balki tizimli ravishda, so‘zlar ifodalaydigan tushunchalarga ko‘ra va fikrga eng mos keladigan so‘z yoki so‘zlarni topishga qaratilgan. Taxminan bir asr o‘tgach, bu atama ma’lumot qidirishda ishlatiladigan lug‘at ro‘yxatlariga nisbatan qo‘llanila boshlandi[Abdurakhmonova, N.2020,2021,2022].

Ma’lumot qidirish ma’nosida tezaurus hozirgi xalqaro standart ISO 2788da “boshqariladigan indekslash tilining lug‘ati” bo‘lib, unda tushunchalar o‘rtasidagi munosabatlar aniq ifodalab, indekslovchi ishlatadigan lug‘atni qidiruvchi tiliga moslashtirish va shu bilan tegishli hujjatlarni qidirishni yaxshilaydi. Dastlab, tezaurus kordinatalarini indekslash va kordinatalarini indekslashdan foydalanib qidirishda qo‘llash, shuningdek, qo‘lda postkordinatalar tizimi yordamida indekslash Batten, Calvin Mooers va boshqalar tomonidan ishlab chiqilgan. Postkordinatalar tizimida qidiruvchi aniq bir so‘zni topish uchun so‘zlarning bir qancha kombinatsiyalaridan foydalanishi mumkin [Abdurakhmonova, N.2020,2021,2022].

Shuningdek, tezaurus 1950-yillar davomida turli shakllarda, shu jumladan, Taube tomonidan ishlab chiqilgan Uniterm tizimida rivojlandi. Bunda hujjat matnidan erkin saylash yo‘li bilan, bir marta bo‘lsa ham ishlatilgan so‘zlar olingan. Albatta, sinonimlar, omonimlar, turkumlarni qidirish va boshqa muammolarni hal qilish uchun faqat bitta so‘zli atamalar mavjud bo‘lgan vaziyatda qiyinchiliklar yuzaga keldi. Unitermlar ko‘pincha kontekstga, qo‘llanilishiga qarab turli xil ma’nolarda ishlatiladi, biroq filtr yo‘qligi sababli ular bu o‘ziga xos farqlarni ifoda eta olmaydi. Misol uchun, izolyatsiya termini orqali ifodalangan “issiqlik izolyatsiyasi” tushunchasi “elektr izolyatsiyasi” dan juda katta farq qiladi va bir so‘zdan iborat “izolyatsiya” atamasi ularni ajratish uchun yetarli bo’lmaydi. Unifikatsiyalashgan terminlar yetarli emasligi sababli, sof shakldagi tizim uzoq davom etmaydi. Oxir-oqibat, uning o‘rnini ko‘p sonli “assotsiatsiyalangan” yoki qo‘shma terminlarni o‘z ichiga olgan lug‘atlar egalladi. Bu ilk tezauruslar sinonimlar va omograflarni nazorat qila boshladi, atamalar o‘rtasidagi iyerarxik va assotsiativ munosabatlarni ko‘rsata boshladi.[J.Aitchison, SD.Clarke. 2004:]

Shu bilan birga, Adam Kilgarriff va Colin Yallop ham o‘zining maqolasida [Adam Kilgarriff, Colin Yallop. 2000:5] tezaurusning to‘rt turini tafsiflab, ular uchun umumiy bo‘lgan xususiyat ularning iyerarxik yoki tarmoq tashkil etilishini asoslab berishgan.

O‘zbekiston Milliy Ensiklopediyasida [O‘zbekiston Milliy Ensiklopediyasi, 2004: 329] tezaurus terminiga quyidagicha izoh berilgan: tezaurus (yun. thesaurus – xazina, boylik) – 1) muayyan tildagi barcha so‘zlarni qamrab oladigan, ularning matnda qo‘llanish holatlarini to‘la-to‘kis aks ettiradigan lug‘atdir.

Avvalambor, rezaurus termini tilning lug‘at fondini maksimal darajada to‘liqlik bilan ifodalaydigan xazina sifatida talqin etiladi. Matnlarning adekvat talqini va ma’lumotlari uning sistem ma’nolarini o‘zida mujassam etgan va soha vakillari tomonidan e’tirof etilgan tezaurusda aniqlanadi. [Valitova, 2012: 19]. Tezaurus – bu umumiy ma’noda maxsus bilim sohasi yoki faoliyat sohasining tushunchalari, ta’riflari, atama yoki terminlarini to‘liq qamrab oladigan maxsus terminologiya, yanada qat’iy va obyektiv ravishda lug‘at, ma’lumotlar to‘plami, korpus yoki kod demakdir.

Jumladan, tezaurus atamasini L.V. Shcherba “Опыт теории общей лексикографии” maqolasida [Л.В. Щерба,1940:10] tezaurus deyishganda, hozirgi paytda ko‘pincha “Thesaurus linguae latinae”ni nazarda tutishini, aslida u beshta nemis akademiyasining hamkorlik faoliyati bo‘lib, 1900-yilda boshlanganini aytib, muhim xarakterli xususiyati xususida ma’lum bir tilda ro‘y bergan barcha muhim so‘zlarni o‘z ichiga olib, har bir so‘z ostida, ushbu tilda mavjud bo‘lgan matnlardan misollar keltiriladi. Olimlar A.N. Baranov va D.O. Dobrovolskiy [Добровольский, Баранова.1965: 4] tezaurusga quyidagi ta’rifni bergan: “Maxsus tur lingvistik materialni tartibga solish usuli bilan boshqalardan (xususan, tushuntirish, ikki tilli va hokazo) farq qiladigan lug‘atdir. Tezaurusda til birliklari oddiy lug‘atda bo‘lgani kabi alifbo tartibida berilmagan, lekin ularning ma’nosiga qarab guruhlangan”. Shuningdek, axborot-qidiruv tezauruslarini qo‘llash sohasida ikkita asosiy yo‘nalish ajratiladi: matnning semantik ma’nosidan foydalanadigan ma’lum qoidalardan foydalangan holda hujjatlarni indekslash va foydalanuvchilarning qidiruv so‘rovlarini qayta ishlashda ierarxik, assotsiativ va sinonimik) foydalanishdir. Bundan tashqari, tavsiflovchilar orasidagi semantik aloqalar :

-hujjatlarni tasniflash va turkumlashtirish;

-so‘rov bilan bog‘liq so‘zlar ro‘yxatini tuzish:

-ma'lumotlarni qidirish vazifalari uchun ishlatilishi mumkin. Uning atamalar to'plami matnda qo'llaniladigan barcha terminologik iboralarni o'z ichiga olishi kerak. Ushbu usulni amaliy amalga oshirish uchun tezaurusda ko'pchilik atamalarning borligi yetarli bo'lib, har qanday mavzu sohasining tabiiy til tavsifi matn konvolyutsiyasi uchun ishlatiladigan ba'zi konstruksiyalar bilan tavsiflanadi, masalan, bir xil atamalar, anaforistik va elliptik konvolyutsiyalar. Omonimiya qo'shimcha muammo yaratadi hamda bir xil so'zlar anaforistik va elliptik konvolyutsiyalar bilan bog'liqdir. Ba'zi muammolarni asl tezaurusga sinonimlarning keng bazasini kiritish orqali juda samarali hal qilish mumkin, omonimiya masalalarini hal qilish bilan bog'liq muammolarni faqat statistik tahlil yordamida samarali hal qilish mumkin. [Aktaev, 2015:4]

Xulosa qilib shuni aytish mumkinki, tezaurus tushunchasi haqidagi ilmiy yondashuvlar tahlilidan tezauruslarni yaratish usullari va vositalari sohaning shu kungacha yaxshi o'rganganligini, tadiqoqtchilar amaliyotda yaxshi natija beradigan yangi modellarni, yangicha yondashuv va qarashlarni ilgari surganligi ko'rinadi.

Foydalanilgan adabiyotlar ro'yxati:

1. Kilgarriff, C. Yallop. What's in a Thesaurus? International Conference on Language Resources and Evaluation, 2000.
2. O'zbekiston Milliy Ensiklopediyasi.- Toshkent: 2004.
3. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
4. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
5. Abdurakhmonova, N., & Ismailov, A. S. (2022). APPLYING WEB CRAWLER TECHNOLOGIES FOR COMPILING PARALLEL CORPORA AS ONE STAGE OF NATURAL LANGUAGE PROCESSING. In СОВРЕМЕННАЯ ФИЛОЛОГИЯ. СОЦИАЛЬНАЯ И НАЦИОНАЛЬНАЯ ВАРИАТИВНОСТЬ ЯЗЫКА И ЛИТЕРАТУРЫ (pp. 22-27).
6. Abduraxmonova, N., & Abduvaxobov, G. I. (2021). O 'quv lug 'atini tuzishning nazariy metodologik asoslari. СЎЗ САНЪАТИ ХАЛИҚАРО ЖУРНАЛИ, 103.
7. Sulevmanov, D., Gatiatullin, A., Prokopyev, N., & Abdurakhmonova, N. (2020, November). Turkic morpheme web portal as a platform for turkology research. In 2020 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-5). IEEE.
8. E. Aktaev. O'quv va axborot jarayoniga avtomatlashtirilgan tezaurusning zamonaviy usullarini joriy etish.Yosh olim- 2015.- B. 3-6.
9. Л.В. Щерба.Опыт теории общей лексикографии. 1940.
10. В. Добровольский. От редакции. С. Баранова. От составителя. - М.,1965. С.4-6
11. N. Valitova. Формирование профессионалного тезауруса у студентов физкултурнова вуза. Автореферат на соисканме учёной ступени канд.пед.наук.-Омск, 2012. -В.19.

ISSUES, THEORIES, AND APPROACHES IN THE PRESENT-DAY LEXICOGRAPHIC SEMANTIC STUDIES

Saparova Mohira
Ma'mun university, Faculty of Philology,
Department of Romance-Germanic Philology an English teacher.

Annotation: This article is about "word meaning/lexical semantics" reflects the multidisciplinary nature of current research on the analysis and description of lexical semantic information in dictionaries. The main aim of this paper is to illustrate several issues about the relationship between theory and practice that still plague traditional lexicography and influence its effective participation in this multidisciplinary enterprise, and to place the reported research within the context of the lexical semantic problems, theories, methodologies, techniques, and lexical data that underlie current research on lexicographic semantics.

Keywords: lexical semantics, computational lexicography, sociolinguistic, lexicographer, mutual influence

Аннотация: Данная статья на тему «Значение слова/лексическая семантика» отражает мультидисциплинарный характер современных исследований по анализу и описанию лексико-семантической информации в словарях. Основная цель данной статьи — проиллюстрировать некоторые проблемы взаимоотношений между теорией и практикой, которые до сих пор преследуют традиционную лексикографию и влияют на ее эффективное участие в этом междисциплинарном предприятии, а также поместить представленные исследования в контекст лексико-семантических проблем, теорий, методологии, приемы и лексические данные, лежащие в основе текущих исследований лексикографической семантики.

Ключевые слова: лексическая семантика, компьютерная лексикография, социолингвистика, лексикограф, взаимное влияние.

Annotatsiya: Ushbu maqola "so'z ma'nosi/leksik semantika" mavzusidagi lug'atlardagi leksik-semantik ma'lumotlarni tahlil qilish va tavsiflash bo'yicha zamonaviy tadqiqotlarning ko'p tarmoqli xususiyatini aks ettiradi. Ushbu maqolaning asosiy maqsadi - nazariya va amaliyot o'rtasidagi munosabatlardagi an'anaviy leksikografiyaning hanuzgacha qiyinchiliklarga duchor bo'lishi va uning ushbu fanlararo munosabatlarda samarali ishtirokiga ta'sir qiladigan ba'zi muammolarni ko'rsatish va taqdim etilgan tadqiqotni leksik-semantik muammolar kontekstida joylashtirish, leksikografik semantika bo'yicha joriy tadqiqotlar asosida yotgan nazariyalar, metodologiya, uslublar va leksik ma'lumotlar haqida fikr yuritilgan.

Kalit so'zlar: leksik semantika, kompyuter leksikografiyasi, sotsiolingvistika, leksikograf, o'zaro ta'sir.

Introduction

All publications about "word meaning/lexical semantics" center on the challenges encountered during the examination, characterization, and lexical semantic information is acquired. Like most other lexicographic research, their ultimate goal is to improve the lexical

semantic information in lexicographic resources, such as natural language processing (NLP) systems lexica or conventional dictionaries intended for human users. Nonetheless, these works also demonstrate the interdisciplinary contribution of recent lexicographic semantics research. The emergence of computers has not only fundamentally altered the process of dictionary construction, but it has also additionally given rise to new lexicographic approaches and techniques, new dictionary types, and a host of new disciplines that address the theoretical and practical issues associated with dictionary compilation (in the broadest sense of the word). Despite significant overlap and mutual influence, professionals in the domains of computational linguistics, theoretical linguistics, computational lexicography and traditional lexicography. According to the particular objectives of their lexica, computational lexicography, and knowledge engineering each apply a unique combination of theoretical and practical competence to address the language semantic issues encountered in their respective domains [Abdurakhmonova, N.2020,2021,2022].

This paper's primary section will tackle the challenging challenge of situating the presented lexical meaning research within this multidisciplinary framework of lexical semantic difficulties, ideas, and approaches. The information and methods used in Euralex 1994 form the basis of the present lexicographic lexical semantics study. [2;87]

The metalexico-graphic study is being hampered by a variety of common misconceptions and assumptions that are holding back the discipline. Among these misconceptions, the most significant one is the purported "theorylessness" of practical lexicography. The most harmful misconception is that of the metalexico-grapher as the cynical critic of dictionaries and the people who create them, preoccupied with theoretical issues and unconcerned with the "harsh" realities of dictionary creation.

Methods and level of study

The lexicographer and semanticist Anna Wierzbicka gained notoriety for her statements that there is no theoretical foundation for practical lexicography and that "even the best lexicographers, when pressed, can never explain what they are." doing, or why" (Wierzbicka 1985:5) and that "it is the lexicographers' achievements, not their failures, which are truly remarkable" (Wierzbicka 1990:366) in light of the lack of support from semantic theory in particular.

Atkins (1992/1993:5) stated "Many people in contemporary lexicography deal with theoretical linguists by keeping their heads down below the barricades and getting on with writing dictionary entries." In an interview published in the inaugural edition of the Cambridge Language Reference News ((1) 1994), renowned computational linguist Ted Briscoe states that lexicography, as opposed to linguistics, "is more atheoretical - a practical activity pursued to sell dictionaries." [2;78]

It is not always the case that lexicography has no empirical descriptive issues because theoretical linguists overlook these issues in their data selection and theory building. theoretical foundation. Lexicographers are inevitably meta-lexicographers since linguistic principles and presumptions have always informed and directed their actual descriptive operations. These presumptions may be eclectically constituted, that is, they may not form a systematic or strictly coherent body of hypotheses on lexical semantic matters, but that does not lessen their status as guiding principles for lexicographic praxis. It also may not always be articulated, or if

articulated, not strictly adhered to (which gives rise to accusations of the "gulf between theory and practice"). [3;66]

Lexical semantic issues, prevalent theories of semantics, and methodologies Following this overview, I'd like to look more closely at various some of the more well-liked lexical semantic theories that scholars rely on, an overview of the lexical-semantic issues addressed in the articles in this subject cluster, and an explanation of how they relate to the findings presented in some recently released literature.

Results and discussion

In the discipline of lexicography, a research approach that includes any two or more of the following phases has emerged and been established in the last few years:

1. lexicographical (semantic) problem identification;
2. problem analysis within the context of one or more linguistic frameworks (hypotheses, techniques, data);
3. a lexicographic perspective on the problem, i.e. an analysis of current and past approaches (assumptions, methods, procedures, techniques) to the problem within lexicography itself; a lexicographic perspective on the problem, i.e. an analysis of current and past approaches (assumptions, methods, procedures, techniques) to the problem within lexicography itself;
4. presentation of a solution to the research problem from (ii) and/or (iii) and
5. presentation of a motivation for the proposed solution either in terms of the theoretic assumptions, methods and data presented in (ii) and/or (iii) or in terms of any other relevant considerations such as the dictionary type, constraints on space, user's information needs or research findings within other relevant fields (e.g. information science or cognitive psychology).

It is possible to see two aspects of lexicography, which are conditionally connected with each other, scientific-theoretical and practical. LP Stupin distinguishes lexicography as a field in three aspects:

1) a special branch of linguistics that studies the principles of creating various dictionaries;
2) practical activity dealing with vocabulary;
3) vocabulary of a specific language

The material selected for the content of such dictionaries is directed to the didactic purpose.

Vocabularies serve as a learning tool in training. Its educational character consists in the fact that the content of the dictionary, the arrangement of articles, the forms of providing linguistic information, the method of presenting the understood meaning or content, the size, etc., are aimed at teaching or learning a language [1]. Another feature of these dictionaries is the minimization of all elements of the language within a certain level. The units included in the dictionary are based on the principle of lexical minimum. [4; 90]

Also, the sociolinguistic character of the word and the explanation of the lexicon content selected for monolingual dictionaries show that it is oriented towards the educational goal. Monolingual educational dictionaries may include word explanations, phraseological, orthographic aspects, word combinations, antonyms, synonyms, and artificial words. In bilingual educational dictionaries, together with the fields mentioned above, the translation and explanation of words are also provided. [1; 23-24]

The dictionary content may differ to a certain extent depending on whether the educational dictionaries are intended to be used in education as a native language, a second language, or a foreign language. The dictionary also differs from one another in terms of whether the articles are presented in alphabetical order, flat or grid. According to the audience of users, it is adapted to the general needs and interests of preschool, school, higher education or language learners. Summarizing the above-mentioned differences in the creation of educational dictionaries, it should be noted that this type of dictionaries should serve as a communication tool for language learners.

Vocabulary selection is one of the most important issues in educational vocabulary [Abdurakhmonova, N.2020,2021,2022]. In this regard, the creation of electronic corpora and linguistic studies have greatly served the educational dictionary in determining the active and passive lexicon. As a result of mastering the vocabulary of active words, the answer to the question of what types of active words should be included in the teaching of the mother tongue or a foreign language is found. In the 30s of the 20th centuries, ideas were developed in English lexicography to create a dictionary of active words in teaching the language as a foreign language.

Charles Ogden's dictionary Basic English (1930) notes that this term is used in three different meanings [5]: 1) the lexical minimum consisting of the basic vocabulary in connection with the teaching of the English language; 2) system of teaching English as a foreign language; 3) the main tool of international communication. It should be noted that L.P. Stupin reduces the main vocabulary in the educational dictionary compiled by Charles Ogden to 7500, and later notes that he "rationalized" it to 1500 and finally to 850 English words. Of these, there are 200 types of subject names, more than 400 nouns, 100 types of conjunctions (12 of which are verbs, the rest are non-independent word groups). The lexicographer selects the words from the minimum lexicon that mean the concept and leaves the necessary words from among them depending on the frequency of use.

A.S. Trifonov [5] classified dictionaries according to different parameters: 1) according to language participation (monolingual, bilingual, trilingual); 2) according to the order of providing the material (alphabetical or ideographic); 3) according to the direction of reception or reproduction (active and inactive words); 4) area of limitation of lexicon (general and special dictionaries); 5) according to the content of the dictionary (linguistic, encyclopedic). As mentioned above, educational dictionaries can be monolingual and multilingual. According to the data, monolingual educational dictionaries are more important than bilingual dictionaries. Bilingual educational dictionaries are used to learn the language as a foreign language. The lumping x splitting problem is addressed in a work by De Regt ("The description of multiple meaning in some Biblical Hebrew lexicographical projects"; this volume). The so-called "linguistic" and "contextual" meanings of words are not distinguished in Hebrew dictionaries, and contextual usages are noted as independent senses. He suggests grouping contextual senses that are part of the same semantic domain and providing more information about a word's syntagmatic properties in each of its recorded senses.

Conclusion

As a conclusion, it can be said that lexicography (lexicology) is not a product of a purely theoretical field of knowledge, it is a product of the need that arose in the past in the process of communication between people for trade and exchange of ideas. As the French lexicographer Alan Ray rightly stated, "modern civilization is a civilization of dictionaries" [1;24]. The improvement

of computer technologies from year to year and various software programs created to further improve the process of natural language processing in language technology play an important role in the development of computer educational lexicography.

References:

1. Abduraxmonova N, G‘. Abduvaxobov “O‘quv lug‘atini tuzishning nazariy metadologik asoslari” СЎЗ САНЪАТИ ХАЛИҚАРО ЖУРНАЛИ, 103
2. Apresjan, Jurij, D. 1992/1993. "Systemic Lexicography as a Basis of Dictionary-making". Dictionaries 14:79-87.
3. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
4. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
5. Abdurakhmonova, N., & Ismailov, A. S. (2022). APPLYING WEB CRAWLER TECHNOLOGIES FOR COMPILING PARALLEL CORPORA AS ONE STAGE OF NATURAL LANGUAGE PROCESSING. In СОВРЕМЕННАЯ ФИЛОЛОГИЯ. СОЦИАЛЬНАЯ И НАЦИОНАЛЬНАЯ ВАРИАТИВНОСТЬ ЯЗЫКА И ЛИТЕРАТУРЫ (pp. 22-27).
6. Abduraxmonova, N., & Abduvaxobov, G. I. (2021). O ‘quv lug ‘atini tuzishning nazariy metodologik asoslari. СЎЗ САНЪАТИ ХАЛИҚАРО ЖУРНАЛИ, 103.
7. Sulevmanov, D., Gatiatullin, A., Prokopyev, N., & Abdurakhmonova, N. (2020, November). Turkic morpheme web portal as a platform for turkology research. In 2020 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-5). IEEE.
8. Atkins, B.T.S. 1991. "Building a Lexicon: The Contribution of Lexicography". International Journal of Lexicography 4.3:167-204.
9. Dik. S.C. 1989. The theory of Functional Grammar Parti: The structure of the clause. Dordrecht: Foris.
10. Evens, M.A.(Ed.) 1988. Relational models of the lexicon. Cambridge: Cambridge University Press.

TERMINOLOGIK ONTOLOGIK BILIMLAR BAZASINI MODELLASHTIRISH BOSQICHLARI

Qodirova Zebo Gulboyevna
O‘zMU Kompyuter lingvistikasi va
amaliy tilshunoslik kafedrasi o‘qituvchisi.

Jo‘raqulova Madina Subhonovna
O‘zMU Kompyuter lingvistikasi va
amaliy tilshunoslik kafedrasi o‘qituvchisi.

Annotatsiya: Mazkur maqolada terminologik ontologik bilimlar bazasini yaratish bosqichlari, xususan, ontologiyani ishlab chiqish jarayoni: Spetsifikatsiya, kontseptualizatsiya, formallashtirish va amalga oshirish bosqichlari haqida nazariy ma’lumotlar o‘rin olgan. Bugungi kunda mavjud bo‘lgan eng mashhur Ontologiya tillari (OWL va RDF), Protégé muharririda terminologik ontologik bilimlar bazasini yaratish, terminologik bazani shakllantirish xususida so‘z boradi.

Kalit so‘zlar: Spetsifikatsiya, kontseptualizatsiya, formallashtirish, ontologik bilimlar bazasi, terminologiya.

Annotation: This article contains theoretical information about the stages of creating a terminological ontological knowledge base, in particular, the ontology development process: specification, conceptualization, formalization and implementation stages. The most popular Ontology languages available today (OWL and RDF), the Protégé editor is about creating a terminological ontological knowledge base, forming a terminological base.

Keywords: Specification, conceptualization, formalization, ontological knowledge base, terminology.

Аннотация: В данной статье содержатся теоретические сведения об этапах создания терминологической онтологической базы знаний, в частности о процессе разработки онтологии: этапы спецификации, концептуализации, формализации и реализации. Наиболее популярные на сегодняшний день языки онтологий (OWL и RDF), редактор Protégé предназначен для создания терминологической онтологической базы знаний, формирования терминологической базы.

Ключевые слова: Спецификация, концептуализация, формализация, онтологическая база знаний, терминология.

Ma’lumki, model bevosita kuzatish imkoni bo‘lmagan hodisani o‘rganish uchun yaratilgan sun’iy mexanizmdir. Lingvistik model tushunchasi ilmiy muomalaga struktur tilshunoslikning E.Sepir, L.Blumfild, R.Yakobson, X.Chomskiy, Z.Xarris, Ch.Hokket kabi namoyandalari asarlaridan kirib kelgan. Uning taraqqiyoti XX asrning 60 – 70-yillari (matematik va kibernetik lingvistika rivojlana boshlagan davr)ga to‘g‘ri keladi[Abduraxmonova N.Z., 2018: 165].

Ontologiyalarni yaratish va tadqiq qilishning ikkita muqobil yondashuvi mavjud. Birinchisi (rasmiy) mantiqqa asoslanadi, ikkinchisi (lingvistik) tabiiy tilni (xususan, semantikani) o‘rganishga va korpus deb ataladigan katta matn massivlarida ontologiyalarni qurishga asoslangan [Sharipbay A. A., Mukanova A. S., Niyazova R.S., Razakhova B.SH., Omarbekova A. S., Abdurakhmonova N. Z., 2021: 55].

Ontologiyani ishlab chiqish jarayoni 4 bosqichni o‘z ichiga oladi.

Spetsifikatsiya - ontologiyalarni yaratish maqsadlariga, uning maqsadli qo‘llanilishiga va potensial foydalanuvchilarga bo‘ysunadi. Ma’lumotlar bazasiga kiritilgan elementlar ro‘yxatini o‘z ichiga olgan loyiha shakli, dizayni deyish mumkin.

Kontseptualizatsiya - aniq modelda fan bilimlarini tuzilishini ta’minlaydi. Konseptualizatsiya – tushunchalar to‘plamini aniqlash jarayonini belgilab beruvchi yo‘l demakdir.

Rasmiylashtirish bu – formallashtirishdir. Obyekt yoki hodisaning mohiyatini ko‘rsatishda belgilangan shartlarni qayta ishlash jarayoni. Tanlangan dastur tilida ba’zi bir algoritm yordamida ma’lum modelni yaratish bosqichi.

Amalga oshirish – amalga oshirish bosqichida ontologiyaga mos keladigan va bilimlarni ifodalash tillaridan birida ontologiya modeli yaratiladi.

Konseptualizatsiya ikki bosqichni o‘z ichiga oladi, ular doirasida muammolar hal qilinadi:[Назаров Д.А., 2013]

1) terminlar lug‘atini shakllantirish
2) kontseptlarning tasniflash iyerarxiyalarini qurish.

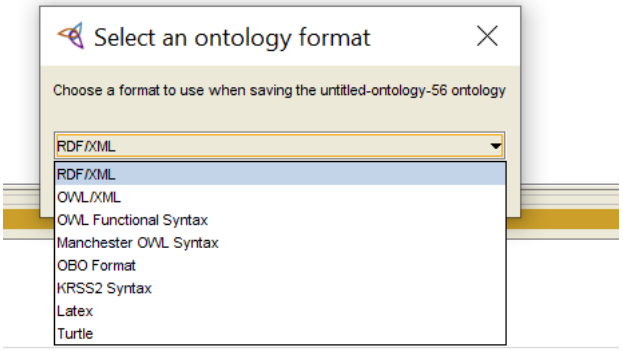
Konseptualizatsiyaning tabiati uning qanday paydo bo‘lganiga yoki qanday yaratilganiga bog‘liq. Konseptualizatsiya bu – tegishli ma’lumotlarni olish va umumlashtirishga olib keladigan jarayon. Konseptualizatsiya – bu muayyan ma’lumotlarning o‘zi va muayyan vaziyatlardan, vakillik tillaridan mustaqildir. Ikkinchi bosqich spetsifikatsiya, chunki ontologiya vakillik tilida konseptual izlanishni belgilaydi. Murakkablik va aniqlik darajasidan tashqari, hal qiluvchi narsa shundaki, ontologiyalar kontseptualizatsiyaning tilga bog‘liq spetsifikatsiyalari sifatida, muloqotning asosi bo‘lib, umumiy tushuncha o‘rnatiladigan ko‘prikdir [Huang, C. R., Calzolari, N., Gangemi, A., Lenci, A., & Oltramari, A., 2010].

Bugungi kunda mavjud bo‘lgan eng mashhur OWL va RDF muharriri Protégé texnologiyasi hisoblanadi. Protégé OWL ontologiyalarini yaratish uchun qulay muharrir bo‘lib, uning imkoniyatlari ham bir qadar kengroq.

Tibbiyot terminlarining ontologik lug‘atini yaratishda, dastlab bizga ma’lumotlar bazasi kerak bo‘ladi. Ma’lumotlar bazasi deganda ma’lum bir strukturada saqlanadigan ma’lumotlar to‘plami tushuniladi. Boshqacha qilib aytganda MB - bu ma’lum berilgan aniq bir strukturaga ega bo‘lgan ma’lumotlarni o‘z ichiga oluvchi maxsus formatga ega bo‘lgan fayldir.

Shuni qayd qilish lozimki, **MB**ni yaratishda ikkita muhim shartni hisobga olmoq zarur: Birinchidan, ma’lumotlar turi va ko‘rinishi ularni qo‘llaydigan dasturlarga bog‘liq bo‘lmasligi lozim, ya’ni **MB**ga yangi ma’lumotlarni kiritganda yoki ma’lumotlar turini o‘zgartirganda, programmalarni o‘zgartirish talab etilmasligi lozim. Ikkinchidan, MBdagi kerakli ma’lumotni bilish yoki izlash uchun biror programma tuzishga hojat qolmasin. Shuning uchun ham **MB**ni tashkil etishda ma’lum qonun va qoidalarga amal qilish lozim. Bundan tashqari, **axborot** so‘zini **ma’lumot** so‘zidan farqlashimiz kerak ya’ni ***axborot*** so‘zini umumiy tushuncha sifatida qabul qilib, ***ma’lumot*** deganda aniq bir belgilangan narsa yoki hodisa sifatlarini nazarda tutamiz [Qarang: https://tami.uz/php]. Ma’lumotlar bazasi quyidagilardan iborat: tibbiy terminlar

izohli lug‘ati, sinonimlar, omonimlar jadvali va boshqa. Ontologiyani qurishda leksik birliklar o‘rtasidagi munosabatlarni o‘rnatish bosqichida umumiy ma’lumotlar bazasiga murojaat qilinadi va leksik birliklarni munosabatlar bo‘yicha bog‘lab chiqiladi. Dastlabki bosqichda maqbul bo‘lgan ontologiya tili tanlanadi. Bir qancha ontologiya tillari mavjud bo‘lib, ular quyida ifodalangan.



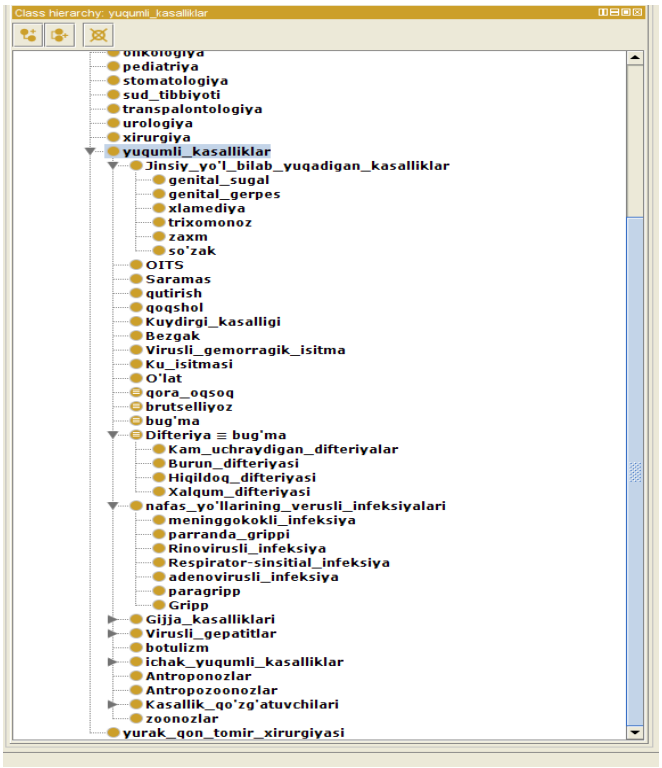
1-rasm. Ontologiya tavsif tillari

Bilamizki, World Wide Web – bu online tizim orqali kirish mumkin bo‘lgan giperbog‘langan hujjatlar tizimi. Bu oddiy hujjatdan farq qiladi, chunki u nafaqat matnni, balki boshqa ma’lumotlarni ham o‘z ichiga olishi mumkin: rasmlar, musiqa va video fayllar va boshqalar. Bunday hujjatlarni maxsus dasturlar yordamida ko‘rish mumkin. Olimlarning fikrlariga ko‘ra, World Wide Webning vorisi sifatida sun’iy intellekt yoki semantik veb deya e’tirof etilmoqda. Semantik veb butunjahon online tarmoq tizimining avlodi bo‘lib, yuqoridagilarga qo‘shimcha sifatida ma’lumotlar bazasi semantikasi tavsiflarini taqdim eta olish imkoniyatiga ega hisoblanadi. Bunday tavsiflar nafaqat foydalanuvchilar uchun balki, dasturiy ta’minot uchun ham tushunarli bo‘ladi. Aqlli tizimlar tomonidan ma’lumotlar bazasi mazmunini tushunish online tizimda ma’lumotni aniqroq qidirish imkonini beradi, qolaversa, bu ma’lumotlarni avtomatik tarzda tartibga solish hamda boshqa dasturiy ta’minotlar bilan almashinish imkonini beradi[Abdurakhmonova, N.2020,2021,2022].

Veb-ma’lumotlar semantikasini tavsiflash uchun RDF (Resurs tavsifi ramkasi) tilidan foydalaniladi. Bu til veb hujjatlarning resurslarini tavsiflash uchun mo‘ljallangan, ammo bu to‘liq ontologiyalarni ifodalash uchun yetarli emas. Shu sababli, Web ontologiyasini tavsiflash uchun maxsus til, OWL tili ham taklif qilindi. Ushbu til RDF tiliga mos keladi, ya’ni har bir OWL hujjati ham RDF hujjatidir. RDF XML sintaksisi asosida qurilgan bo‘lib, u internetda ma’lumotlar almashinuvi uchun standart sifatida ishlatiladi [Eshonqulov H. I., 2022: 128].

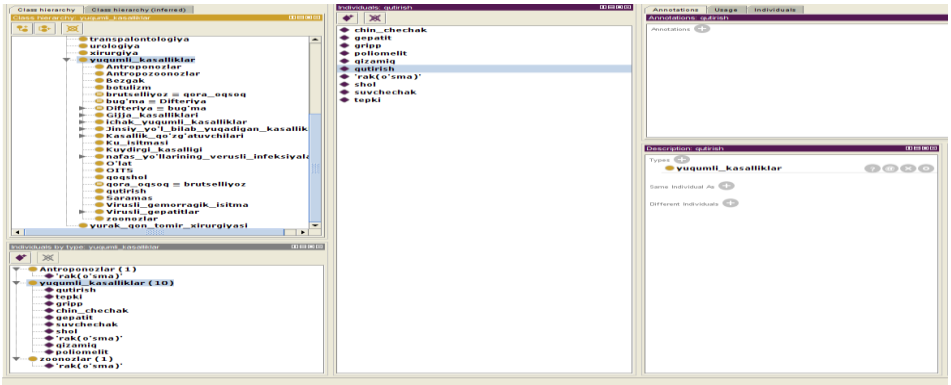
OWL veb-ontologiya tili shunchaki ma’lumot berib qolmasdan, ma’lumotlar mazmunini qayta ishlashga muhtoj bo‘lgan ilovalardan foydalanish uchun mo‘ljallangan. OWL XML, RDF tillaridan ko‘ra mashina talqinini osonlashtiradi. RDF (Resource Description Framework) bu online tizimda joylashgan ma’lumotlarni tavsiflash imkonini beruvchi tildir. RDF tavsif tili bo‘lish bilan birgalikda shu tilda yozilgan ma’lumotlarni to‘liq o‘qish va almashtirishni ta’minlaydigan turli xil qo‘shimcha dasturiy modellar ma’nosida ham qo‘llaniladi.

Ma’lum ontologiya tavsif tili tanlangandan so‘ng, keyingi bosqichda asosiy subdomainlar va ular tarkibidagi subkategoriyalar aniqlab sinf iyerarxiyasi shakllantiriladi.



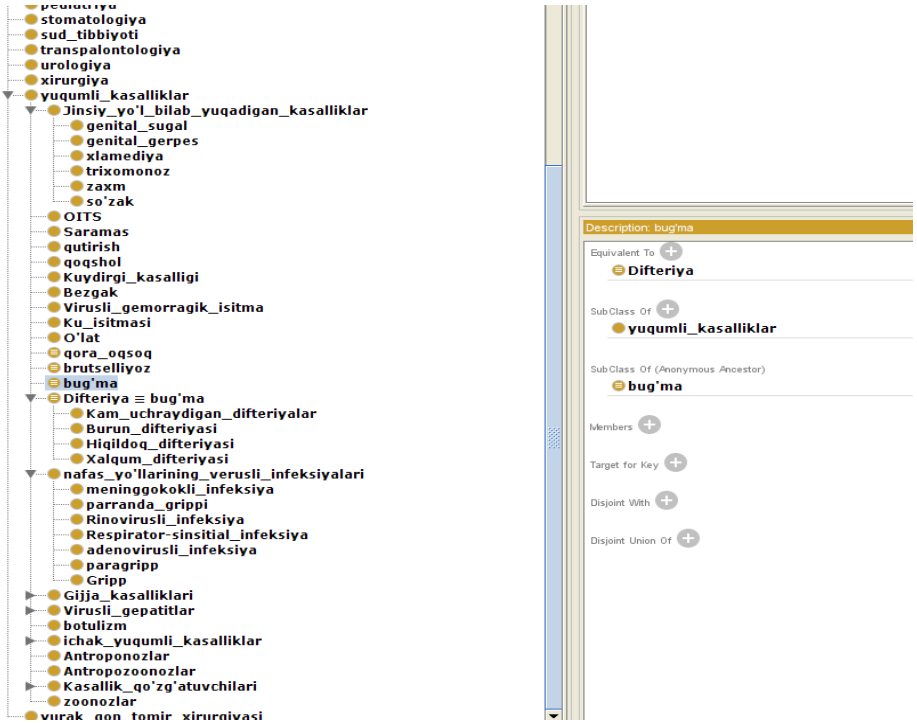
2-rasm. Tibbiyot birliklarining sinf iyerarxiyasi

Bu yerda tibbiyotning ayrim sohalari bo‘yicha sinf iyerarxiyasini ko‘rishimiz mumkin bo‘ladi.



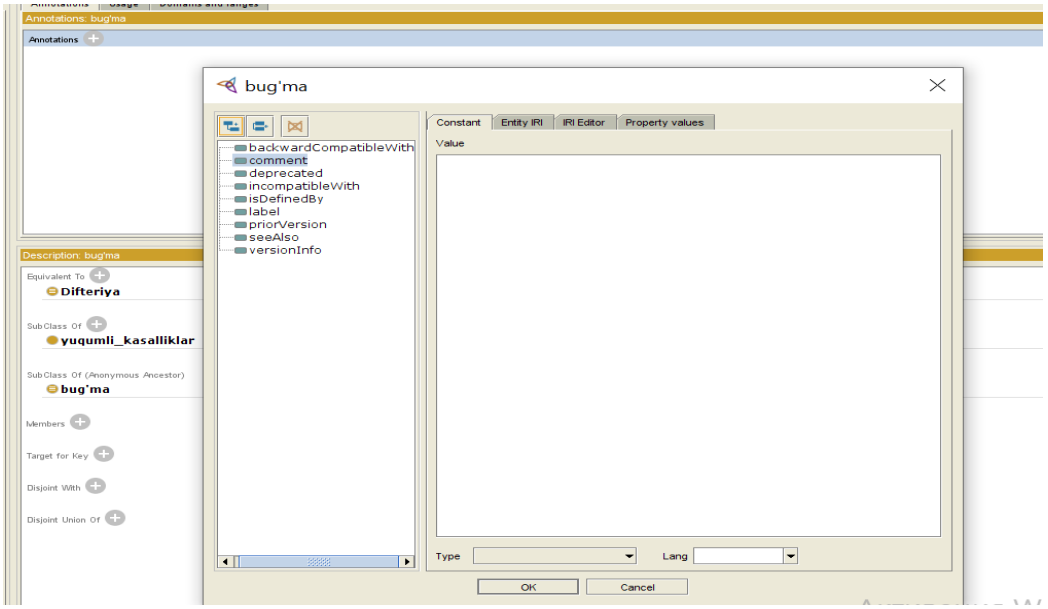
3-rasm. Sinf elementlarining individuals oynasida ifodasi

Har bir sinfga tegishli bo‘lgan elementlar *individuals* oynasida ifodalanadi va tegishli sinflarga bog‘lab chiqiladi. Bu yerda *poliomelit, qizamiq, rak, shol, suvchechak, gepatit, chinchechak, gripp, tepki, qutirish* kabi kasalliklar *yuqumli kasalliklar* sinfiga mansubligi keltirilgan. *Individuals oynasi* sinf misollarini aniqlash imkonini beradi. Yorliq yordamida ikki turdagi misollarni ko‘rish mumkin: 1) foydalanuvchi tomonidan kiritilgan misollar; 2) tizim tomonidan ontologiyada belgilangan qoidalarga asoslangan holda taxmin qilingan misollarni ko‘rish mumkin. Masalan, agar biror munosabat boshqasiga teskari bo‘lsa va bu munosabat bilan bog‘langan elementlar juftligi kiritilgan bo‘lsa, tizim avtomatik ravishda teskari munosabat uchun mos keladigan juftlikni kiritadi.



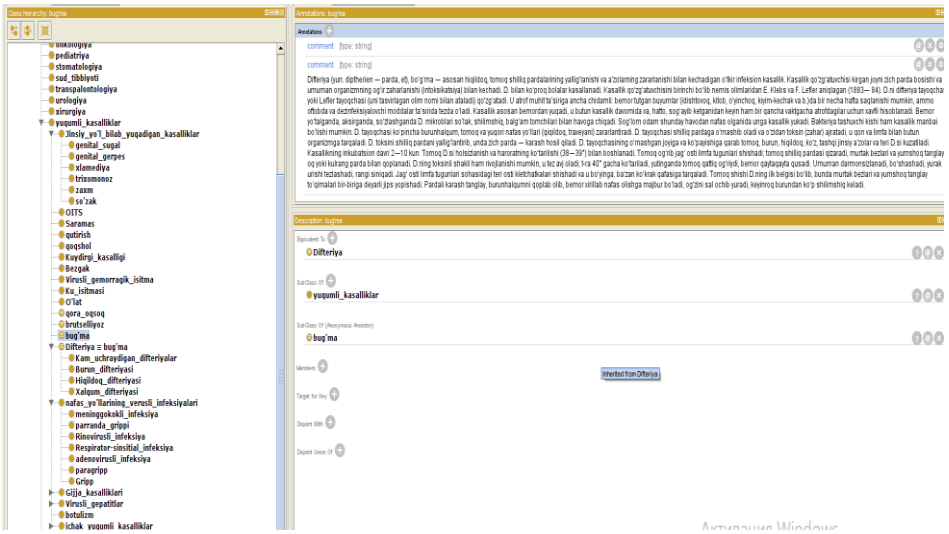
4-rasm. Sinf iyerarxiyasi ifodasi

Keyingi bosqichda esa, elementlar o‘rtasidagi munosabatlarning ayrimlarini ko‘rib o‘tamiz. Masalan, *bug‘ma – difteriya* bilan sinonim bo‘lganligi uchun *Equivalent* sifatida belgilangan.



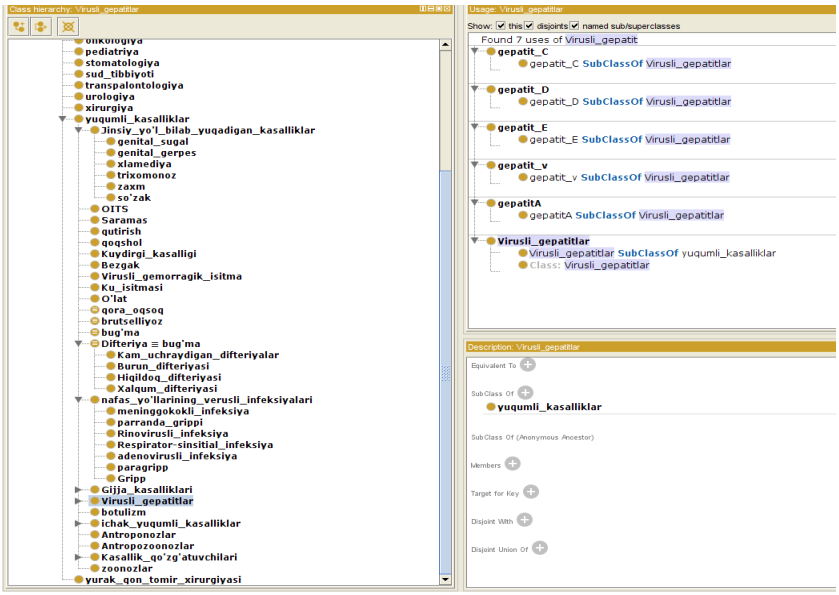
5-rasm. Elementlar tavsifining matn tipi ifodasi

Har bir sinf hamda uning elementiga *comment* yozilayotganda matn turi tanlanadi. Bu bosqichda biz *string:* tipidagi matndan foydalanamiz va *language*ga esa *ing:* tili tanlanadi.



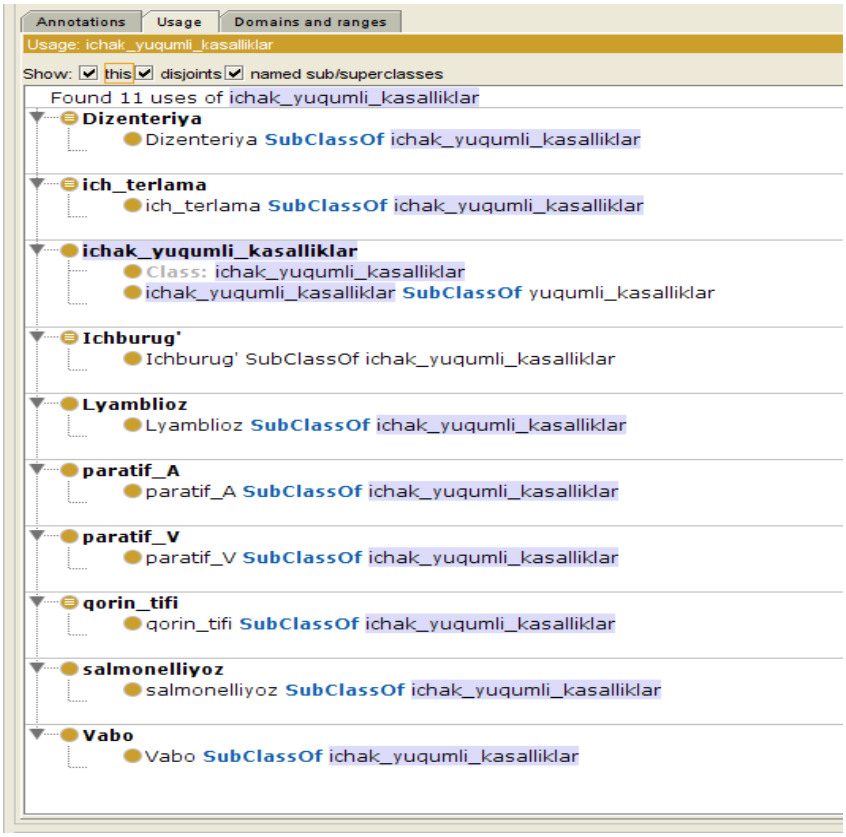
6-rasm. Sinf elementlarining annotation qismida ifodasi

Annotation qismida tanlangan sinf yoki uning elementi uchun *comment* e'lon qilinadi. Commentda tanlangan matn tipi hamda tili ham ifodalanadi.



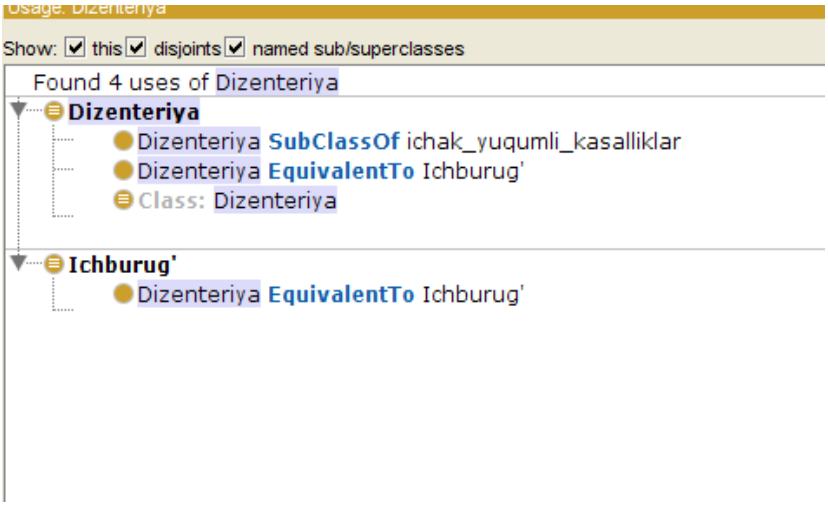
7-rasm. Sinf iyerarxiyasida tanlangan sinf tarkibida subkategoriyalar ifodasi

Sinf iyerarxiyasida tanlangan sinf(*virusli gepatitlar*)ning *description* oynasida qaysi *SubClassOf* ga tegishli ekanligi (*yuqumli kasalliklar sinfi*) ko'rinib turadi. *Usage* qismida esa *Subclass* (*virusli gepatit*) sinfiga mansub bo'lgan *Superclasses(gepatit C, gepatit D, gepatit E, gepatit V, gepatit A)*lar ifoda etilgan va barcha *Class: virusli gepatitlar SubClassOf*ga mansub deya ifoda etiladi.



8-rasm. Sinf hamda elementlarning Usage qismida ifodasi

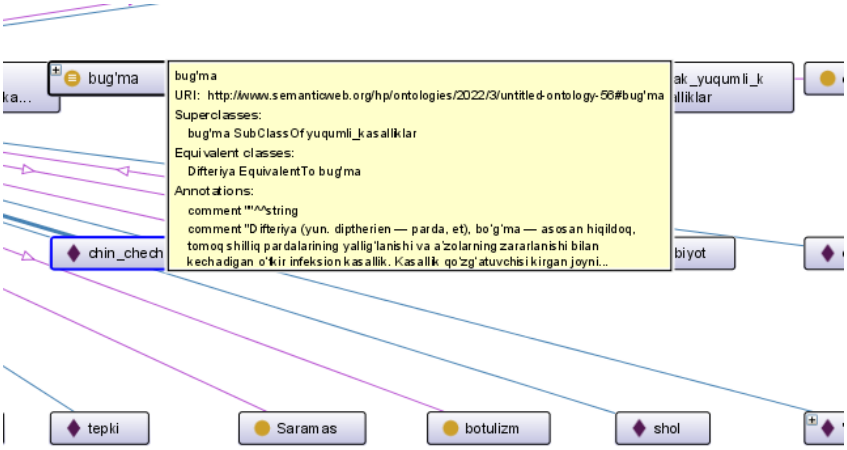
Yuqorida ham sinf iyerarxiyasida tanlangan sinfning *Usage* oynasida ifodasi berilgan. Bu yerda, *SubClassOf: yuqumli_kasalliklar*ga mansub *Class: ichak_yuqumli_kasalliklar* tarkibida mavjud elementlar qayd etilmoqda ya'ni *dizenteriya=ichburug', ich_terlama=qorin_tifi, lyamblioz, paratif_A, paratif_V, salmonelliyoz, vabo;* quyida esa *dizenteriya va ichburug'* elementlari *EquivalentTo* sifatida ifodasi berilmoqda.



9-rasm. Sinf aksiomalari

Sinf aksiomalarida sinf misollari bir-biri bilan bog'lanishi ifodalanadi. Sinfning asosiy aksiomasi esa sinf nomini e'lon qilishdir. Bu aksioma shunday nomli sinf mavjudligining

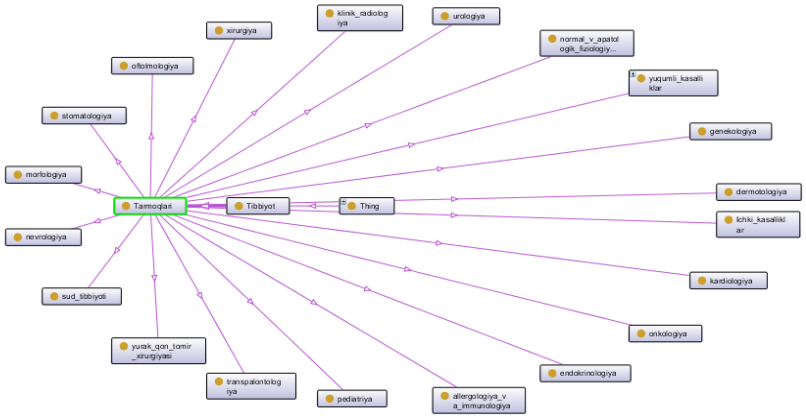
ifodasidir. Bu esa sinflar orasidagi oddiy iyerarxik munosabatlarni ifodalash uchun ko‘pincha yetarli emas. Shuning uchun, OWL tilida quyidagi sinf aksiomalarini ham qayd etish mumkin. SubClassOf aksiomasi, sinfni boshqasining kichik sinfi sifatida tavsiflaydi. Bundan ma’lum bo‘ladiki, agar klass boshqa klassning quyi sinfi bo‘lsa, uning barcha misollari ham yuqori SubClassning elementlari hisoblanadi va yana bir aksioma EquivalentClass aksiomasi, bu sinfga mansub sinf yoki elementlar namunalari bir-biriga, mos kelishini ko‘rishimiz mumkin bo‘ladi[Abdurakhmonova, N.2020,2021,2022]. Bu esa sinonimlarni aniqlashning bir usuli hisoblanadi.



10-rasm. OntoGrafda belgilangan elementning metadata ifodasi

*OntoGraf*da (10-rasm) belgilangan elementning *metadatasi* keltirilmoqda: ya’ni
Bug‘ma
URI: http://www.semantikweb.org/hp/ontologies/2022/3/untitled ontology.56#bug‘ma
Superclasss:
Bug‘ma SubClassOf yuqumli kasalliklar
Equvalent classes:
Difteriya EquivalentTo bug‘ma
Annotation:
*Commen *string*
Comment: ***Difteriya*** *(yun. diptherien — parda, et), bo‘g‘ma — asosan hiqildoq, tomoq shilliq pardalarining yallig‘lanishi va a’zolarning zararlanishi bilan kechadigan o‘tkir infeksion kasallik. Kasallik qo‘zg‘atuvchisi kirgan joyni zich parda bosishi va umuman organizmning og‘ir zaharlanishi (intoksikatsiya) bilan kechadi. Difteriya bilan ko‘proq bolalar kasallanadi va bosh.*

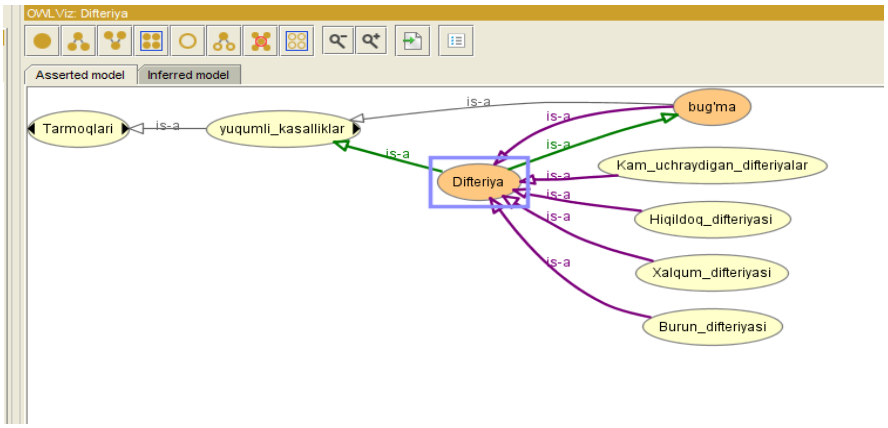
Metadata yorlig‘i ontologiya haqidagi metama’lumotlarni tahrirlash imkonini beradi. Bu ontologiyaning o‘zida tavsiflangan ma’lumotlarga bevosita aloqador bo‘lmagan, balki ontologiyaning o‘zini tavsiflovchi ma’lumotlar bazasidir. Bu oynadan foydalanib ontologiya nomini o‘zgartirish, ontologiya nomiga qo‘shimcha qo‘shish mumkin bo‘ladi, bu ontologiyadan foydalanadigan boshqa dasturlar tomonidan ham, ontologiyaning o‘zida ham qisqa nom sifatida ishlatilishi mumkin. Nomni o‘zgartirish ontologiyada aniqlangan nomlarni boshqa ontologiyalarning nomlaridan ajratish imkonini beradi.



11-rasm. OntoGrafda sinf iyerarxiyasi ifodasi

OntoGrafda sinf iyerarxiyasi ifodasini quyidagilarda ko‘rish mumkin. **OntoGraf yaratilayotgan OWL ontologiyalari o‘rtasidagi munosabatlarni interaktiv ravishda ifodalash imkonini beradi. Ontologiyalarning tuzilishini avtomatik tartibga solish uchun turli xil sxemalar mavjud**[Abdurakhmonova, N.2020,2021,2022]**. Qolaversa, turli xil munosabatlarni ham ifodalash mumkin: pastki sinf, individual, domen/diapazon obyekt xususiyatlari va equivalentlik va boshqa.**

OWLViz Ontologiyada sinf iyerarxiyasini ko‘rish va bosqichma-bosqich navigatsiya qilish imkonini beradi, bu esa yaratilgan sinf iyerarxiyasi va maqsad qilingan sinf iyerarxiyasini solishtirish imkonini beradi.OWLViz Protégé bilan foydalanish uchun mo‘ljallangan. OWLViz bir xil rang sxemasidan foydalanib, Protégé-OWL muharriri bilan birlashadi, shuning uchun sinf iyerarxiyasidagi o‘zgarishlarni aniq ko‘rish mumkin. OWLViz sinf iyerarxiyasini PNG, JPEG va SVG kabi turli xil aniq grafik formatlarga saqlash imkoniyatiga ega. Quyida tibbiyotning ayrim sohalari sinf sifatida keltirilib, uning ayrim sinflari tavsifi OWLVizda ifoda etilgan.



12-rasm. Sinflar tavsifining OWLvizda ifodasi

OWLVizda sinonim sifatida belgilab olingan elementlar o‘rtasida ikki tomonlama bog‘lanish kuzatiladi. Masalan, *difteriya=bug‘ma*;

Xulosa o‘rnida shuni aytish mumkinki, Ontologik lug‘atlar hozirgi kunda kompyuter lingvistikasida yangi tadqiqotlardan biri sifatida, korpuslar tarkibida sifatli qidiruvni amalga

oshirish, qolaversa, foydalanuvchi va dasturiy ta’minot o‘rtasida vositachilik maqsadida yaratilmoqda. Keyingi vaqtlarda yaratilayotgan tezaurus va ontologiyalar funksional jihatdan semantik lug‘atlar tarkibiga kiradi. Bundan tashqari, bunday lug‘atlar ideografik lug‘atlar tipiga ham mansub, chunki bu lug‘atlarni tuzishda semantikaga alohida e’tibor beriladi va mavzu maydoni muhim ahamiyatga ega hisoblanadi. Terminlarning kundan-kun ko‘payishi axborotdan foydalanish jarayoniga salbiy ta’sir etishi mumkin, shu jihatdan, ontologik bilimlar bazasi matnli ma’lumotlarning ma’nosi bo‘yicha samarali qidiruvni tashkil qilish, fikrlash mexanizmlarini amalga oshiradigan tizimlarni yaratish, turli xil intellektual bilimlar uchun umumiy bilim bazalarini qurish va ulardan foydalanish tizimlarini yaratishda ahamiyatlidir.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N.Z. Inglizcha matnlarni o‘zbek tiliga tarjima qilish
2. dasturining lingvistik ta’minoti (sodda gaplar misolida): Filol. fan. bo‘yicha falsafa dokt. (PhD)…disc. –Toshkent, 2018. – B. 165.
3. Sharipbay A. A., Mukanova A. S., Niyazova R.S., Razakhova B.SH.,
4. Omarbekova A. S., Abduraxmonova N. Z. Ontologiya va semantik texnologiyalar. Elektron darslik. Nur-Sulton. 2021. – B. 55.
5. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).
6. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
7. Abdurakhmonova, N., & Ismailov, A. S. (2022). APPLYING WEB CRAWLER TECHNOLOGIES FOR COMPILING PARALLEL CORPORA AS ONE STAGE OF NATURAL LANGUAGE PROCESSING. In *СОВРЕМЕННАЯ ФИЛОЛОГИЯ. СОЦИАЛЬНАЯ И НАЦИОНАЛЬНАЯ ВАРИАТИВНОСТЬ ЯЗЫКА И ЛИТЕРАТУРЫ* (pp. 22-27).
8. Abduraxmonova, N., & Abduvaxobov, G. I. (2021). O ‘quv lug ‘atini tuzishning nazariy metodologik asoslari. *СЎЗ САНЪАТИ ХАЛҚАРО ЖУРНАЛИ, 103*.
9. Sulevmanov, D., Gatiatullin, A., Prokopyev, N., & Abdurakhmonova, N. (2020, November). Turkic morpheme web portal as a platform for turkology research. In *2020 International Conference on Information Science and Communications Technologies (ICISCT)* (pp. 1-5). IEEE.
10. Назаров Д.А. Интеллектуальные информационные технологии. Методы
11. представления знаний. Редакция: октябрь 2013.
12. Huang, C. R., Calzolari, N., Gangemi, A., Lenci, A., & Oltramari, A. (Eds.).
13. (2010). Ontology and the lexicon: A natural language processing perspective. Cambridge University Press.
14. Eshonqulov H. I. Ontologiyalar va bilimlarni ifodalash, avtomatlashtirilgan
15. axborot tizimlarining ontologiyasini qo‘llash asoslari [Matn] / Eshonqulov H.I. Buxoro: Durdona, 2022. – 128 b.
16. Лукашевич Н.В. Тезаурусы в задачах информационного поиска. –

17. Москва, 2010. – 388 с.
18. Никоненко, А. А. (2009). Обзор баз знаний онтологического типа.
19. Соловьев В.Д., Добров Б.В., Иванов В.В., Лукашевич Н.В. Онтологии и
20. тезаурусы, Учебное пособие. Казань/Москва. 2006. 157 с.
21. Sharipbay, A., Shirinova, R., Abdurakhmonova, N., & Qodirova, Z. (2023).
22. ONTOLOGICAL KNOWLEDGE BASED MODELS REPRESENTING MEDICINE TERMINOLOGY. *Advances in Social Sciences and Economics*, *1*(1), 4.
23. Dilrabo, X. (2023). BADIIY MATN VA UNGA LINGVOPOETIK
24. YONDASHUV. UNIVERSAL JOURNAL OF SOCIAL SCIENCES, PHILOSOPHY AND CULTURE, 1(8), 80-94.
25. Ibodullayeva, Q. R. Q. (2023). BADIIY MATNDA RUHIY KECHINMA VA
26. OBRAZLILIK. Academic research in educational sciences, 4(10), 168-175.
27. Kuvondikova, G. I., & To‘rayeva, S. (2024). O ‘ZBEK TILIDAGI
28. KO‘CHMA MA’NOLI SO ‘ZLARNI FRANSUZ TILIGA TARJIMA QILISH XUSUSIYATLARI. Educational Research in Universal Sciences, 3(2), 254-257.
29. Ismailov A., Abduraxmonova N., Shamsiyeva G. Statistical machine
30. translation proposal for Uzbek to English. Science and Education 2(12), 212-219, 2021
31. Shamsiyeva G. O‘zbek ingliz tillari uchun parallel Korpusga asoslangan
32. Mashina tarjimasi texnologiyasi. INTERNATIONAL SCIENTIFIC CONFERENSES WITH HIGHER EDUCATIONAL INSTITUTIONS 1…, 2023
33. Shamsiyeva G. State of syntactic analysis of text in machine translation.
34. INTERNATIONAL SCIENTIFIC CONFERENSES WITH HIGHER EDUCATIONAL INSTITUTIONS 1…, 2022
35. Shamsiyeva G.A. O‘zbek tili uchun tarjima texnologiyasini
36. avtomatlashtirishning lingvistik asoslari. INTERNATIONAL CONFERENCE ON LEARNING AND TEACHING 1 (8), 520-528, 2022
37. https://tami.uz/php

BILINGVAL TEZAURUS YARATISHDA O ‘ZBEK VA INGLIZ TILLARINING VALENTLIK MUNOSABATI

Babajanova Lobar
Urganch davlat universiteti Xorijiy filologiya fakulteti
“Tarjima nazariyasi va amaliyoti” kafedrasi tayanch doktoranti.

Annotatsiya: Tezaurus lug‘atlari, so‘zlar va ularning o‘zaro aloqalarini tahlil qilish uchun yaratilgan ma’lumotlar bazasi hisoblanadi. Ular bir-biriga yaqin aloqada bo‘lgan so‘zlar to‘plamidir. Bu lug‘atlarning asosiy maqsadi, bir so‘zning ma’nosi bo‘yicha boshqa so‘zlarni va ularga o‘xshash va aloqador yana boshqa so‘zlarni topishdir. Ushbu tezisda tezaurus, aniqrog‘i, bilingval tezaurus yaratishda o‘zbek va ingliz tillarining valentlik munosabatlari haqida fikr-mulohazalar yuritiladi.

Kalit so‘zlar: oktant, argument, bilingval tezaurus, valentlik munosabati, valentlik, tezaurus, tilshunoslik.

Abstract: Thesaurus is a database created for the analysis of dictionaries, words and their interactions. They are a special set of related words. The main purpose of these dictionaries is to find other words in the meaning of a word and to find similar words. This thesis discusses the valence relations of Uzbek and English languages in creating a thesaurus, a bilingual thesaurus.

Keywords: actant, argument, bilingual thesaurus, valence relation, valency, thesaurus, linguistics.

Аннотация: Тезаурус – это база данных, созданная для анализа словарей, слов и их взаимодействий. Они представляют собой набор родственных слов. Основная цель этих словарей – найти другие слова по значению и найти похожие слова. В данной тезисе рассматриваются валентные отношения узбекского и английского языков при создании тезауруса, двуязычного тезауруса.

Ключевые слова: актант, аргумент, валентное отношение, валентность, двуязычный тезаурус, тезаурус, лингвистика.

Tilshunoslikda valentlik tushunchasining paydo bo‘lishi XIX asr oxirlariga borib taqaladi ya’ni o‘sha paytlarda Charlz Sanders Pirs ilk marta o‘z asarlarida “valentlik metaforasi” iborasini qo‘llagan. Lekin mazkur atamani izohlagan va tilshunoslikka alohida atama sifatida olib kirgan shaxs Lusiyen Tesniere hisoblanadi. Olimning 1950-yillardagi faoliyati tilshunoslik sohasidagi valentlik munosabatlari tushunchasini rivoj topishida muhim qadamlardan bo‘ldi. Xususan, boshqa tilshunoslar ham valentlik munosabatlarini tushunishga sezilarli darajada hissa qo‘shgan, misol sifatida Kellogg Allertonni olishimiz mumkin. Tilshunos lingvistikada semantik va sintaktik valentlik tushunchalarini bir-biridan farqlagan. [1] Tilshunoslikda valentlik tushunchasi bir fe’lning grammatik jihatdan to‘liq bo‘lishi va gap tuzilishi uchun talab qiladigan so‘zlar (elementlar yoki argumentlar) miqdorini va turini anglatadi. Shuningdek, bu atama fe’lning gap sifatida to‘liq shakllanishi va ma’no ifodalashi uchun qancha ishtirokchisi yoki subyekti bo‘lishi kerakligini tasvirlaydi. Yana ham soddalashtiradigan bo‘lsak, valentlik bu -fe’lning boshqa so‘z turkumlari bilan aloqa o‘rnata olish qobiliyatidir. Valentlik tushunchasi nima ekanligini anglab

yetish gap tuzilishini chuqurroq o‘rganishga va fe’llarning boshqa so‘z turkumlari bilan munosabatini tahlil qilishga yordam beradi. Bu tushuncha sintaktik tahlil metodining markaziy tushunchalaridan biri hisoblanadi. Valentlik nazariyasining asoschisi bo‘lmish Lusiyen Tesniereni bu tushunchani quyidagicha turkumlaydi:

Fe’llar atomlar sifatida: Tesniere fe’lni atom bog‘lanishlariga qiyoslaydi. Atomlar ma’lum bir miqdordagi elektronlarning muvozanatda bo‘lishini talab qilganidek, fe’llar ham to‘liq jumla va yoki gap hosil qilish uchun aniq sonli birikmalarni (u bu elementlarni oktantlar deb ataydi) talab qiladi. Bu analogiyada gap tuzilishiga e’tibor qaratgan holda valentlikning sintaktik tomoni ta’kidlanadi.

Aktantlar va adyunkt (ravish): Tesniere farqlagan oktant tushunchasi fe’lning ma’nosi uchun juda muhim hisoblanib, u fe’lga bog‘lanib kelib, uni izohlashga va ma’nosini to‘ldirishga yordam beradi. Masalan “John is cold” va “John became chairman” jumlalaridagi “cold” va “chairman” so‘zlari aktant hisoblanadi. Uning fikricha oktantlarni jumlada qo‘llash majburiy, ravishni bo‘lsa qo‘llash ixtiyoriy, u faqat fe’lga qo‘shimcha ma’no beradi. Misol uchun, “They arrived on Sunday” gapida on Sunday so‘zi adyunkt hisoblanadi, buni gapda qo‘llash ham qo‘llamaslik ham mumkin ya’ni gapning ma’nosiga katta ta’sir ko‘rsatmaydi. [2]

Tesniere fe’llarning valentligiga qarab, ya’ni qancha miqdorda oktant qabul qilishiga ko‘ra quyidagi kategoriyalarga bo‘ladi:

Avalent (0 oktant): Bu fe’llar o‘zlari mustaqil ravishda to‘liq jumlalarni hosil qilishlari mumkin, misol uchun, “It rained” (yomg‘ir yog‘di) kabi.

Monovalent (1 oktant): Bu turdagi fe’llar bir dona actant qabul qiladi, masalan, "sleep" (uxlamoq) fe’li, "John sleeps" (Jon uxlaydi) gapida John so‘zi aktant va sleep fe’li bir dona aktant bilan bog‘lanib kelyapti.

Divalent yoki ditranzitiv fe’llar (2 oktant): Bu fe’llar ikkita oktant qabul qiladi, masalan, "give" (bermoq) fe’li "Mary gave a book to John" (Mariya Jonga kitobni berdi.) gapida “book” va “John” so‘zlari give fe’liga bog‘langan oktantlardir.

Trivalent (3 oktant oladigan fe’llar): Bu fe’llar 3 ta oktant talab qiladi, misol qilib, "tell" (aytmoq, demoq) fe’lini oladigan bo‘lsak, "The teacher told a story to the children" gapida tell fe’li “teacher”, “story”, “children” so‘zlari bilan bog‘lanib, o‘zaro valentlik munosabatlarini o‘rnatgan.

Tesnierening valentlik nazariyasiga oid ushbu g‘oyalari gap tuzilishi va fe’llarga bog‘langan elementlarni qanday bog‘lanishi va ularning o‘zaro munosabatlari bo‘yicha olib boriladigan kelgusidagi tadqiqotlarga asos bo‘ldi desak hech ham mubolag‘a bo‘lmaydi. Bilingval tezaurus yaratish borasida ham fe’llarning valentlik munosabatlari hisobga olinadi, shuning uchun ham bu tushunchaga alohida yondashish talab etiladi.

Bilingval tezaurus yaratishda fe’llarning valentlik munosabatlarini chuqur anglab yetish juda ham ahamiyatlidir. Eng avvalo, bilingval tezaurus nima va undan qanday maqsadlarda foydalanilishini bilib olishimiz zarur. Tezauruslar, odatda, ikki xil bo‘ladi: monolingual (bir tilli) va bilingval (ikki tilli). Bilingval tezaurus bu ikki tilda mavjud so‘zlar va unga bog‘liq boshqa so‘zlarni sinonimlari bilan taqdim qiladigan maxsus, murakkab lug‘at turidir. Bu turdagi tezaurus tarjima jarayonida ikki tildagi ibora va atamalarning ekvivalent va adekvatlarini topishga yordam beradi. Xo‘sh, bilingval tezaurusga kimlar murojaat qiladi:

Til o‘rganuvchilari: til o‘rganish jarayonida o‘rganuvchilarga o‘rganayotgan tilida sinonimlar va unga bog‘liq so‘zlarni topishga va o‘rganishga yordam beradi, so‘z boyligini kengaytiradi va o‘z fikrlarini aniq ifodalash qobiliyatini oshiradi.

Tarjimonlar: bilingval tezauruslar tarjimonlarga o‘girilayotgan tilda so‘z va iboralar uchun eng aniq va eng mos keluvchi ekvivalentlarni topishda yordam berishi mumkin va tilmochlarning eng yaqin hamrohlaridan biri hisoblanadi.

Qo‘shimcha qilib aytadigan bo‘lsak, bilingval tezauruslar madaniyatlararo tushunchalarning turli tillarda qanday ifodalanganini tushunish va madaniyatlararo tafovutlarni anglashda va chuqurroq tushuncha hosil qilishda ko‘maklashishi mumkin. Tarjima jarayonida so‘zlarni boshqa tilga o‘girishda, ayniqsa fe’llarni tarjima qilishda so‘zlarning valentlik munosabatlarini inobatga olish juda ahamiyatlidir. Bir tildagi mavjud fe’l boshqa tilda umuman boshqacha oktantlar(argumentlar)ni talab qilishi mumkin. Bilingval tezauruslar valentlik munosabatlarini to‘g‘ridan-to‘g‘ri hisobga olmasa-da, o‘xshash valentlik munosabatdagi so‘zlar, bir xil oktant qabul qiladigan fe’llar va ularning sinonimlarini topishda juda ham samaralidir.

Tilshunoslikda turli xil ko‘nikmalarga asoslangan holda bilingval tezaurus yaratish munosabati mavzusini chuqurroq tushunishga yordam beruvchi metodlar mavjud, ya’ni so‘zlarni aniq topishni osonlashtirish, til o‘rganuvchilar va tarjimonlar uchun ma’lumotlar to‘plamini mustahkamlashda yaxshi qo‘llanma sifatida ishlatilishi mumkin bo‘lgan materiallar keltirib o‘tilgan. Tezaurusda "valentlik munosabati" tushunchasi har bir so‘zga o‘ziga xos ma’nolar va atrof-muhit bilan bog‘liq bo‘lgan misollar va izohlar qo‘shilgan. Tezaurusning bilingval xususiyati, o‘zbek va ingliz tillarini integrativ ravishda bir-biriga bog‘lab, munosabatlarni eng to‘liq ifodalovchi va o‘rganishga oson bo‘lishiga imkoniyat yaratadi. Bilingval tezaurus yaratish jarayonida, o‘zbek va ingliz tillarining "valentlik munosabati" mavzusiga oid so‘zlar va ularning ma’nolari, izohlarini birlashtirish uchun quyidagilarni amalga oshirdik:

Mavzuni belgilash: bilingval tezaurus yaratishda o‘zbek va ingliz tillarining valentlik munosabatiga oid so‘zlar to‘plamini tuzishdan oldin, bu mavzuga oid yirik ko‘lamli mavzu nomini belgilash lozim. Mavzuning boshqa aspektlari ham ko‘rsatilgan bo‘lsa, ularga ham alohida qism tuziladi.

Ma’nolar va izohlar: Har bir so‘zga o‘ziga xos ma’no va izoh qo‘shish kerak. O‘zbek va ingliz tillaridagi to‘g‘ri so‘zlarni tanlab, ularni ma’no va izohlari bilan birga qo‘shish lozim.

Misollar va gaplar: Har bir so‘zga misollar va gaplar keltirib o‘tiladi. Bu so‘zning qanday holatda ishlatilishi mumkinligini tushuntiradi va izohlashni osonlashtiradi.

Integrativ bog‘lanishlar: O‘zbek va ingliz tillaridagi o‘zaro mosliklarni ko‘rsatuvchi so‘zlar tanlanadi. Bu tezaurusdan foydalanish jarayonida birmuncha qulayliklar yaratib beradi.

Qo‘shimcha mavzular: Mavzuga oid boshqa mavzularni ham qo‘shish orqali tezaurusni kengaytirish lozim.

O‘zgartirishlar va yangilanishlar: Tezaurusni rivojlantirish uchun foydalanuvchilar, til o‘qituvchilari, yozuvchilar va boshqa so‘zlashuvchilar bilan hamkorlik qilinadi. O‘zgartirishlar va yangilanishlar uchun to‘g‘ridan-to‘g‘ri mutaxassislar bilan maslahatlashish lozim [6, 78-b].

Odatda, biror-bir so‘zning ma’nolarini emas, balki so‘zning o‘zini qidirilganda, tezauruslarga murojaat qilinadi [3, 234-b]. Ya’ni bu yerda tushuncha ma’lum bo‘lsa-da, biroq shu tushuncha ostidagi so‘zlar guruhi yoki so‘z shakllari haqida ma’lumotlar olinadi. Bu tizim ichida turgan bo‘lishi mumkin-u, ammo so‘zning o‘zi nimaligini bilmasligimiz mumkin. Shu bois, manbalarda ta’kidlanganidek, tezaurus – bu lug‘at bo‘lib, so‘zlar uchun omborxona vazifasini o‘taydi[Abdurakhmonova, N.2020,2021,2022]. Tezaurus umumiy ma’noda – maxsus terminologiya, ya’ni lug‘at, ma’lumotlar yig‘indisi, korpus yoki jamlanma, ma’lum sohadagi

bilimlar yoki faoliyat sohasidagi tushunchalar, ta'riflar va atamalarni to'liq o'z ichiga oladi [5, 3-b].

Tezaurusning bilingval shakli yaratilishi, o'zbek va ingliz tillarini integrativ ravishda bir-biriga bog'lash orqali til o'rganuvchilari va tarjimonlar uchun qulay, tushunarli va ma'lumotlarni to'plagan eng muhim manba bo'lishi mumkin.

Bilingval tezaurus yaratishda o'zbek va ingliz tillarining valentlik munosabati uchun so'zlar va iboralardan foydalanilgan qulayroq. Bu so'zlarni barcha bo'limlarga ajratib, ularning ma'nolari va izohlarini qo'shishingiz mumkin. O'zbek va ingliz tillarida tezaurus yaratishda boshlang'ich va qo'shimcha so'zlar qo'shish, izohlarni tafsilotlar bilan qo'shish va ularga misollar keltirish bu tezaurusni mukammal va yanada samarali qila oladi. Bu misollar bilingval tezaurusda "valentlik munosabati" mavzusini o'zbek va ingliz tillarida tushuntirishga yordam beradi.

Tezaurus mualliflar, muharrirlar va tarjimonlarga yordam berishi mumkin. Biroq, tezaurusdagi ma'lumotlardan foydalanish ehtiyotkorlik bilan va to'g'ri bajarilishi kerak [Abdurakhmonova, N.2020,2021,2022]. So'zlarning ma'nolari va qo'llanishlarini to'liq tushunmaslik ularning noto'g'ri ishlatilishiga olib kelishi mumkin. Shuning uchun, so'zlarni ishlatishdan oldin, so'zning ma'nosi va ishlatilishini to'liq tushunish muhimdir. Tezaurus til o'rganish, matn tasnifi, sinonimlarni topish, konsepsiyani xaritalash va ma'lumot olish kabi sohalarda muhim o'rin tutadi. Tezaurus so'z ma'nolari va ishlatilishini yaxshiroq tushunish uchun ishlatilishi mumkin. Xususan, yozuvchilar, muharrirlar va tarjimonlar so'z tanlash va imloni tekshirish uchun aforizmlardan foydalanishlari mumkin [4, 5-b].

Yuqoridagi munozara va ilmiy tahlillardan xulosa qilishimiz mumkin bo'ladi-ki bilingval tezaurus yaratishda o'zbek va ingliz tillarining o'zaro leksik, grammatik, sintaktik, morfologik munosabatlarini va aloqadorligini qiyoslash va o'rganish bilingval tezaurus yaratishda eng samarali va qulay uslublardan biri hisoblanadi. Buning uchun esa, avvalo, valentlik tushunchasini sinchiklab o'rganish va tahlil qilish talab etiladi.

Foydalanilgan adabiyotlar:

1. Höllein D. Valency Theory. Oxford Bibliographies, 2023.
2. Abduhamidova L. Tilshunoslikning yangi yo'nalishi: kompyuter lingvistikasi. -Toshkent, 2015.
3. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).
4. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
5. Abdurakhmonova, N., & Ismailov, A. S. (2022). APPLYING WEB CRAWLER TECHNOLOGIES FOR COMPILING PARALLEL CORPORA AS ONE STAGE OF NATURAL LANGUAGE PROCESSING. In *СОВРЕМЕННАЯ ФИЛОЛОГИЯ. СОЦИАЛЬНАЯ И НАЦИОНАЛЬНАЯ ВАРИАТИВНОСТЬ ЯЗЫКА И ЛИТЕРАТУРЫ* (pp. 22-27).

6. Abduraxmonova, N., & Abduvaxobov, G. I. (2021). O ‘quv lug ‘atini tuzishning nazariy metodologik asoslari. *СЎЗ САНЪАТИ ХАЛИҚАРО ЖУРНАЛИ*, *103*.
7. Sulevmanov, D., Gatiatullin, A., Prokopyev, N., & Abdurakhmonova, N. (2020, November). Turkic morpheme web portal as a platform for turkology research. In *2020 International Conference on Information Science and Communications Technologies (ICISCT)* (pp. 1-5). IEEE.
8. Taylor A. The Proverb. Harvard University Press, 1985. –234 p.
9. Mohira, S. (2023). TEZAURUS LUG ‘ATLARINING TARJIMA SOHASIDA TUTGAN O ‘RNI. Innovations in Technology and Science Education, 2(8), 1038-1043.
10. Suyunov, B. (2021). Tezaurus va kompyuter texnologiyalariga doir. Computer Linguistics: Problems, Solutions, Prospects, 1(1).
11. Лукашевич Н. В. Тезаурусы в задачах информатsіонного поиска. М.: Изд-во МГУ, 2011.]

O‘ZBEK TILIDA ELEKTRON LUG‘ATLAR YARATISH ZARURATI

Abduvaxobov G‘iyosiddin G‘ayratjon o‘g‘li
Farg‘ona davlat universiteti dotsenti, PhD.

Annotatsiya: Maqolada an’anaviy lug‘at va elektron lug‘atdan foydalanishining ijobiy hamda salbiy tomonlari mavjudligi haqida so‘z boradi. Amaliy tilshunoslikda elektron lug‘atga bo‘lgan ehtiyoj dinamikasi o‘sib borar ekan. Ularda ma’lum konsepsiyalarni shakllantirishga talab kuchayadi. Bu, ayniqsa, chet tillarini o‘rganish jarayonida etolon tilda mavjud bo‘lgan so‘z shakllarining to‘g‘ri qo‘llanilishiga xizmat qiladi. Shuningdek, maqolada lug‘atlarni alohida sohalarga ajratish, ularni elektron lug‘at shakliga keltirish jarayonlari aks ettirilgan.

Kalit so‘zlar: elektron, Elektron leksikografiya, denotatsiya, kompyuter lingvistikasi, algoritim, lug‘at, korpus, analizator, kompyuter leksikografiyasi, dastur.

Аннотация: В статье речь будет идти о наличии положительных и отрицательных сторон использования традиционных и электронных словарей. По мере роста динамики потребности в электронном словаре в прикладной лингвистике возрастает спрос на формирование его определенных концепций. Это способствует правильному использованию словоформ, существующих в эталонном языке, особенно в процессе изучения иностранных языков. Помимо этого, в статье отражены процессы разделения словарей по отдельным сферам применения, приведения их к форме электронного. словаря.

Ключевые слова: электронный, электронная лексикография, денотат, компьютерная лингвистика, алгоритм, словарь, корпус, анализатор, компьютерная лексикография, программное обеспечение.

Abstract: The article will focus on the presence of positive and negative aspects of using traditional and electronic dictionaries. As the dynamics of the need for an electronic dictionary in applied linguistics grows, the demand for the formation of certain concepts of it increases. This contributes to the correct use of word forms that exist in the reference language, especially in the process of learning foreign languages. In addition, the article reflects the processes of dividing dictionaries into separate areas of application, bringing them to the electronic form. dictionary.

Keywords: electronic, electronic lexicography, denotation, computer linguistics, algorithm, dictionary, corpus, analyzer, computer lexicography, software.

Kirish

Bugungi kunda dunyoda turli madaniyatlarning vujudga kelishi, tillar rivojlanishi va ularni o‘zlashtirishga oid texnik imkoniyatlarning kengayishi jamiyat kishilari dunyoqarashini tubdan o‘zgartirdi. Jamiyat tomonidan tilga alohida talab va ehtiyoj paydo bo‘ldiki, bu ko‘proq elektron lug‘atlar tuzish masalasiga aloqadordir. Bu talab va ehtiyojlar leksikografik nazariyaning yangi bir bosqichga ko‘tarilishiga olib keldi.

Yigirma yildan ko‘proq vaqt mobaynida elektron lug‘at tushunchasi lingvistika sohasida tor ma’noda, faqatgina kompyuter lingvistikasi doirasida tushunilib kelindi, soha muammolari faqat elektron texnologiyalar joriy etilishi bilan aloqador ekanligi haqidagi o‘ta tor ilmiy qarashlar natijasida kompyuter leksikografiyasi tilshunos olimlar tomonidan yetarlicha e’tiborga olinmadi.

Aslida mazkur atama va u bilan bog‘liq tushunchalar tilshunoslik sohasi rivoji va uning bugungi kundagi jamiyat ehtiyojlariga to‘laqonli javob berishiga xizmat qiladi. Bunda amaliy tilshunoslikning leksikografiya sohasi katta o‘zgarishlarga – takomillashuvga uchradi. Elektron lug‘at tushunchasi tilshunoslikka bir qancha jiharlar bilan turlicha ta’sir ko‘rsatdi.

Tilshunoslik qadimdan matematika fani bilan bog‘liq holda rivojlanib kelgan: ularning tub mohiyati bir o‘zakka – mantiqqa borib taqaladi. Amaliy tilshunoslikning rivoji bu fikrni yana bir bor quvvatlash bilan birga yana-da samaraliroq qildi. Yangi paydo bo‘lgan texnologiyalar matnlarni to‘g‘ridan-to‘g‘ri elektron formatga tarjima qilishni ta’minlash bilan jarayonni osonlashtirgan bo‘lsa, har bir tilning tabiatidan kelib chiqib, matnlarning o‘ziga xos shakllantirilishini anglab yetishda foydalanuvchi mas’uliyatini oshirdi.

An’anaviy lug‘atlar bilan elektron lug‘at o‘rtasida farqlar bo‘lishi tabiiy. Bu farqlar u yoki bu lug‘at shaklining ustunligini ko‘rsatadi. Ya’ni bir jihat bilan an’anaviy lug‘atlar afzallik qilsa, ikkinchi bir jihatdan elektron lug‘atlardan foydalanish samaraliroq bo‘ladi. Masalan, an’anaviy lug‘atlardan foydalanish uchun moddiylik, unga erishish uchun lozim bo‘lgan sharoit va amallar (lug‘at olish uchun do‘konga yoki kutubxonaga borish zarurati, uni topish, kutubxona bazasining quvvati yoki sotib olish uchun sarflanadigan moddiy qiymat) elektron lug‘atlarda mutlaqo talab qilinmasligi pandemiya sharoitida foydalanuvchilarga juda qo‘l keldi. Etalon til sir-asrorlarini bilish va chet tillarini o‘rganishda elektron lug‘atlardan foydalanish qulayligi yana bir bor o‘z isbotini topdi. Shu o‘rinda leksikografiya sohasi va uning yana-da samarali ishlash prinsiplariga to‘xtalmoqchimiz[Abdurakhmonova, N.2020,2021,2022].

Leksikografiyada ro‘y berayotgan tub o‘zgarishlar tilshunoslikning boshqa fanlar bilan integratsiyaga kirishuvi, hamkorlikda rivojlanishini taqozo etadi. Bunda yuzaga kelgan vaziyat va unda mavjud muammolarga yechim topishda tilshunoslikning texnika fanlari bilan aloqasi talab qilinadi. Tilshunoslikning texnika fanlari bilan aloqasi asosida yangi injenerlik lingvistikasi, hisoblash lingvistikasi, elektron lingvistika atamalari paydo bo‘ldi. Ushbu konsepsiya ostida tilni o‘rab turgan bir qator muhim muammolar hali ham dolzarb bo‘lib qolmoqda. Kompyuterlar orqali tizimli til o‘rganish, mashina tarjimasi va leksikografiya masalalari bu sohadagi muhim masalalar sirasiga kiradi. Kompyuter leksikografiyasiga doir yangi ilmiy qarashlar filologiya fanlari doktori N.Abduraxmonovaning “Mashina tarjimasining lingvistik ta’minoti” monografiyasida o‘z aksini topgan. Manbada hozirgi o‘zbek tilida kompyuter leksikografiyasi yutuqlari bilan bog‘liq ko‘plab lug‘atlar yaratilgani, ammo ushbu lug‘atlar haligacha kitob shaklida bo‘lib, elektron lug‘at holatiga keltirilmaganligi ta’kidlanadi [2, 59].

Tilshunos va informatik olim A.Pedron o‘zining “Elektron leksikografiya” deb nomlangan asarida an’anaviy lug‘atlar bilan yangi avlod lug‘atlar o‘rtasidagi farqni Venn diagrammasi asosida tushuntirib bergan[4,238]. Muallif X.Bergenxolzning elektron leksikografiya haqidagi qarashlariga murojaat qilib, yuqoridagi savollarga aniqlik kiritildi. Uning ta’kidlashicha, kompyuter leksikografiyasida elektron lug‘atlarni ishlab chiqish jarayonida lug‘atlardan foydalanuvchilar o‘z maqsadlariga tezroq erishadilar va elektron lug‘atlarni yaratishda har ikki soha vakillari, ya’ni tilshunos va axborot texnologiyalari soha vakillari birdek ishtirok etadi deb hisoblaydi[Abdurakhmonova, N.2020,2021,2022].

Avtomatik ma’lumot berish, mashina tarjimasi, inson-mashina suhbati tabiiy til materialining dasturlash tillari asosida kodlanishiga asoslanadi. Faqat ular orasida muayyan farqlanishlar mavjud, masalan, tabiiy tilda amalda bo‘lgan polisemantik so‘zlar dasturlash tilida alohida qiymat bilan kodlanadi. Rivojlanish shuni ko‘rsatadiki, yaqin kelajakda an’anaviy

tilshunoslikda mavjud bo‘lgan bir qator munozarali masalalar kompyuter lingvistikasi orqali hal qilinishi mumkin bo‘ladi.

Zamonaviy tilshunoslikda lingvistik masalalarni avtomatlashtirishga ham katta urg‘u berilgan. O‘zbek tilshunosligida bu haqdagi ilmiy mulohazalar tadqiqotchi A.Rahimovning “Kompyuter lingvistikasiga kirish” o‘quv-uslubiy qo‘llanmasida uchraydi[3]. Tilshunoslikning kompyuter dasturlari bilan yaqin aloqadorlikdagi tadqiqotning yangi predmeti – bu mashinaning kibernetik miyasi aks ettira oladigan til strukturasini shakllantirishdir. Bu yerda talabning izchil yechilishi uchun hisoblash tizimi (algoritm) ishlab chiqiladi. Algoritm tili kompyuterlarda har qanday dasturlashni amalga oshirish uchun ishlatiladi. Bunday dasturiy ta’minotlarni ishlab chiqish va uni kompyuter lingvistikasi orqali tilshunoslikning barcha bo‘limlariga aloqadordir, shu sababli uning talqinida tishunoslikning barcha bo‘limlari tadqiqini qayta ko‘rib chiqish zarurdir.

Kompyuter lingvistikasi sohasida o‘zbek tilshunosligining tadrijiy taraqqiyoti mashina tarjimasi, matnlarni rasmiy tahlil qilish, turli lug‘atlarni tuzish va boshqalar bilan bog‘liq. Alohida ta’kidlash lozimki, yangi davr tilshunosligi soha masalalari yechimiga aniq fanlar mutaxassislarining yana-da ko‘proq jalb qilinishi bilan xarakterlidir[Abdurakhmonova, N.2020,2021,2022]. Til materialini tadqiq etish bilan bir paytda filologiya sohasida ilmiy daraja oluvchi mutaxassislar soni ham ortib bormoqda. Jumladan, 2021-yilda N.Abduraxmonova “O‘zbek tili elektron korpusining kompyuter modellari” mavzusidagi filologiya fanlari doktori (DSc) ilmiy darajasini olish uchun yozgan dissertatsiyasida kompyuter texnologiyalarining til rivojiga va, o‘z navbatida, til hodisalarining kompyuter texnologiyalariga ijobiy ta’siri orqali qator ilmiy yutiqlar qo‘lga erishilishi alohida ta’kidlaydi. Shuningdek, 2022-yilda Sh.Gulyamova tomonidan “O‘zbek tili semantik analizatorining lingvistik asoslari” mavzusidagi filologiya fanlari doktori (DSc) dissertatsiyasida o‘zbek tili omonimlar lug‘atini takomillashtirish va uning semantik analizatordagi o‘rni masalasi ilmiy asoslandi.

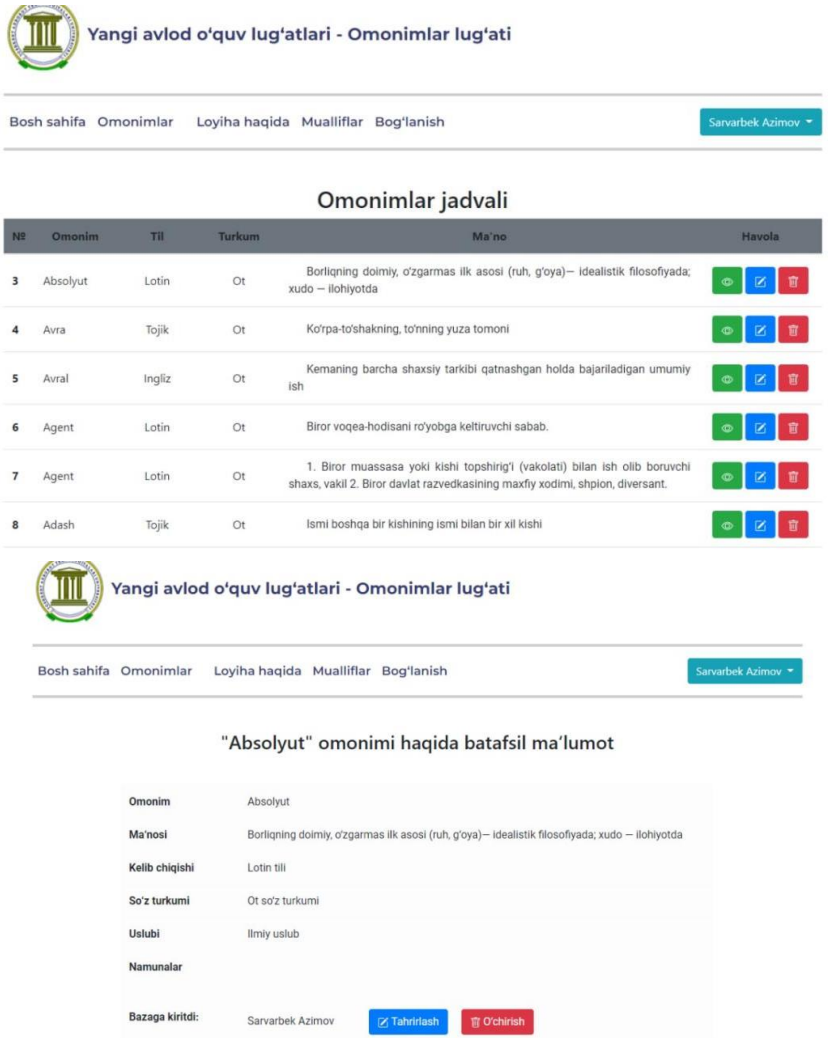
Umuman, kuzatishlar shuni ko‘rsatadiki, tilshunoslikning boshqa fanlar bilan aloqasi unga zarracha zarar yetkazmaydi, aksincha, uning jadal rivojlanishiga olib keladi.

“Elektron lug‘at” tushunchasi tilshunoslik fanida so‘nggi yigirma yil mobaynida paydo bo‘lgan atamadir. “Kompyuter leksikografiyasi” atamasi esa nisbatan ilgari paydo bo‘lgan. Mazkur tushunchaning ilmiy talqini Y.N.Marchukning 1976-yilda nashrdan chiqqan “Hisoblash leksikografiyasi”, E.V.Vertelning 1984-yilda nashr qilingan “Mashina leksikografiyasi”, O.A.Kazakevichning 1985-yildagi “Avtomatlashtirilgan leksikografiya”, V.V.Markovkinning 1990-yilda yaratilgan “Kompyuter leksikografiyasi” nomli darslik va o‘quv qo‘llamalarida bayon etilgan. Xususan, Y.N.Marchuk ushbu atamalarni “Kompyuter lingvistikasi” sarlavhasi ostida umumlashtirdi va tilshunoslikning dolzarb muammolarini hal qilish zamonaviy dasturiy yo‘naltirilgan kompyuterlar yordamida bevosita muhokama qilinishini ta’kidladi. Q.M.Mandrikova esa “hisoblash leksikografiyasi”, “hisoblash mashinasi” va “avtomatik leksikografiya” atamalarining bir tushuncha ekanligini va ularning mohiyatida bitta vosita – kompyuter mavjudligini ta’kidlaydi. U “kompyuter leksikografiyasi” tilshunoslikning alohida sohasi sifatida, leksikografiyaning yangi ko‘rinishi sifatida ko‘rib chiqish lozim, degan xulosaga keladi. Ushbu fanning mazmuni har xil turdagi kompyuter lug‘atining nazariy va amaliy masalalarining mavjudligi, shuningdek, ushbu muammo bilan bog‘liq barcha natijalarni nashr etishdir deb hisoblaydi[Abdurakhmonova, N.2020,2021,2022].

O‘zbekiston Respublikasi Prezidentining 2020–yil 20-oktabrdagi PF-6084-son farmoniga ko‘ra yangi avlod o‘quv lug‘atlari yaratish samaradorligi tanqidiy-tahliliy o‘rganib chiqilmoqda.

Maskur amaliy jarayonda muayyan konsepsiyaning mavjud bo‘lishi va amal qilinishi sohaga oid muammolarni hal qilish va tizimli ijobiy natijalarga erishishni ta’minlaydi. Shu ma’noda, tanlangan mavzu yangi avlod lug‘atlarini yaratish ishlari shu yo‘nalishdagi konsepsiyalar ishlab chiqish bilan boshlanishi maqsadlidir.

O‘zbek tilining yangi avlod lug‘atlarini yaratilishi, bir so‘z bilan aytganda, yuqorida ko‘rsatilgan Farmonda nazarda tutilgan maqsadlarga xizmat qiladi. Tadqiqodning amaliy natijasi Vazirlar Mahkamasi huzuridagi O‘zbek tilini rivojlantish jamg‘armasi tomonidan moliyalashtirilgan “Yangi avlod o‘quv lug‘atlari va ularning mobil ilovalarini yaratish” loyihasi doirasida amalga oshirilayotgan o‘zbek tilining omonim, sinonim va paronimlar lug‘ati mobil versiyalari ishlab chiqilganligi bilan belgilanadi. Ushbu lug‘atlarning web shakli bugungi kunda tilshunos.com sayitida mumtazam to‘ldirib borilmoqda.



Global Internet leksikografiya faniga o‘z ta’sirini o‘tkazib, elektron lug‘atlar jarayoni uchun qulay sharoit yaratdi. Kompyuter texnologiyalari lug‘at tuzish ishlarini tezlashtirdi, lug‘atlar bilan ishlashni osonlashtirdi. Internetning ta’siri natijasida lug‘atlar o‘z shaklini butunlay o‘zgartirib yubordi. Lug‘atlar kompyuter lug‘atlari va elektron lug‘atlar nomi bilan ishlay boshladi. Kompyuter lug‘atlari va elektron lug‘atlarning farqi juda katta. Bu asosan qidiruv funksiyasi cheklanganligida, so‘rovga javob bermasligi, foydalanuvchi bilan o‘zaro aloqada bo‘lmasligi va texnik funksiyalarga ega emasligida ko‘rinadi. Elektron lug‘atlar asosan an’anaviy lug‘atlardan foyda ko‘radi, ammo bu yerda ishlash bosma lug‘atlarga qaraganda ancha oson. Elektron

lug‘atlarda qidirish cheklanmagan, so‘zlar alifbo tartibida joylashtirilgan va foydalanuvchi talabiga to‘la javob beradigan qabul qiluvchiga ega. Elektron lug‘atlarning paydo bo‘lishi an’anaviy lug‘atlarning shakli va mazmunini o‘zgartirdi. Katta hajmdagi lug‘atlar allaqachon elektron vositalarga (CD, Fleshka) o‘tkazilgan. An’anaviy lug‘atlardan farqli o‘laroq, elektron lug‘atlarda asosiy ish nafaqat so‘zlarning alifbo tartibida, balki elektron lug‘atlarda so‘z va iboralarni tabaqalashtirilgan sanab o‘tish, ularni misollar bilan tushuntirish, keng lug‘at yozuvlari, eng so‘nggi multimedia elementlari bilan zamon talablari asosida taqdim etilgan. Ya’ni so‘zlarning elektron lug‘atlarda ifodalanishi rangli, obrazli, grafik, tovushli va videotasvirlar bo‘lishi mumkin. Elektron lug‘atlarda so‘zlarning izohi til me’yorlariga muvofiq berilishi kerak. Lug‘atda so‘zning ma’nosini ochib berishda kontekst va situatsion ma’nolar, omonimlar, sinonim va antonimlar, frazeologik birikmalar, idiomalardan foydalanish mumkin. Lug‘at imkoniyatlari keng bo‘lib, matnlarni turli tillarga tarjima qilish imkoniyati mavjud.

Adabiyotlar ro‘yxati:

1. Abduraxmonova N. Mashina tarjimasining lingvistik ta’minoti [Linguistic database of machine translati]. – T.:2018.
2. Pedro A. Fuertes-Olivera., Bergenholtz H. E-Lexicography. The Internet, Digital initiateves and Lexicography. London/New York: Continuum. 2011. – bet 341
3. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
4. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
5. Abdurakhmonova, N., & Ismailov, A. S. (2022). APPLYING WEB CRAWLER TECHNOLOGIES FOR COMPILING PARALLEL CORPORA AS ONE STAGE OF NATURAL LANGUAGE PROCESSING. In СОВРЕМЕННАЯ ФИЛОЛОГИЯ. СОЦИАЛЬНАЯ И НАЦИОНАЛЬНАЯ ВАРИАТИВНОСТЬ ЯЗЫКА И ЛИТЕРАТУРЫ (pp. 22-27).
6. Abduraxmonova, N., & Abduvaxobov, G. I. (2021). O ‘quv lug ‘atini tuzishning nazariy metodologik asoslari. СЎЗ САНЪАТИ ХАЛИҚАРО ЖУРНАЛИ, 103.
7. Sulevmanov, D., Gatiatullin, A., Prokopyev, N., & Abdurakhmonova, N. (2020, November). Turkic morpheme web portal as a platform for turkology research. In 2020 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-5). IEEE.
8. Rahimov A. Kompyuter lingvistikasi asoslari [Basics of computer linguistics]. – T.: Akademnashr, 2011.
9. Boboyev A.M. Tilshunoslikka kirish [Introduction to Linguistics]. – Baku.: Maorif, 2004.
10. Rizayev S. Kibernetika va tilshunoslik [Cybernetics and Linguistics]. – T., O‘zbekiston, 1976

O‘ZBEK TILI TEZAURUSINI YARATISHDA STANDART ME’YORLAR

Abduraxmonova Nilufar
O‘zbekiston milliy universiteti
Kompyuter lingvistikasi va amaliy tilshunoslik
kafedrasi mudiri, filologiya fanlari doktori, professor.

Xoliyorova Gulruhsor
O‘zbekiston milliy universiteti
Kompyuter lingvistikasi va amaliy tilshunoslik
kafedrasi o‘qituvchisi.

Annotatsiya: Mazkur maqolada sun’iy intellektning so‘z ma’nolarini aniqlash va ma’nolarni bog‘liq shaklarda tushuntirishdagi muammolar tahlilga tortilgan. Maqola tadqiqotchilar tomonidan olib borilgan WordNet va strukturasi, ulardagi mavzuiy bog‘lanishlar, ma’lumotlarni izlash tizimlarida tezaurusning ahamiyati kabi muhim masalalarni ilmiy jihatdan taqqoslashga asoslangan. Tezaurus yaratish uchun lug‘atning mikro va makro sxemasini loyihalashtirishda ingliz tili va boshqa tillar uchun yaratilgan lingvistik me’yorlar o‘zbek tili tezaurusi uchun namuna bo‘lib xizmat qilishi mumkin. Olimlar tomonidan tadqiq etilgan tezaurus turlari va ular uchun tan olingan lingvistik me’yorlar tavsifi bayon etilgan.

Tayanch so‘zlar: tezaurus, WordNet, konseptual masofa, ideografik lug‘at, monolingval tezaurus.

Abstract: This article analyzes the problems of artificial intelligence in determining the meanings of words and explaining the meanings in related forms. The article is based on a scientific comparison of important issues such as WordNet and its structure, thematic links in them, the importance of the thesaurus in information retrieval systems, conducted by researchers. Linguistic standards created for English and other languages can serve as a model for the thesaurus of the Uzbek language when designing the micro and macro scheme of the dictionary for creating a thesaurus. The types of thesauruses researched by scientists and the linguistic norms recognized for them are described.

Keywords: thesaurus, WordNet, conceptual distance, ideographic dictionary, monolingual thesaurus.

Аннотация: В данной статье анализируются проблемы искусственного интеллекта при определении значений слов и объяснении значений в родственных формах. Статья основана на научном сопоставлении таких важных вопросов, как WordNet и его структура, тематические связи в них, значение тезауруса в информационно-поисковых системах, проведенном исследователями. Лингвистические стандарты, созданные для английского и других языков, могут служить моделью тезауруса узбекского языка при проектировании микро- и макросхемы словаря для создания тезауруса. Описаны виды тезаурусов, исследованные учеными, и признанные для них лингвистические нормы.

Ключевые слова: тезаурус, WordNet, концептуальная дистанция, идеографический словарь, одноязычный тезаурус.

Sun’iy intellekt uchun bilimlarni aniqlash, qolaversa matnning og‘zaki va yozma shakldagi mazmunini tushunishda wordnet- tezaurus juda muhim ahamiyat kasb etadi. WordNET tezaurusning bir turi hisoblanadi. E.Agirre WordNet tipidagi lug‘atlardagi har bir so‘z mavzuiy belgilariga ko‘ra ma’no jihatdan sohaviy vektorlarga bog‘lanishini ta’kidlaydi. Uningcha, har bir lug‘atdagi so‘zlar mavzuiy vektor hisoblanib, ularning og‘irligi ikkinchi so‘z ma’nosiga bog‘liqdir [Eneko Agirre1, Enrique Alfonseca2, and Oier Lopez de Lacalle Approximating Hierarchy-Based Similarity for WordNet Nominal Synsets using Topic Signatures -P. 15 / GWC 2004 Second International WordNet Conference, GWC 2004 Brno, Czech Republic, January 20–23, 2004 Proceedings (CD-ROM version)]. Tadqiqot mualliflari ikki metod orqali so‘z ma’nosiga bog‘langanligini ko‘rsatib o‘tishadi:

1. Asosiy so‘z ma’nosining umumiy munosabatlarda foydalanish (sinonim, giponim);
2. Asosiy so‘zning faqay monosemantik munosabatlarda foydalanish.

Yuqoridagi dastlabki metod orqali polisemantik semalar bo‘yicha asosi so‘z ma’nosiga u darajada aniq bo‘lmagan namunalarni to‘plasa, ikkinchi metoddda yanada to‘liqroq namunalarga erishish uchun bu muammolarni chetlab o‘tadi. Ushbu tadqiqotda dastlab so‘zlar lemmatizatsiya qilinadi, so‘ng kontekstga kalit so‘z sifatida ishlatilgan har bir so‘z chastotasiga ko‘ra vektor hosil qilinadi (bag of words).

Biroq shuni ham aytish o‘rinliki, kontekstda uchragan har bir so‘zning chastotasi bo‘yicha xulosa chiqarish to‘liq xulosani bermaydi. Chunki tabiiy til turli nutq pozitsiyasida turli leksik va grammatik qo‘shimchalarni olishi natijasida kutilgan natijaga olib kelmasligi mumkin.

E.Agirre tezaurus tarkibidagi so‘zlarning o‘xshashlik me’yorlarini aniqlash uchun WordNetda mavjud bo‘lgan asosiy standartni iyerarxik bog‘lanish uchun asos qilib oladi [Eneko Agirre1 , Enrique Alfonseca2 , and Oier Lopez de Lacalle Approximating Hierarchy-Based Similarity for WordNet Nominal Synsets using Topic Signatures -P. 15 / GWC 2004 Second International WordNet Conference, GWC 2004 Brno, Czech Republic, January 20–23, 2004 Proceedings (CD-ROM version)]. - Synsetning informatsion kontentiga asoslangan Resnik masofasi metrikasi;

Konseptual masofa deb nomlanuvchi (Conceptual Distance) WordNet iyerxiyasidagi ikki sinset o‘rtasidagi giporonimik bog‘lanishlarning minimal miqdordagi inversiyasi. Yuqorida nomi zikr qilingan olimlarning asosiy yondashuvi shundan iboratki, o‘xshashlik mezoni va uch bosh mezon o‘rtasidagi korrelyasiya sifat o‘lchovi sifatida hizmat qiladi. Kerrelyatsiya quyidagi bosqichlardan iborat bo‘ladi: birinchi bosqichda har bir ot so‘z turkumi uchun simmetrik o‘xshashlik matritsasi mavuiy belgilariga ko‘ra hisoblanadi. Ikki matritsa o‘rtasidagi korrelyatsiya vektorlarga transforma qilinib hisoblanadi. Agar o‘xshashlik juda yaqin bo‘lsa 1, teskarisi bo‘lsa, 0 raqam qo‘yiladi. Shu usul orqali so‘zlarning ma’no jihatdan yaqinligi aniqlanadi[Abdurakhmonova, N.2020,2021,2022].

Tezaurus borasida dunyo tillari bo‘yicha juda ko‘plab ishlar amalga oshirilgan. Har qaysi tilda yaratilgan WordNetni olmaylik ingliz tilida yaratilgan WordNet asosida amalga oshirilgan. Tezaurusni modellashtirishda konseptlarning o‘rni juda muhim sanaladi. Prinston WordNeti turli tillar uchun namuna vazifasini o‘taydi. Ayniqsa, ko‘p tilli semantik resurs sifatida quyidagi jihatlari bilan ahamiyatlidir [Russian WordNet From UML-notation to Internet/Intranet Database Implementation Valentina Balkova 2, Andrey Sukhonogov 1, and Sergey Yablonsky – P.31.]:

- Informatsion aniqlovchi instrumentariy;
- Mualliflarni aniqlovchi instrumentariy;

- Tilni o‘rganishga mo‘ljallangan instrumentariy;
- Tarjima instrumentariysi;
- Qisqartirish;
- Semantik web.

Tezaurus ideografik lug‘at sifatida tilning barcha jihatlarini inobatga oladi. Masalan, rus tili korpusi uchun quyidagi lug‘atlar kiritilgan:

1) rus tili gGrammatik lug‘ati;
2) kompyuter lug‘ati;
3) geografik nomlar lug‘ati;
4) kishi nomlari, ptaronim va familiyalar lug‘ati;
5) biznes lug‘ati;
6) huqushunoslik lug‘ati;
7) jargon so‘zlar lug‘ati
8) izohli lug‘at.
9) Rus tili Russicon tezaurusi
10) Rus tili Orfografik lug‘ati

Lingvistik ma’lumotlar bazasi sifatida barcha lug‘atlar *text* fayl sifatida kiritilgan bo‘lib, har bir leksik birlik so‘z t*urkumi, kelishik shakli, jinsi, son, zamon, shaxs-son, daraja, nisbat, mayl, aspect, shakl, tur, o‘timli-o‘timsizligi, jonli yoki jonsiz* ekanligiga ko‘ra teglangan. Mualliflar rus tili tezaurusini yaratishda EuroWordNet tizimidagi namunadan faydalangan:

- Birlashgan yondashuv (merge approach): monolingval leksik resursdan taksonomiyani qurish, so‘ngra bilingval lug‘atdan foydalanib, moslashtirish jarayonini amalga oshirish (mapping);
- Kengaytirilgan yondashuv: bilingval lug‘atlardan foydalanib, ingliz tili sinsetiga so‘zlarni bevosita moslashtirish.

2007-yildagi holatiga rus tili tezaurusidagi (WordNet.ru) [Russian WordNet From UML-notation to Internet/Intranet Database Implementation Valentina Balkova 2, Andrey Sukhonogov 1, and Sergey Yablonsky – P.31.] so‘zlar statistikasi:

Russian WordNet Word Report					
Total	Noun	Verb	Adj	Adv	Other
111749	44751	27997	20736	4997	13268

Olimlar tezaurusning konseptual modelini ishlab chiqishda bir nechta yondashuvlar bo‘yicha ishlarni ilmiy-amaliy jihatdan amalga oshrishga harakat qilganlar.

Axborot qidiruv tizimi uchun tezaurusning ahamiyati katta. Foydalanuvchi tomonidan yuboriladigan so‘rov orqali olinadigan ma’lumotlar semantik to‘rga, ya’ni tezaurusga bog‘langan bo‘ladi. So‘zlar va konseptlar o‘rtasidagi leksik va semantik munosabatlarni formal vositalar orqali ifodalash mumkin [Edward A.F. , Nutter J.T. etc. Building a large thesaurus for information retrieval // - P. 101]. Lingvistik lug‘atlarda so‘zlarning leksik sematik jihatdan munosabatlari sinonim, antonim, omonim kabi munosabatlari bilan cheklanib qoladi, biroq axborot qidiruvi uchun so‘zlarning taksonomik va iyerarhik munosabatlarining ham berilishi katta ahamiyatga ega. J. Edward va boshqalar tomonidan olib borilgan tadqiqotlarda axborot qidiruv tizimi uchun katta hajmdagi tezaurus yaratishda mashina o‘qiy oladigan lug‘atlar uchun matnlarni qayta ishlash metodidan foydalanib, maqolalarga berilgan izohlar analiz qilingan. Semantik tarmoqdagi

munosabatlar bilan olingan ma’lumotlar birlashtirilgan. So‘ng Smart va CODER eksperimental qayta ishlash tizimidan foydalanib, tezaurusdagi birliklar tekshirilgan.

J.Aitchison, A.Gilchrist va D.Bawden kabi olimlar tomonidan tezaurus modellari bo‘yicha izlanishlar olib borilgan. Ularning “Thesaurus construction and use: a practical manual” [J.Aitchison, A.Gilchrist va D.Bawden Thesaurus construction and use: a practical manual (4-edtition), Taylor & Francis e-Library, 2005. -202 p.] nomli qo‘llanmasida tezaurusga doir ilmiy tadqiqotlar tahlilga tortilgan. Unga ko‘ra axborot qidiruv tizimida tezaurus 4 xil imkoniyatga ega bo‘lishi mumkin:

- tezaurus qidiruv va indeksatsiya uchun foydalaniladi;
- tezaurus indeksatsiyada foydalaniladi, qidiruvda foydalanilmaydi;
-tezaurus qidiruvda foydalaniladi, indeksatsiyada foydalaniladi;
-tezaurus har ikkisida ham foydalanilmaydi.

Tezaurus yaratishning bir nechta usullari mavjud. Yuqorida tadqiqotchilar tomonidan keltirilgan tezaurus turlariga qarab ularning qurilish arxitekturasi va kontenti ham talablari ko‘p tilli yoki bir tilli tezaurus uchun standartlar yaratiladi. Dastlab 1974-yilda Xalqaro standartlashtirish tashkiloti tomonidan bir tilli tezaurus uchun yaratilgan. Keyingilari 1985-yillarda nashr qilingan. 1967-yilda nashr qilingan Injineriya va ilmiy terminlar Amerika tezaurusi alifbo tartibida tuzilgan. Monolingval tezaurus uchun milliy standartlar Britaniya standartini o‘z ichiga olgan. 1959-yildan 1993-yilgacha bo‘lgan evolyutsion yo‘riqnomasi va tezaurus tuzish tamoyillarining shakllanish bosqichlari Krooks va Lankaster tomonidan tadqiq etilgan [J.Aitchison, A.Gilchrist va D.Bawden Thesaurus construction and use: a practical manual (4-edtition), Taylor & Francis e-Library, 2005. – P. 25.].

J.Aitchison fikricha tezaurus yaratishning dastlabki bosqichlaridan biri bu uning tarkibiga kiritiladigan birliklarning grammatik shakli, talaffuzi, ko‘plik shakli, qisqartma yoki qo‘shma shaklda ekanligi tashkil etadi. Ikkinchi bosqichda u yoki bu konseptga tegishli konseptni ifodolovchi ikki yoki undan ortiq sinonimlarning mosini tanlash masalalari o‘rganiladi. Uchinchi bosqichda muayyan turga kiruvchi terminlarni, jumladan o‘zlashma so‘zlar, slenglar, tijorat nomlar va atoqli otlarni aniqlash kabilar bajariladi[Abdurakhmonova, N.2020,2021,2022]. To‘rtinchi, tezaurusning xususiy maqsadidan kelib chiqqan holda terminlarning turli kontekstda u yoki bu ma’noda qo‘llanishini cheklash uchun ayrim qoidalar yaratiladi. Ushbu cheklov omonimlarni aniqlash imkonini beradi. Ingliz tili terminologiyasida controlled language nazorat qilingan tilde indekslangan termin tanlangan (preffered yoki non preferred term) termin sifatida qo‘llanadi. Tanlangan termin berilgan konseptni ifodalashda ishlatiladi. U deksriptor yoki kalit so‘z deb ham nomlanadi. Tanlanmagan termin sinonim yoki kvazi sinonim sifatida ham qo‘llanadi. Ushbu birliklar indekslash uchun emas, balki tanlangan terminga mos birlikni aniqlashga yordam beradi.

Mualliflarning qayd etishicha, indekslangan terminlar orqali ifodalangan konseptlar umumiy kategoriyalarge tegishli bo‘ladi. Ushbu kategoriyalar fasatlar analizida fundamental kategoriyalardan iborat bo‘ladi. Ular konkret birliklar, mavhum konseptlar va atoqli otlardan iborat bo‘ladi.

Konkret birliklar:

Narsalar va fizik qismlar:

BINOLAR

QAVATLAR

OROLLAR
Materiallar
SEMENT
YOG‘OCH
ALYUMIN
Abstrakt konseptlar
Faoliyat va voqealik
BOSHQARUV
URUSHLAR
MAROSIMLAR
Abstrakt birliklar, narsalar, materiallar, faoliyatlar
QONUN
NAZARIYA
KUCHLANISH
SAMARADORLIK
Sohalar va fanlar
FIZIKA
METEROLOGIYA
PSIXOLOGIYA
O‘lchov birliklari
KILOGRAM
METR
Atoqli otlarga tegishli Individual birliklar
NIGERIA
INSON HUQUQLARI KOMISSIYASI

Shu tarzda standarrt birliklarning kategoriyalari keltiriladi. Shuningdek, indekslangan terminlarga ot va otli frazalar keltiriladi. Otli frazaga kiruvchi umumiy shakllar sifat frazadan iborat bo‘ladi.

Masalan,
AYOL ISHCHILAR
BOSMA NASHRLAR
SHAHAR JAMOAT JOYLARI

Standartlarga ko‘ra, garchi ayrim tezauruslarda vaqt, holat, o‘lcham, hajm, vaziyat va xususiyat kabil sifatlar kiritilgan bo‘lsa-da, indekslanuvchi terminlar sirasiga sifatlar qabul qilinmagan. ANSI standartlarida sifatlarni alohida holatda ko‘rsatish tavsiya qilinmaydi.

Olimlar ushbu fikrga namuna sifatida quyidagilarni keltirishadi:
<PORTABLE=> portable typewriter
USE TYPWRITER and AUXILIARY>
RECTANGULAR=> Rectangular windows
USE WINDOWS AND RECTANGULAR>

Shuningdek, ayrim standartlarda ularni otli birikmalarning tarkibi sifatida keltirish o‘rinli deb qaralgan:
MINIATURE SIZE
PORTABE DEVICES kabi.

Ravish so‘z turkumiga tegishli birliklar, ayniqsa very, highly kabilar tezaurusga kiritilmaydi, ular faqat muayyan birikmaning tarkibida keltiriladi: very large scale integration

Tezaurusda fe’l so‘z turkumi ham kiritilmaydi, uning o‘rniga u hosil qilgan ot, fe’l ishtirokidagi otli birikma, gerundiy kabilar kiritiladi: COMMUNICATION (COMMUNICATE emas), ADMINISTRATION (ADMINISTER emas), WALKING (WALK emas).

Ingliz tilida yaratilgan tezauruslarda artikllar ham kiritilmaydi.

Xulosa o‘rnida shuni e’tirof o‘rinliki, o‘zbek tili tezaurusi uchun yaratiladigan ma’lumotlar bazasi tabiiy tilni qayta ishlash jarayoni uchun, qolaversa, axborot qidiruv tizimi uchun moslanganligi kelgusida sun’iy intellekt texnologiyalarida tadqiqotlar uchun muhim manba bo‘lib xizmatt qiladi.

Foydalanilgan adabiyotlar:

1. Eneko Agirre1, Enrique Alfonseca2, and Oier Lopez de Lacalle Approximating Hierarchy-Based Similarity for WordNet Nominal Synsets using Topic Signatures -P. 15 / GWC 2004 Second International WordNet Conference, GWC 2004 Brno, Czech Republic, January 20–23, 2004 Proceedings (CD-ROM version)
2. Edward A.F., Nutter J.T. etc. Building a large thesaurus for information retrieval // - P. 101
3. J.Aitchison, A.Gilchrist va D.Bawden Thesaurus construction and use: a practical manual (4-edtition), Taylor & Francis e-Library, 2005. -202 p.
4. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).
5. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
6. Abdurakhmonova, N., & Ismailov, A. S. (2022). APPLYING WEB CRAWLER TECHNOLOGIES FOR COMPILING PARALLEL CORPORA AS ONE STAGE OF NATURAL LANGUAGE PROCESSING. In *СОВРЕМЕННАЯ ФИЛОЛОГИЯ. СОЦИАЛЬНАЯ И НАЦИОНАЛЬНАЯ ВАРИАТИВНОСТЬ ЯЗЫКА И ЛИТЕРАТУРЫ* (pp. 22-27).
7. Abduraxmonova, N., & Abduvaxobov, G. I. (2021). O ‘quv lug ‘atini tuzishning nazariy metodologik asoslari. *СЎЗ САНЪАТИ ХАЛҚАРО ЖУРНАЛИ*, *103*.
8. Sulevmanov, D., Gatiatullin, A., Prokopyev, N., & Abdurakhmonova, N. (2020, November). Turkic morpheme web portal as a platform for turkology research. In *2020 International Conference on Information Science and Communications Technologies (ICISCT)* (pp. 1-5). IEEE.
9. Russian WordNet from UML-notation to Internet/Intranet Database Implementation Valentina Balkova2, Andrey Sukhonogov1, and Sergey Yablonsky – P.31.
10. Russian WordNet from UML-notation to Internet/Intranet Database Implementation Valentina Balkova2, Andrey Sukhonogov1, and Sergey Yablonsky – P.34.

THE PROBLEMS OF REPRESENTING NEAR-SYNONYMS IN A COMPUTATIONAL LEXICON

Hasanova Gulhida
PhD student, Uzbekistan State World Languages University.

Abstract: The given thesis makes an attempt to show the difficulties of representing near synonymy in computational lexicon base highlighting the dimensions of variation that near synonyms can present. In particular, denotational, stylistic, structural and expressive variations of near synonyms that are vital determining the differences of near synonyms are briefly overviewed to draw the conclusion on the topic. Based on the differentiation of near synonyms and conclusions drawn from them six main problems of representing near synonyms in a computational lexicon are stated. Knowledge base model that is supposed to be sufficient enough to solve the problems mentioned in the article is going to be a main topic of our further studies.

Keywords: near synonyms, dimensions of variation, denotational variation, stylistic variation, structural variation and variations in expressiveness.

Аннотация: В данной работе делается попытка показать трудности представления близкой синонимии в вычислительной лексиконе, подчеркивая размеры вариаций, которые могут представлять близкие синонимы. В частности, кратко рассмотрены денотационные, стилистические, структурные и экспрессивные варианты близких синонимов, имеющие существенное значение для определения различий близких синонимов, чтобы сделать вывод по теме. На основе дифференциации близких синонимов и сделанных на их основе выводов сформулированы шесть основных проблем репрезентации близких синонимов в вычислительном лексиконе. Модель базы знаний, которая должна быть достаточной для решения упомянутых в статье задач, станет основной темой наших дальнейших исследований.

Ключевые слова: околосинонимы, виды различий, денотационные различия, стилистические различия, структурные различия и различия в выразительности.

A word's core "dictionary" definition is not the only meaning, it has also its connotations, and attitudes. Also, a word frequently contains near-synonyms that are different from it only in above mentioned subtle semantic differences. Therefore, one must carefully weigh the distinctions between all of the possibilities in order to determine the appropriate word to employ in each given situation—that is, the phrase that properly expresses the desired meaning while avoiding undesired implications. Even for humans, selecting the correct term can be challenging, let alone for modern computer systems. For instance, with so many possible translations that are similar but somewhat different, how can a machine translation (MT) system choose the optimal English word for the French *bevue*? *Error, mistake, blunder, slip, lapse, boner, faux pas, boo-boo*, and so forth are among the options that the system may select; however, the best option depends on the context in which the word is used as well as the semantic differences between it and the other English terms. The algorithm needs to ascertain the subtleties that *bevue* conveys in the specific situation that it has been employed in. Alongside selecting other words simultaneously, it must also determine which English word (or words) most nearly express the same subtleties. Since *bevue* is most likely

as distinct from each of its potential translations as it is from each other, an exact translation is probably not achievable. That is to say, every translation option will, on average, either communicate or withhold some potentially unpleasant detail. As a result, accurate translation necessitates a complex lexical-choice procedure that can identify the closest or most suitable near-synonym for a word in another language among those offered by one language in a given circumstance.

More generally, a sophisticated lexical-choice mechanism is also necessary for a natural language generation (NLG) system to be fully articulate. The system needs to be able to consider the possible outcomes of each choice. Think about the possibilities of a brand-new thesaurus for a word processor that works with the writer by rating the terms based on how well they fit the context and the writer's overall preferences, rather than just providing a list of synonyms. Many writers would profit immensely from such an intelligent thesaurus, which would be a clear upgrade over the basic thesauri found in word processors today.

A thorough computational model of fine-grained lexical knowledge is required. However, while being a basic linguistic phenomenon that shapes lexicon structure, synonymy has received significantly less attention in linguistics, psychology, lexicography, semantics, and computational linguistics than polysemy, which is also a fundamental and well-studied phenomenon.

Whether for philosophical, pragmatic, or expedient reasons, synonymy is frequently regarded as a "non-problem": either there are synonyms, but they have exact same meanings and are therefore simple to handle, or there are none, in which case each word can be treated like any other. However, our research on near-synonymy reveals that it is an as intricate phenomenon as polysemy, and that it has an inherent impact on the organization of lexical knowledge.

Even native language speakers may find it challenging to distinguish between near-synonyms well enough to utilize them consistently or to communicate those distinctions when they are understood. Furthermore, using the incorrect word could imply something you don't intend to. As a result, lexicographers have produced a large number of reference volumes that clearly distinguish between members of near-synonym groups (sometimes referred to as "dictionaries of synonyms").

Any study of near-synonyms must start with the idea of difference since, if two apparent absolute synonyms aren't actually the same, then they must be distinct. Difference, according to Saussure[31] (1916, page 114), is essential to the formation and delineation of meaning:

In a given language, all the words which express neighboring ideas help define one another's meaning. Each of a set of synonyms like redouter ('to dread'), craindre ('to fear'), avoir peur ('to be afraid') has its particular value only because they stand in contrast with one another.... No word has a value that can be identified independently of what else is in its vicinity.

Changes between near-synonyms are frequently remarkably complex. The near-synonyms vary not only in how they express different ideas and concepts, like blameworthiness and misconception, but also in how the ideas are implied, suggested, expressed, connoted, and stressed, how frequently they are expressed (common, occasionally, not always), and how strongly they are expressed (in strength).

[31] Saussure F. Cours de linguistique g ´en ´erale. Translated by Roy Harris as Course in General Linguistics. − London: G. Duckworth, 1983.

Dimensional variations. Generally speaking, any feature of the meaning of near-synonyms can have following variations[32] (Cruse 1986):

- denotational variations, in a broad sense, including propositional, fuzzy, and other peripheral aspects
- stylistic variations, including dialect and register
- expressive variations, including emotive and attitudinal aspects
- structural variations, including collocational, selectional, and syntactic variations

Variations in Denotation. Denotation, in its broadest meaning, is involved in several types of variation. While certain differentiae are easily described in terms of discrete abstract (or symbolic) qualities, like continuous/binary, many are not, according to DiMarco, Hirst, and Stede[33] (1993). Actually, the majority of differences in denotational variation are related to full-fledged concepts or ideas rather than basic facts; these distinctions pertain to roles and components of a scenario. For example, the notions that make up an *enemy* and a *foe* are not the same; the former emphasize animosity, while the latter emphasizes active warfare instead of emotional reaction[34] (Gove 1984). These indirect connotations are typically incidental to the major meaning that a word conveys, and it is typically difficult to determine definitely whether or not they were ever intended to be given by the speaker. Denotational differences can also be ambiguous as opposed to distinct.

Stylistic differences. Differences in a finite set of dimensions, on which all words may be compared, are what stylistic diversity entails. Stede (1993)[35], Nirenburg & Defrise (1992)[36], Hovy (1988)[37], and others have all put out a number of stylistic dimensions. Two of the most popular measurements are formal and informal.

Variations that are expressive. Numerous near-synonyms have different markings depending on the speaker's perspective on whether their denotation is positive or negative. Thus, the same person may be labeled as thin or slender if the speaker intended to be more favorable, thin or deprecating, or thin if he wanted to be neutral.

Structural differences. The final category of differences between near-synonyms consists of limitations on how they can be used as well as limitations that other utterance elements place on how they can be used. In either scenario, the limitations are unrelated to the definitions of the terms. The limitations can be syntactic, collocational, or selectional, meaning they rely on other ideas indicated or on other words or components in the speech.

Collocational variety refers to the concepts or words that a term might be paired with, sometimes in an idiomatic way. For instance, the collocational patterns of *task* and *job* are

[32] Cruse D. Lexical Semantics. − Cambridge University Press. 1986.
[33] DiMarco, Chrysanne, Hirst G., and Stede M. The semantic and stylistic differentiation of synonyms and near-synonyms. //In AAAI Spring Symposium on Building Lexicons for Machine Translation. − Stanford: 1993pages 114–121..
[34] Gove, Philip B. Webster's New Dictionary of Synonyms. Merriam-Webster. − Springfield, 1984.
[35] Stede, Manfred. Lexical choice criteria in language generation. //In Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics. – 1993.pages 119- 124.
[36] Nirenburg, Sergei and Christine D. Application-oriented computational semantics. In Michael Rosner and Roderick Johnson, editors, Computational Linguistics and Formal Semantics. − Cambridge University Press, 1992. pages 223–256.
[37] Hovy, Eduard. Generating Natural Language Under Pragmatic Constraints. − Lawrence Erlbaum Associates. 1988.

different: a person can encounter a difficult task, but not a difficult job. This is a lexical constraint; in contrast, the class of eligible objects is determined semantically, not lexically, in selectional restrictions (or preferences). Unlike *die*, for instance, *pass away* can only refer to people (or anthropomorphized pets); it cannot refer to plants or animals.

Different syntactic subcategorizations lead to variations in syntactic limitations. It is assumed that a group of words belong to the same syntactic category if they are synonyms or nearly synonyms. On the other hand, some near-synonyms may fall into a different subcategory than others. For instance, the adjective *open* can be used in any position, whereas the adjective *ajar* can only be used predicatively (The door is ajar; the ajar door).

Multilingual near synonymy. The rule in lexical transfer in translation is near-synonymy rather than synonymy: the word in the target language that is most similar to that in the source text may be a near-synonym rather than an exact synonym. For instance, the German word *wald* and the English word *forest* have similar meanings, but *wald* can also refer to a smaller and more urbanized area of trees than *forest*; in other words, *wald* includes some of the meaning of the English word *woods*, and in some cases, *woods* would be a more accurate translation of *wald* than *forest*. Therefore, we can talk about the distinctions in a group of cross-linguistic near-synonyms just as we can with a group of near-synonyms from a single language. In addition, reference books exist that offer guidance on cross-linguistic near-synonymy to translators and advanced second-language learners, much as they do for single-language near-synonym groups.

In summary, to account for near-synonymy, a model of lexical knowledge will have to incorporate solutions to the following problems:

• The four main types of variation are qualitatively different, so each must be separately modeled.
• Near-synonyms differ in the manner in which they convey concepts, either with emphasis or indirectness (e.g., through mere suggestion rather than denotation).
• Meanings, and hence differences among them, can be fuzzy.
• Differences can be multidimensional. Only for clarity in our above explication of the dimensions of variation did we try to select examples that highlighted a single dimension.
• Differences are not just between simple features but involve concepts that relate roles and aspects of the situation.
• Differences often depend on the context.

References:

1. Cruse D. Lexical Semantics. − Cambridge University Press. 1986.
2. DiMarco, Chrysanne, Hirst G., and Stede M. The semantic and stylistic differentiation of synonyms and near-synonyms. //In AAAI Spring Symposium on Building Lexicons for Machine Translation. − Stanford: 1993 pages 114–121.
3. Gove, Philip B. Webster's New Dictionary of Synonyms. Merriam-Webster. − Springfield, 1984.
4. Hirst G. Near-synonymy and the structure of lexical knowledge. //In AAAI Symposium on Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity. – Stanford: 1995, pages 51–56.
5. Hovy, Eduard. Generating Natural Language Under Pragmatic Constraints. − Lawrence Erlbaum Associates. 1988.

6. Inkpen D., Hirst G. Building a lexical knowledge-base of near-synonym differences. // Workshop on WordNet and Other Lexical Resources, Second Meeting of the North American Chapter of the Association for Computational Linguistics. – Pittsburgh: 2001, 47–52 p
7. Nirenburg, Sergei and Christine D. Application-oriented computational semantics. In Michael Rosner and Roderick Johnson, editors, Computational Linguistics and Formal Semantics. − nCambridge University Press, 1992. pages 223–256.
8. Saussure F. Cours de linguistique g ´en ´erale. Translated by Roy Harris as Course in General Linguistics. − London: G. Duckworth, 1983.
9. Stede, Manfred. Lexical choice criteria in language generation. //In Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics. – 1993.pages 119- 124.

TEZAURUSLAR VA ULARNI TAHLIL QILISH TAMOYILLARI

Oybekova Sarvinoz Oybekovna
O‘zbekiston Milliy universiteti
Kompyuter lingvistikasi magistratura mutaxassisligi
2-bosqich magistranti.

Annotatsiya: Mazkur maqolada tezauruslar, ularning lug‘atlar yaratishdagi ahamiyati hamda tasnif qilish tamoyillari haqida gap boradi. Tezaurus lug‘atlarining afzalliklarini misollar yordamida izohlanadi.
Kalit so‘zlar: tezaurus, lug‘at, semantic bog‘lanish, WordNet.

Abstract: This article discusses thesauruses, their importance in creating dictionaries, and principles of classification. The advantages of thesaurus dictionaries are explained with the help of examples.
Keywords: thesaurus, dictionary, semantic connection, WordNet.

Аннотация: В статье рассматриваются тезаурусы, их значение при создании словарей и принципы классификации. Преимущества тезаурусных словарей поясняются на примерах.
Ключевые слова: тезаурус, словарь, семантическая связь, WordNet.

Tezaurus lug‘atlari, so‘zlar va ularning o‘zaro aloqalarini tahlil qilish uchun yaratilgan ma’lumotlar bazasi hisoblanadi. Ular bir-biriga yaqinligi bo‘lgan so‘zlar to‘plamidir. Bu lug‘atlarning asosiy maqsadi, bir so‘zning ma’nosi bo‘yicha boshqa so‘zlarni topish va ularga o‘xshash so‘zlar topishdir. Bu matnlar yoki turli xil ma’lumotlar bazalarida ishlatilishi mumkin. Tezaurus lug‘atlari, matnlarni tahlil qilishda va tarjimada yordam berishda keng ishlatiladi. Ular matnda o‘xshash yoki o‘zaro bog‘liq so‘zlar ro‘yxatini ko‘rsatadi. Bu matnda qidirayotgan so‘zni topishga yordam beradi. Shuningdek, tezaurus lug‘atlari, matnlarni taqsonlarga bo‘lishda va taqsonlarnishakllantirishda ishlatiladi. Tezaurus lug‘atlari, boshqa lug‘atlarga nisbatan bir necha foydali imkoniyatlar taqdim etadi. Ular so‘zlarning ma’nolarini tahlil qilish, ularning o‘zaro aloqalarini aniqlash, so‘zlar va ularning aloqadorligini tushuntirish va matnlarni taqsonlarga bo‘lishda yordam beradi. Bu, matnlar va turli ma’lumotlar bazalaridagi ishlarimizda juda foydali bo‘ladi. Tezaurus lug‘atlari, dastlabki qarashda, lug‘atlar bilan bir xil deb o‘ylansa ham, o‘zaro farqlanuvchi xususiyatlarga ega.

Tezauruslarni tasnif qilish muammosi bir necha o‘n yillar davomida bir qator chet el tilshunoslari (V.V. Morkovkin, L.P.Stupin, K.Marello va va boshqalar.) tomonidan atroflicha o‘rganilgan. Ushbu tasniflarning asosini quyidagi mezonlar tashkil etadi:

1. Orasida semantik bog‘lanishlar mavjud bo‘lgan lug‘at birliklari.
2. Lug‘at hajmi.
3. So‘z boyligini umumlashtirish.
4. Leksemalarning ma’nosini rivojlantirish.
5. Grammatik-stilistik leksema malakasi.
6. Taqdim etilgan tillar soni;

7. Leksemaning semiotik vosita turi, semantizatsiya uchun ishlatilishi.

Tezaurus lug‘atlari, shu jumldan, bir so‘z yoki ifodaning o‘ziga xos ma’nolarini va ulardan boshqa yuzlab o‘zgaruvchilarni ko‘rsatadi. Bu yuzlab o‘zgaruvchilar, asosan, Giperonim, Giponim, Meronim,Partonim va Halonimlardir.

Giperonim, bir so‘zning umumiy yoki katta ma’nolarini ifodalovchi so‘zdir. Misol uchun, "hayvon" so‘zi "it", "sigir", "qo‘y" va boshqa hayvonlar bilan bog‘liq so‘zlar uchun giperonim bo‘lib hisoblanadi.

Giponim, esa, bir so‘zning kichik yoki to‘g‘ri ma’nolarini ifodalovchi so‘zdir. Misol uchun, "it" so‘zi "dalmachiq", "pudralik", "pishiq" va boshqa kichik hayvonlar bilan bog‘liq so‘zlar uchun giponim bo‘lib hisoblanadi.

Meronim, bir obyektni tarkibiy qismlarini ifodalovchi so‘zdir. Misol uchun, "avto" so‘zi "motor", "shassi", "faralar" va boshqa avtomobilning qismlari bilan bog‘liq so‘zlar uchun meronim bo‘lib hisoblanadi.

Holonim, esa, bir obyektni tarkib qiladigan butunliklarni ifodalovchi so‘zdir. Misol uchun, "uy" so‘zi "xona", "o‘zi", "qavat" va boshqa uyning qismlari bilan bog‘liq so‘zlar uchun holonim bo‘lib hisoblanadi.

Partonim, esa, bir obyektni tarkib qiladigan qismlarni ifodalovchi so‘zdir. Misol uchun, "kitob" so‘zi "sahifa", "qog‘oz", "qop" va boshqa kitobning qismlari bilan bog‘liq so‘zlar uchun partonim hisoblanadi. Halonim, esa, bir so‘zning o‘zining o‘zini tushunish uchun zarur bo‘lgan so‘zdir. Misol uchun, "qalam" so‘zi "qog‘oz", "yozuv", "to‘qima" va boshqa qalamni tushunish uchun zarur bo‘lgan so‘zlar uchun halonim bo‘lib hisoblanadi.

Tezaurus lug‘atlari, so‘zlar orasidagi yaqin munosabatlarni aniqlashda va matnlarni taxmin qilishda yordam beradi. Ularning asosiy manbalari WordNet va Roget's Thesaurus kabi kompyuter dasturlari hisoblanadi. Tezaurus lug‘atlari, matnlar tarjimasida ham ishlatiladi, chunki ular so‘zlar orasidagi munosabatlarni aniqlashda yordam beradi. [5, 78-79]

Tezaurus lug‘atlari shunday afzalliklarga egaki, ular:

1. So‘zlar orasidagi munosabatlarni aniqlashga yordam beradi: Tezaurus lug‘atlari, belgilangan so‘zning boshqa so‘zlar bilan bo‘lgan yaqin munosabatlarini aniqlashda va ularni taxmin qilishda yordam beradi. Bunday qo‘llanmalar matn tuzish, tarjima va tafsilotlar yozish uchun juda foydali bo‘ladi.

2. Matnlarning ko‘rsatilgan ma’nolarini kengaytirishga imkon beradi: Tezaurus lug‘atlari, berilgan so‘zning ko‘rsatilgan ma’nolarini kengaytirish uchun qo‘shimcha so‘zlar va ularning sinonimlari va antonimlari bilan birga ko‘rsatuvlar beradi.

3. Ishtirokchilar uchun ma’nolarni aniqlashda va tarjima qilishda yordam beradi:

4. Tezaurus luġatlari, matnlar tarjimasida yoki ishtirokchilar o‘zining so‘zlari bilan gaplashishda yordam beradi. Ular so‘zlarning ko‘rsatilgan ma’nolarini aniqlash va ularning sinonimlari va antonimlari bilan birga ko‘rsatuvlar beradi.

5. Matndagi so‘zlar orasidagi munosabatlarni tahlil qilishga imkon beradi: Tezaurus lug‘atlari, matndagi so‘zlar orasidagi munosabatlarni tahlil qilishga yordam beradi. Ular misol uchun, bir so‘zning giperonimlari, giponimlari, meronimlari, holonimlari, partonimlari va halonimlari bilan birga ko‘rsatuvlar beradi.

6. O‘zlashtirilgan so‘zlar va maxsus terminologiyani qamrab oladi: Tezaurus lug‘atlari, o‘zlashtirilgan so‘zlar va maxsus terminologiyani ko‘rsatishda ham yordam beradi. Bunday qo‘llanmalar, maxsus terminologiyaga ega so‘zlarni aniqlash uchun juda foydali bo‘ladi. Tezaurus

lug‘atlaridan biri WordNet so‘zlarning leksik- semantik, antonimlik ma’nolarini aniqlashda eng muhim vosita hisoblanadi. [2, 45-48]

WordNet - bu Prinston universitetida ishlab chiqilgan inglizcha so‘zlarning semantik tarmog‘ini o‘z ichiga olgan elektron lug‘at va tezaurus tizimi. WordNet ingliz tilidagi so‘zlarning semantik tuzilishini tavsiflash uchun ishlatiladigan juda ko‘p turli atamalar (sinonim, antonim, gipernim, giponim, meronim, holonim va boshqalar) ro‘yxatini o‘z ichiga oladi. Ushbu atamalar so‘zning ma’nosini yaxshiroq tushunish va so‘zlar guruhlari o‘rtasidagi munosabatlarni ko‘rsatish uchun ishlatilishi mumkin [4, 69] . WordNet, ayniqsa, tabiiy tillarni qayta ishlash, kompyuter lingvistikasi va sun'iy intellekt sohalarida keng qo‘llaniladi. WordNet so‘zning ma’nosini yaxshiroq tushunish va iboralar o‘rtasidagi munosabatlarni ko‘rsatish uchun ishlatilishi mumkin, bu iboralar orasidagi munosabatlarni ko‘rsatish uchun ishlatiladigan daraxt tuzilishi tufayli.

Masalan, “mushuk” so‘zi “uy hayvonlari” iborasining kichik to‘plami va “sutemizuvchilar” iborasining yuqori to‘plamidir. WordNet so‘z ma’nolari va ishlatilishini aniqroq va izchil aniqlashga yordam beradi, shuning uchun uni matn tasnifi, ma’lumotlarni qidirish va mashina tarjimasi kabi ko‘plab sohalarda qo‘llash mumkin. WordNet - bu ingliz tilidagi so‘zlar uchun asosiy leksik tizim hisoblanadi. Bu tizimda so‘zlar taqsonlarni shakllantirish uchun ishlatilgan. Bu degani, bitta so‘zning o‘zining o‘rniga, uni ifodalaydigan boshqa so‘zlar ham mavjud. Misol uchun, "cat" so‘zi "feline" yoki "animal" deb ataladigan taqsonlarga tegishli bo‘lib, "claw" so‘zi esa "cat" so‘zining meronimi (qismini ifodalaydi), yoki "tool" so‘zining holonimi (uchlashuvchisi) bo‘lib ifodalaydi. WordNet o‘zida ko‘p til uchun ishlatiladi va bu tizimning turli dasturlarga integratsiyasi mavjud.

Tezaurus - so‘zlarning ma’nolari va so‘z turkumlarining bir-biri bilan munosabatlari tizimli ravishda tartibga solingan lug‘at turi. Tezaurus til o‘rganish, matnlarni tasniflash, sinonimlarni topish, kontseptsiyani xaritalash va axborot qidirish kabi sohalarda keng qo‘llaniladi. Tezaurus tarkibida so‘zning sinonimlari (sinonimlari) va antonimlari (antonimlari) bo‘lishi mumkin. Shuningdek, ularda so‘zning qaysi kontekstda qo‘llanishi va uning boshqa so‘z turkumlari bilan aloqasi ko‘rsatilgan. Shu tarzda, ular so‘zning ma’nosi va ishlatilishini yaxshiroq tushunish uchun ishlatilishi mumkin.

Tezaurus mualliflar, muharrirlar va tarjimonlarga yordam berishi mumkin. Biroq, namoyishlardan foydalanish ehtiyotkorlik bilan va to‘g‘ri bajarilishi kerak. So‘zlarning ma’nolari va qo‘llanishlarini to‘liq tushunmaslik ularning noto‘g‘ri ishlatilishiga olib kelishi mumkin. Shuning uchun, ko‘rinishlarni ishlatishdan oldin, so‘zning ma’nosi va ishlatilishini to‘liq tushunish muhimdir. Tezaurus til o‘rganish, matn tasnifi, sinonimlarni topish, kontseptsiyani xaritalash va ma’lumot olish kabi sohalarda muhim o‘rin tutadi. Tezaurus so‘z ma’nolari va ishlatilishini yaxshiroq tushunish uchun ishlatilishi mumkin. Xususan, yozuvchilar, muharrirlar va tarjimonlar so‘z tanlash va imloni tekshirish uchun aforizmlardan foydalanishlari mumkin.

Bundan tashqari, tezislar ma’lumot olishda muhim rol o‘ynaydi. Xususan, qidiruv tizimlari va axborotni boshqarish tizimlari foydalanuvchilarga tezaurus yordamida o‘zlari qidirayotgan ma’lumotlarga tezroq va aniqroq kirish imkonini beradi. Tezaurus so‘z ma’nolari va qo‘llanishlarini aniqroq va izchil aniqlashga yordam beradi, shu bilan ma’lumot olish samaradorligini oshiradi. Natijada, dissertatsiya til o‘rganish, matnlarni tasniflash, axborotni boshqarish va ma’lumot qidirish kabi ko‘plab sohalarda muhim o‘rin tutadi. Tezaurusdan foydalanish so‘zlarning ma’nolari va qo‘llanishini yaxshiroq tushunish, to‘g‘ri va izchil yozish,

ma’lumotlarga kirish samaradorligini oshirish uchun juda foydali. Tezaurus lug‘atlaring tarjima sohasida tutgan o‘rni beqiyosdir.

Foydalanilgan adabiyotlar ro‘yxati:

1. Abdurakhmonova N, Tuliyev U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
2. Abdurakhmonova N, Urdishev K. Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (JFLTAL).
3. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016;2 (38):12-7.
4. Abduraxmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref." (2018).
5. Aripov, M., Sharipbay, A., Abdurakhmonova, N., Razakhova B.: Ontology of grammar rules as example of noun of Uzbek and Kazakh languages. In: Abstract of the VI International Conference “Modern Problems of Applied Mathematics and Information Technology - Al-Khorezmiy 2018”, pp. 37–38, Tashkent, Uzbekistan (2018)
6. Atkins, B. T. S. and Beth Levin, 1991. Admitting impediments. In Uri Zernik (ed.), Lexical Acquisition: Exploiting On-Line Resources to Build aLexicon. Hillsdale,
7. Atkins, Sue, 1993. Tools for computer-aided corpus lexicography: the Hector project. Acta Linguistica Hungarica, 41:5–72.
8. Baeza-Yates, Ricardo and Berthier Ribeiro-Neto, 1999. Modern Information Retrieval. ACM Press and Addison Wesley.
9. Buitelaar, Paul, 1998. CORELEX: Systematic Polysemy and Underspecification. Ph.D. thesis, Brandeis University.
10. CIDE, 1995. Cambridge International Dictionary of English. CUP, Cambridge, England. New Jersey: Lawrence Erlbaum, pages 233–262.

VI SHO‘BA
ZAMONAVIY TILSHUNOSLIK MASALALARI

SAYOHAT TURIZMIGA OID MATNLARDA MILLIY-MADANIY BIRLIKLARNING LEKSIK-PRAGMATIK XUSUSIYATLARI

Adizova Nilufar Istamovna
Teacher of the Department of Uzbek
Linguistics and Journalism, BuxSU.

Annotatsiya: Insonning o‘y-fikrlari, ruhiyati, ma’naviy olami uning nutqi orqali namoyon bo‘ladi. Borliqdagi har bir narsa, voqea-hodisaga munosabatini ham inson faqat nutqi orqaligina ifoda eta oladi. Semantik bilimlar kishiga qandaydir voqea yoki narsa haqida muayyan fikrlar, ma’lumotlarni berishi mumkin. Biroq milliy-madaniy jihatlarni anglash uchun faqat shuning o‘zigina yetarli bo‘lmaydi. Shuni inobatga olgan holda mazkur maqolada tilshunoslikda turizm sohasiga oid ayrim milliy-madaniy birliklar va matnlarga doir tahlil amalga oshiriladi. Bundan tashqari maqolada turizm sohasi va unga e’tibor haqidagi ayrim fikrlar ham keltirib o‘tiladi. Turizmga oid milliy-madaniy birliklarni o’rganish jarayonida nazariy yondashuvning ahamiyati va tatbiq etilayotgan ma’nosini nutqda qo’llash maqsadga muvofiqligi, ularni alohida tizimga solish masalalari to’g’risida ham fikr boradi.

Kalit so‘zlar: turistik matnlar, turistik birliklar, nutqiy akt, lingvokulturologiya, madaniy obyektlar, turizm, ziyorat turizmi, sayohat, ichki turizm, ark, masjid, madrasa.

Abstract: A person's thoughts, psyche, and spiritual world are revealed through his speech. A person can express his reaction to everything in existence and events only through his speech. Semantic knowledge can give a person certain thoughts and information about some event or thing. However, this alone is not enough to understand the national and cultural aspects. Taking this into account, this article analyzes some national-cultural units and texts related to the field of tourism in linguistics. In addition, the article also mentions some points about the tourism sector and attention to it. In the process of studying the national-cultural units related to tourism, the importance of the theoretical approach and the appropriateness of using the applied meaning in speech, as well as the issues of their separate system, are discussed.

Keywords: touristic texts, touristic units, speech act, linguistic culture, cultural objects, tourism, pilgrimage tourism, travel, domestic tourism, arch, mosque, madrasa.

Аннотация: Мысли, психика, духовный мир человека раскрываются через его речь. Свою реакцию на все сущее и события человек может выразить только через речь. Семантические знания могут дать человеку определенные мысли и информацию о каком-то событии или вещи. Однако одного этого недостаточно для понимания национально-культурного аспекта. Учитывая это, в данной статье анализируются некоторые национально-культурные единицы и тексты, относящиеся к сфере туризма в лингвистике. Кроме того, в статье также упоминаются некоторые моменты, касающиеся туристического сектора и внимания к нему. В процессе изучения национально-культурных единиц,

связанных с туризмом, обсуждаются важность теоретического подхода и целесообразность использования прикладного значения в речи, а также вопросы их отдельной системы.

Ключевые слова: туристские тексты, туристские единицы, речевой акт, лингвокультура, объекты культуры, туризм, паломнический туризм, путешествие, внутренний туризм, арка, мечеть, медресе.

Lingvomadaniyatshunoslik bugungi kunda muhim sohalardan biri sifatida taraqqiy etayotganligi ma’lum. Uning obyekti madaniyat va tilning o‘zaro aloqasi jarayonida, lisoniy va nutqiy birliklarning madaniyat belgisi sifatida tutashgan nuqtasida yuzaga chiqadi. U tilni madaniyat fenomeni, tashuvchisi sifatida hamda nutq jarayonining o‘ziga xos milliy-madaniy xususiyatlari va qonuniyatlarini inson tafakkuri, ma’naviy-madaniy qiyofasi va amaliy faoliyati asosida o‘rganadi. Har bir til ma’lum bir xalqning azaliy qadriyatlari, an’analari, udumlari, tarixi bilan bog‘liq muhim jihatlarni mujassam etishi bilan o‘ziga xos ahamiyat kasb etadi. Ha, “…har qaysi xalq yoki millatning tafakkuri, turmush tarzi, ma’naviy qarashlari o‘z-o‘zidan, bo‘sh joyda shakllanib qolmaydi. Ularning vujudga kelishi va rivojlanishida aniq tarixiy, tabiiy va ijtimoiy omillar asos bo‘lishini hammamiz yaxshi bilamiz. Shu ma’noda, xalqimizning turmush va tafakkur tarziga nazar tashlaydigan bo‘lsak, boshqalarga hech o‘xshamaydigan, ming yillar davomida shakllangan, nafaqat o‘zaro muomala, balki hayotimizning uzviy bir qismi sifatida namoyon bo‘ladigan bir qator o‘ziga xos xususiyatlarni kuzatamiz…”[Karimov,2008;8]. Haqiqatan ham, tilimizda aks etadigan shunday ifoda va birliklar borki, ular shunchaki chiroyli nutq vositasi emas, balki xalqimizning asriy qarashlari, madaniy va ma’naviy hayotining asosi, tarixiy xotirasi sifatida yuzaga kelgan, tasavvurimizda muhrlangan, millatimizga xos bo‘lgan qadriyatlar, an’analarning lisoniy yoki nutqiy ifodasidir.

Lingvistik madaniyat obyekti lingvistika va kulturologiya, etnografiya, onomastika, toponimika kabi qator fundamental sohalar kesishmasida o‘rganiladi. Ayniqsa, turizmga oid matnlarda o‘z ifodasini topgan muayyan voqelik, shaxs, narsa-buyumlar, inshoot va obyektlar haqida ma’lumot beruvchi lingvokulturemalar o‘tmish manzaralarini, milliy-madaniy tafakkur tarzi, millat qiyofasini ifoda etishi bilan ahamiyatli. Bugungi kunda bu kabi milliy-madaniy jihatlarga, tarixiy manbalar, osori atiqalarga e’tibor qaratish har qachongidan ham muhimdir. Zotan, har qanday davlat, millat o’zining boy o‘tmishi, farovon turmushi, mustahkam iqtisodiyoti, taraqqiy etayotgan sanoati bilan dunyo hamjamiyatida bo‘y ko‘rsatishga intiladi. Jumladan, O‘zbekiston ham zamonaviy turizm industriyasini rivojlantirish uchun barcha zarur manbalarga ega. Buyuk Ipak yo‘li ustida joylashgan vatanimiz qulay tabiiy iqlim sharoitlariga, boy tarixiy, madaniy merosga va, ayni paytda, ham ichki, ham xalqaro turizmni rivojlantirish uchun yuqori salohiyatga ega.

Shularni inobatga olgan holda, O‘zbekiston Respublikasida shu paytga qadar turizm sohasiga oid qator farmon va qarorlar qabul qilindi: dastlabki me’yoriy hujjat bu – O‘zbekiston Respublikasi Prezidentining 1992-yil, 27-iyuldagi “O‘zbekturizm” MK (Milliy Kompaniya) ni tashkil etish to‘g‘risida”gi Farmoni va O‘zbekiston Respublikasi Vazirlar Mahkamasining shu yil 20-oktabrdagi “O‘zbekturizm” MK faoliyatini tashkil etish masalalari haqida”gi qaroridir[Mirzayev,2011;29]. Ushbu farmon va qarorning mazmun-mohiyati mamlakatimizda jahon andozalariga javob beradigan turizm tizimini barpo etish, soha boshqaruvini takomillashtirish, uning iqtisodiy samaradorligini oshirish hamda milliy turizmni shakllantirishga qaratilgan edi.

Shundan so‘ng O‘zbekiston Respublikasi Prezidentining 1995-yil, 2-iyundagi “Buyuk Ipak yo‘li”ni qayta tiklashda O‘zbekiston Respublikasi ishtirokini avj oldirish va Respublikada xalqaro turizmni rivojlantirish borasidagi chora-tadbirlar to‘g‘risida”gi Farmoni hamda mazkur farmonni amalga oshirish maqsadida e’lon qilingan O‘zbekiston Respublikasi Vazirlar Mahkamasining shu yil 3-iyundagi “O‘zbekiston Respublikasida xalqaro turizmning zamonaviy infratuzilmasini barpo etish chora-tadbirlari to‘g‘risida”gi qarori [Mirzayev,2011;30]. mamlakat turizm sohasini rivojlantirishda yana bir muhim qadam bo‘ldi.

O‘zbekiston Respublikasi Prezidentining 1999-yil, 15-aprelda millatlararo aloqalarni kengaytirish va mustahkamlashga ko‘maklashish hamda xalqaro turizm bozoriga uzviy qo‘shilish, O‘zbekistonning madaniy-tarixiy va ma’naviy merosini jahon hamjamiyatiga targ‘ib qilish, shuningdek, turistlarga xizmat ko‘rsatish sifatini xalqaro andozalar darajasiga ko‘tarish maqsadida “2005-yilgacha bo‘lgan davrda O‘zbekistonda turizmni rivojlantirish Davlat dasturi to‘g‘risida”gi Farmoni, turizm sohasidagi munosabatlarni tartibga solish, mamlakatimizga tashrif buyurayotgan sayohatchilar sonini yanada oshirish maqsadida O‘zbekiston Respublikasi Qonunchilik palatasi tomonidan 1999-yil, 20-avgustda “Turizm to‘g‘risida”gi Qonun [Mirzayev,2011;33], 1996-yil, 21-noyabrda O‘zbekiston Vazirlar Mahkamasining “Xorijiy fuqarolarning va fuqaroligi bo‘lmagan shaxslarning O‘zbekiston Respublikasiga kelishlari, ketishlari, bu yerda bo‘lishlari va transit o‘tishlari tartibi to‘g‘risida”gi Qarori [Mirzayev,2011;32]. va boshqa shu kabi qator muhim hujjatlar mamlakatimiz turizm sohasini yaxlit bir tizimda taraqqiy etishini belgilab berdi.

O‘zbekiston Respublikasi “Turizm to‘g‘risida”gi Qonunida turizm so‘ziga quyidagicha ta’rif berilgan: ***turizm*** *– jismoniy shaxsning doimiy istiqomat joyidan sog‘lomlashtirish, ma’rifiy, kasbiy – amaliy yoki boshqa maqsadlarda borilgan joyda haq to‘lanadigan faoliyat bilan shug‘ullanmagan holda uzog‘i bilan bir yil muddatga jo‘nab ketishi (sayohat qilishi)dir* (3-modda).

Xalqaro doirada 1991-yil Ottava (Kanada)da bo‘lib o‘tgan Butunjahon turistik tashkiloti sayohat va turizm statistikasi bo‘yicha konferensiyada “turizm”ga *“odatdagi muhitdan tashqarida dam olish, ishga aloqador va boshqa maqsadlarda bir yildan ko‘p bo‘lmagan davrda qilingan sayohat va muayyan joylarda bo‘lish orqali amalga oshirilgan shaxs faoliyati”,* deb ta’rif berilgan[Mamatqulov, 2009,5].

Ma’lumotlarga qaraganda, dunyoda turizmning ilk uchqunlari dastlab, antik va undan keyingi davrlardan boshlangan bo‘lib, asosiy sayohatchilar olimlar, savdogarlar, badavlat kishilar, elchilar bo‘lgan. Ular yangi-yangi yerlarni o‘zlashtirish, moddiy manfaatdorlik, to‘kin va farovon hayotni izlash bahonasida dunyoning eng rivojlangan sivilizatsiya markazlari, iqlimi, tabiati go‘zal bo‘lgan joylari, yuksak madaniyatga ega davlatlarga sayohatlarni amalga oshirganlar [Abdurakhmonova, N.2019,2021,2022,2023].

Yurtimiz ham turizm uchun azaldan qulay shart-sharoit va imkoniyatlarga boy bo‘lgan. Buning dalilini biz Z.M.Boburning “Boburnoma” memuar asari, Muqimiyning “Sayohatnoma” asari, ispan elchisi Klavixo xotiralari, Herman Vamberi yozishmalari va boshqa qator manbalar orqali ham ko‘rishimiz mumkin. Biroq mamlakatimizda turizm sohasiga milliy mustaqilligimizdan so‘ng alohida e’tibor qaratildi, milliy turizmni rivojlantirish maqsadida muhim chora-tadbirlar belgilandi. Zotan, mamlakatimizning boy tarixi, madaniy merosi, milliy qiyofasi uzoq yillardan buyon dunyo ahlini o‘ziga jalb qilib kelayotgani hech kimga sir emas.

Turistik sayohatlar, asosan, ikki maqsad bilan amalga oshiriladi: 1. *Insonning o‘z ixtiyori bilan sayohatga chiqishi.* Uning zamirida kishining ruhan yengillikni his etishi, turli joylarni

ko‘rishga qiziqishi, inson qo‘li bilan bunyod etilgan go‘zallikdan, tabiatning so‘lim manzaralaridan zavqlanish maqsadi yotadi. 2. *Jamiyat taraqqiyoti, ijtimoiy-iqtisodiy hamkorlik, mamlakatlar nufuzini oshirish maqsadida amalga oshiriladigan sayohatlar.* Ular xalqaro yoki ichki doirada bo‘lib, xalqaro turizm – kirish (O‘zbekiston Respublikasida doimiy yashamaydigan shaxslarning mamlakat hududidagi sayohati), chiqish (O‘zbekiston Respublikasida doimiy yashovchi shaxslarning boshqa mamlakatga sayohati) asosida yo‘lga qo‘yilsa, ichki turizm O‘zbekiston Respublikasida doimiy yashovchi shaxslarning mamlakatimiz hududidagi sayohatini o‘z ichiga oladi [Abdurakhmonova,N.2019,2021,2022,2023]. Ushbu sayohatlar o‘ziga xosligi, maqsadi, davomiyligi, harakatlanish usullari va boshqa xususiyatlariga ko‘ra *madaniy-tarixiy, ziyorat, ekologik, ma’rifiy, etnografik, gastronomik, ishbilarmonlik, ijtimoiy, sport, tibbiy, yoshlar turizmi, agroturizm* kabi qator turlar va yo‘nalishlar asosida amalga oshiriladi.

Turizmga oid matnlarni yaratish sayohatchilar uchun muayyan obyektlar, tarixiy voqealar, mintaqalar, madaniy xususiyatlar, milliy kalorit bilan yaqindan tanishish imkonini beradi. Ayniqsa, O‘zbekistonning Buxoro, Samarqand, Xiva, Qo'qon kabi boy o'tmishga ega bo'lgan tarixiy shaharlari to'g'risida keng tasavvurlarga ega bo'lish va shu orqali sayyohlar tashrifini yanada kengaytirish mamlakat iqtisodiyoti, moddiy bazasining o‘sishi uchun ham muhim ahamiyatga egadir.

Ingliz faylasufi J. Ostin nazariyasi asosida til faqat voqelikni tasvirlabgina qolmay, balki harakat bajarilishini ham ifodalaydi. Jon Ostin Solonning mashhur qadimiy “So‘z - bu o‘ziga xos ishdir” iborasiga tayangan holda quyidagi fikrlarni bildiradi: “Ma’lum bir gaplarning ma’lum bir sharoitda aytilishi qandaydir harakat, ishning bajarilishi bilan barobardir”-deydi va nutq aktini uchta komponentga bo‘ladi: lokutiv, illokutiv, perlokutiv. Lokutiv – bu aslida gapirish, nutqiy harakat; Illokutiv – bu adresatga ta’sir qilish maqsadida aytilgan axborot; bu bosqichda so‘zlovchining kommunikativ niyati amalga oshadi; Perlokutiv – bu nutq harakatini bajarish natijasida erishilgan ta’sir[J.Ostin,1986;93-95]. Turizmga oid matnlarda nutqiy aktning har uchala komponenti ham amalga oshirilishi nazarda tutiladi. Bu kabi matnlarda sayohatbop joylar, obyektlar, noyob, milliy madaniyatni aks ettiruvchi til birliklari, milliy tushunchalar ifodasi, tarixiy voqeliklar haqidagi ma’lumotlar aks etadi[Abdurakhmonova,N.2019,2021,2022,2023]. Masalan, Buxoro ko‘hna va tarixiy shahar sifatida madaniy-tarixiy sayohat uchun juda qulay imkoniyatlarga ega. Sayyohlar Buxoroning milliy-madaniy qiyofasini tasavvur qilishi uchun obyektlar haqida nafaqat kerakli ma’lumotni, balki madaniy jihatlarni ham yaxshi anglashlari lozim. Quyida shu kabi milliy-madaniy, tarixiy birliklarning ayrimlari asosida ba’zi tahlil namunalari keltirildi: *Arki oliy (Ark), “Mag‘oki attor” masjidi, Ismoil Somoniy maqbarasi, Toqi telpakfurushon, Toqi zargaron, “Namozgoh” masjidi, Minorai kalon, Timi Abdulloxon, Sitorai Mohi Xossa, minora, sarhovuz, Masjidi kalon, xonaqo, madrasa* va boshqalar shular jumlasidan.

“Ark” so‘zi forscha ichki qal’a, poytaxtdagi qasr, xon yoki amirning atrofi baland devor bilan o‘ralgan saroyi degan ma’noni ifodalaydi:

Arki Oliy *– Buxoro shahrining g‘arbiy chegarasida hamon qaddini viqor bilan rostlab turgan bu qasr minglab yillar mobaynida hukmdorlarning saroyi bo‘lgan. Ko‘p yillar avval uning Sharq tomonida ham darvozasi bo‘lib, uni Siyovush darvozasi deb atashgan…*

Qasr atrofida katta shahar vujudga kelib, savdo-sotiq, hunarmandchilik rivojlanibdi. Kuyovining bunyodkorliklaridan xursand bo‘lgan Afrosiyob qarib qolgach, Turon yurtini unga topshirishini aytibdi. Taxtni ko‘zlab turgan g‘animlar shahzodani badnom qilish maqsadida shoh xotiniga yomon ko‘z bilan qaragani to‘g‘risida mish-mishlar tarqatishibdi. Noxush xabarlar

qulog'iga yetib borgan Afrosiyob Siyovushni o'zi qurgan ***Arki oliy*** *darvozasi ostonasida qatl ettiribdi…****Arki oliy*** *atrofidagi keng maydon azaldan shaharning eng gavjum joyi bo'lgan.*

Masjid – musulmonlarning jam bo'lib, namoz o'qiydigan, sajda qilinadigan joy (ibodatxona).

"Mag'oki attoron" masjidi *– Buxorodagi tarixiy, XII-XVI asrlarga oid ustunli gumbazli masjid. Yer sathidan 4,5 metrdan ko'proq pastda joylashgan. O'ymakor bezaklari bilan mashhur. Buxoro yodgorliklarining bugungi kungacha saqlanib qolganlari orasidagi eng qadimiysi. Mag'okda (chuqurlikda) hamda attor bozorlari yaqinida joylashgani sababli "Mag'oki attor" masjidi deb nomlangan. Shahar markazida, Shahriston va Shohrud kanali oralig'ida, Toqi Sarrofon savdo gumbazi hududida, islomgacha bo'lgan davrdan 1930-yillargacha faoliyat ko'rsatgan bozorlar hududida joylashgan. Tarixchi olim Narshaxiy qadimda Buxoro hokimi Moh savdoning adolatli kechishini nazorat qilish uchun taxtini bozorga ko'chirib kelgani haqida aytib o'tgan.*

Maqbara – ar. qabriston, go'riston\qabr ustiga qurilgan me'moriy inshoot, sag'ana.

"Ismoil Somoniy maqbarasi – *O'rta Osiyoda Somoniylar davri me'morchiligining nodir namunalaridan biri bo'lgan bu tarixiy obida Ismoil Somoniy tomonidan bunyod etilgan. Ismoil Somoniy Buxoroda Somoniylar davlatiga asos solgan yirik siyosiy arbob hisoblanadi. Ismoil Somoniy 848-yil Buxoro shahrida tug'ilgan, 874-yilda somoniylarning Buxorodagi noibi, 888-yildan esa butun Movarounnahrga hokim bo'lgan. Ismoil Somoniy markaziy hokimiyatni mustahkamlash siyosatini olib borib, turli yerlardan ulamolar, adiblar, usta va hunarmandlarni Buxoroga to'plagan. Madaniyat rivojiga katta hissa qo'shgan. Bag'dod xalifalaridan ibrat olib, o'ziga ulkan maqbara qurdirgan. 893-yilda Tarozga yurish qilib shimoliy chegaralarni mustahkamlagan. 900-yilda Movarounnahr va Xurosonni birlashtirgan.*

Toq – ar. gumbaz, qubba, arka\ eshik va binolarning tepasidagi yarim doira shaklida ishlangan qismi.

Toqi telpakfurushon – *Buxoroda bosh kiyimlar sotish va tayyorlash uchun qurilgan savdo markazlaridan biri bo'lgan. Toqi Tepakfurushon XVI asrda Buxoro xonligini idora qilgan Shayboniy hukmdor Abdullaxon II davrida bunyod etilgan va chorrahaga yaqin qurilgan. O'rta asrlarda bu yerda tulki va qunduz telpaklar, misqoli salla sotilgan.*

Toqi zargaron – *Buxoroda zargarlik buyumlari sotish va tayyorlash uchun qurilgan savdo markazlaridan biri bo'lgan.*

Minora – ar. baland qilib qurilgan inshoot, yorug'lik manbai bo'lgan joy.

Minorayi kalon – *Katta minora. Buxoro hukmdori Arslonxon tomonidan bunyod etilgan. Poyi Kalon me'moriy ansamblining bir qismi. Eski Buxoroning eng baland inshooti, balandligi 47.5 metr. O'z vaqtida Rossiya imperiyasining hujumi natijasida bomba (26 ta) borib tekkan, lekin qulab tushmagan. Mazkur inshoot me'mor Usta Baqo loyihasi asosasida qoraxoniy hukmdor Arslonxon topshirig'iga binoan 1127-yilda bunyod etilgan. Minorayi kalonda jami ko'tarilish uchun zinalar soni 105 ta bo'lgan.*

Sitorai Mohi Xossa *– Buxorodagi mang'itlar sulolasining 3 avlodiga mansub saroy – bog'. Buxorodan 4 km shimolda. Dastlabki binolar Nasrullaxon (1826-1860) va Muzaffarxon (1860-1885) hukmronligi davrida qurilgan. Abdulahadxon hukmronligi davri (1885-1910)da saroy va xushmanzara bog' barpo etilgan. Amir Olimxon hukmronligi davri (1910-1920)da yangi saroy qurilgan. Maydoni 6,7 ga bo'lib, eski saroy, g'isht to'shalgan 3 hovli va ko'pgina xonalardan iborat.*

Tim – fors. bozor yaqinidagi karvonsaroy\ usti yopiq bozor\ savdo rastasining baland tomi.

Timi Abdulloxon *– usti yopiq eng katta savdo rastalaridan biri hisoblangan. Abdullaxon timi Katta tim, ya'ni Timi Kalon kabi nom bilan ham atalgan. Ushbu tim shayboniylardan Buxoro xonligi hukmdori Abdullaxon II davrida 1577-yilda gazlamalar sotish maqsadida qurilgan. Mazkur tim asosiy savdo ko'chasida joylashgan va 6 ta ravoqli darvozadan iborat bo'lgan. Bu inshootda sakkiz gumbazli xona bo'lib, savdo do'konlari tutashgan. Abdullaxon timida asrlar davomida gazlamalar sotilgan. XX asr boshlarida timda shoyi va boshqa turdagi gazlamalar sotilgan. Gazlamalar maxsus sandiqlarda saqlangan.*

Sarhovuz **–** ar-fors. asosiy, bosh hovuz\ katta hovuz.

Xonaqo – fors. xilvatxona, uzlat; darvishlar makoni\ masjidning namoz o'qiladigan keng va katta xonasi\ so'fiylar mashg'ulotlarini o'tkazadigan katta xona va uning atrofidagi hujralar.

Xulosa o'rnida aytish mumkinki, madaniyatni tildan foydalanuvchi kishilar yaratadi. Ba'zi millatlarda odatiy me'yor, madaniy xususiyatni ifodalovchi etalonlar boshqa bir millat mentalitetiga mos kelmasligi mumkin. Shuning uchun ham muayyan bir millatda madaniyat belgisi deb qaraladigan ifoda birliklari boshqasi uchun oddiygina muloqot vositasi tarzida namoyon bo'ladi. Zotan, milliy-madaniy birliklar har bir xalqning milliy qarashlari asosida o'ziga xos shakldagina namoyon bo'ladi.

Foydalanilgan adabiyotlar:

1. Каримов И.А. Юксак маънавият-енгилмас куч. – Тошкент: Маънавият, 2008. – В.8.
2. Adizova, Nilufar. "Poetic works and rhetorical questions.(On the example of Erkin Vahidov's epic" Nido")." Центр научных публикаций (buxdu. uz) 1.1 (2020).
3. Дж Остин. Слово как действие// НЗЛ. Вып .XVII.-М.,1986. Стр.93-95
4. Istamovna, A. N. (2022). Occasional derivatives in creative epics. Galaxy International Interdisciplinary Research Journal, 10(5), 146-152.
5. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 1790-1793). IEEE.
6. Абдурахмонова, Н., & Бойсариева, С. (2023). Tabiiy tilni qayta ishlashda (nlp) okkazionalizmlarning morfem tahlili. Международный журнал искусство слова, 6(3).
7. Abdurakhmonova, N., Tuliyev, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
8. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
9. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
10. M.X.Mamatqulov, A.B.Bektemirov, I.S.Tuxliyev, A.N.Norchayev. Xalqaro turizm. Toshkent: O'zbekiston faylasuflari milliy jamiyati, 2009. B.5.
11. M.A.Mirzayev. Turizm asoslari\o'quv qo'llanma. Toshkent:O'zbekiston faylasuflari milliy jamiyati,2011. B.29

BOSHLANG‘ICH SINF O‘QUVCHILARINI ORFOEPIK ME’YORLARGA O‘RGATISH VA UNING AMALIY AHAMIYATI

Jumayeva Dilnoza Baxshulloyevna
Navoiy davlat pedagogika instituti
Maktabgacha va boshlang‘ich ta’lim fakulteti
Boshlang‘ich ta’lim yo‘nalishi katta o‘qituvchisi f.f.f.d (PhD).

Usmonova Sevinch Muzaffar qizi
Navoiy davlat pedagogika instituti
Maktabgacha va boshlang‘ich ta’lim fakulteti
Boshlang‘ich ta’lim yo‘nalishi 3-kurs talabasi.

Bizning asosiy maqsadimiz-yoshlarning sifatli ta’lim olishi mkoniyatiga ega bo‘lishga erishish ularning qobilyati va iste’dodini ro‘yobga chiqarish uchun barcha zarur shart-sharoitlarni yaratib berishdan iborat.

SH. Mirziyoyev.

Annotatsiya: Ushbu maqolada orfoepik me’yorlar haqida so‘z yuritilib, boshlang‘ich sinf o‘quvchilarini to‘g‘ri yozish me’yorlariga o‘rgatish usullari bayon etilgan. Keyingi yillarda dunyo tilshunosligida tilning ijodiy funksiyasiga qiziqishning amaliy ahamiyati yoritilgan.

Kalit so‘z: orfoepiya, boshlang‘ich ta’lim, to‘g‘ri yozish meyorlari, dunyo tilshunosligi, fonetika, grafika, talaffuz

Abstract: This article talks about orthoepic standards and describes the methods of teaching elementary school students to the standards of correct writing. In recent years, the practical importance of interest in the creative function of language has been highlighted in world linguistics.

Keyword: orthography, primary education, standards of correct writing, world linguistics, phonetics, graphics, pronunciation

Аннотация: В данной статье говорится об орфоэпических нормах и описываются методы обучения учащихся младших классов нормам правильного письма. В последние годы в мировой лингвистике подчеркивается практическая значимость интереса к творческой функции языка.

Ключевые слова: орфография, начальное образование, нормы правильного письма, мировое языкознание, фонетика, графика, произношение.

Ma’lumki, O‘zbekiston mustaqillikka erishgandan so‘ng hamma sohalarda tubdan o‘zgarishlar davri boshlandi. Jumladan, ta’lim sohasida ham 2020-yil 23-sentabrda qabul qilingan “Ta’lim to‘g‘risida”gi qonunda ta’lim sohasida hozirgi davrda amalga oshirilgan islohotlar o‘z aksini topadi. O‘zbekiston Respublikasi Birinchi Prezidenti Islom Karimov shunday degandi: “Ma’lumki,o‘zlikni anglash, milliy ong va tafakkurning ifodasi, avlodlar o‘rtasidagi ruhiy-

ma’naviy bog‘liqlik til orqali namoyon bo‘ladi. Jamiyki ezgu fazilatlar inson qalbiga, avvalo, ona allasi, ona tilining betakror jozibasi orqali singadi. Ona tili-bu millat ruhidir”.

Keyingi yillarda dunyo tilshunosligida tilning ijodiy funksiyasiga qiziqish sezilarli darajada kuchaydi. Tilning ijodiy funksiyasini aks ettiruvchi omillardan biri “Til o‘yinlari”dir. “Til o‘yini” termini tilshunoslikda birinchi bo‘lib nemis olimi Lyudvig Vitgenshteyn “Filofskil issledovaniy” asarida qo‘lladi. A.Gridina til o‘yinlari haqida fikr yuritar ekan, bu usul nutq o‘stirish va til o‘zlashtirishning eng samarador usuli ekanligini ta’kidlaydi. Ayrim mualliflar bunday o‘yinlarni, “fonetik trenajyorlar” deb ataydilar. Ular, odatda, jurnalistlar, notiqlar, yuristlar, nutqida nuqsoni bor bolalar va boshlang‘ich sinflarda savod o‘rgatish jarayonida qo‘llash uchun tavsiya etiladi.

Boshlang‘ich sinflarda fonetika va grafikani, qolaversa, orfoepiyani o‘rgatish orqali o‘quvchilarda o‘qish, yozish, gapirish, ko‘nikmalarini shakllantiradi. Fonetik trenajyorlarni yaratishda so‘zdagi tovushlar va harflarning sonini aniqlash, urg‘u qaysi bo‘g‘inga tushishini aniqlash kabi holatlarga e’tibor qaratish lozim. Ma’lum belgilari bilan o‘zaro zidlanuvchi unli yoki undoshlar yonma-yon kelganda talaffuzda qiyinchiliklar yuzaga keladi. Talaffuzdagi ana shu qiyinchiliklarni bartaraf etish maqsadida, fonetik trenajyorlardan foydalanish maqsadga muvofiq bo‘ladi. Ular nutqidagi nuqsonlarning oldini olishda fonetik trenajyorlarning o‘rni beqiyosdir. Demak, aynan til o‘yinlari orqali ham bolalarni orfoepik me’yorlarga o‘rgatish mumkin.

Ajdodlar tomonidan yillar davomida sayqallanib kelingan tilni kelajak avlodga o‘rgatish, tabiiyki, o‘qituvchining vazifasidir. Ayniqsa, maktab ostonasiga endi qadam qo‘yayotgan bolajonlarga dastlabki saboqni berishda boshlang‘ich sinf o‘qituvchilaridan yuksak bilim va mahorat talab etiladi. Hali o‘qish va yozishni bilmaydigan bolajonlar uchun dastlabki saboqlar og‘zaki shaklda olib boriladi. Bunda o‘qituvchi o‘quvchilarda adabiy til me’yorlarini rivojlantirib borishi, darsda ma’lum bir hudud shevasida emas, adabiy tildanishi, bir so‘z bilan aytganda,notiqlik san’atini chuqur anglashi kerak. Mashhur notiq Sitseron: “Tarixda yaxshi harbiy sarkarda yoki yaxshi notiq bo‘lish kerak”, - degan edi. Bu gapda katta hikmat bor. G‘arbiy sarkardaning itoatida katta lashkar bo‘lishini tasvvur qilsak, bu so‘z tasodifiy o‘xshatish emasligiga guvoh bo‘lamiz. O‘qituvchi orfoepik me’yorlarga amal qilsagina o‘quvchilarda shevaga oid so‘zlardan foydalanish,har xil orfoepik me’yorlarni buzish holatlarini oldini oladi. Shundagina respublika hududida yashovchi aholi nutqining bir xil bo‘lishiga erishamiz.

Orfoepiya-so‘zlarni, uning tarkibidagi tovushlarni to‘g‘ri talaffuz qilish me’yorlar tizimidir.

Orfoepik me’yorlar aslida, xalq xalq jonli faktlari asosida yaratiladi –jonli tildagi turlicha talaffuz ko‘rinishlaridan adabiy til an’alariga taraqqiyot traditsiyalariga mos keladiganlari tanlanadi. Masalan,o‘zbek shevalarida bir so‘z turlicha talaffuz etiladi: yo‘q-jo‘q; men-man; aka-oka kabi. Orfoepiya esa adabiy talaffuz me’yori (orfoepik norma)ni o‘rgatadi.

Ba’zi orfoepik me’yorlar:

1)i unlisi:

a)bir bo‘g‘inli so‘zlarda qisqa va ingichka talaffuz qilinadi:til, tish kabi.

b) q,g‘,x undoshlari bilan yondosh qo‘llanganda yo‘g‘on talaffuz etiladi:qish, g‘isht kabi.

d) bu unli so‘z oxiridagi ochiq bo‘g‘inda bir oz kengayadi (“i”dan kengroq, “e”dan torroq) tepki, bordi, xuddi kabi.

2)u unlisi:

a) ko‘p bo‘g‘inli so‘zlarning urg‘usiz bo‘g‘inida qisqa (tush, pul) urg‘uli bo‘g‘inda esa kuchliroq (bulut, yetuk, popuk kabi)

b) k, g, y undoshlaridan keyin ingichka (old qator unli tarzida), q, g‘, x undoshlaridan so‘ng yo‘g‘on (orqa qator unli tarzida)talaffuz etiladi.

d) -uvchi,-uv affikslari tarkibida cho‘ziqroq aytiladi :yozuvchi ,o‘quvchi ,oluv kabi

Boshqa tillardan o‘zlashtirilgan ayrim so‘zlar orfoepiyasi:

1. Turkiy tillarda,ma’lumki,f undoshi bo‘lmagan, shuning uchun jonli so‘zlashuvda arab, fors, rus tillaridan o‘zlashgan so‘zlarda “f”ni “p” deb talaffuz etish hollari uchraydi: ulfat>ulpat, faner>paner kabi. Hozirgi o‘zbek tilida ularni “f” holida talaffuz etilishi me’yor holiga keltirilgan.

2. Arab tilidan o‘zlashtirilgan ta’na, da’vo,ma’no ,me’yor kabi so‘zlarning birinchi bo‘g‘inidagi unlilar bo‘g‘iz artikulatsiyasi bilan cho‘ziqroq talaffuz etiladi.

3. Rus tilidan o‘zlashgan so‘zlarda “i” unlisi o‘zbek tilining “i”unlisidan torroq va cho‘ziqroq talaffuz qilinadi: til(o‘zb)-ti: r(rus)

Xulosa qilib shuni aytib o‘tish kerakki, to‘gri talaffuz nutq madaniyatining muhim belgilaridan biri sanaladi. Adabiy tilda to‘g‘ri yozish qanchalik muhim bo‘lsa to‘g‘ri gapirish ham shunchalik ahamiyatlidir.

Yosh avlodni tarbiyalashdek o‘ta mas’uliyatli ishni o‘z zimmasiga olayotgan talaba o‘z fanini chuqur o‘rganishining o‘zi yetarli emas. Chunki o‘qituvchning nutqi o‘tmas, nochor bo‘lsa, uning bilimi qanchalik tugal bo‘lmasin, o‘ziga ham o‘quvchiga ham azob. O‘qituvchi o‘zini ham, so‘zni ham qiynamaydigan, ravon ifodalarga nutqi boy bo‘lib, shevalardan foydalanmay, orfoepik me’yorga amal qilsagina, o‘quvchilarni mahliyo etib, bermoqchi bo‘lgan bilimini yosh inson shuuriga osonlik bilan olib kiradi.

Foydalanilgan adabiyotlar:

1. S.Matchonov, H.Baniyeva, X.G‘ulomova, Sh.Yo‘ldosheva, G.Xolboyeva. Boshlang‘ich sinflarda ona tili o‘qitish metodikasi darslik. Toshkent.2021.
2. M.E.Toirova. Ona tili 4-sinf 1-qism. Tosh.2023
3. Жумаева Д.Б. Антропоним мурожаат бирликлари ва уларнинг нутқий-мавзуий турлари // Сўз саньати. – Тошкент, 2022. – № 6. – В. 180-185 (10.00.00. № 31).
4. H.A. Jamolxonov “Hozirgi o‘zbek adabiy tili”

O‘ZBEK AYOLLAR NUTQIDA KONNOTATSIYA VA PRAGMATIK TO‘SIQ

Mirobidova Nilufar Murodjon qizi,
O‘zbekiston Milliy universiteti
O‘zbek tilshunosligi kafedrasi 1-kurs tayanch doktoranti.

Annotatsiya: Jahon tilshunosligida so‘ngi yillarda ilmiy paradigmalar almashinuvi yuz berib, endi tilshunoslar diqqat markaziga antroposentrik paradigma chiqdi. O’tgan asr oxirlarida antroposentrik tilshunoslikning yo’nalishlaridan biri sifatida pragmalingvistika rivojlandi. Pragmalingvistika til birliklarining nuqt vaziyati, konteks, til egalarining nutq o’bekti haqidagi umumiy bilimlari kabi qator omillar bilan yuzaga chiqadigan imkoniyatlarini o’rganadigan zamonaviy tilshunoslik yo’nalishidir. Dunyo tilshunosligida tilshunoslikning bu kabi muammolari qatorida ayollar nutqini o’rganish, pragmatik to‘siq va konnotativ ma’noning ifodalanishidagi o‘ziga xosliklarning o‘rganish dolzarbdir. Ushbu maqolada shu masalalar badiiy adabiyot namunalari asosida yoritib berilgan.

Kalit so‘zlar: Pragmalingvistika, gender lingvistikasi, ayollar nutqi, konnotatsiya.

Abstract: In recent years, there has been a change of scientific paradigms in world linguistics, and now the anthropocentric paradigm has come to the center of attention of linguists. At the end of the last century, pragmalinguistics developed as one of the directions of anthropocentric linguistics. Pragmalinguistics is a branch of modern linguistics that studies the possibilities of language units arising from a number of factors, such as situation, context, general knowledge of language speakers about the object of speech. Among such problems of linguistics in world linguistics, it is important to study women's speech, to reveal its pragmatic features, to study the peculiarities of the expression of connotative meaning in women's speech. These issues are discussed in this article based on examples of fiction.

Keywords: Pragmalinguistics, gender linguistics, women’s speech, connotation.

Аннотация: В последние годы в мировой лингвистике произошла смена научных парадигм, и теперь в центр внимания лингвистов попала антропоцентрическая парадигма. В конце прошлого века прагмалингвистика развивалась как одно из направлений антропоцентрической лингвистики. Прагмалингвистика — раздел современной лингвистики, изучающий возможности языковых единиц, возникающие из ряда факторов, таких как ситуация, контекст, общие знания носителей языка об объекте речи. Среди подобных проблем языкознания в мировой лингвистике важное значение имеет изучение женской речи, выявление ее прагматических особенностей, изучение особенностей выражения коннотативного значения в женской речи. Эти вопросы рассматриваются в данной статье на основе примеров художественной литературы.

Ключевые слова: Прагмалингвистика, гендерная лингвистика, женская речь, коннотация.

Tilshunoslik alohida, mustaqil fan sifatida XIX asrning birinchi choragida fanlar olamiga falsafa fanidan ajralib chiqqan holda kirib keldi. Ungacha boshqa fanlar kabi lingvistika ham falsafa fani tarkibida o`rganib borilgan. Oradan ma`lum davr o`tgandan so`ng shak-shubhasiz,

tilshunoslik ham boshqa qator fanlar kabi o`z tadqiqotlar, tajribalar, izlanishlar, g`oyalar, kuzatishlar natijasisifatida o`zining alohida o`rganish obyektiga va tekshirish metodiga ega bo`ldi. Tilshunoslik ildizlari qadimgi hindlar, yunonlar, arablarga borib taqaladi. Jahon tilshunosligi fani ham shu kungi taraqqiy etgan holatiga yetib kelguncha uzoq tarixiy taraqqiyot davrini, bir qancha rivojlanish bosqichlarini o`tkazib bordi. Sohada turli maktablar, oqimlar, ta`limotlar paydo bo`ldi. Tilshunoslar tomonidan lisoniy hodisalar haqida aytilgan fikrlar qarama-qarshiligiga, til hodisalari haqida yaratilgan juda ko`plab manbalarga ega bo`lindi. Shunday ekan, tilshunoslik fani ham jahon fanlari tizimida o`ziga xos doimiy mustahkam o`rin egalladi.

XX asrda lingvistikaning tilni shunchaki aloqa quroli va tafakkur natijasi sifatida emas, millatning madaniy aksi sifatida tadqiq etadigan sohalari yuzaga kelmoqda. Bu qarashlarning paydo bo`lishiga V. Gumboldt, A.A. Potebnya va boshqa olimlarning ilmiy ishlarida tamal toshi qo`yilgan. Jumladan, V.Gumboldt “Millatim tilining chegaralari mening dunyoqarashim chegaralarini belgilaydi” , - deb aytgan edi. [В.Гумбольдт, 1984: 23] Til nafaqat borliqni aks ettirish funksiyasini bajaradi, balki, inson yashayotgan atrof-muhitni uning ongida aks ettirib beradi. Shu sababli ham keyingi asrlarda G`arb falsafasi tildan foydalanish jarayoniga asoslanib kelinmoqda. Tilshunos olim A.M.Xaydegger tilni “mavjudlik uyi” deb atagan. [Хайдеггер М., 1997:11] Shu bois ham tilshunoslik har qanday ijtimoiy fanlar tizimida yetakchi xususiy mavqega ega.

Shunday qarashlar majmui sifatida antroposentrik nazariya maydonga keladi. Antropotsentrik nazariya – tadqiqot obyektining subyektda o`rganilishidir, yanada aniqroq aytadigan bo`lsak, til insonda ya`ni nutq egasida, inson esa til bilan birgalikda tahlilga tortiladi. I.A.Boduen de Kurtenening fikricha, “til faqat yakka shaxsning ongida, ruhida, qalbida bo`ladi, bu shaxs til jamoasini tashkil qiladi”. [И.Б.Куртенэ., 2004: 89] Tilni anropotsentrik paradigmadan turib o`rganish g`oyasi –hozirgi zamon tilshunosligining asosiy yo`nalishi hisoblanadi.

Yaqin vaqtlargacha lingvistlar faqat odamlarning psixik, ijtimoiy, millatga yoki jamiyatga xos bo`lgan ma`lum guruhlari o‘rtasidagi farqlar bilan bog‘liq bo‘lgan tildagi farqlarni tadqiqi etib kelishgan. Ammo negadir, hech kim nutq va muloqotning jinsga qarab farqlanishiga e'tibor bermadi. So`nggi yillarda tilshunoslar erkak va ayol nutqining o'ziga xos xususiyatlariga e'tibor berishni boshladilar. Bu boradagi ilk ilmiy tadqiqotchilar sifatida E.A. Zemskaya, M.V. Kitaygorodskaya va N.N. Rozanovlarni e`tirof etishimiz mumkin.

Sharq ayollarining nutqi G’arb ayollari nutqidan ba’zi xususiyatlari bilan ajralib turadi. Kattalarga bo`lgan hurmat, shirin muomalalilik,andisha,hayo va hamiyatlilikning beqiyos darajasi ularning nutqining o`ziga xos xususiyatlari hisoblanadi. O`zbek ayollariga xos yana bir xususiyat shundan iboratki, ular qarshi jins vakillariga nisbatan bo’ysinuvchanlik xarakteriga egadirlar.

O`zbek ayollari nutqi erkaklar nutqidan keskin farq qiladi. Unda konnotativ xususiyatlar, illokuliv aktning aks ettirilishi, ijobiy va salbiy baho munosabatlari, presuppozitsiya, tagma`no, kesatish, kinoya erkakalanikidan ancha kuchli. Har bir ayol o`z nutqida milliy xususiyatlarini, umumiy dunyoqarashi, madaniy holati, diniy bilimlarini aks ettirgan holda so`zlaydi. Shunga ko`ra o`qimishli ayollar, olima ayollar nutqi va uy Bekasi bo`lgan ayollar nutqi ham o`zaro farqlanadi. Ayollarimiz nutqidagi ibo-hayo boshqa millat, xalq vakillarida ajralib turadi. Ushbu parchada ham yuqoridagi sifatlar o`z aksini to`laligicha namoyon qilgan desak bo`ladi:

Enaxon kelinga qaradi. Kelin Enaxonning bu qarashidan “Qani gapiring nima deysiz?” deganini anglagan bo`lsa-da, qaynana oldida birinchi bo`lib gap ochishniep ko`rmasdan jim qoldi. Shundan keyin Enaxon bir narsa demakka majbur edi.

- Mayli, siz o`zingiz bilasiz, oyi. Chaqirish kerakligiga “yo`q” demadingiz. Mehmon deganni non qoqiga chaqirib bo`lmaydi. Muni o`zingiz bilasiz.

Kelin shu yerga kelganda, bir og`izgina so`z bilan aralashib qo`ydi:

- Albatta, ayniqsa, shuncha joydan keladigan mehmonni. [Cho`lpon Abdulhamid, 2021: 42]

Yuqoridagi parcha olingan badiiy asar XX asr 1-choragi ijod mahsullaridan biri hisoblanadi. Parchada o`zbek oilalaridagi qaynana kelin munosabatlarining bir ko`rinishi o`z aksini topgan. Milliy xususiyatlarimiz qahramon nutqiga ham ko`chmoqda. Ya`ni kelinning qaynana oldida qaynonadan oldin gapirmasligi, qaynanadan iymanish, fikrini ochiq va dadil bayon etolmaslik kelin ayollarimiz nutqining o`ziga xos xususiyati sanaladi. Bunga “o`zbekcha” milliy sharm-hayo yo`l bermaydi. Yozuvchi qahramonlar nutqida bu xususiyatlarni ko`rsatib ketadi. Bu bilan u o`sha davr urf-odatlarini, kaloritini yorqin aks ettirgan.

Ayollar nutqiga xos bo`lgan xususiyatlardan yana biri shundaki, ular juda diqqat bilan tinglaydi. Ayollar nutqi diqqat signallarini qo'llash bilan tavsiflanadi ya`ni ularda ko`lab emotsiya ifodalovchi so`zlar ishtirok etadi. (masalan, "aha", "uh-huh", "ha"). Ular ayollar nutqida muhim funksiya bajaradi: suhbatni jonlantiradi va rag'batlantiradi. Suhbatlashganda ayollar tabassum qiladilar va o`z nutqiy fikrlarini ko'z bilan ham ifodalay oladilar. Ayollar kamdan-kam duduqlanadilar. Bir suhbatga tushsa, to`xtatish uchun ancha vaqt ketadi. Yuqorida adabiyot namunalari asosida misollarni ko`rdik. Shunday misollardan yana biri Abdulhamid Cho`lponning “Kecha va kunduz” asarida ko`zga tashlanadi

“Qizim Bahridan so`rayman, bir og`iz gapni eplab gapirolmaydi. Kecha-kechasi o`sha yerdaydi. Shaharga ham birga borib edi. Gap so`rasam, xuddi tili yo`q soqovday “i, i…” qiladi, xolos… yuragim “tars” eta yorilguday bo`ldi. Shundan keyin, bu yerga chopib keldim, aylanay.” [Cho`lpon Abdulhamid, 2021: 125]

Qahramon tilidan olingan ushbu jumlada, ayollar ma`lum psixologik holatlar ta`sirida duduqlanishlarini, fikrlarini ochiq ifoda etib bera olmasliklarini ko`rishimiz mumkin. Ko`p holatlarda ularga qo`rquv ta`sir qiladi va bu ularning nutqini o`z ko`rinishini beradi. Garchi parchada yolg`on ma`lumot berib, to`qima fikr aytilgan bo`lsa ham, u kitobxon tomonidan xuddi tabiiydek qabul qilinadi. Bundan tashqari ayollar nutqida undovlardan, undalmalardan foydalanish faol ko‘rinish beradi.

Ayollar kommunikativ muloqotda so‘z qo‘llash me’yorini buzilishi natijasida to‘siqqa duch kelishadi. So‘zlovchi so`zlagan so`zning ma’no-mazmuni, tinglovchi ongida dekodlanamaydi. Suhbat haqida to`liq tasavvur yuzaga kelmaydi.

Ba’zan matnda shunday so`zlar qo`llanadiki, ular matndan tashqarida o‘z ma’nosida ishlatiladi, matn tarkibida esa umuman boshqa ma’noda qo‘llanadi. Ba’zan o`xshatishlarda ham shunday hodisani ko`rishimiz mumkin.

Havli yuzida to‘p o‘ynag‘uvchi “qirq qizlar”dan yetti-sakkiztasi Gulshan oldig`a yugurdilar.

- *Esonmisiz!*
- *Yelib yugurib hormang!*
- *Eringiz o`lmadimi!..*

O`zbek xalqida o`zaro salomlashishning o`ziga xos shakllari bor. unda suhbatdoshlar tirik odamdan to o`tib ketgan insonlargacha bir-biridan so`rab so`rashishadi. Bu o`zbek millatiga xos bo`lgan o`ziga xos xususiyatlardan biri desak ham bo`laveradi. Deyarli, har bir milliy adabiyot namunalarimizda salomlashish sahnalarini uchratishimiz munkin. Yuqoridagi parcha shulardan biri bo`lib, so`zning noo`rin qo`llanishiga guvoh bo`lamiz. Ya`ni salomlashishda hamma ezgu

niyat va maqsadlarni aks ettirsa, parchada “eringiz o`lmadimi?” kabi so`roq beriladi. Bu o`ziga xoslik o`quvchida pragmatik to`siqni yuzaga keltiradi. Qaysi maqsad, qanday niyat bilan ushbu so`roq yangragani tinglovchi ongida bo`shliqni, mavhumlikni yuzaga keltiradi.

Pragmatik to`siq quyidagi shakllarda namoyon bo`ladi :

1) nutq subyektining faoliyati bilan bog`liqlikda ;
2) nutqiy uslublar noo`rin qo`llanadi. Nutq egasi boshqa vaziyatda ushbu vaziyatga mos tushmaydigan uslub tanlaydi va shu asosida fikr ifodalaydi ;
3) so`zlovchining mujmal, noaniq javobi ;
4) so`zlovchining me`yorsiz nutqiy harakatlari ;
5) talaffuz me`yorining buzilishi ;
6) so`z va birikmalarni o`z o`rniga nomuvofiq qo`llash ;
7) nutqdagi mantiqiy nomuvofiqlik.
8) so`zlovchining psixologik holatlari bilan bog`liq shaklda yuzaga chiqish ;
9) so`zlovchi va tinglovchi o`rtasida hududiy o`ziga xoslik sabab tushunmovchilik aks etishi ;
10) adresat va adresant orasidagi o`zaro turli ijtimoiy munosabatlar ;
11) nutq obyekti haqida yetarli axborotga ega bo`lmaslik ;
12) xotiradagi muammolar bilan bog`liq shakllarda pragmatik to`siq hosil bo`lishi.

Yuqoridagi shakllardan kelib chiqib o`zbek tilshunosligida pragmatik to`siqni quyidagi omillar ta`sirida yuzaga chiqishini belgilaymiz : 1. Pragmatik to‘siqni yuzaga keltiruvchi lingvistik omillar, 2. Pragmatik to‘siqni yuzaga keltiruvchi ekstralingvistik omillar.

Tishunoslikda konnotatsiya pragmalingvistika yo‘nalishining tadqiqot obyekti sifatida o‘rganila boshlandi. Konnotativ ma’no o‘z ichiga bir qancha jihatlarni birirktirib oladi. Umuman olganda, so‘zlovchining nutq jarayonidagi hissiy-emotsional, ekspressiv munosabati ba’zi so‘zlar, birikmalar, jumlalarda namoyon bo‘ladi. Denotativ ma’no ustiga qo‘shimcha konnotativ ma’no ham yuklanadi. Bunda so‘zlovchi o‘zining ijobiy yoki salbiy munosabatini nutq jarayonida ko‘rsatib ketadi. Konnotatsiya ko‘plab tilshunoslar tomonidan o‘rganilgan.

M.Mamadaliyevaning fikricha, konnotatsiya pragmatik hodisa sifatida baho munosabati bilan uzviy bog‘liq, unga so‘z (va boshqa birlik)larning barcha ma’nolari asosida obyektivlik va subyektivliklarni bog‘lovchi vosita sifatida munosabatda bo‘lish lozim. [Мамадалиева М., 2000: 40-42]

Tilshunos M. Qurbonova bolalar nutqining pragmatik xususiyatlarini o‘rganar ekan, “Umuman, konnotatsiyaga umumiy tarzda til birliklariga hamroh bo‘lib keluvchi ekspressiv-emotsional ma’nolar, ma’no ottenkalari sifatida qaralishi, bizningcha ham, uning birlamchi mohiyatini to‘g‘ri belgilashga xizmat qiladi” deb ta’kidlaydi. [Qurbonova M., 2018:120] Olima doktorlik dissertatsiyasida konnotatsiya turli til sathlariga daxldor bo‘lishini aytib o‘tadi. Fonetik, leksik, morfologik, sintaktik sathlarda uchraydigan konnotatsiyalarni ko‘rib chiqqan.

Ayollar nutqida konnotatsiyani o‘rganar ekanmiz, fonetik, leksik, morfologik usullarda yuzaga kelgan konnotatsiyalarni uchratishimiz mumkin. Badiiy adabiyot namunalarinin ko‘zdan kechirar eknmiz, konnotatsiyaga ega, qo‘shimcha ekspressiv ma’nolar ifodalashga xizmat qiladigan bir qator leksik, morfologik, sintaktik birliklarga duch kelamiz.

Konnotatsiya tilning sintaktik sathida ham faol uchraydi: “ Xax, qiliqlaring ordona qolsin-a! Otangga nima deyman?”

Yuqorida konnotatsiya subyektning obyekt haqidagi emotsional-ekspressiv munosabatini ifodalashini ta’kidlab o‘tgan edik. Bu baho munosabati nafaqat ijobiy, balki salbiy munosabat

ottenkasiga ega bo‘lishi mumkinligini ham kuzatishimiz mumkin. “ – Juvongina o‘lgur, noinsof! Bolaning uvoli tutgur, iymonsiz! [Hoshimov O‘., 2015: 36]

Ushbu parchada so‘zlovchi salbiy semaga ega “juvon” so‘ziga –gina lug‘aviy shakl yasovchisini qo‘shish orqali konnototsiya hosil qilgan. O‘zbek tilida erkalash ma’nosini hosil qiluvchi affikslar sifatida -jon, -xon, -oy, -boy, -loq (-aloq), -gina, (-kina, -qina), -toy kabilar ajratiladi. Bular ichida -jon, -gina, (-kina, -qina) affikslari ayollar nutqida nisbatan faol qo‘llanib, nutqiy konnotatsiyani hosil qiladi. O‘zbek tilida egalik affikslarining faxrlanish, hurmat, tanqid, kesatiq ma’nolarini ham hosil qilishi ilmiy asoslangan.[Абдуллаев А., 1983:75] “ – Voy, Mosh**ing** o‘lsin! – “Qora ammam” tandirga sho‘ng‘irkan, savat burchagida turgan jajji kulchaga imo qildi – avval, o‘zing ye”. Ushbu parchada egalik affiksi tegishlilik ma’nosini ifoda etmagan. Bunda obyektning subyektga bo‘lgan tanqidiy munosabatini aks ettiradi va bu, o‘z navbatida konnotatsiyani yuzaga chiqargan.

Xulosa qilib aytadigan bo‘lsak, ayollar nutqida konnotatsiya faol uchraydigan pragmatik hodisa hisoblanadi. U tilning barcha yaruslarida ko‘zga tashlanadi. Fonetik hodisalar, bo‘g‘inlarning takrori, buzib talaffuz qilish, unlini cho‘zish, undoshni ayrilib talaffuz qilishda ham konnotatsiya yuzaga chiqadi. Bunsan tashqari o‘xshatishlar, metonimiya, sinekdoxa kabi ma’no ko‘chishlari, leksik bo‘yoqdor so‘zlar ham ushbu ma’no yuzaga chiqishida yetakchi hisoblanadi. Emotsional-ekspressiv boyoqdor so‘zlar nutning ta’sirchanligini oshirishga xizmat qiladi.

Foydalanilgan adabiyotlar ro‘yxati:

1. В.Гумбольдт. Избранные труды по языкознанию : сборник работ. Прогресс.1984.
2. Хайдеггер М. Бытие и время. — М.: Ad Marginem, 1997.
3. И.Б.Куртенэ. Введение в языковедение. Едиториал УРСС. 2004.
4. Cho‘lpon Abdulhamid. Kecha va Kunduz.Toshkent.Adabiyot uchqunlari. 2021.
5. Мамадалиева М. Номловчи бирликларнинг ички маънолари // Ўзбек тили ва адабиёти. – Тошкент, 2000. – № 6.
6. Qurbonova M. O‘zbek bolalar nutqining pragmatik xususiyatlari. Doktorlik diss.- Toshkent ,2018.
7. Hoshimov O‘. Ikki eshik orasi. Meriyus. T- 2015.
8. Абдуллаев А. Ўзбек тилида экспрессивликнинг ифодаланиши. – Тошкент: Фан, 1983.

XALQONA QASAMLARDA FONATSION VOSITALAR

Nurmonova S.D.
Filologiya fanlari bo'yicha falsafa doktori(PhD)
Andijon davlat universiteti.

Annotatsiya: Nutq janrlaridan biri bo‘lgan qasam nutqiy janrining tarlaridan biri xalqona qasamlardir. Xalqona qasam matni emotsional ekspressivligi bilan boshqa qasam turlaridan ajralib turadi.

Ushbu maqolada xalqona qasamlarning emotsionalligini oshirishda fonatsion vositalarning o‘rni ko‘rsatilgan.

Kalit so‘zlar: Intonatsiya, urg‘u, nutq melodikasi, pauza, neytral intonatsiya.

Аннотация: Клятва, являющаяся одним из жанров речи, является одним из родов речевого жанра народной клятвы. Текст народной клятвы отличается от других видов клятв своей эмоциональной выразительностью.

В данной статье показана роль средств фонации в повышении эмоциональности народных клятв.

Ключевое слово: Интонация, акцент, мелодия речи, пауза, нейтральная интонация.

Annotation: Oath, which is one of the genres of speech, is one of the genres of the speech genre, folk oaths. The folk oath text is distinguished from other types of oaths by its emotional expressiveness.

This article shows the role of phonation tools in increasing the emotionality of folk oaths.

Keywords: Intonation, accent, speech melody, pause, neutral intonation.

Nutqiy janrlar inson faoliyati bilan bog'liq ravishda vujudga keladi. Ta'kidlash kerakki, inson faoliyatining turli jabhalarini o'rganuvchi fan sohalarining o'zida ham nutqiy janrlarning ichki subjanrlari kuzatiladi. Nutqiy janrlardan biri bo'lgan qasam nutqiy janri ham o'z ichida kichik subjanrlarga ajralib ulardan biri xalqona qasamlardir. Xalqona qasamlar o‘zining emotsionalligi, uslubiy jihatdan boy ekanligi va tinglovchiga kuchli ta'sir ko'rsatish kuchiga ega ekanligi bilan ajralib turadi.

Xalqona qasamlar matni oldindan tayyorlanmaydi, emotsionallik kuchli ifodalanib, shakliy va mazmuniy jihatdan tugal butunlikka ega bo'ladi.

Xalqona qasamlarda fonatsion vositalar faol ishtirok etadi. Shu o'rinda atoqli o'zbek tilshunosi A.G'ulomovning fonatsion vositalar haqidagi quyidagi fikrlari diqqatga sazovor: “Intonatsiya so'zlashda ohangning o'zgarishi – tonning baland-past bo'lib o'zgarishi – nutq melodikasining o'zgarishidir. Bunga ritmika – urg'u va pauzalar haqidagi ta'limot ham kiradi. Intonatsiya tushunchasi o'z ichiga nutq tempini ham oladi. Nutqni fonetik jihatdan shakllantiradigan, bir butun qiladigan intonatsion vositalar – melodika, urg'u va pauzalar, asosan, sintaktik funktsiyani bajaradi, ya'ni sintaktik holatlar: gapdagi tugallikni bildirish, gapning ichidagi logik-grammatik bo'linishlar, qismlarning o'zaro munosabatini ko'rsatish va boshqalarni ifodalaydi”. A.G'ulomov ayni paytda intonatsiyaning logik-grammatik va emotsional vazifalari haqida ham to'xtaladi va har bir gap tipi (darak, so'roq, buyruq, undov kabi) o'ziga xos intonatsion

modelga ega ekanligini ta'kidlaydi. A.N.Baskakov turkiy va boshqa tillarda ham intonatsiyaning birlamchi – logik-grammatik va ikkilamchi – emotsional-ekspressiv vazifalari mavjudligini e'tirof etadi . Professor N.Mahmudov intonatsiyaning ahamiyati haqida quyidagilarni aytadi: “Nutqda intonatsiyaning alohida o'rni mavjud. Masalan, bahor so'zi gap intonatsiyasi bilan aytilmasa, u so'zligicha turaveradi. Ammo tegishli intonatsion modelga muvofiq aytilsa, darak (Bahor), so'roq (Bahor?), undov (Bahor!) gaplarga aylanishi mumkin. Intonatsiyaning o'zgarishi bilan gapning mantiqiy mazmuni ham tamoman o'zgarishi, kommunikativ maqsad boshqa yo'nalishga o'tishi ham qonuniyatlardan biridir. Aytaylik, o'qituvchi o'quvchiga murojaat qilarkan, “Jim o'tirasizmi?” gapini bir intonatsiya bilan, ya'ni “Ji-im o'tirasizmi?” tarzida aytsa, “Jim o'tirmang, javob bering” mazmuni anglashiladi, yana boshqa intonatsiya bilan, “Jim o'tirasizmi?!” tarzida aytilganda esa, “Jim o'tiring, bo'lmasa, jazolayman” mazmuni ifodalanadi. Ha/yo'q javobini talab qiladigan mazkur so'roq gapning o'qituvchi nutqidagi ikkala intonatsion shakliga ham o'quvchi yo'q javobini berolmaydi (chunki mazmun boshqacha, o'qituvchining kommunikativ maqsadi so'rash emas, balki buyurish). Bu gaplardagi intonatsiyalar faqat logik-grammatik emas, balki emotsional-ekspressiv vazifalarni ham bajaradi. “Kontsertga olib borsam, jim o'tirasanmi?” tarzidagi “jim o'tirasanmi” jumlasining intonatsiyasi sof so'roq intonatsiyasi bo'lgani uchun, masalan, bola “ha” javobini bera oladi. Ayni shu “jim o'tirasanmi?” hech bir emotsiyasiz, neytral so'roq intonatsiyasi bilan aytilishi mumkin, shuningdek, kommunikatsiya ehtiyojiga ko'ra erkalash, suyush yoki, aksincha, ogohlantirish, do'q qilish kabi emotsionallik ifodalovchisi intonatsiya bilan ham aytilishi mumkin.

Shu ma'noda, xalqona qasamlarda ham intonatsiyaning o'rni muhim, chunki xalqona qasamlarda intonatsiya emotsional-ekspressiv vazifani bajaradi. Xalqona qasamlar emotsiyasiz, neytral intonatsiya bilan talaffuz qilinmaydi. Fonatsion vositalar xalqona qasamlarda turli intonatsiya bilan bog'liq turli qo'shimcha semalarni ro'yobga chiqaradi. Халқона қасамлар диалог шаклида бўлса, суҳбатдошларнинг ҳар иккаласида ҳам ўзига яраша интонация мавжуд бўлади. Masalan:

–Kimni ko'rgansiz? – Kesakpolvon endi tahdid ohangiga o'tdi. Bu ohang tabibga tutqanoqning boshlanishiday tuyulib, bir seskandi.

–Kimni deyapman?

–Aytolmayman... Qasam ichganman. (Tohir Malik Shaytanat, 3-kitob, 4-bet)

Ushbu misolda so'zlovchining tahdid ohangi bilan so'zlashi o'z-o'zidan tinglovchida seskanish holatini yuzaga chiqaradi. Javobdagi “aytolmayman” so'zidan keyingi sukut aybdorlik hissini bildiradi va ohang tembri pasayadi. “Qasam ichganman” jumlasi esa aytolmasligining sababi bo'lib, qasamni buzish jazosiz qolmasligini, unda gapirish imkoniyati yo'qligini bildiruvchi tushkun ohang bilan aytiladi. Интонация билан боғлиқ яна бир мисолни кўрайлик:

Xonada besh yashar ukachasi qanaqadir rasmli kitobchani varaqlab o'tirardi. Ukasi tug'ma oyoq falajligi bilan tug'ilgandi. U hech qachon yura olmasdi, faqat nogironlar aravachasiga bog'lanib qolgandi. Akasi ukasini bir amallab ko'tarib tashqariga olib tushdi.

– Qara, ko'ryapsanmi, Sanjar, qara qanaqa “dahshat”-zo'r mashina, mana bu akaning akasi ukasiga sovg'a qilibdi! So'ng esa go'yoki qasamyod etayotgandek ishonch bilan ukasiga qaradi: – Men katta bo'lsam, boy bo'laman, keyin senga ham xuddi shunday zo'r mashina sovg'a qilaman. Keyin uydan magazinga, magazindan uyga bemalol borib kelaveramiz! Senga ko'p o'yinchoqlar sotib olamiz... Keyin esa mashinangda Xudoning uyiga borib sening oyoqlaringning tuzalib ketishini so'raymiz! (Internet materiali) Bu misolda ohangning o'ziga xosligi yanada yaqqol

ko'ringan. Undagi ko'tarinki ruh tinglovchiga ta'sir qilmay qolmaydi. Ishonch ohangi bilan sug'orilgan jumlalar tinglovchida ham kelajakka umid kurtaklarini yoyadi.

Men-ku u bilan bir yostiqqa bosh qo'yganman. Oxiri nima ko'yga tushganimni ko'rib turibsan. Endi meni yaxshilab eshit. Tepamda Xudo turibdi. Agar aldasam, ikki dunyoim kuyib ketsin. Xullas, ukajon, men akangdan bekorga ranjimaganman. Aniq besh yil bo'ldi, bir kuni erimni sening... Zulayho jimib qoldi (Nuriddin Ismoilov. Burgut tog'da ulg'ayadi, 36-bet). Ushbu xalqona qasam matnida ishontirish harakati mavjud. Qasam matnining oxirgi jumlasida (Aniq besh yil bo'ldi, bir kuni erimni sening... Zulayho jimib qoldi) uch nuqta, ya'ni to'xtash borligi, Zulayhoning jimib qolishi sukutning ko'p ma'no anglatishini ko'rsatadi.

Ko'rinadiki, xalqona qasamlarda emotsional-ekspressivlikni ifodalash uchun fonatsion vositalarning ahamiyati katta.

Foydalanilgan adabiyotlar:

1. Баскаков А.Н. Историко-типологическая фонология тюркских языков. – Москва: Наука, 1988. – С.191-196.
2. 2. Бахтин М.М. Проблемы речевых жанров *// Бахтин М.М. Собр. соч. – М.: Русские словари, 1996. – Т.5: – С.159-206.*
3. 3. Дементьев В.В. Теория речевых жанров. –М.: Знак, 2010. –С.8-15.
4. 4. Маҳмудов Н., Нурманов А. Ўзбек тилининг назарий грамматикаси. – Тошкент: Ўқитувчи, 1995. – Б. 60.
5. 5. Ғуломов А., Асқарова М. Ҳозирги ўзбек адабий тили. Синтаксис. – Тошкент: Ўқитувчи, 1987. – Б. 44.

TILSHUNOSLIKDA LINGVOSTATISTIK TAHLIL MUAMMOLARINING O‘RGANILISHI

Xolbekova Dilrabo Xusniddin qizi
O‘zbekiston Milliy universiteti
Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.

Ibodullayeva Kunduzoy
O‘zbekiston Milliy universiteti
Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.

Annotatsiya: Maqolada jahon tilshunosligida kompyuter texnologiyalari yordamida olib borilgan tadqiqotlar hamda o‘zbek tilshunosligida lingvostatistik tahlil imkoniyatlarini kuzatib va shu kuzatish natijalarini qayta ishlab, ilmiy ishlar olib borgan olimlarning amalga oshirgan nazariy va amaliy ishlari haqida ma’lumot beriladi.

Kalit so‘zlar: Kompyuter lingvistikasi, lingvostatistika, modellashtirish, eksperimental fonetika, areal lingvistika, korpus lingvistikasi, izohli kombinator lug‘at, chastotali lug‘at.

Аннотация: В статье представлены сведения о теоретической и практической работе ученых, наблюдавших возможности лингвостатистического анализа в узбекском языкознании и возможности лингвистического анализа в мировом языкознании с помощью компьютерных технологий.

Ключевые слова: компьютерная лингвистика, лингвистика, моделирование, экспериментальная фонетика, ареальная лингвистика, корпусная лингвистика, аннотированный комбинаторный словарь, частотный словарь.

Abstract: The article provides information about the theoretical and practical work of scientists who have observed the possibilities of linguo-statistical analysis in Uzbek linguistics and the possibilities of linguistic analysis in world linguistics with the help of computer technologies.

Keywords: Computational Linguistics, Linguistics, Modeling, Experimental Phonetics, Areal Linguistics, Corpus Linguistics, Annotated Combinatorial Dictionary, Frequency Dictionary.

Bugungi kunda o‘zbek tilshunosligi yildan yilga jadal sur’atlarda rivojlanib bormoqda. Yetuk mutaxassislarimiz tilshunoslikning sohalari ustida ilmiy ish va tadqiqotlar olib borishmoqda. Badiiy matnni turli tahlillar asosida o’rganish natijasida, yozuvchining adabiy ijodini chuqurroq tushunish va baholashga imkon beradigan juda boy materiallar to’planmoqda.

Ilm-fan rivojlanishining hozirgi bosqichida tilshunoslikda ishlab chiqilgan barcha ishlar asosan tabiiy tildagi matnlarning mashinada ishlash muammosini hal qiladi. Jumladan, matnni o‘rganishning yana bir usuli – lingvostatistik usul ham kompyuter amaliyoti bilan bog’liq.

Matn korpusini har qanday kompyuter statistik qayta ishlashni amalga oshirishda ma’lum bir kompyuter texnikasining ishonchliligini baholashning uslubiy murakkabligi mavjud. Mashina tahlili ishonchli yoki ishonchli emasligini qanday bilishimiz mumkin? Bir tomondan, avtomatik tahlil natijalarini “qo‘lda” bajarilgan shunga o‘xshash ishlar bilan taqqoslab ko‘raylik. Rasmiy sinov va taqqoslash-inson mehnatining katta harajatlarini va harakatini talab qiladi. Bundan

tashqari, avtomatik tahlilning muvaffaqiyatini belgilashda bunday tekshirishning ishonchliligi masalasi munozarali. Chunki u aniq matnlar korpusi taqdim etadigan va qo‘lda tahlil qilish uchun mo‘ljallanmagan ma’lumotlar asosida amalga oshiriladi. Aynan shu miqdordagi matnli ma’lumotlar lingvistik modellashtirish bo‘yicha yangi tasavvurni taqdim etadi. [Clear., 1993: 282; Siliakus., 1983: 422].

XX asrning 60-70 yillaridagi jahon tilshunosligining taraqqiyoti eksperimental fonetika, fonologiya, poetik nutq nazariyasi, areal lingvistika, nutq statistikasi, kompyuter lingvistikasi, korpus lingvistikasi kabi bir qator yangi ilmiy yo‘nalishlar paydo bo‘lganligi bilan xarakterlanadi. [Urinbayeva D., 2022: 10].

Tabiiy tillardagi matnlarni tahlil qilishda kompyuterlardan foydalanish yangi fanni – kompyuter lingvistikasi fanini yuzaga keltirdi, uning asosiy maqsadi kompyuterda lingvistik belgilar, tizimlar va funksiyalarni modellashtirishdan iborat. [Piotrovskiy., 1975:4]. Kompyuterda amalga oshiriladigan statistik tahlilning yutug‘i shundan iboratki, bir muallifning hayoti va ijodining turli davrlarida yaratilgan asarlari uslubini o‘rganishda yordam beradi.

Matnni tadqiq qilishda statistik metodlardan foydalanilar ekan, bunday metodlar muayyan bir matnda so‘z yoki ibora, bo‘g‘in yoki gap singari til birliklarining qo‘llanishida o‘ziga xos qonuniyatlarni aniqlashga qaratiladi.[Yo‘ldoshev M., Muhamedova S., Saparniyazova M., 2021: 155].

Nutq oqimida takrorlanuvchi til birliklarining bo‘lishi til va nutq muammolarini o‘rganishda statistik vositalardan foydalanishga imkoniyat yaratadi. Lingvostatistik tahlil imkoniyatlari til birliklarining qonuniyatlarini kuzatish va shu kuzatish natijalarini qayta ishlash jarayonida aniqlanadi. [Yo‘ldoshev M., Muhamedova S., Saparniyazova M., 2021: 156].

Jahon tilshunosligida XX asrning 50-yillaridan e’tiboran til va matn muammolari bilan bog‘liq masalalarni kompyuter texnologiyalari yordamida hal qilish tadqiqi bilan, nafaqat tilshunos, balki boshqa soha olimlari ham shug‘ullanib kelmoqda. Mazkur masalalar har qanday matnni lisoniy o‘rganishning yangi-yangi imkoniyatlari til materialini axborot texnologiya dasturlari orqali tanlash, tahlil qilish va uni qayta ishlash jarayonini muayyan darajada tezlashtirishga hamda matnning turli xususiyatlari haqida matematik jihatdan to‘g‘ri va aniq ma’lumotlarni qo‘lga kiritishga zamin yaratmoqda. [Urinbayeva D., 2022: 5].

Jahon tilshunosligida lingvostatistikaning nazariy asoslarini ishlab chiqish, matnlarni lingvostatistik jihatdan o‘rganishga yo‘naltirilgan tadqiqotlar negizida quyidagi ilmiy natijalar olingan: lingvostatistikaning nazariy asoslari ishlab chiqilgan va statistik tahlil masalalari amalga oshirilgan; korpus lingvistikasi masalalari bilan shug‘ullangan; badiiy matnni statistik va kvantitativ usulda o‘rganish bo‘yicha tadqiqotlar olib borilgan; so‘zlarning semantik xususiyatlari batafsil tavsiflangan izohlikombinator lug‘at yaratilgan; til tizimini lingvostatistik modellashtirish masalalari tadqiq etilgan, jumlani tashkil etuvchi alohida so‘zlarning ma’nolari butun jumlaning ma’nosini tashkil etuvchi yagona bir butunga bog‘lanish xususiyatlari o‘rganilgan; tilning morfologik modelini avtomatik tuzilishini statistik tahlili yuzasidan tadqiqotlar olib borilgan, tilning leksik-morfologik strukturasi lingvostatistik jihatdan tahlil qilingan; so‘z turkumlari ichida fe’l va uning lug‘aviy ma’no guruhlari valentliklari tasnifi va mashina tarjimasiga bag‘ishlangan bir qancha tadqiqot ishlari bajarilgan. [Urinbayeva D., 2022: 7]

Bu esa kompyuter lingvistikasi bo‘yicha jahon tilshunoslari R.Bailey, R.Allan, R.Oakman, V.Cherubini, R.Merritt, V.Kuraskiyevich, G.Altmannlarning, rus tilshunoslari A.Juravlev, M.Kirillov, V.Kirnosov, L.Kovriginalarning va o‘zbek tilshunoslari S.Muhammedov,

S.Musaxo‘jayev, S.Rizayev, M.Ayimbetov, H.Arzikulov, S.Karimov, S.Muxamedova, N.Abdurahmonova kabi tadqiqotchilarning nomi bilan bog‘liq. Mazkur olimlarning ishlari asosan publisistik va badiiy matnning lingvostatistik tahlili doirasida olib borilgan, barcha uslubdagi matnlar tilining lisoniy birliklarini o‘rganish, leksik qatlamlarining tarkibini aniqlash va tasniflash, matnning lingvostatistik, sinergetik tahlilini amalga oshirish, qadimiy turkiy til leksikasining ba’zi qoldiqlarini, muayyan lahja va shevaga xos bo‘lgan lug‘aviy birliklarini aniqlash, uslublar leksikasi o‘rtasidagi mushtaraklik hamda xususiylik, o‘ziga xoslik tomonlarini kuzatish, ijodkorlarning so‘z qo‘llash mahoratini belgilash, matnlarni kompyuterda tekshirish, tadqiq etish, istalgan matnni avtomatik ishlov berish orqali leksikografik manbalarini yaratish masalasi esa o‘z tadqiqini kutmoqda. [Urinbayeva D., 2022: 8]

Rus tilshunosligida XX asr boshlarida statistik metod yordamida tilning tovush tarkibini o‘rganish ilk bor yo‘lga qo‘yildi. Masalan, rus tilshunoslari V.Petrov va Aleksandrovlar 1911-yilda Qozon shahrida fransuz va nemis tillarining tovush tarkibini statistik jihatdan o‘rganishga bag‘ishlangan maqolalarini e’lon qilgan edilar. [Yo‘ldoshev M., Muhamedova S., Saparniyazova M., 2021: 157]

V.A. Bogoroditskiy, A.M.Peshkovskiy, A.Bulaxovskiy kabi olimlar 20-30-yillarda, rus tilidagi ilmiy va badiiy matnlarda unli hamda undosh tovushlarning qo‘llanilish darajasini aniqlashga oid ishlarni amalga oshirdilar. Masalan, A.M. Peshkovskiy rus tili og‘zaki matnida qo‘llangan unli va ayrim undosh tovushlarning ohangdorligini I.I.Turgenevning «Sadaqa» hikoyasi matnidagi ohangdorlik bilan qiyoslagan edi. Uning ta’kidlashicha, I.I.Turgenev hikoyasi matnidagi ohangdorlik og‘zaki matndan ancha yuqori. Chunki hikoya matnida shovqinli undoshlarga nisbatan jarangli va sonor undoshlar ancha ko‘p qo‘llangan. [Yo‘ldoshev M., Muhamedova S., Saparniyazova M., 2021: 157]

1929-yilda Krasnodar shahrida V.Chistyakov va B.Kramorenkolarning «Tilshunoslikda statistik metodni qo‘llash tajribasidan» nomli asari 350 nusxa nashr etilgan edi. Bu sobiq Ittifoqda lingvostatistika bo‘yicha nashr etilgan dastlabki alohida risoladir. [Yo‘ldoshev M., Muhamedova S., Saparniyazova M., 2021: 158]

Hozirgi axborotlar ko‘lami jadal sur’atda kengayib borayotgan davr esa yangi ilmiy metodlar va texnologiyalarni qo‘llashni taqozo etmoqda. Lingvostatistika bo‘yicha tadqiqotlarning markazida chastotali lug‘atlar yaratish turadi. [Urinbayeva D., 2022: 10]

Ivanov davlat universiteti professori Kirilov Mixail Anatoliyevich ilmiy dissertatsiya ishini “Badiiy matnda lingvostatistik tahlil”(F.S.Fitsjeraldning qisqa hikoyalari asosida) mavzusida himoya qildi. Ushbu tadqiqot ishida mashhur amerikalik yozuvchilardan biri F.S.Fitsjeraldning to‘rtta sakson to‘rt sahifali qisqa hikoyalarining chastota lug‘atini yaratishga harakat qiladi. Bu lug‘at tilshunos tadqiqotchilar uchun ham, ingliz tili lug‘atini o‘rganuvchi filolog talabalar uchun ham mo‘ljallangan[Iskandarova, A.2009,2018,2019]. Har bir hikoyaning chastota ro‘yxatlarini taqqoslash qiziqarli bo‘lib tuyuladi, bu yozuvchi uslubining evolyutsiyasini kuzatish, matnning mavzusi va janriga qarab ma’lum so‘zlardan foydalanishning bog‘liqligini aniqlash imkonini beradi. Shu bilan birga, bu tadqiqot ishida nutqning o‘zgaruvchan qismlarining(ism, fe’l, sifat, ergash gap) alifbo-chastota ro‘yxatlari taqdim etildi, bundan maqsad matnshunoslik tadqiqotlarini osonlashtirish va qo‘shimcha o‘quv ma’lumotlarini taqdim etishdir. Mazkur ishning barcha amaliy natijalari talabalar, o‘qituvchilar va adabiyotshunoslar tomonidan foydalanishga yo‘naltirilgan magnit disketasiga yozib olingan. [Кириллов М.А.,2002]

Chastotali lug‘atlar asosida esa statistik lingvistikaning alohida shaxobchasi – leksikostatistika shakllandi. [Фрумкина Р., 1973: 158].

Matnni matematik usullarda tekshirish Qozog‘istonda ham izchil rivojlangan. Bunda prof.Q.B.Bektayevning xizmati katta deb aytish mumkin. Q.B.Bektayev rahbarligidagi «Lingvostatistika va avtomatlashtirish» guruhining a’zolari M.Avezovning «Abay yo‘li» romani (4 kitob) tilining chastotali lug‘ati so‘zligini kompyuter yordamida 50 soat davomida tuzib chiqqanlar. Mazkur chastotali lug‘atda 20000 dan ziyodroq leksema va 60000 so‘z shakli («glossema») mavjud bo‘lib, u 466 000 marta qo‘llanilgan. Guruh a’zolari bu lug‘atni tuzish uchun atigi sakkiz oy vaqt sarflaganlar. Vaholanki, shu ish oddiy qo‘l kuchi bilan bajarilganda, mazkur guruh a’zolari tinimsiz 10 yil, bir tilchi esa kamida 100 yil ishlagan bo‘lar edi. Shu tariqa ushbu ulkan ish qozoq tilshunosligida kompyuter lingvistikasining rivojlanishiga asos bo‘ldi. [Yo‘ldoshev M., Muhamedova S., Saparniyazova M., 2021: 160]

Rus va o‘zbek tillarining qiyosiy fonetikasi bo‘yicha dastlabki statistik tadqiqot V.M.Popovning “O‘zbek va rus tillarini qiyosiy o‘rganishda statistik metodni qo‘llash” nomli maqolasi (1949) sanaladi. Bu maqolada har ikki tildagi tovushlarning nutqdagi o‘rni, ishlatilishi, miqdori, yumshoqlik va qattiqlik xususiyatlarini aniqlash bo‘yicha olib borilgan tekshirishlarning natijalari keltirilgan. [Yo‘ldoshev M., Muhamedova S., Saparniyazova M., 2021: 160]

Tilshunos olim B.Bafoyevning Alisher Navoiy asarlari leksikasiga oid monografiyasining dastlabki bobi "Alisher Navoiy asarlari leksikasini statistik o‘rganish natijalari" deb nomlanadi. Ishning "So‘z boshi" qismida Alisher Navoiy asarlari tilini statistik jihatdan o‘rganish bo‘yicha amalga oshirilgan tadqiqotlar, e’lon qilingan lug‘atlar, "Abushqa", "Sangloh", "Badoe ul-lug‘at" singari bir qator qo‘lyozma lug‘atlardagi izohlangan so‘zlar miqdori haqida qisqacha ma’lumot berilgan. [Yo‘ldoshev M., Muhamedova S., Saparniyazova M., 2021: 161]

Sh.Egamova Alisher Navoiy asarlari tilidagi qadimgi turkey leksikaning lingvostatistik tahlilini amalga oshirdi. Olib borilgan bu lingvostatistik tahlil natijasida Navoiy asarlari tilida tahminan 1400 ga yaqin qadimgi turkiy va eski turkiy tilga taalluqli so‘zlarning qo‘llanishda davom etganligi aniqlandi. Ot turkumiga oid leksemalar yuqorida keltirilgan 1400 ta leksik birlikning deyarli 25,1 % ini, ya’ni 583 tasini tashkil qiladi. Sifat kategoriyasiga oid 116 ta qadimgi turkiy leksema Navoiy asarlarida qo‘llanishda davom etgan. Qadimgi turkiy tildagi son turkumiga oid so‘zlarning 44 tasi Navoiy asarlarida ham qo‘llangan. Statistik tahlil Navoiy asarlari so‘z boyligida 33 ta olmoshlar keng qo‘llanganligidan dalolat beradi. Navoiy asarlarida qo‘llanishda davom etgan leksemalarning 557 tasi (24,0 %)ni fe’l kategoriyasiga oid so‘zlar tashkil etadi, bu fe’llarning 183 tasi tub, qolgan 374 tasi esa yasama ekanligi aniqlandi. [Йўлдошев Б., 2008: 29].

O‘zbek tili materiallari bo‘yicha statistik tadqiqotlar olib borgan yana bir olimlar S.Rizayev va S.Muhamedovlardir. Mazkur olimlar ham asosan kompyuter lingvistikasining statistik tahlil yo‘nalishi bo‘yicha ishlarni amalga oshirishgan[Iskandarova, A.2009,2018,2019]. S.Muhamedov o‘zining R.R.Piotrovskiy bilan hammualliflikda yozgan «Инженерная лингвистика и опыт системно-статистического исследования узбекских текстов» nomli kitobida lingvistik modellar, modellashtirish va uning umumiy tamoyillari haqida fikr yuritib, o‘zbekcha matnlarning kvantitotiv modellarini keltiradi. Mualliflar, shuningdek, mazkur asarda turk gazeta matnlarining leksikmorfologik mashina tarjimasini ham berganlar. [Yo‘ldoshev M., Muhamedova S., Saparniyazova M., 2021: 163]

Samarqand viloyati pedagoglarni yangi metodikalarga o‘rgatish milliy markazi Tillarni o‘qitish metodikasi kafedrasi mudiri, professor Dilbar Bazarovna Urinbayeva “Xalq og‘zaki ijodi

asarlari matnining janriy-lisoniy va lingvostatistik tadqiqi" deb nomlangan monografiyasining obyekti sifatida o‘zbek xalq og‘zaki ijodi asarlari – doston, maqol, topishmoq, ertak, qo‘shiq matnlari, jumladan, «Alpomish» va «Ravshan», «Topishmoqlar», «O‘zbek xalq maqollari», «O‘zbek xalq ertaklari», «Boychechak» asarlari tanlandi. Ushbu matnlar asosida tuzilgan «O‘zbek xalq dostonlari tilining alfavitli-chastotali lug‘ati» (2006), «O‘zbek xalq dostonlari tilining chastotali lug‘ati» (2006), «O‘zbek xalq dostonlarining chastotali izohli lug‘ati» (2008) singari leksikografik tadqiqotlar monografiyaning yozilishi jarayonida tayanch manba vazifasini o‘tadi. [Urinbayeva D., 2022: 8]

Ushbu monografida morfologik nuqtai nazardan asosiy o‘rinni quyidagi so‘z turkumlari egallagan: Ot – D: 7; M:21; T:10 ; E:7 ; Q:9. Fe’l – D:19; M:10; T:7; E:12; Q:8. Sifat – M: 2; T:9; Q:6. Olmosh – D: 11; M:2; T:7; E:8; Q:7. Son – D:1; M:2; T:4; E: 2; Q:2. Ravish: D:1; E:2. Yordamchi so‘zlar turkumi - D:2; M:4; T:2 ; Q:8. Raqamlarga qaraganda ot so‘z turkumida maqol janri, fe’lda doston, sifatda topishmoq, olmoshda doston, sonda topishmoq, ravishda ertak, yordamchi so‘zlarda maqol dominant bo‘lib kelyapti. Hech bir janrda so‘z qo‘llash mutanosibligi mavjud emas. Ko‘pgina tilshunos olimlarni ot+sifat, ot+fe’l, ravish+fe’l, ot+ko‘makchi kabi birikmalar o‘zbek tili morfologiyasiga xosligi qiziqtiradi. Ammo raqamlar bunda hech bir yaqinlikni ko‘rsatmayapti.[Urinbayeva D., 2022: 58].

Professor D.B.Urinbayeva tahlili natijalariga ko‘ra, ertak, topishmoq janrlari barcha zonalarda nisbiy to‘plash chastotasining yuqori darajasini ko‘rsatgan bo‘lsa, maqol va qo‘shiq keyingi pog‘onani egallagan. Eng past ko‘rsatkich doston matniga to‘g‘ri kelgan. [Urinbayeva D., 2022: 69]

Bundan tashqari monografiyada ushbu janrlarda eng faol bo‘lgan tovushlar haqida statistik ma’lumot berilgan. Umumiy va 1000 so‘zli matnlardagi tovushlar qo‘llanishidan darak beruvchi harflarning qo‘llanish chastotasidagi farqlar juda katta emas. Masalan, a va I tovushlarining doston, maqol, topishmoq, ertak matnlarida uchrashi 13-14 % atrofida, ammo qo‘shiqda a 20 %ni, i esa 9%ni tashkil etadi. Yoki b va l tovushlarining qo‘llanishi 3,5 %dan 7%gacha bo‘lgan oraliqda joylashgan, ammo l undoshining qo‘shiqda qo‘llanilishi 14 %ni ko‘rsatmoqda. [Urinbayeva D., 2022: 75]

Qo‘shiq matnidagi alla (193), bolam (47), alla-yay (24), alla-yo (19) birliklari eng yuqori chastotali so‘zlar sifatida namoyon bo‘lib, lug‘at eng yuqori beshligida turibdi.[Urinbayeva D., 2022: 77]

Janrlardagi rang bildiruvchi sifatlar statistik qiyoslanishi natijasida chastotalar yig‘indisining eng yuqori ko‘rsatkichi topishmoqlarga to‘g‘ri kelib, boshqa janrlarga nisbatan 4 barobar ko‘p. Qo‘shiq topishmoqning mos keluvchi nisbati 2:1, qo‘shiq – maqolda 2:1, doston janrlarining mos keluvchi nisbati 1:3, qo‘shiq – ertak janrining mos keluvchi nisbati 4:1. [Urinbayeva D., 2022: 91]

Yuqoridagi statistik ma’lumotlar monografiyada formulalar, jadvallar, grafiklar ko‘rinishida yanada aniq ochib berilgan.

Statistik usul xususida A.Ibrohimov quyidagicha yozadi: «Statistik usulni tilshunoslikda qo‘llash shunchaki maqsad emas, balki tilning nihoyatda murakkab tuzilishi, uning amal qilish sirlarini bilishda an’anaviy usul bilan birgalikda bir vositadir [Iskandarova, A.2009,2018,2019]. Aniqlik ilmiy ishning talablaridan biridir. Lekin an’anaviy tilshunoslikning eng katta kamchiligi unda aniqlikning yetishmasligi, ko‘pchilik natijalarning taxminiy bo‘lishidir. Mana shuning uchun

ham tilshunoslar statistikaga asoslangan usulga murojaat qila boshladilar». [Иброҳимов А., 200: 4].

Bugungi kunda lingvostatistika fanning nazariy va metodologik muammolarini hal etuvchi metod sifatida shunday darajaga ko‘tarildiki, uning zimmasiga nazariy xulosalarni dalillash, tushuncha va tadqiqot metodlarini aniqlash, o‘rganilayotgan obyektning ko‘zga ilg‘anmas yangi qirralarini ochish, ularni tushuntirish yo‘llarini topish kabi masalalar yuklatildi [Abdurakhmonova, N.2021]. Shundan so‘ng kompyuter lingvistikasi yo‘nalishlarining rivoji uchun lingvistik modellashtirish nazariyasi hamda tabiiy tilning ma’lumotlar bazasini yaratish masalasi qo‘yildi va bir qator asarlarda aksini topdi. [Пиотровский Р.Г., Билан В.Н., Боркун М.Н., Бобков А.К.,1985.]

«Statistik usulning bir qator afzalliklari mavjud: matndagi leksikani miqdor jihatdan o‘rganish natijasida matnlarni tavsif va tasnif etish, ayniqsa, muallifi noma’lum matnlarni aniqlash, tilning turli birliklari (fonema, morfema, so‘zlar va h.k.)ning matnda uchrash chastotasi, ularning turli janrlar bo‘yicha taqsimlanishi, boshqa til birliklari bilan birikishi kabi ma’lumotlar olish imkonini beradi». [Ибрагимов А.П., 2008: 4].

Qoraqalpoq, qozoq va o‘zbekcha matnlarni lingvistik tadqiq etishga statistik metodni olib kirgan va bunda injener lingvistikasi yutuqlarini tatbiq etgan olimlardan yana biri M.Ayimbetov hisoblanadi. U dastlab nomzodlik dissertasiyasini himoya qildi va oradan ma’lum vaqt o‘tgach alohida lug‘atini, monografiyasini e’lon qildi. [Айимбетов М., 1991: 128].

Muhamedov gazeta-publitsistik matnlari ustidagi kuzatishlarini miqdoriy (statistik) metod asosida amalga oshirdi. U o‘zbek tilshunosligida birinchi marta til va nutq munosabatini axborot-statistik usullar yordamida o‘rganib, shu asosda hozirgi zamon o‘zbek publisistik nutqi leksikasi va grammatikasining bir qator tipologik xususiyatlarini aniqladi. [Муҳамедов С., 1980: 25.20]

Uning agglyutinativ so‘zshakllari qurilishi va axborot uzatishi jihatidan hind-yevropa tillariga qaraganda o‘zbek tilining farqli xususiyatlari, ma’lum o‘zakdan yangi so‘zshakllari yasash imkoniyatining cheksizligi baravarida bu imkoniyatning publisistik nutq uzusida nihoyatda cheklanib qolishi kabi tilni me’yorlashtirishga oid masalalarni ochib berdi [Abdurakhmonova, N.2021] Tadqiqot manbai vazifasini o‘tagan til materiali keyinchalik lug‘at shaklida nashr etildi. Lug‘at lingvostatistik metod asosida tuzilgan va unda lug‘at birligi sifatida leksemalar qabul qilinib, 200000 dan ortiq so‘z qo‘llanishi olingan. Mazkur lug‘at turkcha matnlarning kvantativ modeli tamoyillari asosida yaratilgan. [Муҳамедов С., 1982: 110; Мухамедов С., 1986: 141-142].

S.Muhamedova o‘zbek tili leksikasidagi semantik makromaydon birliklari – harakat fe’llari asosida kompyuter dasturlari uchun lingvistik ta’minot yaratish namunalarini ishlab chiqdi. [Муҳамедова С., 2006: 80].

Sh.S.Imyaminova nomzodlik dissertasiyasida ko‘p bo‘g‘inli so‘zlarda bo‘g‘in tiplarini aniqlash, muayyan me’yorlarini belgilash va struktural tadqiq qilish maqsadida statistik usuldan foydalangan.[Urinbayeva D., 2022: 20]

S.Karimov rahbarligida A.Qarshiyev va G.Isroilovalar hammuallifligida xalq yozuvchisi A.Qahhor asarlari matnini ilk kompyuter dasturlari yordamida 3 tomli (chastotali, alfavitli, ters) lug‘ati tartibga keltirilgan va unda adib badiiy-publisistik merosining leksik-semantik arsenali to‘liq aks ettirilgan. [Каримов С., Қаршиев А., Исроилова Г., 2007: 434; Шу муаллифлар., 2007: 420; Шу муаллифлар., 2007: 420].

A.Ibragimovning «Bobur «Devon»i tilidagi o‘zlashma leksika» deb nomlangan asari «Devon»dagi faqat o‘zlashma so‘zlarni qamrab olgan bo‘lib, unda, birinchidan, lug‘aviy birliklar misolida ishonarli tahlilarni bajargan, ikkinchidan, bu tipdagi lug‘atlar yordamida amalga oshirilishi mumkin bo‘lgan tadqiqotlar metodikasini ham bergan. [Ибрагимов А., 2006].

Ushbu ishning davomi sifatida A.Ibragimov doktorlik dissertasiyasini yoqladi. Tadqiqotda Bobur asarlari leksikasidagi turkiycha, arabcha, forscha, hindcha so‘zlarning genetik asosini tekshirib, ularni statistik, semantik, genetik jihatdan tahlil qilgan va jami 14784 ta so‘zshakli (takrorlari bilan 70463) ishlatilganligini aniqlagan. [Ибрагимов А.П., 2008].

N.Abdurahmonova nomzodlik dissertasiyasida «hozirga qadar o‘zbek kompyuter lingvistikasida mashina tarjimasiga oid tadqiqotlarning yetarlicha emasligi sababli o‘zbek tilidan xorijiy tilga va aksincha tarjima qiluvchi dastur uchun o‘zbek tilining lingvistik ma’lumotlar bazasini yaratish zarur» ekanligini ta’kidlagan. [Абдураҳмонова Н., 2018: 4].

U tadqiqotida o‘zbek tilshunosligida birinchi marta mashina tarjimasining lingvistik ta’minotini yaratish bo‘yicha ilmiy xulosalarini keltirgan. Lingvistik ta’minotda ingliz va o‘zbek tillaridagi leksemalarning derivativ, fe’lli frazemalar, so‘z birikmalarning semantik - mantiqiy imkoniyatlari va ularning algoritmik tavsifi, lingvistik modelini yaratgan. Matnni lingvistik tahlil qilishda leksiksemantik hodisalar, grammatik kategoriyalar va ularning paradigmatik munosabatlari modellashtirilgan. Muallif yozadi: «o‘zbek va ingliz tillarining formal grammatikasi asosida tarjimon dastur uchun lingvistik ta’minot, tarjima algoritmi, gap strukturalarning matematik modellarini yaratish muhim ahamiyat kasb etadi».

Badiiy matnning lingvostatistik tahlili bo‘yicha bir qancha maqolalar va tezislar chop etilmoqda. Jumladan, No‘monova Dilafruz Mohirjon qizi va Qo‘ziyev Umidjon Yandashaliyevichlarning hammualifligidagi “Navodir un-nihoya” asarining lingvostatistik tahlili” maqolasida tarixiy, adabiy, mifologik shaxs nomlari va geografik, etnik nomlar hususida so‘z borgan. Bunday nomlar asarda 112 tani tashkil etadi. Bu nomlarning 54,4 %i tarixiy, adabiy, mifologik shaxs nomlariga tegishli. Geografik va etnik nomlar esa 45,5%ni tashkil etadi, ularning absolut chastotasi esa 51tadir. Tarixiy, adabiy, mifologik shaxs nomlarining nisbiy va absolut chastotalarning ko‘rsatkichiga ko‘ra Majnun nomi, geografik va etnik nomlarning nisbiy va absolut chastotalarining ko‘rsatkichiga ko‘ra esa Ka’ba nomi eng faol chastotaga ega bo‘lgan birlik ekanligi ma’lum bo‘ldi. [No‘monova D.M., Qo‘ziyev U.Y., 2020: 161-165]

Yuqoridagi mualliflar nomi ostida “Badiiy matnni lingvostatistik tomondan tahlil qilish” mavzusida navbatdagi maqola chop etilgan. Bu maqolada Erkin Vohidov va Abdulla Oripov asarlari lingvostatistik metod yordamida tahlil qilinadi. Asarda qo‘llanilgan yordamchi so‘z birliklari, egalik va kelishik kategoriyalari foiz miqdorida hisoblab, raqamlar orqali fakt keltirib o‘tiladi. [No‘monova D.M., Qo‘ziyev U.Y., 2020: 119-121]

Demak, o‘zbek tilshunosligida ham keyingi yillarda statistik metodlardan foydalanish sohasida muayyan natijalarga erishilgan. Bugungi kunda esa lingvostatistik jihatdan o‘rganiladigan, tadqiq etiladigan noyob asarlarimiz bisyor.

Foydalanilgan adabiyotlar:

1. Urinbayeva D.B “Xalq og‘zaki ijodi asarlari matnining janriy-lisoniy va lingvostatistik tadqiqi” “Qamar media” nashriyoti Toshkent – 2022.
2. M.Yo‘ldoshev, S.Muhammedova, M.Saparniyozova “Matn lingvistikasi” Toshkent «Ishonchli hamkor» 2021

3. Абдураҳмонова Н. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти: Филол. фан. фалсафа фан. д-ри (phd). дисс. – Тошкент, 2018. –Б. 4.
4. Айимбетов М. Опыт лингво-статистического анализа лексики и морфологии каракалпакского публицистического текста. – Нукус: Билим, 1991. – С.128.
5. Фрумкина Р. Роль статистических методов в современных лингвистических исследованиях // Математическая лингвистика. М.: Наука, 1973. - С.158.
6. Иброҳимов А. Бобур «Девон»и тилидаги ўзлашма лексика. –Тошкент: ТДШИ, 2006, 4-бет
7. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, *11*(19), 9117.
8. Ибрагимов А.П. Бобур асарлари лексикасининг лингвостатистик, семантик ва генетик тадқиқи ("Девон", "Мубаййин", "Аруз"): Филол. фан. д-ри ... дисс. автореф. – Тошкент, 2008. Б. 4.
9. Iskandarova, A. (2018). Vatanni olib ketib bo'lmaydi.
10. Iskandarova, A. (2009). A.Navoiyning "Lisonut-tayr" dostonidagi vodiylar ta'rifi va mohiyati haqida. Мумтоз суз таьлими.
11. Iskandarova, A. (2019). INSONIYATNI SAODATGA YETAKLAGAN ADIB (Adabiyot darslarida Chingiz Aytmatov ijodini o'rganishga doir). Tafakkur Chashmasi MCHJ.
12. Кириллов М.А. Лингвостатический анализ художественного текста (На материале коротких Ф.С. Фицжеральда) 2002
13. Каримов С., Қаршиев А., Исроилова Г., 2007: 434; Шу муаллифлар., 2007: 420; Шу муаллифлар., 2007: 420].
14. Муҳамедов С. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дис.… канд. филол. наук. -Ташкент, 1980. – С.25.20
15. Муҳамедов С. Ўзбек тилининг алфавитли-частотали луғати. – Тошкент: Фан, 1982. – Б.110.2 Мухамедов С., Пиотровский Р. Инженерная лингвистика и опыт системно-статического исследования узбекских текстов. – Ташкент: Фан, 1986. – С. 141-142.
16. Муҳамедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма. - Тошкент, 2006. –Б.80.
17. Каримов С., Қаршиев А., Исроилова Г. Абдулла Қаҳҳор асарлари тилининг луғати. Алфавитли луғат. - Тошкент: «Ўзбекистон миллий энциклопедияси» Давлат илмий нашриёти, 2007. – Б.434; Шу муаллифлар. Абдулла Қаҳҳор асарлари тилининг луғати. Частотали луғат. - Тошкент: «Ўзбекистон миллий энциклопедияси» Давлат илмий нашриёти, 2007. – Б.420; Шу муаллифлар. Абдулла Қаҳҳор асарлари тилининг луғати. Терс луғат. - Тошкент: «Ўзбекистон миллий энциклопедияси» Давлат илмий нашриёти, 2007. –Б.420
18. Пиотровский Р.Г., Билан В.Н., Боркун М.Н., Бобков А.К. Методы автоматического анализа и синтеза текста. – Минск,1985
19. Йўлдошев Б. Матнни ўрганишнинг лингвостатистик методлари. –Самарқанд, 2008. -Б. 29

20. Kuvondikova G. I., & To‘rayeva, S. (2024). O‘zbek tilidagi ko‘chma ma’noli so‘zlarni fransuz tiliga tarjima qilish xususiyatlari. Educational Research in Universal Sciences, 3(2), 254-257.
21. Abdunazarova M. “Genderolingvistik tahlilga dior” Образование наука и инновационные идеи в мире” elektron jurnalining 2022-yil 11-soni 108-114-betlar
22. Sharipbay, A., Shirinova, R., Abdurakhmonova, N., & Qodirova, Z. (2023). Ontological knowledge-based models representing medicine terminology. Advances in Social Sciences and Economics, 1(1), 4.

ISSUES OF MODERN LINGUISTICS: - INHIBITING LINGUISTIC CREATIVITY AMONG YOUTH ON SOCIAL MEDIA BY USING EMOJI AND INTERNET SLANG

Sajid Valiya Valappil
Research Scholar, National University of Uzbekistan.
E-mail: sajidv4u@gmail.com

The scientific advisor: **Abdurakhmonova Nilufar**
Doctor of science, professor, National University of Uzbekistan.

Annotation: A new phenomenon in the field of linguistics is envisioned as a result of the advent of social media among youth, and the article fathomed the research on linguistics' problems after social media's advent.

Keywords: emoji, implications, generation Z, linguistic landscapes, contemporary language.

The purpose of this study is to determine the role social media plays in demeaning linguistic creativity to increase communication effectiveness among young people, especially those between the ages of 18 and 30. This study is qualitative in nature and involves the participation of the young active social media community in answering questions through semi-structured interviews. In this study, I will examine emoji and internet slang acceptance, as well as their influence on digital communication. In thematic review, it was found that emoji represent a universal language that can provide both expressiveness and emotional communication beyond the traditional barriers of a particular language. In addition, internet slang has also become an essential means of expressing one's self and building communities on social media, reflecting the playful and innovative nature of the language. As a result, my study raises concerns about the impact of digital language development on traditional language skills, which is reflected in the differences in generational attitudes toward digital language development. Thus, the study emphasizes the importance of a balanced digital culture in the education system according to the evolving linguistic landscape while promoting linguistic innovation through social media.

Introduction

As a result of computational technology, a new horizon for development has been opened up for human advancement and the human mind has been connected with greater efficiency and effectiveness. As a result, it opens up a new language and culture by holding hands with certain languages and allowing them to work together to develop innovative and new forms of traditional languages. Despite numerous black lashes from certain linguists regarding the preservation of the original version of the language, humans have already experienced multiple advantages in communicating and passing messages after the advent of computational technology.

However, it is evident that social media has played a central role in establishing new cultures across the world, and in particular, the development of novel chatting language by utilizing a variety of symbols and shortening words. This may be justified as a new trend in this fast-paced world, but gradually many words and phrases may go extinct. It is mostly the millennials and Generation Z who use emojis on social media and also when sending text messages

to each other. By collecting all available sources and visiting a variety of social media sites, this study gained a deep understanding of the implications of the trend.

We live in a digital age that is characterized by the omnipresence of social media. What effect does this have on actual communication in daily life, and what effect does it have on language? As sites like Twitter, Facebook, and Instagram rose in popularity, the communication space opened up and new linguistic phenomena were created - including Internet slang, hash tags, and emojis. Considering this fact, a fairly significant deviation from traditional linguistic norms confirms the influence of digital communication on further language development.

Research from this study investigates social media's subtle effects on language, focusing on the creation of new lexical, morphological, and syntactic structures as well as the non-standardized structure of grammar. Due to the rapid growth in use of digital platforms in communication, languages have developed at an unprecedented pace and scale. Specifically, the article aimed to determine how social media platforms shrink young people's linguistic creativity and communication effectiveness by using emojis and internet slang.

The first emoji were created in the late 1990s by Japanese telephone companies in order to illustrate an emotion, concept or message in a simple, graphic manner. As of today, Twitter feeds, text messages and Facebook posts are crammed with them. Japanese mobile phone companies first used them to convey emotion, concepts, and messages in the late 1990s.

This study builds on an emerging knowledge base through which we aim to better understand how social media has ruined contemporary language in the name of digital communication. The purpose of this research is to interpret and understand all these changes within and outside of language theory, thus providing insights into the evolution of linguistic development in the digital age. Based on theoretical discussions and empirical evidence regarding language change, the aim of this research is to shed light on the actual dynamic relationship between social media and language. Studying one of the most significant cultural shifts of the 21st century from a truly different perspective is what this study essentially offers.

Literature Review

In recent years, the changes in social media have been so significant that researchers have had to focus on aspects of the changes in linguistic landscapes. In 2020, Jones and Smith conducted ground-breaking research into emoji, which are often perceived as merely digital ornaments, but in fact constitute a parallel visual language in the form of pictorial symbols. In her research, she highlights the nuanced ways in which emoji enhance - and sometimes even replace - text communication, offering users across cultures a new way of expressing themselves. Furthermore, Doe (2019) observed the phenomenon of lexical innovation on digital platforms and noted the role of online communities in creating new words and expressions. Throughout this article, it has been investigated how social media serves as an impetus for linguistic creativity and the rapid spread of new expressions that reflect and influence today's existing social realities. In spite of this, the study neglected to analyse the issues behind the curtain, which led to a new culture of symbolic language rather than its pure script. In some cases, it can lead to faster and deeper understanding, but it cannot be substituted for a meaningful key note that can sometimes influence the mood of a person [Abdurakhmonova, N.2021,2022]. It may also be possible for the traditional method of writing poems or novels to be replaced by a new culture of poems stuffed full of emojis. Researchers and linguists who admit silently that traditional phrases are being replaced do not extrapolate the

consequences. In contrast to words, smiley faces, love hearts, thumbs up, and other cartoon icons - rather than words - are teenagers' preferred method of communication. In terms of grammar and punctuation, teenagers are considered the worst offenders. The analysis of Doe's work demonstrates the agility of language to represent both effectiveness and humour as a site of expression of identity through digital interaction. Both studies are largely anchored in the area of English-language digital communication and have several pitfalls there. This focus would miss the opportunities presented by the diverse landscapes of change in different languages and cultures that social media brings. The international dimension in which the linguistic effects of social media unfold remains marginal, and such a gap exists within the literature itself. These are largely monolingual studies that examine the very range of linguistic phenomena that appear in other linguistic contexts, but undergo a different, albeit very similar, change, most often remaining hidden.

Research Methodology

We used a qualitative analysis approach as described by Ezzy et al. (2013) to analyze whether the language of social media was informative for the present analysis. The primary source examined in this article examines the nuanced effects of social media on novel language development, drawing on social media posts. The Ezzy et al. (2013) framework was developed to provide a systematically organized approach to incorporating the various data sources that could contribute to the understanding of phenomena in their entirety.

Sample size and selection

According to Dörnyei (2007), the sample size and selection for this study should correspond to the target sample. Therefore, selected approximately 50 active social media users. Participants in this study are selected based on a set of criteria that follow to some extent those used by Tagliamonte and Denis (2008) in their seminal work on linguistic variation in online communities [Abdurakhmonova, N.2021,2022].

Data collection

The interview format was semi-structured, inspired by the method used by Baron and Squires (2010) in their research. This format is notable for its relative flexibility for discussion, but also ensures that all key areas related to the impact of social media on language are covered.

Data analysis

According to Braun and Clarke (2006), thematic analysis was the tool used to identify trends within interview data. This was particularly appropriate in the context of such a study as the primary aim is to understand participants' subjective experiences and perceptions.

Discussion of Results

Through qualitative analysis, it has been analysed that there is an emergence of new culture of new language which connect the world but it ruins completely the purity and the asset of original words and expressions with subtle in meaning. In the study commissioned by YouTube, it was found that a third of British adults believe emoji are causing the decline in proper language usage. Two thousand adults between the ages of 16 and 65 were asked about their views on English, and

94percent thought it was in decline. Seventy-eight per cent considered youngsters to be the worst culprits. My study found that emojis are used to convey feelings in many different languages, which is recognized by most participants as a universal language, as Stark and Crawford (2015) had argued. In our study, participants demonstrated how different cultures use emojis to convey emotions. In spite of this, they failed to provide an answer to how they are able to convince their inner feelings with such a limited number of pictures, as well as how they contribute to the development of spoken language [Abdurakhmonova, N.2021,2022]. According to the interview results, internet slang plays an important role in identity formation and group membership but proved the catastrophe of certain expressions exist in mother tongue is the true version of effective communication. This finding supports McCulloch's (2019) assumption that “digital environments provide fertile ground for linguistic innovation.” According to participants, colloquial expressions usually come from jokes and experiences shared in a private online community, but the associated data is then made public. Thus, this fundamental generational perception of these digital forms of language has also been highlighted through qualitative research.

The majority of participants see the use of emojis and internet slang as flexible and liberating, but most of them expressed concerns about possible risks to traditional languages. This is related to the tension that Barons (2008) highlights in his observation between the positive effects of digital communication and its harmful effects on languages.

Conclusion

This study deeply examined how young people's minds are saturated with emojis rather than handsome quotes and expressions and how it created a new trend in digital language on social media that completely replaced the discipline, rules and regulations of effective communication. A study, published in Social Psychological and Personality Science, advised that people should avoid emoji at work altogether, especially the first time you talk to someone. Overall, the results of our qualitative study show clearly that these digital forms of language put a gap between old and new and replace scripts with symbols even a few argue that these digital forms of language represent an essential tool for improving communication as they open up new channels of expression and understanding in an increasingly globalized and interconnected world. With a pure traditional script, you will be able to express highly specific ideas and emotions succinctly, thus creating a sense of community and belonging.

Limitations of the Study

However, there are certain limitations to the generalization of these results to the group of young adults aged 18 to 30 years. First, studying young adults ages 18 to 30 may not provide enough information to account for other generations of social media users and may lack the subtlety to explain generational differences in digital communication. Second, relying on self-assessments from interviews could also introduce bias, as social media users who participate in interviews would be more likely to emphasize the positive aspects of social media use. Finally, the qualitative nature of the study, although detailed, presents a problem in generalizing the results to a broader population without adding additional quantitative data. Future researchers will include studies with a broader age range of participants and may adopt a mixed research methodology to generalize or triangulate these results.

References:

1. Androutsopoulos, J. (2020). Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism*, 24(2), 149-164.
2. Baron, N. S. (2008). *Always On: Language in an Online and Mobile World*. Oxford University Press.
3. Baron, N. S., & Squire, L. (2010). Instant messaging and the future of language. *Communications of the ACM*, 53(7), 41-43.
4. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
5. Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage publications.
6. Crystal, D. (2011). *Internet Linguistics: A Student Guide*. Routledge.
7. Danet, B. (2001). Cyberpl@y: Communicating Online. *Berg Publishers*.
8. Doe, J. (2019). *Lexical Innovation in Online Communities*. Journal of Digital Linguistics, 4(1), 34-56.
9. Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford University Press.
10. Ezzy, Douglas. *Qualitative analysis*. Routledge, 2013.
11. Jones, A., & Smith, B. (2020). *Emojis as a Parallel Visual Language: An Analysis of Their Linguistic Potential*. Visual Communication Quarterly, 27(2), 103-117.
12. McCulloch, G. (2019). *Because Internet: Understanding the New Rules of Language*. Riverhead Books.
13. McCulloch, G. (2019). *Because Internet: Understanding the New Rules of Language*. Riverhead Books.
14. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, *11*(19), 9117.
15. Abdurakhmonova, N., Alisher, I., & Sayfulleyeva, R. (2022, September). MorphUz: Morphological Analyzer for the Uzbek Language. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 61-66). IEEE.
16. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
17. Stark, L., & Crawford, K. (2015). The Conservatism of Emoji: Work, Affect, and Communication. *Social media + Society*, 1(2).
18. Tagliamonte, S. A., & Denis, D. (2008). Linguistic ruin? LOL! Instant messaging and teen language. *American Speech*, 83(1), 3-34.
19. Thurlow, C., & Brown, A. (2003). Generation Txt? The sociolinguistics of young people's text-messaging. *Discourse Analysis Online*, 1(1), 30.

O‘ZBEK BOLALAR NUTQINING KORPUS ASOSIDA PSIXOLINGVISTIK TADQIQI

Abduraxmonova Sohiba Aktam qizi
Mirzo Ulug‘bek nomidagi O‘zMU
Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi o‘qituvchisi.
E-mail: abduraxmonovasohiba16@gmail.com

Annotatsiya: Ushbu maqola o‘zbek bolalar nutqining o‘rganilishi va uning ilmiy, amaliy ahamiyatini qamrab oladi.Maqolada korpus asosida bolalar nutqini o‘rganish, psixolingvistik tahlil qilishga alohida to‘xtalib o‘zbek bolalar nutqining antroposentrik jihatdan o‘rganilishi tahlil qilingan.

Kalit so‘zlar:korpus lingvistikasi, eksperiment, psixolingvistika, nutq ontogenizi.

Аннотация: В статье рассмотрено изучение речи узбекских детей и ее научно-практическое значение. В статье анализируется изучение детской речи на основе корпусного, психолингвистического анализа и антропоцентрического изучения речи узбекских детей.

Ключевые слова: корпусная лингвистика, эксперимент, психолингвистика, онтогенез речи.

Abstract: This article covers the study of Uzbek children’s speech and its scientific and practical significance. The article analyzes the study of children’s speech based on the corpus, psycholinguistic analysis, and the anthropocentric study of Uzbek children’s speech.

Keywords: corpus linguistics, experiment, psycholinguistics, speech ontogeny.

Kirish. XX asrning o‘rtalariga kelib psixolingvistikaning mustaqil ilmiy fan sifatida shakllanishi natijasida nutq ontogenizini o‘rganishga talabni kuchaytirdi. Buning natijasida G‘arbda bola nutqi rivojini kuzatgan holda turli ilmiy tadqiqotlar yuzaga keldi. Xorijiy manbalardan tajriba sifatida foydalangan holda o‘zbek tilshunosligida O.Umarxo‘jayeva[38], N.Sayidiraximova[39], G.Iskandarova[40], M.Qurbonova[41] va O.Saidaxmedovalar[42] tomonidan bolalar nutqini qator tadqiqotlarda kuzatilgan va o‘zbek tilining birinchi til sifatida fonetik, grammatik xususiyatlari lingvistik, sotsiolingvistik, pragmalingvistik va psixolingvistik jihatdan tadqiq etilgan. O‘zbek bolalar nutqida kuzatiladigan ayrim fonetik o‘zgarishlar va grammatik xususiyatlar, shuningdek, bolalar nutqining sotsiopsixolingvistik talqiniga bag‘ishlangan.

[38]Умархўжаева О. Болалар нутқининг баъзи фонетик хусусиятлари. Ўзбек тили ва адабиѐти. – Тошкент, 1997. – № 5. – Б. 52-53; Искандарова Г. Болалар мулоқот хулқининг гендер ва лингвокультурологик хусусиятлари ЎзМУ хабарлари. – Тошкент: Университет, 2015. – № 1 2. – Б. 200-203.
[39]Сайидираҳимова Н. Мактабгача ѐшдаги ўзбек болалар нутқининг лингвистик хусусиятлари: Филол. фан. номз. …дисс. – Тошкент, 2004. – 126 б.; Қурбонова М. Ўзбек болалар нутқи лексикасининг социопсихолингвистик тадқиқи: Филол. фан. номз. … дисс. – Тошкент, 2009. – 154 б.
[40]Искандарова Г.Болалар нуткига оид баъзи мулохазалар//Ўзбек филологиясига оид тадқиқотлар:Илмий-амалий анжуманлар материаллари – Тошкент:ЎзМЎ,2009–Б.200-203.
[41] Қурбонова М,Ўзбек болалар нутқининг прагматик хусусиятлари:Филол.фан.ф-ри….дисс.-Тошкент,2018.-238 б
[42] Saidaxmedova O. Tabiiy tartib gepotezaasining o‘zbek tili uchun talqini. Toshkent – 2021

O‘zbek tilshunosligida bolalar muloqotida nutqiy akt turlarining voqealanishi, deyksis va presuppozitsiyaning leksik, morfologik va sintaktik vositalar orqali ifodalanishi, konnotatsiyaning yuzga kelishi masalalari M.Qurbonovaning doktorlik dissertatsiyasida tadqiq etildi.[43] Bu ishda bolalar nutqi o‘ziga xos fonetik, leksik, grammatik xususiyatlarga ega bo‘lib pragmatik omillarga tayangan[Abdurakhmonova, N.2021,2022].

Ontolingvistikaga doir tadqiqotlarda bolalar nutqini bevosita kuzatish hamda ta’lim-tarbiya muassasalarida o‘tkazilgan maxsus tajribalarda to‘plangan daliliy materiallar tahlilga tortilgan.Ularda bolalar tomonidan ona tilining fonetik tizimi, grammatik imkoniyatlari, shuningdek, ikkinchi tilning o‘zlashtirilishi jarayonidagi muammoli jihatlar, bolalar nutqida kuzatilagan o‘ziga xos leksik-semantik birliklar lingvistik, linvokulturologik hamda genderologik aspektlarga o‘rganilgan[44]. O‘zbek tilshunosligida bolalarda qo‘shimchalar o‘zlashtirilish tartibini aniqlash maqsadida O.Saidaxmedova tadqiqot olib borgan. Olima tadqiqotida til o‘zlashtirish jarayonini tavsiflab, qo‘shimchalar o‘zlashtirish tartibiga ta’sir qiluvchi xususiyatlarni aniqlagan. Tadqiqot ikki turdagi tajriba kuzatuv asosida o‘tkazilgan. Tajriba-kuzatuvga jismoniy va ruhiy sog‘lom hamda nutqi me’yorda rivojlanayotgan bolalar tanlab olingan.

Boshlang‘ich sinf o‘quvchilari uchun tuzilgan “So‘z sandiqchasi: O‘zbek tilining izohli lug‘atchasi” ni bolalar lisoniy zaxirasi va nutqini o‘stirishga xizmat qiluvchi manba sifatida baholash mumkin[45].

Tilshunosligimizda maktabgacha yoshdagi bolalarning muloqotga kirishish xususiyatlari o‘rganilgan bo‘lib, ushbu mavzuga oid amaliy, nazariy xulosalar G. Iskandarovaning qator maqolalarida o‘z aksini topgan. G.Iskandarova bolalarning 6 oydan 4 yoshgacha bo‘lgan davrini lisoniy jihatdan tahlil qilgan[46].

90-yillarning oxiridan boshlab o‘zbek tilshunosligida L.Vigotskiy, A.Leontyev, A.Soxin, M.Shaxnarovich, I.Negnevitskaya va boshqa olimlar tomonidan yaratilgan qoidalarga tayangan holda M.Qodirovaning ilmiy rahbarligi ostida ona tilida va shu bilan birga chet tillardagi nutq ontogenezi qirralarini tadqiq qilish ishlari olib borilmoqda. Xususan, D.Babayeva tomonidan maktab yoshidagi katta bolalarni atrof-olam bilan tanishtirish jarayonida ularda nutqni (ona tili-o‘zbek tilida) o‘stirish, N.Nurmuhammedova tomonidan ikki tillilik sharoitida bilingval bolalarda ravon nutqni (rus tilida) rivojlantirish, Sh.Sharipova tomonidan maktabgacha yoshdagi katta bolalarda vaziyatli o‘yinlar asosida nutqiy muloqot (rus tilida) madaniyati ko‘nikmalarini tarbiyalash masalalari bo‘yicha tadqiqotlar o‘tkazilgan[47].

Bundan tashqari, L.Mirjalolova, R.Yo‘ldoshev, F.Umarova, G.Muhammedjanova, D.Fattaxova, G.Asilova kabi tadqiqotchilar monografiya va dissertatsiyalarida ta’lim tizimining turli bosqichlarida o‘zga tilli guruhlarda o‘zbek tilini o‘qitish masalasi turli aspektlarda o‘rganilgan[Abdurakhmonova, N.2021,2022].

Jahon tilshunosligida XXI asrning boshlariga kelib muayyan tilni o‘rganilish doirasini kengaytirish, til imkoniyatlarini namoyon etishning asosiy vositasi internet tizimida til

[43] Қурбонова М.Ўзбек болалар нутқининг прагматик хусусиятлари:Филол.фан.ф-ри….дисс.-Тошкент,2018.-238 б.
[44]Умархўжаева О. Болалар нутқининг баъзи фонетик хусусиятлари Ўзбек тили ва адабиѐти. – Тошкент, 1997. – № 5. – Б. 52-53; Искандарова Г. Болалар мулоқот хулқининг гендер валингвокультурологик хусусиятлари.ЎзМУ хабарлари. – Тошкент: Университет, 2015. – № 1 2. – Б. 200-203.
[45] Mahmudov N. So‘z sandiqchasi: O‘zbek tilining izohli lug‘atchasi. – Toshkent: G‘afur G‘ulom, 2016. – 216 b.
[46] Saidaxmedova O. Tabiiy tartib gepotezaasining o‘zbek tili uchun talqini. Toshkent – 2021.
[47] Saidaxmedova O. Tabiiy tartib gepotezaasining o‘zbek tili uchun talqini. Toshkent – 2021.

korpuslarini o‘rganishga alohida e’tibor qaratildi. Bugungi kunda XX asrning buyuk kashfiyoti bo‘lgan kompyuter texnologiyalari boshqa sohalar kabi tilshunoslik fani uchun ham keng imkoniyatlar eshigini ochishi hamda uning zimmasiga kompyuter tili bilan bog‘liq ulkan vazifalarni yuklashi, kompyuter lingvistikasi yo‘nalishining paydo bo‘lishi tabiiy tillarning taraqqiyoti uchun muhim ahamiyat kasb etadi. Kompyuter texnologiyalari vositasida axborot-resurslarni joriy etish uchun zarur dasturiy hamda metodik ta’minotni yaratish, internet tizimida til korpusini va shu asosda milliy til korpusining ilmiy-nazariy jihatlarini ishlab chiqishga alohida diqqat qaratildi.

XX asrning 90-yillari birinchi yarmida korpus tilshunosligi til to‘g‘risidagi fanning alohida qismi sifatida shakllandi. Natijada labaratoriya bilan bog‘liq fan psixolingvistikada ham korpus kirib keldi. Korpus lingvistikasi va psixolingvistika bir-biridan farqli maqsad va metodologiyalarga ega. Korpus texnologiyasi asosan ma’lum nutq jamoasi yoki matn turi bo‘lgan til namunalarida qo‘llanish modellarini aniqlash bilan shug‘ullanadi. U tayyor manbalar bilan ish olib boradi. Psixolingvistika aksincha, tilni tushunish va ishlab chiqarish bilan bog‘liq aqliy jarayonlar bilan shug‘ullanadi. U diqqat bilan nazorat qilinadigan labaratoriya tajribalari orqali nazariy modellarni sinab ko‘rishga qaratilgan. Ushbu turli maqsad va yondoshuvlarni hisobga olgan holda, ikki sohani birlashtirishga urunishlar munozarali bo‘lib kelmoqda. Bundan tashqari, psixolingvistika, odamlarni nutqining o‘ziga xos xususiyatlariga ko‘ra tanib olish, mashina tarjimasi, ma’lumotni kompyuterga og‘zaki kiritish muammolarini hal qilish uchun kerak va shunga mos ravishda bu fan informatika bilan chambarchas bog‘liq. Aynan shu amaliy muammolar psixolingvistikaning korpus bilan bevosita aloqasiga turtki bo‘ldi[Abdurakhmonova, N.2021,2022].

Namuna olingan korpuslar bilan bugungi kunga qadar kam ish qilingan bo‘lsada ularning imkoniyatlari yetarlichadir.Til nuqsonlari bo‘lgan odamlarning tilini va normal til qobiliyatiga ega bolalarning tilini o‘rganishda mavjud korpuslar kam. Biroq so‘nggi o‘n yillikda ushbu sohada o‘sish bo‘ldi. Misol uchun POW (Polytechnic of Wales) korpusi bolalar tili korpusidir. Shuningdek, CHELDES (Child Language Data Exchange System). 1984-yilda Brian MacWhinney va Catherine Snow tomonidan birinchi til ma’lumotlari uchun markaziy ombor sifatida xizmat qilish uchun tashkil etilgan korpusdir. CHILDES asosan yosh bolalarning tilini tahlil qilish uchun ishlatiladi va kattalarning bolalar nutqiga qaratilgan.Yana shunday korpuslardan biri Germaniyaning AIBO korpusi bolalar korpusining yangi turini yaratish maqsadida tuzilgan. Bolalardagi tabiiy va his tuyg‘ularga boy nutqni farqlash maqsadida o‘tkazilgan tadqiqot ikki maktabdan 51 nafar bola (10-13 yosh, 21 erkak 31 ayol) dan foydalanilgan. Nutq tabiiy bo‘lishi uchun xonada faqat nazoratchi va bola bo‘lib bolani AIBO bilan gaplashishga undashgan. Nutq simsiz eshitish vositasi UT 14/20 TPSHURE UHF-seriyasi WH20TQG yordamida yozib olingan. Har bir yozib olish jarayoni 30 daqiqani tashkil etgan. Lug‘at hajmi 1200 so‘zdan iborat bo‘lib 13791ta gapdan iborat[48]. Bu korpus bolalar nutqidagi tabiiy jarayonlarni kuzatish hamda ularning his tuyg‘ulari nutqda qanday ifoda qilinganini aniqlashni osonlashtirish maqsadida tuzilgan.

“Amerika meros korpusi” (“The American Heritage Intermediate Corpus”) tuzuvchilari tomonidan qo‘llangan. 5 million so‘zshakl hamda ingliz tilida 22 turdagi bolalar, o‘smirlar janriga oid matnlar kiritilgan. AQShning 221ta maktabiga qanday matnni tanlash maqsadga muvofiqligini

[48]The PF STAR Children’s Speech Corpus.Anton Batliner1, Mats Blomberg2, Shona D’Arcy3, Daniel Elenius2,Diego Giuliani4, Matteo Gerosa4, Christian Hacker1, Martin Russell3, Stefan Steidl1, Michael Wong3.

aniqlovchi so‘rovnoma yuborilgan. So‘rovnoma natijasi o‘rganilgach, 19 000 nomdagi kitoblar ro‘yxati tuzilgan. Bu asosda har biri 500 so‘z shakldan iborat 1045 matn tanlab olingan[49].

O‘zbek tili korpusi o‘zbek tilshunosligida qilingan ilk amaliy ish bo‘lib, uning sohani rivojlanishida ahamiyati va o‘rni beqiyosdir. O‘zbek tilining elektron korpusi bir necha yillik tinimsiz olib borilgan amaliy tadqiqotlar, ilmiy hamkorlik va loyihalardan olingan natijalarning mahsuli hisoblanadi. Respublikaning turli oliy o‘quv yurtlarining yigirmadan ortiq o‘qituvchi, magistrant va talabalardan iborat loyiha a’zolari yordamida Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti dotsenti, f.f.doktori (DSc) N.Z.Abduraxmonova tomonidan o‘zbek tili korpusi yaratildi.Ushbu korpus tarkibida hosil bo‘lgan o‘zbek bolalar nutqi xorijiy davlatlar bolalar korpusidan namuna olingan holda yaratildi.

Foydalanilgan adabiyotlar:

1. Azimova I. O‘zbek tilidagi gazeta matnlari mazmuniy persepsiyasining psixolingvistik tadqiqi.monografiya,T.:2019
2. Леоньтев А. Основы психолингвистики. – М., 1997.
3. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
4. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
5. Iskandarova, A. (2018). Vatanni olib ketib bo’lmaydi.
6. Iskandarova, A. (2009). A.Navoiyning "Lisonut-tayr" dostonidagi vodiylar ta’rifi va mohiyati haqida. Мумтоз суз таьлими.
7. Iskandarova, A. (2019). INSONIYATNI SAODATGA YETAKLAGAN ADIB (Adabiyot darslarida Chingiz Aytmatov ijodini o’rganishga doir). Tafakkur Chashmasi MCHJ.
8. Сафонова О. Психолингвистика: новые технологии анализа поетического текста. Учебное пособие. – Краснодар, 2012.
9. Усманова Ш. “Психолингвистика” фанидан маъруза курслари. –Тошкент: Университет, 2014.
10. Abdunazarova M. “Gender steriotiplari tahlili” International scientific-online conference: “Principal issues of scientific research and modern education”188-196-betlar
11. Захаров В., Богданова С. Корпусная лингвистика: учебник. 3-е изд., перераб. − СПб.: Изд-во С.-Петерб. ун-та, 2020. − 234 с.

[49]Nursulton Zamon o‘g‘li Shayxislamov.Til korpusi yaratishning umumiy tamoyillari xususida –Toshkent viloyati Chirchiq davlat pedagogika instituti.

SEMANTOGRAFIK TARJIMA YOZUVINING LEKSIK, GRAMMATIK VA STILISTIK XUSUSIYATLARI

Umirov Ixtiyor Ergashevich
Filologiya fanlari nomzodi, dotsent
O‘zbekiston Davlat Jahon tillari universiteti,
Umumiy tilshunoslik kafedrasi dotsenti.
E-mail: ixtiyorumirov0@gmail.com

Annotatsiya: Ma’lumki, madaniyatlararo muloqot masalalariga bo‘lgan qiziqish hamda ketma-ket tarjima faoliyatiga bo‘lgan katta e’tibor tarjimonlik sohasida ham jiddiy islohatlarni bo‘lishligini talab etmoqda. Shundan kelib chiqib, "tarjima san’ati"ni va "tarjimaga o‘qitish jarayoni"ni ham qayta ko‘rib chiqishlik zarurdir. Bular o‘z-o‘zidan o‘rganuvchidan professionallik imkoniyatini yanada oshirishni hamda tarjima muammolarini optimallashtirishga imkon beradigan tarjima mahoratini shakllanishiga yo‘l ochadi.

Bilamizki, ketma-ket tarjima qilish sohasidagi tarjimonning tildan foydalanish qobiliyati uning bevosita shu sohani yaxshi o‘zlashtirganligi haqidagi malakasining tarkibiy qismlaridan biri hisoblanadi. Biroq, undagi bu malaka sohaning professonal egasi degan ma’noni bermaydi, chunki bugun ushbu sohaga bo‘lgan talab kundan-kunga o‘zgarib, rivojlanish bosqichiga yuz tutmoqda. Ushbu maqolada amaliy tilshunoslik fanining muhim sohalaridan biri bo‘lgan tarjimashunoslik sohasining yangi yutuqlari haqida, aniqrog‘i, ushbu faoliyatning ketma-ket tarjima jarayonida qo‘llaniladigan semantografik yozuvi haqida gap boradi. Semontagrafiya o‘ziga xos tarjima yozuvi bo‘lib, u tarjimashunoslik fanining istiqboliga xizmat qiladi. Bu individual tarjima yozuvi tarjimon uchun qulay imkoniyatlarni beradi. Muhimi, ushbu tarjima yozuvi o‘iga xos imkoniyatlarga va ba’zi xususiyatlarga ega bo‘lishi bilan yozuvning boshqa maxsus turlaridan farqlanadi. Xullas, maqola mohiyati tarjimashunoslik sohasining ketma-ket tarjima faoliyatini yanada rivojlantirishga, aniqrog‘i, aynan shu soha vakillari uchun tavsiya etilgan tarjima yozuvining o‘ziga xos imkoniyatlarini, ya’ni leksik, Grammatik va stilistik xususiyatlarini ochib berishga qaratilgan. Bundan tashqari, maqolamizda yangi yozuvni o‘rganish uchun bir qancha amaliy tavsiyalar berishni ham ko‘zda tutganmiz.

Demak, bu tarjima faoliyatidagi tezkor yozuv har qanday uzunlikdagi nutqni yozib olish va uning mazmunni yo‘qotmasdan takrorlashga imkon beruvchi maxsus yozuvga ega bo‘lishdir.

Kalit so‘zlar: Semontagrafiya, semontagramma, markirofka, IT, OAV, TJ, kompozitsion, affiks, segment, indeks, kodlashtirish, modillashtirish, fiksatsiya, sinxronlashtirish.

Abstract: It is known that the interest in the issues of intercultural communication and the great attention to the work of consecutive translation require serious reforms in the field of translation. Based on this, it is necessary to reconsider the "art of translation" and the "process of teaching translation". These allow the self-learner to further develop their professional skills and develop translation skills that allow them to optimize translation problems.

We know that a translator's ability to use language in the field of consecutive translation is one of the components of his competence that he has mastered this field. However, this qualification in him does not mean that he is a professional owner of the field, because today the demand for this field is changing day by day and is entering the stage of development. This article talks about new achievements in the field of translation studies, which is one of the important fields of applied linguistics, and more precisely, about the semantic recording of this activity, which is

used in the process of consecutive translation. Semontography is a kind of translation writing, which serves the perspective of the science of translation studies. This individual translation record provides convenient options for the translator. Importantly, this type of translation record has its own capabilities and some features that distinguish it from other special types of record. So, the essence of the article is aimed at further development of consecutive translation activities in the field of translation studies, more precisely, at revealing the specific possibilities of the translation writing recommended for the representatives of this field, i.e. lexical, grammatical and stylistic features. In addition, in our article we intend to give some practical recommendations for learning a new script.

Hence, a quick note in this translation activity is to record a speech of any length and have a special note that allows you to repeat it without losing the content.

Keywords: Semontography, semontagram, marking, IT, media, TJ, composition, affix, segment, index, coding, modeling, fixation, synchronization.

Аннотация: Известно, что интерес к вопросам межкультурной коммуникации и большое внимание к работе последовательного перевода требуют серьезных реформ в сфере перевода. Исходя из этого, необходимо переосмыслить «переводческое искусство» и «процесс обучения переводу». Это позволяет самообучающимся дополнительно развивать свои профессиональные навыки и развивать навыки перевода, которые позволяют им оптимизировать проблемы перевода.

Мы знаем, что умение переводчика использовать язык в области последовательного перевода является одной из составляющих его компетентности, которой он владеет в этой области. Однако данная квалификация у него не означает, что он является профессиональным владельцем месторождения, поскольку сегодня спрос на это месторождение меняется день ото дня и вступает в стадию развития. В данной статье говорится о новых достижениях в области переводоведения, которая является одной из важных областей прикладной лингвистики, а точнее, о смысловой фиксации этой деятельности, которая используется в процессе последовательного перевода. Семонтография – это разновидность переводческого письма, которая служит перспективе науки о переводе. Эта индивидуальная запись перевода предоставляет переводчику удобные возможности. Важно отметить, что этот тип записи перевода имеет свои возможности и некоторые особенности, отличающие его от других специальных типов записей. Итак, суть статьи направлена на дальнейшее развитие последовательной переводческой деятельности в области переводоведения, точнее, на выявление конкретных возможностей переводческого письма, рекомендуемых представителям этой области, т.е. лексических, грамматических и стилистических особенностей. . Кроме того, в нашей статье мы намерены дать несколько практических рекомендаций по изучению нового сценария.

Итак, небольшая заметка в этом переводческом занятии — записать речь любой длины и иметь специальную заметку, которая позволит вам повторить ее без потери содержания.

Ключевые слова: Семонтография, семонтаграмма, маркировка, ИТ, носитель, ТЯ, композиция, аффикс, сегмент, индекс, кодирование, моделирование, фиксация, синхронизация.

Yangi tarjima yozuvining shakllantirilishi amaliy tilshunoslik sohasining muhim sohalaridan biri hisoblanayotgan tarjimashunoslik fanining ichki rivojlanishidir. Bu fanning rivojlanishini shu sohaga tegishli bo‘lgan barcha imkoniyat va ko‘nikmalarning shakllantirishi

darajasi bilan belgilash lozim. Ko‘nikma va imkoniyatlar - bu muayyan bir sohaning rivojlanishi uchun yuzaga kelgan operatsiyalar va harakatlardir.[6]

Semantografiya, ya’ni tezkor tarjima yozuvi ham aynan shu sohada, aniqrog‘i, yillar davomida to‘plangan tajriba asosida vujudga kelgan sa’yi harakatlar va ko‘nikmalarning mahsuli bo‘lib, u tatjimashunoslik faoliyatining muhim sohalaridan biri bo‘lgan ketma-ket tarjima jaroyonida qo‘l keladigan yangi bir yozuv turidir.

Ilk bor “semantografiya” atamasiga (yunoncha semanto “belgi” va grafo – “xat” dan) Charlz Blisc asos solgan. U universal ramzlar tizimini, ya’ni o‘zining yangi yozuvini yaratgan.[2] Bundan tashqari, tarjima yozuvi bilan bog‘liq dastlabki nazariy qoidalarni Jenevadagi mashhur tarjimonlar maktabi vakili J. Erberning asarlarida ham ko‘rish mumkin. [3]

Ushbu semontagrafiya jarayonida ma’lumotni inson omilining oqilona tuzatish va uzatish qobiliyatini, aniqrog‘I, og‘zaki tarjima matnini o‘ziga xos yaratish jarayonini, shuningdek, uning tildan matematik informasion kodlashtirish qobiliyatini baholash mumkin bo‘ladi.[5]

Tarjimon tezkor tarjimaning har bir bosqichida qanday operatsiyalar yoki harakatlar bajarishi mumkinligiga qarab, ko‘nikmalarning uchta guruhini va tezkor yozuv ko‘nikmalarining ikkita guruhini ajratamiz. Xullas, semantografiya yozuvining taxminiy yoki dastlabki semantogrammalarini quyidagi texnik dizayin va ko‘nikmalar asosida tushuntirishga harakat qilamiz:

1. Lisoniy qoliplar va ko‘nikmalar.

a) leksik ko‘nikmalar:

-so‘zlar va so‘z bloklarining qisqartirilgan yozuvini shakllantirish (matematik modellashtirish);

-aniq ma'lumotlarning raqamli belgilanishi;

- individ qisqartirishlar;

-tushunchalarni ramziylashtirish;

b) grammatik ko‘nikmalar:

-turli xil sintaktik va mantiqiy tuzilmalarning markirovkalari (modallik, predikativlik, rad etish, kuchaytirish, taqqoslash va boshqalar) ni shakllantirish;

-so‘z qoliplarini modellashtirish.

v) tarkibiy va kompozitsion ko‘nikmalar:

-yozuvlarni vertikal ravishda tartibga solish;

-vizual va iqtisodiy yozuvga asos solish.

2. Turli xil faoliyatlarni sinxronlashtirish bilan bog‘liq tezkor yozuv qobiliyatlari:

-bir vaqtning o‘zida faol tinglash, yozish va ifodalash;

-nutqning sur'atidan qat'i nazar, yuqori tezlikdagi yozuvni shakllantirish;

-nutq vaqtini ko‘paytirish bilan fiksatsiya miqdorini kamaytirish;

-nutq tovushi tugashi bilan fiksatsiyaning ham tugashi;

-tarjima shartlariga qarab, o‘tirib yoki tik turgan holatda o‘zlashtirish.

3. Omaviy axborot vositalarini (OAV) semantik tahlil qilish bilan bog‘liq tezkor yozuv ko‘nikmalari.

a) leksik ko‘nikmalar:

- kalit so‘zlarni, tiniq fikrlarni, ya’ni semantik yadro deb topilgan gaplarni ajratib ko‘rsatish va belgilash;

b) tarkibiy va kompozitsion ko‘nikmalar:

-nutqning mavzulariga va qofiyalariga urg‘u berish;

-nutqning tugallangan segmentining mantiqiy va semantik tuzilishini ta'kidlash;

-asosiy, aniqlashtiruvchi ma'lumotni ta'kidlash;

-qisqartirish, o‘zgartirish, umumlashtirish.

Tezkor yozuvning umumlashtirish masalasini quyidagicha tasniflash mumkin bo‘ladi:

1. Yaratilgan matnni rejalashtirish bilan bog‘liq tezkor yozuv ko‘nikmalari:

-axborotni tasniflash, umumlashtirish, ierarxizatsiya qilish;

-tayyor mahsulotning ehtimoliy prognozi;

-matn ulanishini yaratish;

-matn yaxlitligini yaratish;

2. Texnik jarayonni(TJ) dasturlash bilan bog‘liq tezkor yozuv ko‘nikmalari:

-IT (axborot texnologiya) invariantini ta'kidlash va qayd etish;

-kelajakdagi TJ imidjiga muvofiq IT-transformatsiyasini shakllantirish;

-murakkab semantik konstruktsiyalarni almashtiradigan, ya’ni siqiq holatga keltirilgan tezkor yozuv formulalarini favqulodda topish;

-IT ning semantik ortiqcha sonini olib tashlash;

-mazmunli IT ma'lumotlarning rivojlanishini va uni tezkor yozuv matnida aks ettirishni prognozlash;

-tanlangan mavzularni va yozuvlarni yozish jarayonida kengaytirish va ketma-ket joylashtirish.

Endi, yuqorida aytilgan va qabul qilingan semantagrammalarni quyidagi namunalar va jadval asosida ko‘rib chiqishimiz mumkin:

a) so‘z o‘rtasidagi unli tovushlar tushuriladi: qobiliyat - qblyt, raqobat - rqbt, lekin - lkn, strategiya – strtgy;

b) qo‘sh undoshlarni soddalashtirish usuli bilan: programma - prgma;

c) teleskopik birikmalardan foydalanish usuli bilan: mumkin- mm, kerak- kk, informatsiya- info

d) so‘z qismlarini (affikslarning chastotali qo‘shimchalari va boshqalar) indeks belgisi bilan almashtirish usuli bilan: Buni quyidagi jadval asosida ko‘rib chiqamiz:

Belgi	So‘z qismi		
	Ingliz tili	O‘zbek tili	Rus tili
θ	-(t)ion	-siya	-ция
λ	-logy	-logiya	-логия
φ	philo-	filo-	фило-
ψ	psycho-	psixo-	психо-

Masalan, so‘zning o‘zagini almashtirish: *philology – φλ;* qo‘shimchalarni o‘zgartirish : *evolutsiya – evl*$^{\theta}$.

Yana tezkor yozuvda so‘zni qisqartirish usuli ham eng qulay imkoniyat hisoblanadi. Qisqartirish bu bosh harflar yoki bo‘g‘inlarni yozib olish orqali so‘z yoki iboraning grafik qisqartmasidir. Qisqartma atamalarni, tashkilotlarning nomlarini, mamlakatlarni, millatlarni, tillarni va boshqa tushunchalarni belgilash uchun ishlatiladi. Masalan, BMT, NATO tarjima qilish yo‘li bilan uzatiladi, MDH, BAR kabi.

Raqamli belgilar qisqartmas

Individ semontagrammalarga asoslangan raqamlar quyidagi aniq ma'lumotlarni ko‘rsatadi (miqdor sonlar):

15 yoki 15 ' (o‘n besh ming);

15 yoki 15 " (o‘n besh million);

15 yoki 15 " ' (o‘n besh milliard);

Tartib raqamlar, ya'ni sanalar:

31.XII.24 - ikki ming yigirma to‘rt oyning o‘ttiz birinchi kuni;

yilning davri:

'24 ikki ming yigirma to‘rtning boshi;

2'4 ikki ming yigirma to‘rtning o‘rtasi;

24 ' ikki ming yigirma to‘rtning oxiri;

yil fasllari:

(1 qish, (2 bahor, (3 yoz, (4 kuz);

chorak:

I\4 (birinchi chorak); II\4 (ikkinchi chorak); III\4 (uchinchi chorak); IV/4 (to‘rtinchi chorak).

oy nomlari:

I<24 (yanvar), II<24 (fevral), III<24 (mart), IV<24 (aprel), V<24 (may), VI<24 (iyun), VII<24 (iyul), VIII<24 (avgust), IX<24 (sentyabr), X<24 (oktyabr), XI<24 (noyabr), XII<24 (dekabr).

hafta kunlari nomlari:

Dushanba (6/1), seshanba (6/2), chorshanba (6-3), payshanba (6/4), juma (6/5), shanba (6/6), yakshanba (6/7).

Tezkor tarjima yozuvi faoliyatida so‘z turkumlarini grammatik jihatdan texnikalash imkoniyati quyidagi beriladigan qator uslublar bilan (inkor etish, taqqoslash, vaqt munosabatlari, modallik, jins, raqam) turli grammatik hodisalarni belgilash mumkin bo‘ladi. Masalan, **Ot va sifat so‘z turkumlariga doir so‘zlarning sonini yoki jinsini belgilash bilan bog‘liq tavsiyalar:**

Eslatma: Ismlarning ayol jinsi **e** indeksi bilan belgilanadi.

O‘zb – o‘zbek erkagi, RU^{e} – o‘zbek ayoli

Otlarning ko‘pligi **2-indeks** bilan belgilanadi.

□^{2} – *davlatlar* □ - *davlat*

Tegishlilikni belgilash

Eslatma: Gaplarda tegishlilik ‘ belgisi bilan belgilanadi

□ S’ – davlatning maydoni

Sifatlarning darajasi quyidagicha taqqoslanishi mumkin. Sifatlarning darajasini ifodalashda ∧ kichik va ∨ katta belgilari qo‘llaniladi. Ularning o‘rta darajasiga 2 va yuqori darajasiga 3 soni qo‘shiladi.

 – *qimmat*

 – *arzon*

 ^{2} – *qimmatroq*

 ^{2} – *arzonroq*

 ^{3} – *juda qimmat*

 ^{3} – *juda arzon*

Yoki fe’llarda shaxsini aniqlash masalasini ko‘rib chiqishimiz mumkin.

Fe’llar nechanchi shaxsda turganligini ✓ belgisi bilan belgilaymiz

men ✓ 1 biz ✓ 4

sen [2] siz [5]

u [3] ular [6]

[4] – *biz eksport qilyapmiz*

Fe'llardagi ish-harakatning vaqtini ifodalash uchu**n ed** indeksi (ing. O'tgan zamon) ni yoki ? **ni** qo'llash mumkin.

I"[ed] yoki I" – *men gapirdim*

Kelasi zamon **ll** indeksi (ing. Kelasi zamon will) belgisi bilan ifodalanadi. .

I"[ll] yoki I' – *men aytaman*

Fe'llarning passiv formasini **"bo'lmoq'** = va **"ed"** (ing. Sifatdoshning tugashi) indeksi bilan belgilash mumkin.

= %[ed] – *qiziqqan*

= [[ed] –*ochmoq*

Fe'lardagi modallikni quyidagicha ifodalash mumkin."Bo'lishi mumkin, ehtimol, nazarimda" ma'nosini ifodalasa fe'lning yuqorisiga **"p"** (lot. Posse) indeksi qo'yiladi.

[p] – *Ehtimol,bu partiya yutadi*

"Kerak, darkor, xohlamoq" ma'nolarini ifodalashda **"d"** (lot. Debere) indeksi qo'llaniladi.

[d] [4] – *Biz bosim o'tkazishimiz kerak*

Noaniqlik `belgisi bilan ifodalanadi.

['e – Ko'rgazma ochiladimi yoki yo'qmi, noaniq.

Xohish, istak **"v"** (lot. Volere) indeksi bilan ifodalanadi.

[v4] – *Biz kontrakt tuzishni istar edik*

Yoki inkorni ifodalash uchun belgining ustiga chiziq tortish kifoya.

OK – rozilik, ~~OK~~ – norozilik

I OK – men roziman, I ~~OK~~ – men noroziman

Ko'rinadiki, yozuvning bu maxsus turi amaldagi yozuvlardan induvidualligi bilan, ijtimoiyligi bilan, zamonaviyligi bilan keskin farqlanadi. Ayniqsa, semontagrafik yozuvning eng muhim tomoni—uning soddaligi va induvid alifboga asoslanganligidir.

Foydalanilgan adabiyotlar:

1. Аликина Е.В. Переводческая семантография. Запись при устном переводе [Текст]: практ. учеб. пособие: рек. УМО / Е. В. Аликина. - М.: АСТ: Восток - Запад, 2006. - 159 с2
2. Bliss C.K. Semantography. – Sydney, 1956.
3. Herbert J. Manuel de l'interprète. – Genève, 1952
4. Сачава О.С. Методические рекомендации по переводческой скорописи. – СПб.: Изд-во СПбГУЭФ, 2010. – 35 с.
5. Ширяев А.Ф. Синхронный перевод: деятельность синхронного переводчика и методика преподавания синхронного перевода. – М.: Воениздат, 1979. – 183 с
1. Чернов Г.В. Синхрон таржима назарияси ва амалиёти (рус тилидан Х.Ҳамидов таржима қилиб нашрга тайёрлаган) (o'quv qo'llanma). –Т.: ТошДШИ, 2019. – 112 б.

ВОЗМОЖНОСТИ КОРПУСНОЙ ЛИНГВИСТИКИ В ИЗУЧЕНИИ РУССКОГО СЛОВООБРАЗОВАНИЯ

Аблаева Нодира
Соискатель, Университет-Маъмуна.

Джуманиязова Интизор
Соискатель, Университет-Маъмунаю.

Аннотация: Корпусная лингвистика – раздел компьютерной лингвистики, занимающийся разработкой общих принципов построения и использования лингвистических корпусов с использованием компьютерных технологий. Корпусный подход как метод лингвистического исследования ориентирован на прикладное изучение языка. В настоящее время все современные лингвистические исследователи и составители словарей все больше используют объёмные корпусы текстов. Далее описаны способы использования корпусов, в частности, на основе представительного объема корпуса возможность обучения русской грамматике и создания современного словообразовательного словаря русского языка. Многоаспектное применение корпусов свидетельствуют о значительной роли корпусной лингвистики и представления ее в качестве инструмента современного лингвистического исследования.

Ключевые слова: корпусная лингвистика, Национальный корпус русского языка, текст, конкорданс, автоматическая семантическая разметка, словари, словообразовательные характеристики, исследование.

Annotatsiya: Korpus lingvistikasi - kompyuter texnologiyalaridan foydalangan holda lingvistik korpuslarni qurish va ulardan foydalanishning umumiy tamoyillarini ishlab chiqish bilan shug'ullanadigan hisoblash tilshunosligi bo'limi. Korpus yondashuvi lingvistik tadqiqot usuli sifatida amaliy tilni o'rganishga qaratilgan. Hozirgi vaqtda barcha zamonaviy lingvistik tadqiqotchilar va lug'at tuzuvchilari katta matn korpuslaridan tobora ko'proq foydalanmoqda. Quyida korpusdan qanday foydalanish, xususan, korpusning vakillik hajmidan kelib chiqqan holda, rus grammatikasini o'rgatish va rus tilining zamonaviy so'z yasalish lug'atini yaratish imkoniyati tasvirlangan. Korporaning ko'p qirrali ishlatilishi korpus lingvistikasining muhim rolini va uning zamonaviy lingvistik tadqiqotlar vositasi sifatida taqdim etilishini ko'rsatadi.

Kalit so'zlar: korpus lingvistikasi, rus tilining milliy korpusi, matn, muvofiqlik, avtomatik semantik belgilar, lug'atlar, so'z yasalish xususiyatlari, tadqiqot.

Abstract: Corpus linguistics is a branch of computational linguistics that develops general principles for the construction and use of linguistic corpora using computer technology. The corpus approach as a method of linguistic research is focused on applied language learning. At present, all modern linguistic researchers and compilers of dictionaries are increasingly using voluminous text corpora. The following describes ways to use corpora, in particular, based on the representative volume of the corpus, the possibility of teaching Russian grammar and creating a modern word-building dictionary of the Russian language. The multifaceted application of corpora

testifies to the significant role of corpus linguistics and its presentation as a tool for modern linguistic research.

Keywords: corpus linguistics, Russian National Corpus, text, concordance, automatic semantic markup, dictionaries, word-formation characteristics, research.

Введение. Российский лингвист доктор филологических наук Владимир Плунгян в своих рассуждениях о новом взгляде на язык, отмечал, что «если мы хотим назвать такую область лингвистики, которая по определению является суперсовременной, то первое, что приходит в голову, — это как раз лингвистика корпусов» [Плунгян В., 2013].

Корпус языка и корпусная лингвистика – та лингвистическая сфера, которая с конца XX века начала развиваться очень стремительно. Причиной тому послужили формирующиеся современные интеллектуальные программные системы, предназначенные для обработки текстов на естественном языке. В то же время, эти системы требуют большой экспериментальной лингвистической базы. Эти два фактора – технические возможности и массивы текстов, на наш взгляд, и обусловили растущий спрос на корпусные данные. В настоящее время все современные лингвистические исследователи и составители словарей все больше используют объёмные корпусы текстов [Abdurakhmonova N., 2020]

В материалах свободной энциклопедии Википедии дано следующее определение лингвистического / языкового корпуса текстов – «это большой, представленный в машиночитаемом формате, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» [https://ru.wikipedia.org/wiki/]

Под машиночитаемым форматом нужно понимать представление данных в формате, который может читать компьютер. Это означает, что машиночитаемые данные должны быть структурированными.

Также корпусом первого порядка называют любое собрание текстов, объединённых одним признаком, к примеру, язык автора, жанр, время создания текстов.

Основная часть. Лингвистические корпусы текстов появляются во второй половине XX века. В настоящее время корпусы созданы для многих языков мира, в частности, для венгерского, испанского, датского, итальянского, словенского, польского, чешского, эстонского, хорватского, словацкого, японского, русского. Национальный корпус русского языка (НКРЯ) охватывает почти 4.5 млн русских текстов общей длиной более 1,5 млрд слов. В настоящее время в состав НКРЯ вошли 38 корпусов: Основной корпус, Газетный корпус, Региональный корпус, Поэтический корпус, Параллельный корпус, Синтаксический корпус, Диалектный корпус, Обучающий корпус, Устный корпус, Акцентологический корпус, Мультимедийный корпус, Мультимедийный параллельный корпус, Древнерусский корпус, Церковнославянский корпус, Корпус берестяных грамот и Старорусский корпус [2]. Параллельный корпус содержит 22 параллельных подкорпуса общим объемом 165 млн слов, к примеру, английский подкорпус (43 млн), китайский (4,4 млн), башкирский (550 тыс), многоязычный (5 млн) [Национальный корпус русского языка].

Эффективность и целесообразность создания текстовых корпусов обусловлено достаточно большим объёмом данных из реальных контекстов, а также возможностью

многократно использовать однажды созданный корпус для решения самых разнообразных лингвистических задач.

Корпусный подход как метод лингвистического исследования ориентирован на прикладное изучение языка. Например, лексикографический анализ на базе корпусов раскрывает контекстное употребление слов, особенно синонимов, частотную сочетаемость их с другими словами, регулярность в тех или иных стилях, и четко определяет их семантику. Метод анализа на базе корпусов базируется на методах статистического и качественного анализа текста.

Уже существуют и активно задействованы многие типы корпусов текстов. Так, для изучения различных аспектов функционирования языковой системы используются исследовательские корпусы, для подтверждения и обоснования лингвистических фактов – учебные корпусы.

«В настоящее время конкордансы являются современным эффективным инструментом анализа текста, которые следует активно применять в практике преподавания языка и в лингвистических задачах» [Соснина Е.П., 2015].

Конкорданс-программа предназначена для выполнения установленной лингвистической задачи, к примеру, поиск в контексте определенной морфемы, слова или в словосочетания. Конкордансы нашли широкое применение в области прикладного языковедения, эти специальные программы представляют новые возможности для эффективного изучения языка и лексической обработки различных текстов.

В решении сложных лингвистических задач используются также статистические методики, которые особенно эффективны в машинном переводе, распознавании и синтезе речи, в системах автоматической обработки текста. Корпусы, в этом отношении, могут представить внушительную базу данных для исследований по грамматике и составлению словарей. Нужно отметить, что современные исследования по составлению словарей и грамматик в основном используют представительные корпусы текстов. «Развитие современных интеллектуальных программных систем, предназначенных для обработки текстов на естественном языке, также требует большой экспериментальной лингвистической базы. Спрос на корпусные данные совпал с появлением соответствующих технических возможностей» [Захаров В.П., 2005; 3-4].

Национальный корпус русского языка позволяет по любому слову составить конкорданс – это список всех употреблений данного слова в контексте, также в этом списке указаны ссылки на источник. На основе корпуса можно получить данные о частоте словоформ, лексем, грамматических категорий, проследить изменение частот и контекстов в различные периоды времени, получить данные о совместной встречаемости лексических единиц и т.д. Корпусы служат источником и инструментом в подготовке разнообразных исторических и современных словарей. Большой объем данных также используются для построения и уточнения грамматик и в целях обучения языку.

В НКРЯ все словоформы текстов получают автоматическую семантическую разметку, которая основана на наборе дискретных семантических характеристик, приписываемых в словаре. «В настоящее время в Корпусе реализована система поиска по лексико-семантическим признакам, основанная на частичной семантической разметке текстов. При такой разметке большинству слов в тексте приписывается один или несколько семантических и словообразовательных признаков, например, 'лицо', 'вещество',

'пространство', 'свойство человека', 'диминутив', 'отглагольное имя' и т. п.» [Национальный корпус русского языка].

«В основу семантической разметки положена система классификации русской лексики, принятая в базе данных "Лексикограф", которая разрабатывалась с 1992 г. в Отделе лингвистических исследований РАН под руководством Е. В. Падучевой и Е. В. Рахилиной... Словник семантического словаря базируется на морфологическом словаре системы DIALING, представляющим собой расширение Грамматического словаря русского языка А. А. Зализняка» [Национальный корпус русского языка].

Лексико-семантическая информация, приписываемая произвольному слову в тексте, состоит из трех групп помет:

- разряд (например, имя собственное, возвратное местоимение);
- собственно лексико-семантические характеристики (например, тематический класс лексемы, признаки каузативности, оценки);
- деривационные (словообразовательные) характеристики (например, «диминутив», «отадъективное наречие»).

Лексико-семантическая информация имеет различную структуру для разных частей речи. Кроме того, каждый из разрядов существительных - имена предметные, непредметные и собственные - имеет свою структуру помет.

Словообразовательные характеристики включают несколько типов:

- морфосемантические словообразовательные признаки (например, «диминутив», «каритив», «семельфактив»);
- разряд производящего слова (например, отглагольное существительное или отадъективное наречие);
- лексико-семантический (таксономический) тип производящего слова (например, наречие, образованное от прилагательного размера);

морфологический тип словообразования (субстантивация, сложное слово) [Национальный корпус русского языка].

Особо надо отметить, что корпусы используются лингвистами-теоретиками для проверки и доказательства своих теорий.

Компьютерные лингвисты, как особый класс пользователей, используя статистические и лингвистические закономерности в текстах, создают компьютерные модели языков. «Корпусы используются для разработки и настройки различных автоматизированных систем (машинный перевод, распознавание речи, информационный поиск)» [Захаров В.П., 2005; 7-8].

Выводы. Корпусная лингвистика – область лингвистики, в которой создаются, совершенствуются корпусы текстов, а также применяются в качестве инструмента лингвистического исследования.

Под корпусом в современной лингвистике нужно понимать совокупность текстов в электронной форме. Тексты собраны по определенным критериям, в соответствии с конкретной исследовательской задачей, и отражают ту или иную сферу языка.

Работая в корпусе данных, можно составить конкорданс - список всех употреблений слова в контексте со ссылками на источник. Представительный массив языковых данных за определенный период позволяет исследовать динамику изменения лексического состава

языка, анализировать лексико-грамматические характеристики разных жанров и у разных авторов.

Корпусы также служат инструментом в подготовке разнообразных исторических и современных словарей. Данные корпусов используются прикладными лингвистами (преподавателями, переводчиками) для построения грамматик, в обучении языку и для решения своих профессиональных задач. Корпусная лингвистика имеет практические механизмы создания и использования представительных массивов языковых данных, которые предназначены для лингвистических исследований в интересах широкого круга пользователей.

Таким образом, можно сказать, что посредством компьютерных технологий корпус превратился из метода работы с языковым материалом в эффективный инструмент исследования.

Список использованной литературы:

1. Abdurakhmonova N. Komputer lingvistikasi. Учебник – Globe edit publishing, 2020, 395b.
2. Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005. – 48 с.
3. Национальный корпус русского языка. 2003-2022. Доступен по адресу: ruscorpora.ru
4. Плунгян Владимир. Корпусная лингвистика. 14.01.2013. https://postnauka.ru/video/7783.
5. Соснина Е.П. Параллельные корпусы в обучении языку и переводу. - 2015. http://ling.ulstu.ru/linguistics/resourses/literature/articles/corpus_education_translation/
6. https://ru.wikipedia.org/wiki/

O‘ZBEK TILIDAGI MATNLARDA ATOQLI OTLARNI AVTOMATIK ANIQLASH NE VA NER TUSHUNCHASI

Abdullayev Abdulla Quranbayevich
“Urganch innovatsion university” NTM
Oliy ta’lim muassasasi Ta’limni kredit
tizimini boshqarish bo‘lim boshlig‘i.

Elov Botir Boltayevich
Alisher Navoiy nomidagi
Toshkent davlat o‘zbek tili va adabiyoti universiteti.
Texnika fanlari falsafa doktori(PhD), dotsent.

Samatboyeva Madina To‘lqinjon qizi
Kompyuter lingvistikasi mutaxassisligi magistranti
Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili
va adabiyoti universiteti.

Annotatsiya: Ushbu maqolada o‘zbek tilidagi matnlarda atoqli otlar va ularni avtomatik aniqlash metodlari to‘g‘risida fikr yuritilgan. Xususan, atoqli otlarning NLPda ifodalanishi va ularning lingvistik xususiyatlari o‘rganilgan. NE va NER tushunchalariga atroflicha ta’rif berilib, ularning xususiyatlari, aniqlash usullari va matnlar tarkibida kelish holatlari tadqiq qilingan.
Kalit so‘zlar: Ot, turdosh ot, atoqli ot, nomlash, nomlangan obyekt, NE, NER

Abstract: This article discusses proper nouns in Uzbek texts and methods for their determination. It studies the expression of common nouns in NLP and their linguistic features. Detailed definitions of the concepts of NE and NER, their properties, methods of definition and their use in texts are given.
Keywords: Noun, common nouns, proper nouns, naming, named object, NE, NER.

Аннотация: В данной статье рассматриваются собственные имена существительные в узбекских текстах и методы их определения. В нем изучается выражение нарицательных имен существительных в НЛП и их лингвистические особенности. Даны подробные определения понятия НЭ и НЭР, их свойства, способы определения и употребление их в текстах.
Ключевые слова: Существительное, нарицательные существительные, собственные имена существительные, именование, названный предмет, НЭ, НЭР.

Narsa-hodisalarga nom qo‘yish va uni nomlashning asosiy tamoyillari, atalish motivlari masalasi, olimlarni, shu jumladan, tilshunoslarni ham juda qadimdayoq qiziqtirgan. Geraklit, Demokrit, Platon va Aristotellarning shu to‘g‘ridagi qarashlari buning isbotidir. Narsa va hodisalarga nom qo‘yish masalasi umumiy va nazariy tilshunoslikning yirik vakillari bo‘lmish A.A.Potebnya, I.A.Boduen de Kurtene, F.de Sossyur kabi olimlarni ham qiziqtirgan. Ular

o‘zlarining nazariy qarashlari orqali nomlanish muommosini hal etishga katta hissa qo‘shganlar. [Суперанская, 1973:45-90]

Qadimdan insonlar borliqdagi barcha narsani nomlab o‘rganib kelishgan. Tabiatdagi jonzotlarni, voqealarni, hodisalarni, jonli va jonsiz narsalarni, manzillarni, hattoki o‘zlarini ham “nom”lar bilan atashgan. Bu orqali ularni bir-biridan farqlash, toifaga ajratish va yodda tez saqlab qolishga intilishgan.

Atrofimizdagi barcha obyektlar, hattoki subyektlar ham nomlanadi. Jumladan, inson ismlari, joy nomlari, suv ishootlari, hayvonlar, o‘simliklar nomlari bunga misoldir. Bu nomlar juda ko‘pki, ularning barchasini bir qarashda anglab olish qiyin. To‘g‘ri, tilshunoslik nuqtayi nazari bilan qaraganda nomlangan obyektlar matn tarkibida bosh harfda yoziladi, ammo barcha obyektlar ham emas. Bunda matnlarda nomlangan obyektlarni mashinaviy o‘qitish yordamida aniqlash ishimizni osonlashtiradi.

“O‘zbek tilining izohli lugati”da turdosh otni bir turdagi predmetlarning nomini anglatadigan otlar [Ўзбек тилининг изоҳли луғати. II том, 1981:62], “Atoqli ot bir shaxs yoki predmetga xos nom” [Ўзбек тилининг изоҳли луғати. II том, 1981:225], deb izohlangan. Ushbu tushunchalar onomastika nuqtayi nazaridan ham o‘ziga xos izohlariga ega. Masalan, N.V.Podolskaya “Собственное имя” (atoqli ot)ni *“onim”* terminiga havola qiladi va “onim“ termini shunday talqin qilingan: “Onim (onoma - atoqli ot) – o‘zi nomlayotgan obyektni boshqa obyektlardan farqlashga, individuallashtirish va identifikatsiyalashga xizmat qiluvchi so‘z, so‘z birikmasi yoki gapdir”[Подольская, 1988:91]. Ushbu muallif turdosh otni apellyativ terminiga havola qilgan holda quyidagicha tushuntiradi: “Turdosh ot atoqli otning teskarisidir”. Ushbu izohda turdosh ot tushunchasining mohiyati ochilgan emas. Keltirilgan izoh: “turdosh ot – apellyativdir, apellyativ – turdosh otdir” deyilgan aksiomaga aylanib qolgan. Buning ustiga apellyativ tushunchasini faqat turdosh ot, turdosh otlardan iborat, turdosh ot tushunchasiga teng deb bilish ham noto‘gridir. Atoqli otlarning lug‘aviy asoslari faqatgina turdosh otlardan iborat bo‘lmaydi. Chunki atoqli otlarga boshqa so‘z turkumlari ham lug‘aviy asos bo‘lishi mumkin.

Atoqli (antroponimik) va turdosh otlar orasidagi ma’noviy munosabatlar. O‘zbek tilshunosligida atoqli otlarning ma’nosiga oid muammolar, ayniqsa, atoqli ot bilan unga asos bo‘lgan turdosh so‘z (apellyativ)ning ma’nosini farqlash, chegaralash masalasi yetarli tadqiq qilingan emas. Shunga qaramay, turdosh va atoqli ot ma’nolari haqida mulohazalar aytish mumkin.[Бегматов, 2013:90]

1. Atoqli ot turdosh otlardan hosil bo‘ladi, ammo ular ma’nosi bilan assosiativ aloqani yo‘qotib, narsa va hodisalarni bir-biridan farqlash uchun xizmat qiladi.

2. Turdosh ot bir tildan ikkinchi tilga tarjima qilinadi, onimlar tarjima qilinmaydi.

3. Turdosh otlar ongda doimiy turadi, onimlar mexanik ravishda esda saqlanadi, tezda esdan chiqadi.

4. Turdosh ot grammatik jihatdan birlik va ko‘plikda qo‘llanadi. Onimlarda bu xususiyat yo‘q.

5. Turdosh otlar umumiylikni anglatsa, atoqli otlar xususiy bo‘ladi: o‘quvchi – sinfdagi har bir yosh; Tolibjon shulardan biri. *[*Qilichev, 2023:145]

Atoqli otlar – bu ot so‘z turkumining funksional-semantik sinfi bo‘lib, ularning asosiy vazifasi obyektlarni tanlash va aniqlash, ushbu mavzu bo‘yicha yagona va to‘liq tushunchalarni ifodalashdir.[Фонякова, 1990:3]

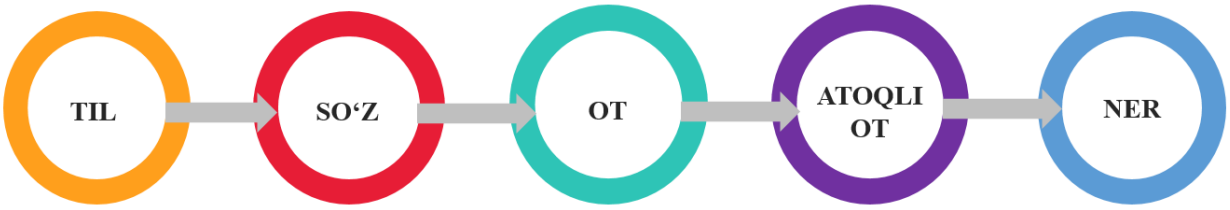
Atoqli otlar bu nomga atab qayta “yangi nom” qo‘yishdir. Atoqli otlarda *aniqlik, xoslik, yagonalik, mazmunlilik* xususiyatlari mavjud bo‘ladi.

Matnlar tarkibidagi “nom” (atoqli ot, nomlangan obyekt)lar esa matnning asosiy mazmunini belgilovchi unsurlardan biridir. Shuningdek, NLP (Natural Language Processing – Tabiiy tilni qayta ishlash)ning muhim masalaridan biri bo‘lgan *nomlangan obyektni tanib olish (NER – Named Entity Recognition)* hodisasidir. Matnlarni avtomatik qayta ishlab, matn tarkibidagi nomlangan obyektlarni aniqlash jarayoni – NER obyektlarini aniqlash deb nomlanadi. NER obyektlari matndagi lug‘atdan tashqari leksika, ya’ni atoqli otlarni ifodalaydi. NER obyektlari, kishi ismlari, joy nomlari, tashkilot nomlari, yo‘nalishlar…kabi bir qancha atoqli otlarni ifodalaydi, nomni aniqlashtiradi, aniqlangan nomlarni to‘g‘ri toifalarga ajratadi. Ilk bora NER va uning vazifalariga doir ilmiy ishlar 1996-yil Message Understanding (MUC-6)” konfrensiyasida e’lon qilingan. Shundan so‘ng jahondagi ko‘pgina kompyuter lingistikasidagi ilmiy ishlar matnlardagi nomlangan obyektlarni aniqlash(NER)ga qaratildi.

Bizni o‘rab turgan har qanday geografik obyektning nomi bor va bu “nom” obyektlarni bir-biridan farq qilish uchun xizmat qiladi. Shu sababdan, tildagi ko‘pgina so‘zlar nomlarga (atoqli otga) aylangan.[Rachna Jain…, 2020:5] Kundalik hayotimizdagi “kutubxona”, “maktab”, “bog‘cha” so‘zlari ham obyektni bildiruvchi so‘zlar, ammo ular turdosh ot(appellyativ – lotincha “atoqli ot aksi”)lardir. Turdosh va atoqli ot tushunchalari o‘zi anglatib kelgan obyektning “nomlanish” xususiyatiga ko‘ra bir-biridan farq qiladi. Ya’ni, nomlangan hamda nomlanmagan obyekt mazmunning kengayishi va qayta nomlanish sifati bilan ajralib turadi. *NE (Named Entity)* – ot so‘z turkumi doirasida nomlangan obyekt – “atoqli ot”ni ifodalaydi.

Nomlangan obyektlarni avtomatik aniqlovchi jarayon – *NER (Named Entity Recognition)* – nomlangan obyektni tanib olish bo‘lib, matndagi barcha nomli obyektlar aniqlanadi. NER *tabiiy tilni qayta ishlash(NLP – Natural Language Processing)* ning ko‘plab sohalarida qo‘llaniladi.[Susan Li, 2018:4]

NER obyektlari – atoqli otlarni ifodalaydi. Atoqli otlar o‘z nomi bilan nomlangan obyektlarga ishora qiladi. Matnlar tarkibidagi nomlangan obyektlar modellar yordamida avtomatik aniqlanganda keng doirada matn tarkibidagi barcha ot so‘z turkumidagi so‘zlar ko‘zdan kechirilib, so‘ngra tor doirada faqat atoqli otlar aniqlanadi. Ot so‘z turkumining eng birinchi til xususiyati bu – uning uch obyekt(kishi (kim?), narsa (nima?), joy (qayer?))ni ifodalashi. Bu xususiyat otlarning boshqa so‘z turkumlari(sifat, fe’l, olmosh…)dan farqlanishini ko‘rsatadi. NER obyektlari esa atoqli otlarga xos eng birinchi xususiyat “bosh harflilik”ga ko‘ra aniqlanadi. NER atamasi ingliz tilidan abbreviatura (qisqartma) shaklida hosil bo‘lgan termin hisoblanadi. Ingliz tilida amalga oshirilgan ilmiy ishlarni kuzatganimizda, NER obyektlari ro‘yxati ko‘p ekanligini kuzatamiz. Ammo biz o‘zbek tili doirasida ushbu ro‘yxatni qo‘llay olmaymiz. Chunki, ingliz tilida NER obyektlari tarkibiga otlardan tashqari, sonlar, sifatlar, miqdor o‘lchovlari, vaqt o‘lchovlari…kabi birliklar ham kiritilgan. Bu esa o‘zbek tili grammatikasi uchun xos emas. Bunga sabab, ikki til (o‘zbek va ingliz) ikki xil til oilasiga mansubligidadir. Xususan, til oilalari turlicha bo‘lgach, ularning grammatikasi (morfologiyasi, leksikologiyasi) tubdan farq qiladi.



1-rasm. Tilning grammatik xususiyatiga ko‘ra NER obyektlarini aniqlash

NER obyektlari matn tarkibiga ko‘proq ma’no yuklashga yordam beradi. Bu orqali esa matndan ma’lumot olish ishi amalga oshiriladi va u bog‘lanib kelayotgan so‘z NER orqali yanada aniqlashtiriladi. NER obyektlarini aniqlash, asosan, izohsiz matn bloklarini tahlil qilishdan boshlanadi. Ya’ni, bu yerda tayyor matn olinadi va uning tarkibidagi atoqli otlar(proper noun)lar ajratiladi.

Lekin atoqli ot tushunchasini faqat o‘zbek tili matnlarida NER obyektlarini topish uchun qo‘llash mumkin. Chet elda mavjud bo‘lgan NER obyektlarini aniqlashga yo‘naltirilgan dasturlar va ushbu yo‘nalishda qilingan ilmiy ishlar o‘rganilganda, mashinaviy o‘qitish tizimida faqat atoqli otlar NER obyekti sifatida olinmagan. Bu ro‘yxatda atoqli otlardan tashqari yillar, miqdorlar, foizlar, pul qiymatlari, fasllar, ranglar, qismlar, dizaynlar, voqealar, tibbiy kodlar ham kiritilgan. Bundan ko‘rinadiki, oz’bek tilidagi matnlar uchun atoqli otlar – NER obyekti, lekin NLP uchun NER obyekti bu faqat atoqli ot emas. Atoqli otlar – NER, NER bu – atoqli otgina emas! NER tushunchasi qamrovi kengdir. Unda atoqli otlardan tashqari yana ko‘pgina obyektlarni qamrab oladi. Ammo biz til oilalari farqliligi, ularning grammatikasi, leksikasi, morfologiyasi turlichaligini hisobga olib, o‘zbek tili uchun faqat atoqli otlarni NER obyekti sifatida olish maqbul yo‘l ekanligiga amin bo‘ldik.

NER obyektlari gap mazmuniga ko‘proq ma’no yuklashga yordam beradi. Matn tarkibidagi otlar ma’no va mazmun jihatdan bir-birini to‘ldiradi va bir-biriga bog‘lanib keladi. NER obyektlari esa gap mazmunini yanada boyitishga va asosiy ma’noni aniqlashtirishga yordam beradi.

Foydalanilgan adabiyotlar:

1. Суперанская А.В. общая теория имени собственного. М., “Наука”,1973, с.45-90; Жаббаров Э. Ўзбек ҳалқ ўйинлари лексикаси – Филол. фанлари номзоди, дисс. автореф., Тошкент, 1998, 8-бет
2. Ўзбек тилининг изоҳли луғати. II том. М.: Рус тили, 1981.225-бет. Уэбек тилининг изоҳли луғати. 1 том. М.: Рус тили, 1981.62-бет.
3. Уэбек тилининг изоҳли луғати. 1 том. М.: Рус тили, 1981.62-бет.
4. Подольская Н. В. Словарь русской ономастической термикологии, М.: Наука, 1988.
5. Э.А.Бегматов. Ўзбек тили антропонимикаси. Ўзбекистон Республикаси Фанлар академияси “Фан” нашриёти Тошкент. 2013. 90б
6. Bayramali Qilichev. Onomastika. O‘quv qo‘llanma. Buxoro. 2023. 115
7. Фонякова О.И. Имя собственное в художественном тексте: Учеб. пособие / О.И. Фонякова. Л.: ЛГУ, 1990. – С. 3.
8. Rachna Jain, Abhishek Sharma, Gouri Sankar Mishra, Parma Nand, and Sudeshna Chakraborty. Named Entity Recognition in English Text. Journal of Physics: Conference Series. 2020
9. Susan Li. Named Entity Recognition with NLTK and SpaCy. 2018.1p

GENDEROLINGVISTIKA ZAMONAVIY TILSHUNOSLIK SOHASI SIFATIDA

Abdunazarova Maftuna Faxriddinovna
O‘ZMU O‘zbek filologiyasi fakulteti
Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasi
stajyor o‘qituvchisi.

Anotatsiya: Ushbu maqola tilshunoslikning zamonaviy sohalaridan biri genderolingvistikaning o‘ziga xosligini ochib berishga asoslangan va unda genderga oid murojaat shakllari, stereotiplar, shuningdek badiiy matnda uchraydigan lingvomadaniy birliklarning gender xususiyatlari ko‘rib chiqiladi.
Kalit so‘zlar: genderolingvistika, antroposentirizm, feminizm, gender steriotiplari, murojaat shakllari.

Abstract: This article is based on revealing the uniqueness of gender-linguistics, one of the modern fields of linguistics, and it examines gender-related forms of reference, stereotypes, as well as gender characteristics of linguistic and cultural units found in literary texts.
Keywords: gender linguistics, anthropocentrism, feminism, gender stereotypes, forms of reference.

Аннотация: Статья основана на раскрытии своеобразия гендерной лингвистики, одной из современных областей языкознания, и рассматривает гендерные формы обращения, стереотипы, а также гендерные характеристики лингвокультурных единиц, встречающихся в художественных текстах.
Ключевые слова: гендерная лингвистика, антропоцентризм, феминизм, гендерные стереотипы, формы референции.

Kirish: XX asr oxiri gender muammosini butun insoniyat taraqqiyoti uchun eng dolzarb masalalardan biri sifatida ilgari surdi. Gender omili shaxsiyatning eng muhim ko‘rsatkichlaridan biri bo‘lib, u jinsni nafaqat biologik subyekt sifatida, balki ijtimoiy va madaniy jihatdan belgilangan aqliy tuzilma sifatida ham o‘z ichiga oladi.

Genderolingvistika tilshunoslikning nisbatan yangi sohalaridan biridir. G‘arb tilshunosligida bugungi kunga kelib gender lingvistikasi sohasida bir qancha ishlar amalga oshirilgan [50]. Bu ishlarda gender terminiga turlicha yondashuvlar, talqinlar kuzatiladi. Jumladan, G.Brandtning qayd etishicha, “gender madaniyat ta’siri ostida bo‘ladigan, ma’naviy jihatdan ayol va erkaklarning nutqiy, xulq– atvoriy tomonlarini baholay oladigan xususiyatlar majmuidir” [51]. Demak, gender konsepti nafaqat biologik jihatdan farqlanishni, o‘z navbatida, jinslarning nutqi, dunyoqarashi, odob– axloqi, kiyinishi, fe’l-atvori, yurish-turishi, oiladagi mavqelari jihatidan

[50]Yo‘ldasheva X. Q. O‘zbek tarixiy romanlarining sosiolingvistik tadqiqi: Filol. fan. bo‘y. fals. d-ri (PhD) ... diss. – Toshkent, 2017.

[51] Brandt G.A. Priroda jenshini kak problema: konsepsiya feminizma // Obshestvennie nauki i sovremennost'. – Moskva, 1998. – №2. – S.167–189.

tafovutlarni ham qamrab oladi. X.B.Samigova bu haqda quyidagilarni yozgan edi: “Biz mazkur terminni erkak va ayollarga xos bo‘lgan dunyoqarash, madaniyat, odobning ularning nutqi hamda tilidagi aksi va ifodasidir, deb talqin qilamiz”[52].

Muhokama va natijalar

Nemis tilshunosi va faylasufi F. Mautner (1913) turli ijtimoiy guruhlarda erkak va ayol nutq xatti-harakatlarini o‘rgangan. U ayol tilining kelib chiqishini o‘zaro qadimiy teatrning tarixiy an’analari bilan bog‘ladi, bu yerda ayol rollarini erkaklar o‘ynagan, ya’ni sahnaga chiqishga ruxsat berilgandagina haqiqiy ayol tili paydo bo‘lgan”[53] deb ta’kidlaydi. Ayollarning lingvistik xususiyatlari me’yordan salbiy og‘ish sifatida qabul qilinadi. Ba’zi iboralarni boshqa mazmunda qo‘llash holatlari ham ayollar nutqida uchrab turadi. Masalan, **ko‘zi uchib turmoq** iborasi nisbatan ayollar nutqida ko‘proq ifodalanadigan ibora. Mazmuni kutmaslik, yoqtirmaslik kabi ma’nolarda ifodalanadi.

-Tug‘ilgan kuningga hammani aytdingmi?

-ha, dadajon. Eng yaqin dugonalarimni?

Onasi biroz xomush tortib, qiziga qarab: “**Ko‘zim uchib turuvdi shu payt o‘zi**”.

Qizaloq onasining gapidan ranjib, uyiga kirib ketdi.

Iboralar nafaqat gender jihatdan balki, uslubi jihatidan ham keskin farqlanadi.

G‘arb tilshunoslarining fikricha, ta’kidli tarkibli so‘roq gaplar ingliz tilida so‘zlashuvchi ayollar nutqining asosiy sintaktik jihatlaridan biridir[54]. Ayollar va erkaklar nutqining o‘ziga xosligini belgilovchi asosiy omillardan biri ayollarning erkaklarga qaraganda ko‘proq nutqiy muloqotda xushmuomala bo‘lishga, odob saqlashga, suhbatdoshiga nisbatan hurmat ko‘rsatishga harakat qilishlaridir. Ayollar va erkaklar nutqidagi farq faqat yuqoridagilar bilan cheklanmaydi. Ushbu farqlarning asosida jinslarning psixologiyasi, tafakkuri o‘rtasidagi tafovutlar, jamiyatdagi gender parametrlarini belgilovchi normalar jamiyatda insonlarning ongi va tafakkuri, ijtimoiy-ma’naviy qarashlari asosida shakllanadi[55].

Masalan, ayollar nutqida qarg‘ishlar ko‘plab uchrasa, erkaklar nutqida so‘kinish holatlari ko‘p kuzatiladi. Iboralarda salbiy ma’noda ifodalanuvchi, bo‘yoqdor iboralar ham erkak nutqida, ham ayol nutqida qo‘llanishi mumkin.

Aravani quruq olib qochma (aldama), besh qo‘lingni og‘zingga tiqma (nokaslik qilmoq), dumini tugmoq (qaytmas qilib haydamoq), tirnoq ostidan kir qidirmoq (o‘ta injiq), yuragini oldirib qo‘ydi (bezillab qoldi), chuchvarani xom sanamoq (adashmoq), katta gapirdi (imkoniyatiga ortiqcha baho berib yubordi), kayfi chog‘ (xursand), bir o‘q bilan ikki quyonni urmoq (amalga oshirmoq) kabi turg‘un birikmalar erkaklar nutqida faol qo‘llanadi. Badiiy asarda ba’zan obrazning xarakterini ochish, saviyasi, dunyoqarashini ko‘rsatuvchi til birligi iboralar nutq vaziyatiga ko‘ra turlicha ifodalanishi mumkin.

Nazar Eshonqulning “Qora kitob” nomli qissasida erkak jinsiga mansub shaxsdan quydagi iboralarni uchratamiz.

[52] Samigova X.B. Ingliz va o‘zbek nutq madaniyati ritorik aspektining chog‘ishtirma tadqiqi: Filol. fan. d-ri (DSc) ... diss. Avtoref. – Toshkent, 2017. – 30 b.

[53] Соболева М. Е. Философия как критика языка в Германии. СПб 2005

[54] Lakoff R Language and woman’s place. N. Y, 1975.

[55] Крилина Л. В. Категория “gender” в языкознание. 1997

“Negadir siz bilan juda suhbatlashgim kelyapti, to‘g‘risini aytsam, hozir kimgadir **ko‘nglimni bo‘shatmasam**, tars yorilib ketadigandayman. Yuzim emas, **yuragim kuyib ketgan**, kuygan kovushga o‘xshab qolgan. Sizga **ko‘nglimni ochyapman.**

Qizim esa hozir **abadiy uyquda** yotibdi. Men **o‘zimni bosib olish**, iqrorim osonroq bo‘lishi uchun kallamni qizitib olishga keldim. Menga boshimdan kechirgan jazo oldida basharti hakamlar chiqaradigan jazo hech gap emas. Yig‘lay-yig‘lay **ko‘nglimni yozaman”.**

Yuqoridagi muallif nutqida ayol nutqiga xos belgilar ham ko‘zga tashlanadi. Badiiy matnning murakkabligi ham shundaki, tildagi vositalar har doim ham o‘z o‘rnida ifodalanmaydi. Qahramon ushbu asarda qiyin kunlarni boshidan kechirgan, hayotda qilgan ishlari mutlaqo noto‘g‘ri ekanligini keyin payqab qoladi. Ko‘nglini ochmoq iborasi ko‘proq ayollar nutqiga xos bo‘lib, dardlashganda ushbu iborani qo‘llaydi ammo erkak jinsiga mansub shaxslar tilda kam uchraydi[*Iskandarova, A.2009,2018,2019*]. Bizningcha, qahramon mast holda, dardlari shu qadar ko‘pki, o‘zini yengil his qilish uchun shu iboradan foydalangan. **Ko‘glini bo‘shatmoq** iborasi erkaklarda ham, ayollarda ham qo‘llanadi.

Badiiy asarlar misolida gender tadqiqot amalga oshirishda erkak va ayolga xos bo‘lgan stereotiplar, turli ma’no ko‘chishlari orqali genderga ishora qilish kabi vositalarni tahlil qilish obrazlilikni ochib berishga xizmat qiladi. Gender stereotiplari – bu ayol yoki erkakka xos barcha xususiyatlar tizimi bo‘lib, o‘ziga xos obrazlar tizimida har bir millat vakillarining lingvomadaniyatida turlicha aks etadi[*Iskandarova, A.2009,2018,2019*]. O‘zbek tiliga xos bo‘lgan gender stereotiplarini badiiy asarlar matnini o‘rganib, ikki asosiy turga ajratib tahlil qilishni lozim topdik:

1. Tashqi ko‘rinish bilan bog‘liq gender stereotiplari;
2. Xarakter xususiyati bilan bog‘liq gender stereotiplari.

Badiiy asarda uchraydigan gender stereotiplarni ikki turga ajratib tahlil qilamiz:

1. Erkaklarga xos gender stereotiplar;
2. Ayollarga xos gender stereotiplar.

Erkaklarga xos tashqi ko‘rinishni ifodalovchi gender stereotiplarga quyidagilar kiradi:

a) yuz a’zolari qismi ifodasi;
b) tana a’zolari qismi ifodasi;
c) xarakter ifodasi;
d)xatti-harakat ifodasi.

Erkaklarning yuz a’zolari qismi ifodasini aks ettiruvchi gender stereotiplariga misol qilib quyidagilarni kiritishimiz mumkin: *burgut ko‘z, qirg‘iy ko‘z, lochin ko‘z, o‘tkir nigohli, burni katta, burgut burun, qirra burun, yuzi yalpoq, tirtiq, yonoqlari bo‘rtib chiqqan, labi do‘ddaygan, qoshlari qalin, jingalak sochli*, *echki soqol, shopmo‘ylov, ingichka mo‘ylov, cho‘qqisoqol, qo‘ng‘iz mo‘ylov, sersoqol* kabi. Buni quyidagi misollarda ko‘rib chiqamiz: *Orqa o‘rindiqda ko‘zoynakli **qirra burun** yigit bilan o‘ttiz yoshlar chamasidagi oppoqqina yigit o‘tirardi*[56].

Erkaklarning tana a’zolari qismi ifodasini aks ettiruvchi gender stereotiplariga quyidagilar kiradi: *yo‘g‘on gavdali, bo‘yni yo‘gon, tepakal, qanshari pastroq, davangirday, hamma yog‘i yung, alpqomat, bahaybat devday, barzangi, gavdali, bilaklarda kuch toshib turgan*

[56] Said Ahmad Ufq. Sano-standart nashriyoti. Toshkent,2016. – B.69

oyoq-qo‘li uzun, mitti, ixchamgina, davangirday, beli baquvvat, xushsurat, ko‘rimsizgina, fildek, polvon, pahlavon, buqaday kabi.

Erkaklarning xarakteri ifodasi orqali aks etuvchi gender stereotiplariga quyidagilar kiradi: *boshi buzuq, badjahlroq, halimgina, xunxo'r, mard, cho‘rtkesar, mishiqi, chapdast, chaynalmasdan erkakchasiga, loqayd, musulmonsheva kabi.*

Erkaklarga xos xatti-harakatlar ifodasini namoyon etuvchi gender sereotiplarga quyidagilar kiradi: *irg‘ib turmoq, mushtlashmoq, qitmirlik qilmoq, yotqizib-turgazmoq, ko‘z qismoq, aysh qilmoq, maishat qilmoq kabi.* Masalan: *Tomlardan chumchuq uyasiga intilib, tik teraklardan yiqilib sinmagan, chiqmagan joyi qolmagan.* ***Mushtlashishga ishqiboz****. Atayin Yakkatut, Toshloq, Zarkent taraflaiga borib basma-basga o‘z tenglari bilan* ***mushtlashardi****. Birini urib, biridan kaltak yeb qaytgan paytlari juda ko‘p bo‘lgan*[57].

—Menga qara, - dedi Tog‘a. - Ishlaringni qo‘y. Men ham bugun idoradan qochaman. Ikkovimiz bir ***maishat qilaylik.***

Demak, erkak kishiga xoslikni bildiruvchi asosiy tashqi belgi xususiyatlarga gavdaning ayolnikiga nisbatan yirik bo‘lishi, yelkalarning keng bo‘lishi, oyoq qo‘lning baquvvat va uzun bo‘lishi, yuz tuzilishi ayolnikidan tubdan farq qilishi, asosan, soqol-mo‘ylovning mavjudligi kabi xususiyatlar kiradi. Bundan tashqari, faqat erkaklargagina xos bo‘lgan xarakter xususiyati hamda muayyan xatti-harakatlar erkaklikning bosh va muhim ifodasi sifatida badiiy asarlarda ifodalanadi. Ana shu o‘ziga xosliklar erkaklarga xos bo‘lgan gender stereotiplar hisoblanadi.

Ayollarga xos tashqi ko‘rinishni aks ettiruvchi gender stereotiplar erkaklarga xos bo‘lgan stereotiplardan katta ajralib turadi. Ularga quyidagilar kiradi:

a) yuz a’zolari qismi ifodasi;
b) tana a’zolari qismi ifodasi;
c) xarakter ifodasi;
d) xatti-harakat ifodasi.

Ayolning yuz a’zolari qismi ifodasini aks ettiruvchi gender stereotiplariga quyidagilarni misol qilib olishimiz mumkin: *qora qosh, qora ko‘z, jodu ko‘zli, xumor ko‘z, ohu ko‘z, bodom qovoq, qoshi kamon, oydek, kulgich, sutga chayqalgandek oppoq yuz, oppoqqina, parichehra, g‘uncha labli, anordek qip-qizil lab, qizil olmaday, popukday, qirmizi yanoqli, gulyuzli, uzun kipriklar, qaldirg‘och qanotidan nusxa olgan qoshlar, nildek qora qosh-ko‘zlar bejirim lablar* kabi. Masalan: . ***Sutga chayqalgandek oppoq yuz****lariga,* ***anordek qip-qizil lab****lariga,* ***nildek qora qosh-ko‘zlar****iga tikilib turib birdan she’r aytib yuborgan:*

Yuz qizil, lablar qizil*,*
Barmoq uchi ham qip-qizil*,*
Qosh qaro, ko‘zlar qaro*,*
Hech bo‘lmasin baxting qaro![58]

Ayollarning tana a’zolari qismi ifodasini aks ettiruvchi gender stereotiplariga quyidagilar kiritishimiz mimkin: *oppoqqina, do‘mboqqina, do‘ndiqcha, bo‘liq ko‘krakli, sarvqomat, sochlari uzun, jamalaksoch, mallasoch, ko‘hlikkina, nozik bo‘yni, lo‘ppigina, oyoq-qo‘li chaqqon, qo‘li gul, durkun ko‘rkrakli, ninachidek ozg‘in badan,* kabi. Masalan*:* ***Oppoq tomoqlarga shamolda tegib o‘ynayotgan o‘rim- ga ko‘nmagan sochlar****. Shamol yoqasini har silkiganda ochilib*

[57] O‘sha asar – B.19
[58] Said Ahmad Ufq. Sano-standart nashriyoti.Toshkent,2016. – B.96

yopilayotgan yelkalar, hamon titrab turgan ko‘kraklar, hozirgina yig‘idan bo‘shagandek nam tortib turgan kipriklar-u ***chimchilab qo‘ygandek bejirim burun****, ayniqsa,* ***oq o‘rik po‘stidek bilinar-bilinmas sepkili bor tomoqlari*** *A’zamjonning yigitlik qonini ko‘pirtirib yubordi*[59].

Zebixon uning oxirgi ilinji edi. Uning jarangli kulgilari hamon quloqlari ostida yangrab turardi. Uning ***san’at asaridek ko‘rkam qomati*** *ko‘z oldida turardi*[60].

Endi Dildor o‘zining ***ko‘hlik****ligini ham,* ***nozik****ligini ham unutdi. Hozir erigami, taqdirigami zarda qilib yeng shimarib maydonga chiqdi*[61].

Ayollarning xarakteri ifodasi orqali aks ettiriluvchi gender stereotiplariga quyidagilarni kiritishimiz mumkin*: mushfiq, shallaqi, jonon, lobar, shirin, noshud, yengiltak, suyuqoyoq, jodugar, xushro‘y, sharmanda, ma’suma, hajiqiz, mute, juvon, onasi o‘pmagan, baodob, farishtaday oqko‘ngil, tili bir qarich* kabi. Masalan, *Ana endi yosh,* ***norasta****, hali uy-ro‘zg‘or nimaligini bilmagan, yalangoyoq yurgan, sochiga tol popuk taqishni tark qilmagan, qo‘g‘irchoq o‘ynashdan chiqmagan* ***ma’suma****ning umrini juvonmarg qilyapti*[62].

Shu o‘tgan davr uni ko‘p narsaga o‘rgatgan, bo‘shashgan qishloq qizini o‘z so‘zli, ***cho‘rtkesar*** *qilib qo‘ydi. Kuchga, g‘ayratga to‘lib gapira boshladi*[63].

Ayollarga xos xatti-harakatlar ifodasini namoyon etuvchi gender sereotiplarga quyidagilar kiradi: *ohu ko‘zlari suzilar, ohista qadam tashlamoq, hurkib yashamoq, lallaymoq, chimirilmoq, tug‘moq, ko‘kragini tutmoq, jon-jon deb tegishi mumkin, bevalikning taxir oshini totmoq, sambitdek nozik bellari egilib, buralib hammani mahliyo qilmoq, chimirilmoq, yuzini chetga burish, cho‘chib tushmoq, jilmaymoq, ohista, noz-u karashma qilmoq, tegmoq, anqaymoq, kaltakdan qo‘rqib yashardi, yengil qadam bosmoq,* kabi. Masala: *Dildor astoydil o‘ynardi. U endi* ***nihoyatda ochilib ketgan****, uni ko‘rgan odam bolali xotin deb o‘ylamasdi. Chaqnagan* ***ohu ko‘zlari suzilar****,* ***sambitdek nozik bellari egilib****,* ***buralib*** *hammani mahliyo qilgandi. Uning qoshi ham, ko‘zi ham, jamiki a’zoyi badani silkinib o‘ynardi*[64].

Demak, biz tahlilga tortgan misollardan ham ko‘rinib turibdiki, lingvistikada genderga xos bo‘lgan steriotiplar mavjud. Ma’lumki, ayol va erkak ko‘rinish xatti-harakat va xarakter, jamiyatdagi roli jihatidan erkaklardan ajralib turadi, ana shu farqlanuvchi omillar badiiy matnda ayol va erkak stereotiplar sifatida o‘z aksini topadi[*Iskandarova, A.2009,2018,2019*]. Badiiy asarlarda ayol obrazini tasvirlashda nozik va sarvqomat, jozibador va latofatli, qosh-ko‘zlarida erkakni sehrlab qo‘yishga qodir kuch yoki jodusi bor, g‘uncha dudog‘ va qirmizi yanoq, mayin va shirali ovoz sohibasi degan go‘zal ta’riflardan foydalanilsa, erkaklarga nisbatan esa basavlat, devqomat, mard, kelishgan, bir so‘zli kabi ta’riflar beriladi. Ana shu belgi va ta‘riflar bevosita erkak va ayollarga xos bo‘lgan gender stereotiplarni shakllantiradi.

Xulosa sifatida shuni ta’kidlash kerakki, erkak va ayolning nutqi, dunyoqarashi, madaniyatidagi tafovutlarning tildagi ifodasi gender tilshunosligining muhim xususiyati hisoblanadi.

[59]Said Ahmad Ufq. Sano-standart nashriyoti.Toshkent,2016. – B.604
[60] O‘sha asar - B.372
[61] O‘sha asar - B.521
[62]O‘sha asar - B.521
[63] O‘sha asar - B.554
[64]O‘sha asar – B.96

Foydalanilgan adabiyotlar ro‘yxati:

1. Qurbonova M.M., Boymirzaeva S. O‘zbek tilshunosligi mamlakatimiz innovatsion taraqqiyoti ko‘zgusida. 2011.
2. Said Ahmad Ufq. Sano-standart nashriyoti. Toshkent,2016.
3. Qo‘chqor Norqobil “Guldurak momo”. https://www.ziyouz.com/portal-haqida/xarita/uzbek-nasri/qo-chqor-norqobil-1968
4. Abdunazarova M. “Gender steriotiplari tahlili” International scientific-online conference: “Principal issues of scientific research and modern education”188-196-betlar.
5. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, *11*(19), 9117.
6. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
7. Iskandarova, A. (2018). Vatanni olib ketib bo’lmaydi.
8. Iskandarova, A. (2009). A.Navoiyning "Lisonut-tayr" dostonidagi vodiylar ta’rifi va mohiyati haqida. Мумтоз суз таьлими.
9. Iskandarova, A. (2019). INSONIYATNI SAODATGA YETAKLAGAN ADIB (Adabiyot darslarida Chingiz Aytmatov ijodini o’rganishga doir). Tafakkur Chashmasi MCHJ.
10. Абдураҳмонова Н. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти: Филол. фан. фалсафа фан. д-ри (phd). дисс. – Тошкент, 2018. –Б. 4.

SOTSIOLINGVISTIKADA O‘RGANILADIGAN MASALALAR XUSUSIDA

Xudayberganova Dildora
O‘zMU katta o‘qituvchisi, PhD.

Kasimova Ziyoda
O‘zMU katta o‘qituvchisi, PhD.

Annotatsiya: Ushbu maqolada sotsiolingvistika fanining kelib chiqishi, maqsad va vazifalari, tadqiq etiladigan predmet va obyekti haqida ma’lumot beriladi. Shuningdek, sotsiolingvistika doirasida o‘rganiladigan variativlik, lisoniy hamjamiyat va uning muloqot vositalari, tillarning sotsiolingvistik tasnifi, lisoniy vaziyat va til siyosati kabi masalalar atroflicha yoritib beriladi.
Kalit so‘zlar: Sotsiolingvistika, variativlik, lisoniy hamjamiyat, argo, jargon, sleng, lisoniy vaziyat, til siyosati, pijin va kreol tillar.

Аннотация. В данной статье представлена информация о зарождении, целях и задачах науки социолингвистики, предмете и объекте исследования. Также подробно освещены такие вопросы, как изменчивость, языковая общность и ее средства общения, социолингвистическая классификация языков, языковая ситуация и языковая политика.
Ключевые слова: Социолингвистика, вариация, языковая общность, арго, жаргон, сленг, языковая ситуация, языковая политика, пиджин и креольский языки.

Annotation: This article provides information about the origin, goals and tasks of the science of sociolinguistics, the subject and object of research. Also, issues such as variability, linguistic community and its means of communication, sociolinguistic classification of languages, linguistic situation and language policy are covered in detail.
Keywords: Sociolinguistics, variation, linguistic community, slang, slang, linguistic situation, language policy, pidgin and creole languages.

Sotsiolingvistika tilshunoslikning tilni va uning mavjud bo‘lgan ijtimoiy sharoitlar bilan bog‘liqligini o‘rganadigan sohasidir. Ijtimoiy sharoitlar deganda, amaldagi til rivojlanayotgan tashqi shart-sharoitlar majmuyi: mazkur tildan foydalanayotgan kishilar jamiyati, bu jamiyatning ijtimoiy strukturasi, til egalarining yoshi, ijtimoiy mavqeyi, madaniyat va bilim darajasi, yashash joyi orasidagi farqlari, shuningdek, ularning muloqot vaziyatiga bog‘liq bo‘lgan nutqiy muomalasidagi farqlar tushuniladi. Demak, sotsiolingvistika tilning jamiyat hayotidagi roli, uning ijtimoiy tabiati, ijtimoiy funksiya(vazifa)lari, ijtimoiy omillarning tilga bo‘lgan ta’sir mexanizmi bilan bog‘liq masalalarni o‘rganadigan tilshunoslik sohasidir. “Sotsiolingvistika” termini fanga birinchi marta 1952-yilda amerikalik sotsiolog G.Karri tomonidan kiritildi. Sotsiolingvistikaning *obyekti* til yoki tillarning ijtimoiy faoliyatidir. Til muayyan ijtimoiy strukturaga ega bo‘lgan jamiyatda faoliyat ko‘rsatadi, shunga ko‘ra, sotsiolingvistikani tilni ijtimoiy kontekstda o‘rganadigan fan deyish mumkin. sotsiolingvistikaning *predmeti*ni qisqacha “inson va jamiyat”, deb belgilash mumkin.

Sotsiolingvistikada o‘rganiladigan masalalar va tushunchalarga to‘xtalib o‘tsak. *Lisoniy hamjamiyat* - ijtimoiy, iqtisodiy, siyosiy va madaniy jihatdan o‘zaro bog‘liq bo‘lgan, bir til yoki bir necha til yordamida kundalik hayotda bir-birlari va turli ijtimoiy institutlar bilan o‘zaro muloqotga kirishuvchi insonlar majmuyi.

Lisoniy hamjamiyat deganda, katta bir davlatdan tortib, kichik ijtimoiy guruhlar (oila, sport guruhlari)ga kiruvchi turli miqdordagi individlar majmuyini tushunish mumkin. Bunday holatlarda ajratish mezoni ijtimoiy jamoa va doimiy kommunikativ aloqalarning mavjudligi bo‘lishi lozim. Muayyan lisoniy hamjamiyatda qo‘llaniladigan, funksional jihatdan bir-birini to‘ldiradigan, o‘zaro aloqada bo‘lgan kodlar yoki subkodlar majmuyi *ijtimoiy-kommunikativ tizimni* tashkil qiladi. Ushbu lisoniy hamjamiyatlarda qo‘llaniladigan muloqot vositalari mavjud. Quyida ular haqida ma’lumot beriladi:

Adabiy til (standart) umumxalq tilining ishlangan, sayqal berilgan, grammatik qonun-qoidalarga rioya qilinadigan, ma’lum me’yorga solingan, umummajburiy va qonuniy to‘g‘ri qo‘llanadigan, xalqning turli madaniy ehtiyojlariga xizmat qiluvchi tildir. *Adabiy til* jamiyatning ilm-fan, ta’lim, adabiyot, diplomatiya, huquqshunoslik, ish yuritish, turli tashkilotlar va ziyolilarning kundalik hayotida ishlatiladigan milliy til ko‘rinishidir. Adabiy tilga muomala jarayonida qo‘llanadigan shevaga, so‘zlashuvga xos bo‘lgan so‘zlarning ishlatilmasligi xosdir.

Ingliz tilshunosligi an’analariga ko‘ra, adabiy til *standard language, standard English* deb yuritiladi. Shuningdek, *standartlashgan adabiy til* termini ham qo‘llaniladi.

Norasmiy nutq modeli hisoblangan dialektlar hududiy va ijtimoiy turlarga bo‘linadi. *Hududiy dialektlar* ma’lum hududning o‘ziga xos hududiy belgilari, binobarin, o‘ziga xos lug‘at, grammatik yoki fonetik xususiyatlariga ega bo‘lgan so‘zlashuv tilidir. Masalan, Toshkent shevasi, Namangan shevasi, Xorazm shevasi kabi.

Ijtimoiy dialektga *sotsiolekt* deyiladi. *Sotsiolekt* termini lingvistikada XX asrning ikkinchi yarmida paydo bo‘lgan. U *sotsio-* jamiyatga bo‘lgan munosabat va *dialekt* so‘zlaridan tashkil topgan bo‘lib, rus tilidagi “социальный диалект” (ijtimoiy dialekt) birikmasining qisqargan shaklidir.

Masalan, askarlar (askarlar jargoni), o‘quvchilar (maktab jargoni) nutqi, jinoyatchilar jargoni, xippilar argosi, talabalar slengi, turli sohalarda ishlovchilarning kasbiy “tili”, argolari va h.k. sotsiolektlarga misol bo‘la oladi. [Беликов В., 2001: 30]

Jargon – biron ijtimoiy guruh vakillarining o‘z nutqi bilan ko‘pchilikdan ajralib turish maqsadida, o‘zicha mazmun berib ishlatadigan so‘z va iboralari. Masalan, *oq* (aroq), *qizil* (vino) – ichuvchilar nutqiga xos; *strelk*a (uchrashuv) yoshlarning nutqida; *ko‘k, ko‘kat* (AQSh dollari), *dodasi* (biron narsaning zo‘ri) – ko‘proq bozorchilarning nutqida qo‘llaniladi.

Argo – jargondan farqli o‘laroq, biron ijtimoiy guruh tomonidan boshqalar uchun atayin tushunarsiz qilib qo‘llaniladigan u yoki bu darajadagi *yashirin, yasama* til. Masalan, o‘g‘rilar argosida *shmon qilmoq, gopstop qilmoq* iboralari “o‘g‘irlamoq” ma’nosida, giyohvandlar argosida *dori* “giyohvand modda” ma’nosida ishlatiladi. Argolar jargonga nisbatan o‘zining *yashirinlik* xususiyati bilan ham ajralib turadi. Shuning uchun ular ko‘proq jamiyatning yashirin tabaqalari hisoblangan josuslar, jinoyatchilar, o‘g‘rilar, giyohvandlar, fohishalar va h.k.lar nutqida uchraydi.

Sleng – G‘arb tilshunosligi an’analariga xos termin sanaladi. “Sleng” terminiga haligacha aniq ta’rif berilmagan. Turli lug‘atlarda va sotsiolingvistikaga oid adabiyotlarda uning “jargon” (guruhlar jargoni, yoshlar jargoni) terminiga yaqin turishi qayd qilingan. [Словарь-справочник лингвистических терминов, 1985: 287]

Sleng so‘zining qachon paydo bo‘lgani noma’lum, biroq u ilk bor yozma ravishda XVIII asrda Angliyada qayd qilingan va o‘sha davrda “haqorat” ma’nosini anglatgan. 1850-yillarga kelib, mazkur termin “noqonuniy” oddiy leksikaning ifodasi sifatida keng qo‘llanila boshladi. Ayni paytda uning jamiyatning quyi tabaqalari qo‘llaydigan *slang-lingo* va “rang-barang” aholi foydalanadigan *argot* singari sinonimlari ham paydo bo‘lgan. [Partridge E., 1979: 3]

Sleng – muayyan kasb egalari yoki ijtimoiy guruhlarning og‘zaki nutqida qo‘llaniladigan, emotsional-ekspressiv bo‘yoqqa ega bo‘lgan so‘z va iboralar. Slenglar ko‘proq yoshlarning nutqida uchraydi. Masalan, *olmaxon* “o‘zini oladigan, nozlanadigan odam”, *g‘isht* “xunuk”, *sindirdi* “lol qoldirdi” va h.k.

Umuman, argo, jargon va sleng – sotsiolektning turlari hisoblanadi. Mazkur til shakllarining har biridagi o‘ziga xos xususiyat u yoki bu guruhlarning kasbiy alohidaligi yoki ularning boshqa jamoalardan ijtimoiy cheklanganligiga bog‘liqdir.

Sotsiolingvistik tushunchalardan yana biri – variativlik. Bir xil ma’noga ega bo‘lgan lisoniy belgilarning shaklidagi, tashqi ko‘rinishidagi nomutanosiblikka, har xillikka *variativlik* deyiladi. Uning lisoniy, ijtimoiy va gender variativik kabi turlari bor.

Sotsiolingvistika fani tillar tipologiyasi muammosi bilan ham shug‘ullanadi. Tillarning sotsiolingvistik tipologiyasi uning jamiyatdagi ijtimoiy mavqeyini inobatga olgan tipologiyadir. Hozirgi zamon tilshunosligi tillarni sotsiolingvistik jihatdan quyidagi 7 tipga ajratadi:

1. Standart tillar. Adabiy til normasiga ega bo‘lgan, jamiyatda adabiy til, davlat tili sifatida xizmat qiluvchi tillarga standart tillar deyiladi. Masalan, o‘zbek, tojik, rus, qirg‘iz, afg‘on, ukrain kabi tillar shular jumlasidandir.

2. Klassik tillar. Klassitsizm davrida takomilga etgan, o‘z yozuviga ega bo‘lgan, o‘sha davrda buyuk asarlar yaratilgan tilga klassik til deyiladi. Masalan, lotin tili, sanskrit tili, qadimgi yunon tili, sug‘d, xorazmiy tili kabilar. Bu tillar hozirgi vaqtda o‘lik tilga aylangan.

3. Mahalliy tillar. Son jihatdan ko‘p bo‘lmagan kichik millatlarga, xalqlarga xizmat qiluvchi, o‘z yozuviga ega bo‘lmagan tillarga mahalliy tillar deyiladi. Masalan, Amerika qit’asidagi mahalliy hindular tili bunga yaqqol misol bo‘la oladi. Afrika qit’asidagi ayrim mahalliy aholi tillari ham sotsiolingvistik tipologiyaga ko‘ra ana shu guruhga mansubdir.

4. Kreol tillar. Mustamlakachilik natijasida Amerika, Afrika, Sharq mamlakatlari xalqlari tillari bilan Yevropa tillari (ingliz, fransuz, ispan, portugal) elementlari aralashuvidan tashkil topgan tillar kreol tillar deyiladi. Masalan, Amerikadagi Gaiti orolida mahalliy aholi bilan fransuz tili aralashuvi natijasida gatti – kreol tili shakllangan. Lotin Amerikasidagi mahalliy aholi tili aralashuvi natijasida ispan-nautl kreol tili yuzaga kelgan. Amerikadagi Aruba, Bonayre orollarida negr va ispan tillari aralashuvi bilan papiyamento kreol tili shakllangan. Bunday kreol tillari grammatik sistemasi soddalashgani, aloqa uchun qulaylashgani bilan ajralib turadi.

5. **Pijin tillari.** Mahalliy tillarning Yevropa tillari bilan chatishuvi natijasida yuzaga kelgan tillardir.

Pijin tillar deganda ham yevropaliklar tili bilan mahalliy tillarning chatishuvidan paydo bo‘lgan tillar nazarda tutiladi. Pijin va kreol tillari o‘xshash va farqli tomonlarga ega. Agar kreol tilida so‘zlashadigan, uni ona tili deb hisoblaydigan aholi bo‘lsa, pijin tili bunday xususiyatga ega emas. Pijin tillaridan ikkinchi aloqa vositasi sifatida katta port shaharlarida foydalaniladi. Umuman olganda, pijin tili kreol tillarining paydo bo‘lishida boshlang‘ich bosqich sanaladi. Hozirgi vaqtda eng ko‘p tarqalgan pijin tillaridan biri «pijin-inglish» deb nomlanuvchi janubiy Xitoy portlarida qo‘llanadigan tildir.

6. Sun’iy tillar. Maxsus yaratilgan xalqaro aloqa tillariga sun’iy til deyiladi. Masalan, esperanto tili 1887-yilda polyak shifokori A.Zamengof tomonidan yaratilgan. Bu til ancha keng tarqalgan. Bu tilda gazeta va jurnallar, badiiy asarlar nashr etiladi. Bunday sun’iy tillar qatoriga yana ido, volyapyuk, oksidental, interlingva, neo kabi tillarni ham kiritish mumkin.

7. Dialekt. Bu so‘z yunoncha bo‘lib, sheva, lahja ma’nosini ifodalaydi. Ma’lum bir qabila, xalq yoki millat tarkibiga kiruvchi kichik hududga tarqalgan aholi tilidir. Mahalliy dialektlar ijtimoiy dialektlar va jargonlardan o‘ziga xos grammatik qurilishga, lug’at fondiga ega ekanligi, kelajakda milliy adabiy tilga asos bo‘lib xizmat qila olishi bilan farqlanib turadi.

Sotsiolingvistikada lisoniy vaziyat tushunchasi ham bor. *“Lisoniy vaziyat”* (language situation) deyilganda, muayyan bir vaqtda, muayyan bir joyda qanday tildan foydalaniladi, tillarning soni, qanday turdagi til ushbu hududda ishlatiladi, qancha odam, qanday sharoitda bu tilda gaplashadi, ushbu jamoa a’zolarining bu tilga bo‘lgan munosabati, umuman, tilning qo‘llanish doirasi haqidagi masalalar tushuniladi. [Ferguson C.A., 1971: 157]

Muayyan lisoniy hamjamiyatning u yoki bu davridagi ijtimoiy-kommunikativ tizim tarkibiy qismlari orasidagi funksional munosabat ushbu hamjamiyatga xos bo‘lgan *lisoniy vaziyatni* shakllantiradi. Bu munosabatlar ko‘pincha barqaror bo‘ladi. Biroq mamlakatdagi siyosiy vaziyatning, davlat tuzilishining o‘zgarishi, iqtisodiy islohotlar, ijtimoiy va milliy siyosatdagi yangi yo‘nalishlar va h.k. ijtimoiy-kommunikativ tizimning holatiga, uning tarkibiga, funksiyasiga va uning qismlari bo‘lgan kodlar va subkodlarga ta’sir qilishi mumkin.

“Lisoniy vaziyat” tushunchasi, odatda, katta til hamjamiyatlariga – mamlakatlar, hududlar, respublikalarga nisbatan qo‘llaniladi. Bu tushuncha uchun zamon omili muhim hisoblanadi, ya’ni *lisoniy vaziyat* ijtimoiy-kommunikativ tizimning muayyan davridagi faoliyatidir.

Til siyosati deganda, davlat, sinf, ijtimoiy guruhlarning mavjud tillar yoki til osti tizimlari vazifalarini o‘zgartirish yoki saqlab qolish, yangi lingvistik me’yorlarni joriy qilish yoki eski me’yorlarni saqlab qolishga qaratilgan tadbirlar majmuyi tushuniladi.

Til siyosatini ichki va tashqi til siyosatiga ajratish mumkin. *Ichki til siyosati* deganda, muayyan davlat doirasida bir yoki bir qancha tillarga nisbatan qo‘llanilgan chora-tadbirlar majmuyi ko‘zda tutiladi. Ko‘p millatli davlatlarda bir qancha tillar faoliyat ko‘rsatadi, shuning uchun ham bunday davlatlarning til siyosati milliy siyosatning muhim tarkibiy qismlaridan biri hisoblanadi. *Tashqi til siyosati* mamlakat tashqi siyosatining bir qismi bo‘lib, u biron bir tilni mamlakat tashqarisida, ya’ni xalqaro miqyosda keng yoyishga qaratiladi. Hozirgi kunda o‘zbek tilini AQSh, Angliya, Germaniya, Janubiy Koreya, Yaponiya, Xitoy kabi rivojlangan davlatlarda o‘qitilayotgani, o‘zbek tilini o‘rganuvchilar va bu tilda gaplashuvchilar sonining ortib borayotgani, o‘zbek tilida kompyuter dasturlarining yaratilayotgani, xullas, o‘zbek tilini dunyoviy tillar darajasiga olib chiqishga qaratilgan sa’y-harakatlar – O‘zbekistonning tashqi til siyosatini belgilaydi.

Sotsiolingvistikada o‘rganiladigan asosiy termin va tushunchalar yuqoridagilar bo‘lib, ularning sotsiolingvistik tadqiqini rivojlantirish zarur. Hozirgi zamon tilshunosligi bu kabi masalalarni dunyo tillari miqyosida o‘rganishi kerak, deb o‘ylaymiz.

Foydalanilgan adabiyotlar ro‘yxati:

1. Ferguson C.A. Language structure and language use. – Stanford, 1971. –P.157.
2. Partridge E. Slang Today and Yesterday. - London: Routledge and Kegan Paul, 1979. – P.3.

3. Qurbonova M. va b. Sotsiolingvistika / o‘quv-uslubiy majmua. – T. 2017.
4. Usmanova Sh. va b. Sotsiolingvistika / o‘quv qo‘llanma. – T. 2013.
5. Беликов В. И., Крысин Л. П. Социолингвистика. –М., 2001. –С. 30.
6. Словарь-справочник лингвистических терминов. – М.: Российская энциклопедия, 1985. –С. 287.

AN ANALYSIS OF PRISON ARGOT AND ITS TYPES

Mahmudova Jasmina
O‘ZMU 2-bosqich talabasi.
E-mail: jasminamakhmudova21@gmail.com

Annotation. In recent years, a lot of research is being carried out in various areas of linguistics, which is the reason for the discovery of unknown aspects in the field. Also, the sociolinguistic terminology is enriched with many examples. In this article, the important lexical-semantic features of prison slang are tried to be revealed more through examples.
Keywords: argot, sleng, jargon, argotism, exclusivity, code.

Аннотация. В последние годы проводится множество исследований в различных областях языкознания, что является причиной открытия неизвестных аспектов в этой области. Также социолингвистическая терминология обогащена множеством примеров. В данной статье важные лексико-семантические особенности тюремного жаргона пытаются раскрыть более подробно на примерах.
Ключевые слова: арго сленг арготизм эксклюзивность код.

Til muloqotning asosiy usuli sifatida ijtimoiylashuv va xulq-atvor kodlarini ishlab chiqish va davom ettirishning asosiy tarkibiy qismidir. “Argo” atamasini ta’riflashda turli xil qarashlar mavjud. Shuningdek, O.S.Axmanovaning “Slovar lingvisticheskix terminov” asarida “Argo aynan jargondir, faqat jargondan salbiy ma’nosi, kamsitish, masxaralash ma’nosi yo‘qligi bilan farq qiladi. San’atkorlik argosi. Harbiy argo. O'g'rilar argosidan iboratligi, ingliz tilidagi sleng terminiga tengligi qayd qilinadi[65]. D.E.Rozental va M.A.Telenkovalarning "Slovar-spravochnik lingvisticheskix terminov" lug‘atida quyidagicha izohlanadi: "Argo-alohida guruh, to‘dalarning boshqalardan lisoniy ajralib turish maqsadida sun’iy yaratilgan tili (ba’zan maxfiy tili) bo‘lib, asosan, o‘sha guruhga begona kishilar tushunmaydigan so‘zlari bilan farq qiladi. Talabalar argosi. Sport argosi. Qartabozlar argosi. O‘g‘rilar argosi"[66]. A.Hojievning “Tilshunoslik terminlarining izohli lug‘ati” da quyidagicha ta’rif berilgan: “Argo” (fran. Argot jargon). Yasama til. Biror ijtimoiy guruh, toifaning (masalan, talabalar, sportchilar, o‘g‘rilar va b.) o‘ziga xos, boshqalar tushunmaydigan lug‘aviy birliklardan iborat yasama tili. Mas., kolxoz –chiptasiz yo‘lovchilar (haydovchilar tilida)[67]. Shunga ko‘ra, biz argo tarkibiga kiradigan so‘z va iboralarni argotizmlar deb olishimiz mumkin. Xuddishunday, A.Hojiyev: “Argotizm. Biror argoga xos so‘z va iboralar. Argotizmlar adabiy tilda biror uslubiy maqsad bilan qo‘llanishi mumkin“, - degan fikrni ilgari suradi.

Argolar kimlar tomonidan qo‘llanilayotganiga ko‘ra, o‘g‘rilar argosi, qartabozlar argosi, mahbuslar argosi kabi turlarga bo’linadi. Shuningdek, qamoqxona argosi boshqa argo turlaridan yuqori darajada sirli bo‘lishi kerakligi bilan ajralib turadi, qamoqxonadagi vaziyat ulardan shuni

[65] Ахманова О.С. Словарь лингвистических терминов / О.С. Ахманова. – 2-е изд., стер. – М : УРСС : Едиториал УРСС, 2004. – 271 с.
[66] Розенталь Д.Э., Теленкова М.А. Словарь-справочник лингвистических терминов:пособие для учителя.— Москва: Просвещение, 1985. — 199 с.
[67]HojiyevA. Tilshunoslikterminlariningizohlilug'ati. Toshkent: O'zbekistonmilliyensiklopediyasi, 2002.— 43 b.

talab qiladi. Mahbuslar argosining leksikasi qisqa umr ko'rishi kerak, chunki uning ikkita asosiy vazifasi: eksklyuzivlik va jamiyatga qarshi turishdir. Aftidan, mahbuslar ko‘pincha qamoqxona xodimlarini ularning nutqini tushunishga yo‘l qo‘ymaslikni va o‘ziga xos qarama-qarshi madaniyatni saqlashni xohlashadi. Shulte ta’kidlaganidek, qamoqxona argosi hayotiyligini saqlab qolish uchun doimiy ravishda o‘zini yangilab turadi. Misol uchun, kich argosi bir vaqtlar “qamoqxona” ma'nosida ishlatilgan, ammo u oxir-oqibat eskirgan va o‘rnini uy argosi egallagan.

Qamoqxona argolarini nimaga nisbatan qo‘llanilishiga ko‘ra 3 guruhga bo‘lishimiz mumkin: qamoqxona buyumlariga nisbatan qo‘llaniluvchi argolar, shaxslarga nisbatan qo‘llaniluvchi argolar va holatlarga nisbatan qo‘llaniluvchi argolar.

Buyumlarga nisbatan qo‘llaniluvchi argolar o‘zida ham sirlilik, ham o‘ziga xoslik xususiyatiga ega bo‘ladi. Quyidagi jadvalda bunday argolarga misollar berilgan:

Argo	**Lug‘aviy ma’nosi**
Asosiy	Qamoqxona stoli
Havo	Pul
Reshka	Panjara
Xapok	Suv ichiladigan idish
Lipa	Qalbaki hujjat
Chefir	Achchiq choy
Bo‘ri bileti	Ozod qilinganlik haqidagi guvohnoma
Katki	Yekatirina surati tushirilgan pul

Mahbuslar qamoqxonaga kelgandan keyin, ularning moslashib ketishi uchun vaqt kerak bo‘ladi. Chunki bu yerdagi muhit tashqi olamdan butunlay farq qiladi. Yangi kelgan mahbus “Salom, mujiklar” deyishi yaramaydi. Chunki bu yerdagi toifalarning har birining o‘z nomi bo‘ladi. Agar “blatnoy”lar o‘tirgan bo‘lsa, ularni “mujik” deya haqorat qilib, obro‘sini pastga urgan bo‘ladi. Shaxslarga nisbatan qo‘llaniladigan argolar:

Go‘sht	O‘g‘ri
Quloq	Haqiqatni aytadigan to‘g‘ri mahbus
Depan	Qamoqxona navbatchisi
Xaza, xata	Aka, uka
Xatadosh	Sherik, bir kamerada yashovchi kishi
Saqoqush	Bezori
Kontingent	Jinoyatchi
Frayer	Erkin odam, ozodlikdagi shaxs
Kent	Do‘st

Qamoqxonadagi holatlar, vaziyatlarga nisbatan qo‘llaniluvchi argolar:

Shuxer	Shovqin
Kipish	Janjal
Volya	Ozodlik
Sizo	Tergov
O‘chirish	O‘ldirish
Shmon	Tintuv

Ayrim talqinlarga ko‘ra, “shmon” so‘zi Sharqiy Yevropa yaxudiylari tili bo‘lmish “idish” (qadimiy ivrit va nemis tillari aralashmasi) tilida “sakkiz” raqamini anglatadi. Chor Rossiyasi davrida qamoqxonalarda har kuni soat sakkizda kameralarda tintuv o‘tkazilgan. Shu tariqa "soat sakkizda o‘tkaziladigan tintuv" o‘rniga “shmon” so‘zi ishlatilgan.

Qamoqxona argolari boshqa argo guruhlari singari o‘ta sirli, yashirin, niqoblangan bo‘ladi. Argolarning yangilanib turishi nafaqat qamoqxona argo lug‘atining zaxirasiga, balki muqarrar ravishda til xilma-xilligining leksik-semantik murakkabligiga ham hissa qo‘shadi.

Foydalanilgan adabiyotlar:

1. Langacker, R. W.2008 "Is slang a word for linguists?”, London, 5-17 p
2. Hojiyev A.2002. Tilshunoslik terminlarining izohli lug‘ati, Toshkent, 128 b.
3. Schulte, M. 2010. “The language of the underworld and its sociolinguistic significance”, Contributions to the Study of Language, Literature and Culture. — 45-60 p.
4. https://www.babbel.com/en/magazine/cants-and-anti-languages-the-secret-history-of-cryptolects
5. Ахманова О.С. 2004. Словарь лингвистических терминов, Москва – 271 с.
6. Розенталь Д.Э., Теленкова М.А. Словарь-справочник лингвистических терминов:пособие для учителя.— Москва: Просвещение, 1985. — 199 с.

SCIENTIFIC DISCOURSE AS AN OBJECT OF STUDY OF SCIENTIFIC RESERCH

Ibodullayeva Kunduz
The teacher of the chair of
Computer linguistics and Practical linguistics
in Uzbekistan National University.

Xolbekova Dilrabo
The teacher of the chair of
Computer linguistics and Practical linguistics
in Uzbekistan National University.

Annotatsiya: Diskus tushunchasi bugungi kunda zamonaviy tilshunoslikning muhim tadqiqot obyekti sifatida o‘rganilmoqda. Ushbu maqolada diskurs haqida tilshunos olimlarning qarashlari, diskurs va matn tushunchalari haqidagi mulohazalar va zamonaviy o‘zbek tilshunosligida diskurs tushunchasi va uning turlarini tadqiq etish haqida so‘z yuritilgan. Shuningdek diskurs analizning kompyuter lingvistikasi va korpus lingvistikasi bilan bog‘liq jarayon ekanligi haqida fikrlar bayon qilingan.

Kalit so‘zlar: Diskurs, diskurs tahlili, matn, nutq, kompyuter lingvistikasi, korpus, ilmiy diskurs

Annotation: In recent times the concept of discourse is being studied as an important research object of Modern Linguistics. In this article the views of linguists on discourse, discourse itself and the comments about the concepts of a text and the idea of discourse in modern Uzbek linguistics and the research its types are discussed. It is also stated that discourse analysis is connected process which is with computer linguistics and corpus linguistics

Keywords: Discourse, discourse analysis, text, speech, computer linguistics, corpus, scientific discourse.

Аннотация: Понятие дискус сегодня изучается как важный объект исследования современной лингвистики. В данной статье рассказывается о взглядах лингвистов на дискурс, комментируются понятия дискурс и текст, а также исследуются понятие дискурса и его виды в современном узбекском языкознании. Также утверждается, что анализ дискурса – это процесс, связанный с компьютерной лингвистикой и корпусной лингвистикой.

Ключевые слова: Дискурс, дискурс-анализ, текст, речь, компьютерная лингвистика, корпус, научный дискурс

The concept of “discourse”, “discourse analysis” has become one of the most used term today as a topic of interdisciplinary research. As a result of the development of language in relation with society, we are witnessing an increase in research on the topic discourse and discourse analysis in the field of humanities. This, in turn, indicates that discourse theory is becoming a subject of interdisciplinary research. In addition, to linguistics, this explanation is on object of study in such fields as literary studies, history, philosophy, sociology, political science, pedagogy

computer linguistics, artificial intelligence, translation theory. Wach of these disciplines approaches the study of discourse in it's own way.

We often hear the term discourse in our lives. Discourse is interpreted differently in the fields of science. Many linguists interpret it (French discourse, English discourse, Latin discourse - action, continuous exchange, conversation and speaking) in terms of the process of speech activity and the manner of speaking.

The concept of discourse was first used in a series of articles published in 1952 by the American linguist Z. Harris, titled Discourse Analysis, it was about his work developing transformational grammar in the late 1930s. According to this research, analysis the speech or writing which are analyzed together is a way of descriptive linguistics at the sometime beyond the limits of the simple sentence. While other scientists have confirmed that it is difficult to find a comprehensive and acceptable definition for discourse analysis.

In terms of modern approaches, the discourse is complex is a communicative phenomenon, additionally the text, includes extra linguistic factors such a (thoughts, knowledge of the world, the aims of the reader, attitudes) that is necessary for understanding the text. [Ro'ziyeva N.,2023: 153]

According to the linguist scientist A. Pardayev, discourse is a process of practical use of linguistic and non-linguistic tools in the form and type that they consider to be the most effective in order to exchange ideas and influence each other. It is a process, itnd it is manifestation of hundreds of linguistic and non-linguistic factors in a common form towards one goal. [Ablokulova O.,2021:1167]

The concept of discourse in Uzbek linguistics by Shio Saforov studied. The linguist scientist studied the issue of text and discourse analysis, and according to his interpretation, "Discourse should be interpreted in a broader context than the text. Discourse is both the process of linguistic activity and its product at the same time. is a phenomenon" (Kubryakova). In order for knowledge to be transmitted and formed anew, it is necessary to create a text, which means that the text begins to be created at the very moment of discursive activity. It is known that these two phenomena are studied in the relationship of "heperonym" - "heponym". Discourse is a specific type and category of human conscious activity, and the text is a manifestation of it. The interpretation of the category of discourse in a general sense in such a broad sense is related to communication system, other fields of science that are interested in human conscious activity - philosophy, sociology, psychology, It is a rule that has already been recognized for cybernetics. [Sh. Safarov 4. 249b]. Text and language are a means of connecting culture, because through the analysis of the text we get linguistic-cultural information about the national-cultural features of the language system. Sh. Safarov states that the discourse and text are a unique field of experience that creates all the necessary conditions for the full manifestation of the possibilities of language system reserves, and E.S. He also cites Kubryakova's opinion that "there is no universally accepted definition of discourse in the period of the current development of the science of linguistics."

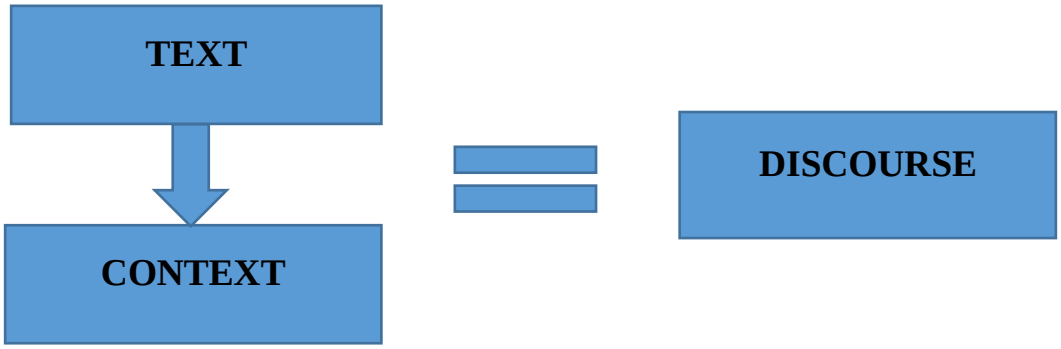
In the field of social humanities, discourse describes a formal way of thinking that can be expressed through language. Discourse is a boundary that defines what statements can be made about a topic. In other words, the chosen speech provides the necessary vocabulary, phrases or style for communication. For example, different guerilla movements may use two separate discourses, describing them as "freedom fighters" or "terrorists".

Therefore, discourse is one of the main problems to be studied in pragmatic and cognitive linguistics, which are recognized as new trends in modern linguistics. Discourse markers are words or phrases that connect, guide, and help organize sentences. They do not change the meaning. These tools connect what is written or written to other things. Discourse is divided into oral and written type according to the means of expression.

Discourse analysis is also of great importance in computer linguistics and corpus linguistics, which are new directions of modern linguistics. Discourse studies include specific analyses. These analyzes are carried out within the framework of corpus linguistics, which is the reason for the emergence of current scientific views. The purpose of discursive analysis is to determine the social content behind oral and written speech, the relationship between language and social processes, and the communicative quality of the text. That is, the texts in the corpus have additional meta tags, which provide information about the author, the period of the text, the genre and other information, and the analyzes carried out on the corpus can be researched according to the nature and nature of the content, giving the researcher a wide perspective. can specify.

In 1998, L. Flaudu focused on the possibilities of corpus linguistics to "implement" discourse analysis. Corpus-based discourse analysis is considered from the following three main approaches proposed by Hyland. They have areas of overlap, each based on their respective areas of overlap. Hyland cites these aspects as follows. A textual approach; critical approach; A contextual approach.

Although the research methods of discourse and corpus analysis are different, the basis that unites them is the context of the text.



If we analyze the above distinctions in general, we can conclude that the main basis in corpus linguistics and discourse analysis studies is text and speech, and such a basis creates a mutual synergistic effect [Nurmatova Go'zal., 2023:463]

Professor N. Abdurakhmonova, doctor of philological sciences, defines discourse as follows in the textbook "Computer Linguistics". "Extended form of sentences made in sequence by one or more people to exchange or transmit information"[Abdurakhmonova, N.2021,2022].

Text Linguistics and Discourse in Computational Linguistics it is necessary to differentiate between the analysis. According to the sources, text is written language, and discourse is formed by oral speech. However, in modern linguistics, any form of information is considered as text. According to Schiffrin, text linguistics studies only the internal elements of the text. Discourse analysis analyzes the issue of text context. A formal approach is reflected in the text, and a functional approach is reflected in the discourse. Laura Alba Juez tries to explain discourse

analysis more fully. Linguistic analysis of naturally occurring written and spoken discourse is interactively interpreted as a form of language use through dialogue between people in a specific social context. [N. Abdurakhmonova., 2021:141]. Based on the above points, since the research object of computer linguistics and corpus linguistics is the text, the natural language processing process is related to discourse analysis is also related. Discourse analysis is important in sentiment analysis and corpus linguistics of computational linguistics. An important aspect of discourse analysis is the use of databases. [N. Abdurakhmonova.,2021:142]

In some sources, it is considered as a discourse-genre. discourse-speech genre, that is, it is interpreted as a speech that is distinguished by its typical structure, communicative purpose, information delivery feature, the uniqueness of using linguistic means, and its focus on a specific addressee. Also, today, the concept of discourse as an object of research is studied by its social nature, divided into such types as political discourse, social discourse, legal discourse, artistic discourse, scientific discourse, philosophical discourse, and it is becoming one of the central topics in the creation of special studies. From this we can see that there are types of discourse that apply to any aspect of human activity.

In the linguistic encyclopedic dictionary, the term discourse is defined as typical of the encyclopedic concept as follows: DISCOURSE - extralinguistic - formed by pragmatic, socio-cultural, psychological and other factors related text in the factual aspect. In it, speech acquires a purposeful social character in connection with the cognitive process of those who create it. Discourse is an "eventful life." For this reason, the term is not used to refer to older texts, unlike the term "text". Because such texts cannot be directly connected with living life. Discourse also includes paralinguistic (mimicry, gesture) features. [AbdurahmanovaN.N., 2023: 745].

Thus, in modern linguistics, discourse is a complex communicative process that occurs between the speaker and the listener in a certain time and space, which is used in the broad sense of text, speech, dialogue. In the narrow sense - oral and written text. Discourse is the sphere of mutual exchange of ideas between the speaker and the listener and their influence on each other, the application of linguistic tools, linguistic and non-linguistic factors for a specific purpose in various tasks during the speech process.

References:

1. Ro‘ziyeva N. “Zamonaviy tilshunoslikda matn va diskurss talqini” LINGUISTICSSCIENTIFIC REPORTSOF BUKHARA STATE UNIVERSITY 2023/1(95),-153 b
2. Ablakulova O. “Diskurs tushinchasi va uning tadqiqi” Academic Research in Educational Sciences. Volume 2 Issue 11, 2021, -1167 b
3. Safarov Sh. Pragmalingvistika (monografiya), Toshkent: 2008, -249 b.
4. Nurmatova G. “Korpus lingvistikasi doirasida diskurs tadqiqotlarning ilmiy-nazariy muammolari” FarDU Ilmiy xabarlar 2023,463-b
5. Abduraxmonova N. Kompyuter lingvistikasi (darslik), Tashkent: Nodirabegim, -141 b.
6. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, *11*(19), 9117.
7. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).

8. Abdurakhmonova, N., Tuliyev, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus. uz. In *2021 International Conference on Information Science and Communications Technologies (ICISCT)* (pp. 1-4). IEEE.
9. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In *2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)* (pp. 1790-1793). IEEE.
10. AbduraxmanovaN. “Zamonaviy til nazariyasida “Diskurs” atamasi va unga ilmiy yondashuvlar” International Jornal of Education, Social Science. Volume 11 Issue 2 february 2023, -745 b
11. Sharipbay, A., Shirinova, R., Abdurakhmonova, N., & Qodirova, Z. (2023). ONTOLOGICAL KNOWLEDGE BASED MODELS REPRESENTING MEDICINE TERMINOLOGY. Advances in Social Sciences and Economics, 1(1), 4.
12. Kuvondikova, G. I., & To‘rayeva, S. (2024). O‘ZBEK TILIDAGI KO‘CHMA MA’NOLI SO ‘ZLARNI FRANSUZ TILIGA TARJIMA QILISH XUSUSIYATLARI. Educational Research in Universal Sciences, 3(2), 254-257

IJTIMOIY MAQOM, IJTIMOIY ROL, LISONIY ISHTIMOIYLASHUV

Iskandarova Aybibi Raximbayevna
O‘zbekiston Milliy universiteti
“Kompyuter lingvistikasi va amaliy tilshunoslik” kafedrasi dotsenti.
E-mail: oybibiiskandarova26@gmail.com

Annotatsiya: Ushbu maqolada Said Ahmadning “Qorako‘z Majnun” hikoyasi tahlili jarayonida sotsiolingvistik atamalar: ijtimoiy maqom, ijtimoiy rol, tilning ijtimoiylashuv hodisalariga diqqat qaratildi. Ijtimoiy maqom, ijtimoiy rol, lisoniy ishtimoiylashuv tushunchalari hikoyadagi Saodat aya, o‘gli Bo‘rixon, Anvar obrazlarining nutqi, xatti-harakatlari bilan bog‘liq tarzda nazariy jihatdanyoritildi.

Kalit so‘zlar: kommunikativ vaziyat, nutqiy akt, ijtimoiy maqom, ijtimoiy rol, lisoniy ijtimoiylashuv.

Аннотация: В данной статье в ходе анализа рассказа Саида Ахмада «Черноглазый Меджнун» были рассмотрены социолингвистические термины: социальный статус, социальная роль, социализация языка. Понятия социальный статус, социальная роль, языковая социализация были теоретически освещены во взаимосвязи с речью и поведением персонажей рассказа Саодат - ая, Борихан и Анвар.

Ключевые слова: коммуникативная ситуация, речевой акт, социальный статус, социальная роль, языковая социализация.

Abstract: This article's examination of Said Ahmed's narrative "Karakoz Majnun" concentrated on sociolinguistic words such as social position, social role, and language socialization processes. The words and actions of the story's characters Saodat, son Borikhan, and Anvar were examined in relation to theoretical explanations of social position, social role, and linguistic communication.

Keywords: communicative situation, speech act, social status, social role, linguistic socialization.

Shaxsning ijtimoiy maqomi, ijtimoiy roli, tilining ijtimoiylashuvi bir biri bilan bog‘liq jarayon bo‘lib, u diaxron sotsiolingvistikada o‘rganiladi. Talabalarni nazariy ma’lumotlar bilan tanishtirganda, Said Ahmadning “Qorako‘z Majnun” hikoyasi tahlilidan foydalanish yaxshi samara beradi. Tahlil jarayonida kommunikativ vaziyat, nutqiy akt, ijtimoiy maqom, ijtimoiy rol, tilning ishtimoiylashuvi atamalarining mohiyati ravshanlashadi, mavzu talabalarning xotirasida oson o‘rnashadi, mustahkamlanadi.

“Jamiyatda yashaydigan va turli guruhlarga mansub bo‘lgan har bir individ jamiyatda bir qancha ijtimoiy o‘ringa ega bo‘ladi… Mazkur o‘rinlarning har biri muayyan qoidalar va vazifalar bilan bog‘liq bo‘lib, u ijtimoiy maqom deb nomlanadi. Ijtimoiy maqomlarning aksariyatiga insonning o‘zi erishadi.”[Sh.Usmanova, N.Bekmuxamedova, G.Iskandarova. 2014:33]

Haqiqatan ham, ijtimoiy maqom shaxsning ma’lum jamiyatda yashab, shu jamiyatning biror bir guruhiga mansub bo‘lishi, egallagan lavozimi, kasbi, shu lavozim, kasbiga doir huquq,

burch, majburiyatlari bilan bog‘liq.Shaxsning shu lavozim va majburiyatdan kelib chiqib bajaradigan vazifalari uning ijtimoiy rollaridir.

“Qorako‘z Majnun” hikoyasi xabar mazmunli matn bilan boshlanadi: “Saodat aya bomdod namozini o‘qib, joynamoz poyida uzoq o‘tirib qoldi. Bundan 3 yil oldin olamdan o‘tgan eri usta Turobga atab Qur’on tilovat qildi. O‘ris shaharlarda daydib qolib ketgan o‘g‘li Bo‘rixonga Xudodan insof tiladi.”[Said Ahmad. 2013: 90]

Bahodir Karim “Ruhiyat alifbosi” asarida sotsiologik metod haqida fikr bildirib: “Sotsiologik metod asosiga e‘tibor qaratgan I. Ten san‘at predmeti jamiyat hayotida mavjud insonlarning xarakteridir, deb ta’kidlaydi. U ilmiy-nazariy tamoyillarni “irq, muhit va vaqt” tushunchalari vositasida izohlaydi. San’at asarini qaysidir irq-xalq-millatga mansubligini, ma’lum bir muhit va muayyan zamonda yaratilishiga alohida urg‘u beradi.” [Bahodir Karim. 2018:157]

Haqiqatan ham, inson qaysi muhitda yashasa, uning nutqida o‘sha muhit so‘zlashuv tili aks etadi, Insonning nutqi u bajaradigan ijtimoiy maqom va ijtimoiy rollari bilan bog‘liq. “Ijtimoiylashuv jarayonining muhim tarkibiy qismlaridan biri lisoniy ijtimoiylashuv bo‘lib, u muayyan jamiyatda qo‘llaniladigan tilni o‘zlashtirish va undan foydalanish qoidalarini u yoki bu ijtimoiy rolni ijrosiga muofiq ishlatishini ifodalaydi”[Sh.Usmanova, N.Bekmuxamedova, G.Iskandarova. 2014:34]

Hikoyada bu holat “Я здоров”, “Ну зачем, зачем плачеш, мама, вот и приехал, хватит, хватит”, “Салям”, “Ведь барана жалько, все равно я столько мясо не ем! У нас баранина не едят” kabi personaj nutqida ishlatilgan sintaktik qurilmalar orqali berilgan. “O‘g‘li Bo‘rixon oltmish yettinchi yili armiyaga ketgan, Harbiy xizmati tugadi hamki, uyga qaytmadi. O‘sha yoqlarda uylanib, bola-chaqali bo‘lib, qolib ketdi. Ba’zi-ba’zida undan “Яздоров” degan ikki enlikkina xat kelardi. Yaqin o‘n besh yildirki, adresni unutib qo‘yganmi, ishqilib, shu o‘rischagina xat ham kelmay qo‘ygan.”

“Personajlar nutqini individuallashtirish zarurati badiiy asar tilidagi differensatsiyalanganlikni yanada orttiradi. Chunki asardagi har bir personajning nutqi uning xarakter xususiyatlariga, dunyoqarashi, muhiti, ma’naviy qiyofasi, madaniy-ma’rifiy darajasi kabi jihatlarga muvofiq bo‘lishi lozim. Sababki, epik va dramatik asarlarda qahramon xarakterini yaratishning asosiy vositalaridan biri personaj nutqi sanaladi. Zero, asardagi har bir personaj o‘zining yoshi, jinsi, ijtimoiy qatlamga mansubligi, fe’l-sajiyasi, madaniy-ma’rifiy saviyasi kabilarga muvofiq gapiradi, demak, nutqi bilan o‘zini tavsiflaydi.”[D. H.Quronov. 2020:164]

“Qorako‘z Majnun” tragik mazmunga ega hikoyadir. Bo‘rixondagi millatsizlik, tilsizlik, Vatansizlik - bu fojiaviy holatlardir: “Bo‘rixon begona uyga kirib qolgan odamdek qovushmay turardi. U ukalariga, singillariga nima deyishni bilardi. Ammo til bilmasa nima qilsin? O‘ylab-o‘ylab, “Салям!” dedi. Jigarlari kulishni ham, yig‘lashni ham bilmay hayron turib qolishdi”…

“Tug‘ilgan yurtiga muhabbat tuyg‘usi uni tark qilganiga ko‘p yillar bo‘lgan. Ona tili qadim-qadim zamonlardayoq unutilib ketgan Shumer tili qatori tumanlar orasida qolib ketgandi,”-deyiladi hikoyada.

“Ijtimoiy-siyosiy terminlarning qisqacha izohli lug‘ati”da ijtimoiy maqom va ijtimoiy rol, ijtimoiylashuv atamalari quyidagicha izohlangan:

Ijtimoiy maqom – shaxsning yoshi, jinsi, kelib chiqishi, kasbi, oilaviy ahvoliga ko‘ra jamiyatda tutgan o‘rni.

Ijtimoiylashuv – inson tomonidan jamiyatda hayot kechirish imkonini beruvchi bilimlar, me'yorlar va qadriyatlar tizimining o'zlashtirilish jarayoni. ["Ijtimoiy-siyosiy terminlarning qisqacha izohli lug'ati: 2020: 95]

Rol – kimsaning ta'sir ko'lami, biror narsada ishtirok darajasi, tutgan o'rni, mavqei. [Ўзбек тилининг изоҳли луғати.2007: 391]

Hikoyada Anvarning "tog'am o'zbekchani esidan chiqarib yuboribdi". "Men bilan o'rischa gaplashdi". "Bitta sog'in echkisi, to'rtta qanor qopdek cho'chqasi, o'ntacha cho'chqachalari bor,", "Qish zabtiga olganda shu mollarini ham uyiga opkirib olisharkan". "Bo'chka-bo'chka samogon aroq yasab, qishi bilan ichisharkan" kabi tavsiflari Bo'rixonning xristian oilasidagi otalik maqomi, shu maqomga mos rollarini ko'rsatmoqda.

"Har qanday maqom huquq, majburiyat va unga mos me'yoriy muomaladan iborat bo'ladi. … Bir maqomga bir necha rollar mos kelishi mumkin." …Shaxsning ijtimoiylashuv jarayoni – yashashi va harakat qilishi kutilayotgan jamiyatga kirishidir. Ijtimoiylashuv individning asta-sekin o'zi uchun yangi bo'lgan guruhlarga kirishi va yangi rollarni o'zlashtirish jarayonidir."[Sh.Usmanova, N. Bekmuxamedova, G.Iskandarova. 2014:34]

Hikoyada shaxsning ijtimoiy maqomi, ijtimoiy roli, shu maqom va roli bilan bog'liq tarzda namoyon bo'lgan nutqiy akti quyidagi jumlalarda ifodalangan: Bo'rixon "o'rmon ichkarisidagi qishloq butxonasi qo'ng'iroqchisining erdan qolgan qiziga oshiq-u beqaror bo'lib qoldi". "Qallig'ining otasi, boshqa dindagi odamga qizimni bermayman, deb turib oldi". "Qiz Bo'rixonni xristian diniga kirishga undadi". "Ishq-muhabbatdan ko'zini parda bosgan Bo'rixon hech ikkilanmay rozi bo'ldi". "Uni cherkovda cho'qintirishdi". "Keyin cherkov oqsoqoli kelin va kuyovga toj kiydirib, nikoh o'qidi".

"Bo'rixon xotini, qaynonasi bilan har kuni cherkovga borib cho'qinadigan bo'ldi." "Qaynotasi o'lgandan so'ng uning o'rniga cherkov qo'ng'iroqchisi qilib qo'yishdi"

"Nimaiki ish bo'lsa, barini u bajaradigan bo'ldi". "Piligi so'xta bo'lgan shamlarning uchini qaychilaydi, yonib tamom bo'lganlarini almashtiradi".

Hikoyadagi bu tavsilotlar Bo'rixonning cherkov qo'ng'iroqchisi jtimoiy maqomi va shu maqomga mos harakatlari - ijtimoiy rolini ko'rsatmoqda. Shaxsning jamiyat ichida bir nechta maqomlari bor. Shu maqomlar bilan bog'liq rollari – harakatlari bor. Demak Bo'rixonnig ijtimoiy maqomi - uning kasbi, kasbi bilan bog'liq xatti-harakatlari esa – uning rollaridir. Bu ijtimoiy maqomning bir ko'rinishi. Oilada ota, er, cherkov qo'ng'iroqchisining kuyovi ekanligiham Bo'rixonning maqomlaridir.

Said Ahmad Saodat ayaning ruhiy holatini vaziyatga mos tarzda tasvirlagan: "Birovga so'zini bermaydigan errayim kampirning shohi sindi, ostona hatlamay uyda muqim o'tirib qoldi". Bo'rixonning bo'ynidagi yaltillagan butni ko'rgandan keyin kampirning "badanidan chayon o'rmalagandek seskandi". "Bir dam uni shuur tark etdi".

Bu o'zbekning oilasi uchun obro' talashadigan, pokiza, xokisor, bolafidoyi onasi - Saodat ayaning onalik maqomi va rollaridir. Hikoyada Saodat ayaning rollari quyidagicha tasvirlanadi:

"Saodat kampir uni chaqaloqligida xuddi shu supada beshikka belab tebratardi." "Uch yoshga kirguncha shu supada bag'rida olib yotgan edi". "Uni yomon ko'zdan asrasin, deb kiyimlariga tumor-u ko'zmunchoqlar taqib qo'yardi". "Sultonimga atab unga kokil qo'ygandi". "Olti yoshga to'lganda uni er-xotin Turkistonga olib borib hazrat Yassaviy maqbarasining shayxiga ataganlarini berib, kokilini kesdirishgan, qo'y so'yib xudoyi qilishgan edi."

Xulosa qilib aytganda, inson qaysi muhitda, jamiyatda yashasa, kasbi, lavozimi bilan jamiyat hayotida ishtirok etsa, uning nutqiy muomalasi, harakatlari ham shu jarayon bilan uzviy bog‘liq bo‘ladi. Bu shaxsning ijtimoiy maqomi, roli, tilining ijtimoiylashivi hodisasidir.

Said Ahmadning “Qorako‘z Majnun” hikoyasi insonning o‘zligi haqidagi hikoya. O‘zlik – tug‘ilgan ona tuprog‘imiz, ichgan suvimiz , yegan nonimiz bilan bog‘liq. Shu tuproq, non va suvga qarzdorlik haqidagi tushuncha.

“Qorako‘z Majnun” hikoyasiXX asr jahon adabiyoti namunalari bilan bemalol bellasha oladigan asardir.

Foydalanilgan adabiyotlar:

1. Sh. Usmanova, N.Bekmuxamedova, G.Iskandartova Sotsiolingvistika. T.2014.- B.84
2. Said Ahmad. Qorako‘z Majnun. T.2013. -B232.
3. Баҳодир Карим. Руҳият алифбоси. T.2018. –B.364
4. D.H. Quronov. Adabiyot nazariyasi asoslari. T.2020. – B.308
5. Ijtimoiy-siyosiy terminlarning qisqacha izohli lug‘ati. T. 2020. – B.308
6. Ўзбектилинингизоҳлилуғати.T. 2007. 3-жилд. – B.688.

TOHIR MALIKNING “ALVIDO BOLALIK” QISSASIDA QO‘LLANGAN SOTSIOLEKTLARNING SINTAKTIK KONSTRUKSIYALARI VA MA’NOVIY JIHATLARI

Iskandarova Aybibi Raximbayevna
O‘zbekiston Milliy universiteti dotsenti.
E-mail: oybibiiskandarova26@gmail.com

Jabborova Gulrang
O‘zbekiston Milliy Universiteti
O‘zbek filologiyasi fakulteti 2-bosqich talabasi.
E-mail: gulrangjabborova@gmail.com

Annotatsiya: Ushbu maqolada Tohir Malikning "Alvido Bolalik" asarida qo‘llangan argo, jargon kabi sotsiolektlarning til qurilishidagi holati, badiiy asar nutqiy aktlarida qo‘llanish etiketlari haqida fikr yuritilgan.
Kalit so‘zlar: nutqiy akt, argo,jargon, nutqiy etiket, tilning ijtimoiylashuvi, til qurilishi

Аннотация: В данной статье рассматривается состояние таких социолектов, как арго и жаргон, используемые в произведении Тахира Малика «Прощай, детство», а также этикет, применяемый в речевых актах художественного произведения.
Ключевые слова: речевой акт, арго, жаргон, речевой этикет, языковая социализация, языковая конструкция.

Abstract: This article discusses Tahir Malik's “Alvido Bolalik” and the sociolects, or slang, utilized in it. It also discusses the manners employed in the creative work's speech acts.
Keywords: speech act, slang, slang, speech etiquette, language socialization, language construction.

Sotsiolekt termini lingvistika fanida XX asrning ikkinchi yarmidan ishlatila boshlagan. U sotsio, yani jamiyatga xos, dialekt - lahja so‘zlaridantashkil topgan bo‘lib ijtimoiy qatlamlarningnutqiy aktlarida ishlatiladigan, kodlashgan (kod-yashirin) sintaktik konstruksiyalardir. Mazkur til birliklari ijtimoiy chegaralangan qatlam vakillarning kommunikativ ehtiyojlariga xizmat qiladi.

Hozirgi rus sotsiolingvistikasida milliy til doirasidagi biron ijtimoiy guruh: kasbiy, tabaqaviy, yosh va hokazolarga xos bo‘lgan lisoniy xususiyatlar majmuyiga sotsiolekt deyiladi, … askarlar,maktab o‘quvchilari jargoni, jinoyatchilar, xippilar argosi, talabalar slengi, turli sohalarda ishlovchilarning kasbiy “tili” sotsiolektlarga misol bo‘la oladi, degan nazariyqarashlar mavjud.[Беликов В.И, Крысин Л.И. 2021:30]

Nemissotsiolingvistikasidayoshlarningayrim guruhlari tomonidan qo‘llaniladigan, turli sabablar ta’sirida vujudga kelgan maxsus shakllarga jargon deyilgan. H. Lofflerning ta’biricha, bu til makon va zamonda chegara bilmaydi. Har bir avlod o‘zining yoshligida o‘ziga xos u yoki bu shakldan foydalanadi deb fikr bildirilgan. [Sh. Usmanova, N.Bekmuxamedova, G.Iskandarova., 2014:17]

Ko‘p hollarda, argo va jargon, slenglar bir-biriga yaqin atamalar kabi aralashtirilib ishlatilmoqda. Bu ma’lum ma’noda notog‘ridir. Argo va jargon, slenglarning ma’noviy chegaralarini belgilab olish kerak.

A.Nurmonov, A.Sobirov, N. Qosimovalar hammuallifligida yozilgan “Hozirgi o‘zbek adabiy tili” darsligida “Ma’lum ijtimoiy guruh doirasidagina ishlatiladigan so‘zlarga sotsial chegaralangan so‘zlar deyiladi. Masalan, “otarchilar” doirasida qo‘llaniladigan so‘zlar: yakan “pul”, danap “qiz bola”, temiryo‘l kuzatuvchilari doirasida qo‘llaniladigan so‘zlar: qaychi “taftishchi”, kolxoz “chiptasiz yo‘lovchi”. Bunday so‘zlar argolar deb yuritiladi, deyilgan. Guvoh bo‘layotganingizdek,nazariy ma’lumotlarda argolarning sirlilik xususiyatiga to‘xtalinmagan.

Shuningdek, “ayrim tabaqa vakillari tilida faol ishlatiladigan leksemalar ham uchrab turadi, bular jargonlar deyiladi. Masalan, olampanoh, volidayi muhtarama, padari buzrukvor, manzirat qilmoq” [A.Nurmonov, A.Sobirov, N.Qosimova., 2010:63]

Olim H. Jamolxonov “Argotizmlar yashrin ma’no ifodalaydi. Masalan “otarchilar orasida – yakan (“pul”), ichuvchilar orasida – novcha (“araq”), qizil (“vino”) kabi.

Jargonizmlarda yashirin ma’no bilan birga tinglovchini yoki o‘zga bir shaxsni kamsitish yoxud uning xislatlarini bo‘rttirib baholash ottenkalari ham mavjud bo‘ladi. Masalan, talabalar jargonida : dum – “topshirilmagan imtihon”, “qarz”: jonli so‘zlashuvda: yorg‘a-“yengil tabiatli”, “buzuq”shaxs” deb izohlangan.

Argotizm va jargonizmlar, asosan, jonli so‘zlashuvda, tor doirada ishlatiladi. Ulardan badiiy asarda obraz va personajlarni tipiklashtirish maqsadida foydalanish hollari ham uchraydi.” [N.Jamolxonov., 2005:202]H. Jamolxonovning“Hozirgi o‘zbek adabiy tili” darsligida argo va jargon atamalarining farqlari chegaralab ko‘rsatilgan.

Tohir Malikning "Alvido Bolalik"asari qahramonlari yosh o‘smirlar bo‘lgani, ularning aksari bilib-bilmasdan jinoyat yo‘liga kirganligi sababli ular nutqidako‘pginasotsilektlar qo‘llanganligini uchratamiz:

“Bu oq shayton Asrorning tushiga ko‘p kiradi.Ajablanarlisi shuki, bu oq shayton dadasiga o‘xshab ketadi”.Ushbu gapda qo‘llangan “oq shayton” birikmasi odamlarni chirkin, jinoyat yo‘liga boshlaydigan qabih niyatli odamlarga nisbatan ishlatilgan.Asarda biz “qora shayton” birikmasini ham uchratamiz. Bu argo insonni o‘g‘rilik va jinoyat ko‘chasiga yetaklaydigan odamlarga nisbatan qo‘llanilgan. Qora shaytonda yashirinlilik xususiyati saqlanib qolgan. Oq shayton birikmasi bugungi kunda ba’zi shevalarimizda aroqqa nisbatan ishlatilmoqda, ya’ni uning ma’nosi ravshanlashgan.

“Sen maktabda professor bo‘lsang men u yoqda akademik bo‘lib ketganman” deydi Qamariddin.Ushbu gapda qo‘llangan “professor” maktabda a’lo baholarda o‘qiydigan,yaxshi xulqli o‘quvchiga nisbatan ishlatilgan - jargon, akademik esa qamoqxonada jiddiy jinoyat sabab bir necha marta ozodlikdan mahrum etilgan shaxsga nisbatan qo‘llanilgan- argodir. Yozuvchiquyidagi gapda jargon ishlatgan: “Katta qarab-qarab qo‘yyapti.Maqsud Soliyev stakanga vinodan quyib simirdi”,“katta” so‘zi unvon jihatidan o‘zidan yuqori lavozimda ishlovchi boshliq ma’nosida qo‘llangan. Bu so‘z bugungi kunda xalq orasida faol ishlatiladi, yashirinlilik qobig‘idan chiqib, jargonga aylangan.

“Kitobni falon joydan,falon odamdan olib yubor desam tushunib ol”. Ushbu gapda jinoiy to‘da boshlig‘i Qamariddin tilidanishlatilgan “kitob”so‘zi pul ma’nosida ishlatilgan argodir.

“Senga bir bir maslahat beray,anavi Marjaga uylan “,– deydi Qamariddin Asrorga.Bu yerda ishlatilgan “Marja” yoqtirgan qizingga uylan ma’nosida qo‘llangan. Bu so‘zning ma’nosi ochiqligi , tushunarliligi uchun jargon deyishimiz mumkin.

“Alvido, Bolalik” qissasidan olingan quyidagi nutqiy aktga diqqat qilamiz:

– Men grafman eshitganmisan?
– Graflar ko‘p qaysi birisan?
– Kurortdan keldim.
– Kurortdan kelganlar ham ko‘p nima demoqchisan?

Dialogda ishlatilgan“graf” so‘zi yaqinda qamoqdan chiqqan mahbus ma’nosida qo‘llanilgan,“kurort” so‘zi esa ozodlikdan mahrum etish muassasalari - qamoqxonani anglatadi. Graf va kurort so‘zida yashirinlilik, sirlilik holati mavjud bo‘lib, bu til birliklari argoga misol bo‘la oladi. Sotsiolektlar orasida sodda argolar ko‘p uchraydi. Keyingi gapda qo‘llangan argo jinoiy shaxs yoki o‘g‘rilar nutqidan emas oddiy yoshlar tilidan aytilgan: “Men har qanday qorayaloqqa qul bo‘lmayman”. Ushbu gapdagi “qoralayoq” so‘zi oddiy ishchi ma’nosida ishlatilgan.Bu so‘zda ekspressiv-emotsional bo‘yoq mavjud bo‘lib” qorayaloq-oddiy, qora ish qiluvchi inson ma’nosini bildiradi. Bu argo shevalarimizda ishlatiladi.

So‘roq mahalida Mamayusupov “tulki,Imomaliyev esa quyon bo‘ldi”. Ushbu gapdagi “tulki” argosi tinmay gapni aylantiradigan,ayyor shaxsni, “quyon” esa haqiqatni aytishdan qo‘rqadigan, shaxs ma’nosida ishlatiladi.“Tulki”, “quyon” argolarining ma’nosi ravshanlashib, xalq tilida ishlatiladigan polisemantik so‘zga aylangan. Argolarning juda oz qismi, uzoq vaqt davomida so‘zlashuv tiliga o‘tishi mumkin. Argolarning sirliligi, mavhumlik belgisi xalq tiliga tez o‘zlashishiga imkon bermaydi. Qissada “katta baliq” argosi ham ishlatilgan:

“Katta baliq” borligini biladi bular.Katta baliq katta ishlarni mo‘ljal qilgan”.“Katta baliq”argosi jinoyatchilarning boshlig‘i ma’nosida ishlatilgan.

“Qamariddin Zayniddinov degan retsidivistni taniysizmi?”.Ushbu gapdagi retsidivist so‘zi harbiylar nutqida qo‘llanadi va bir necha marta qamalgan,ashaddiy jinoyatchi degan ma’noni anglatgan.

“Gap bunda emas,Keldiyorov, hamma front bo‘ylab hujumga o‘tgan”. Bu gapdagi “front” so‘zi o‘zining noqonuniy ishlarini yashirish maqsadida hamma tanish-bilishlaridan foydalanib ish ko‘rish ma’nosida ishlatilgan.“Yana ilonning dumini bosibsan”. Yuqori lavozimda ishlovchi shaxs bilan ziddiyatga duch kelish ma’nosida qo‘llangan ibora shaklidagi argodir.

“Tomsuvoq chiqib qoldi,bormasak bo‘lmaydi “. Ushbu gapdagi “tomsuvoq” bu nozik odamlar, ya'ni pul qistirilmaydigan tóy yoki ziyofat ma’nosida ishlatilgan.

Xulosa qilib aytganda, sotsiolekt -argo va jargonlardanbadiiy asarda obrazlar mohiyatini ochishda ishlatiladi.

Argo-yashirin ma’nodagi yasama til birligi, shuning uchun xalqqa tez o‘zlashmaydi.

Jargonlarning xalq nutqida ishlatilishi argoga qaraganda faol.

Argo mohiyatan salbiy ma’noni ifodalaydigan sirlilik qobig‘idagi so‘z.

Jargonlarda ijobiy ottenka ifodalaydiganshakllari ham ko‘p.

Argolarko‘proq soddashaklda ifodalanadi.

Jargonlarning sintaktik konstruksiyalarida sodda va birikmatarzida shakllarimavjud.

“Alvido, bolalik” qissasidaishlatilgan argolar adashib jinoyat yo‘liga kirgan yoshlarning ong va tafakkuri oynasi. Kitobxon bunday so‘zlarga baho bera oladi, ularni og‘ir hazm qiladi.

Ulardagi salbiy ottenkani sezadi.Kitobxon bunday so‘zlarni ona tilimizdagi qusurlar ekanligini biladi, albatta, hayot davomida argolarni yaratmaslikka harakat qiladilar.

So‘zda buyuk qudrat mavjud, u odamlarni kamolot pillapoyalari sari boshlaydi.

Foydalanilgan adabiyotlar:

1. Беликов В.И., Крысин Л.И. Социолингвистика. – М., 2001. – С. 30.
2. Sh. Usmanova, N. Bekmuhammedova, G.Iskandarova. Sotsiolingvistika. – T. 2013. – B.89.
3. A. Nurmonov, A.Sobirov, M. Qosimova “Hozirgi o‘zbek adabiy tili”.T. 2010.
4. -B. 276.
5. H. Jamolxonov “Hozirgi o‘zbek adabiy tili” Toshkent -Talqin - 2005, B.270.

ERGONIMLARNI MA'NOVIY VA VAZIFAVIY KLASSIFIKATSIYALASH XUSUSIDA

Kultayeva Fazilat
Termiz davlat universiteti o'qituvchisi.
E-mail: fkultayeva@tersu.uz

Annotatsiya: Maqolada **e**rgonimlar tarkibi semantik nuqtai nazaridan hamda ergonimlarning strukturaviy turlari tahlil qilingan. Barcha obyekt nomlarini ham haqiqiy, ham ramziy nomlarga bo'lish mumkinligi tahlilga tortilgan.
Kalit so'zlar: ergonim, fitonim, mifonim, restora, kafe, bar.

Аннотация: В статье анализируется состав эргонимов с семантической точки зрения и структурные типы эргонимов. Все имена объектов можно разделить на реальные и символические имена.
Ключевые слова: эргоним, фитоним, мифоним, ресторан, кафе, бар.

Abstract: The article analyzes the composition of ergonims from the semantic point of view and the structural types of ergonims. All object names can be divided into real and symbolic names.
Keywords: ergonim, phytonym, mythonim, restaurant, cafe, bar.

U yoki bu nomni idrok etishda murojaatchi ongida paydo bo'ladigan tasvir nomi ko'rsatilgan korxona faoliyatidan ko'proq yoki kamroq mavhumdir. Nominator assotsiativ-majoziy ismni «semantika kengligi» va nomda ishlatiladigan so'z yoki so'zlar bilan bog'langan murojaatchi tomonidan qabul qilingan majoziy obyektiv va subyektiv assotsiatsiyalar asosida tanlaydi. S.M.Mezeninning so'zlariga ko'ra: «so'z (leksema) tasvirga nisbatan eng moslashuvchan birlik bo'lib, denotativ va belgilovchi funksiyalarni birlashtirgan semantikaning kengligi tufayli maksimal obrazli salohiyatga ega» [1, B. 49] Shuningdek, u "obyektga nisbatan tasvirning cheklangan, qo'polligi aks ettirishning umumiy xususiyatidir, lekin faqat uning eng yuqori darajasida - inson idrok etish darajasida - tasvirning cheklanganligi va qo'pollashishi ongli jarayondir". Ergonimlarning nomlari tarkibidagi semantika nuqtai nazaridan tahlil qilingan kafe, bar, restoran, o'quv markazlari, savdo inshootlari nomlarini shartli ravishda quyidagi guruhlarga bo'lish mumkin:[2, B, 13]

1. Nominatsiyasi jonsiz nomlarga asoslangan ergonimlar guruhi: "Diydor" kafesi, "Kayfiyat" kafesi, "Cnos" kafesi, "Yoqut" bar, "Samo" kafesi, "Bog" kafesi, "Milliy taomlar" kafesi, "Orzu" kafe-bar, "Yettinchi osmon" kafesi, "Oqshom" karaoke-bar, "Brilliant" to'y liboslari saloni, "Ideal" restoran-mehmonxona, "Ideal plus" kafesi, "Istiqlol"_to'yxonasi, "O'zbegim" kafe-bar, "Tanho", "Afruz" restoran majmuasi, "Yagona" Disco & bar, "Yulduz" bari, "Fayz" savdo majmuasi, "Surxon" to'yxonasi, "Fayz" bari, "Sohil" bari, "Osiyo" restorani, "Ulfat" kafe-bari, "Kushon" restorani, "Uch o'g'lon" kafesi, "Boss" kafe-bari, "Chaqmoq" kafesi, "Sofiya" kafe-bari, "Do'stlar" kafesi, "Dinamo" sport bari, "Samo" restorani, "Bodom" kafesi, "Shashlik" bari, "Chinor" oshxonasi, "Dengiz marvaridi" oshxonasi, "Arslonbek" oshxonasi, "Avazbek" restorani, "Shodlik" kafe va boshqalar. Bu usul ushbu turdagi nomlar uchun an'anaviy hisoblanadi.[3, B. 12] Shu bilan birga, ta'kidlash joizki, ba'zi nominatsiyalar to'g'ridan-to'g'ri

yoki assotsiativ ravishda kafe, bar, restoranlarga xos bo‘lgan xizmatlar turini ko‘rsatadi: “Milliy taomlar” choyxonasi, “Sho‘rchi fast food” tez ovqatlanish kafesi, “Pizza” tez ovqatlanish kafesi, “MEAT BURGERS”, gril bar “BBQ BaR” tez ovqatlanish restorani, “Yess pizza”, “Sho‘rchi zakaz somsa” oshxonasi, “Tilla baliq” oshxonasi, yetkazib berish bilan “avtosushi” kafe tarmog‘i, “KFC” tez ovqatlanish restorani (Kentucky Fried Chicken — “Kentucky Fried tovuq”) bu esa ergonimning ma’lumot berish funksiyasining muhimligini tasdiqlaydi.

Bu guruhda fitonimlardan tuzilgan ergonimlarni alohida ajratish mumkin: “Olma” kafesi, “Limon” kafesi, “Gul” do‘koni, “Manadarin-s” bari, “Sakura” bari, kafesi, “Pomidor fast” oshxonasi.

Shuningdek, maqom bo‘yicha shaxsning xususiyatlarini o‘z ichiga olgan nomlar bilan ifodalanadigan ergonim-xarakteristikalar: kafe-bar “Uch qirol”, kafe-bar “Lady”, “Monarx” bar-restorani, “Qirollar” restorani, “Korolla” restoran bari, “Bek” kafesi, “Malika” maishiy-texnika do‘koni”, “Sultan” maishiy-texnika do‘koni va boshqalar.

Ergonimlar guruhi nominatsiyasi tegishli nomlarga asoslangan. Bu guruhning asosini antroponimlardan hosil bo‘lgan nomlar tashkil etadi: “Narimon hoji ota“ cafe, “Sohib” qurilish do‘koni, “Ozoda” kafesi, “Sogdiana” kafesi, “Abdulla bobo” muzqaymoqxonasi, “Sohib Aralov” kafesi, “Shaxrizoda kafe-snack” bar, “Tura chapag‘on” restorani, “Kushon” restoran majmuasi, “Boybo‘ta” choyxonasi, “Sherzodbek” oshxonasi, “Botir” oilaviy kafesi, “Xojiyev” oshxonasi, “Bahrom” oshxonasi, “Nigora” to‘y saloni, shou-bar “Dilnoz Gold”, “Dildora” to‘y saloni, “Nargiza-81” qandolat do‘koni, “Anvarbek” kostyum-shim do‘koni, “Mustafoxon” kostyum-shim do‘koni, “Jalol ota” savdo majmuasi, “Nora” to‘y saloni, “Milana” go‘zallik saloni, “Shukrona” restoran kafesi, “Dinara” to‘y saloni Nomlar aholining etnik madaniyatini aks ettiruvchi mifonimlardir. Mifonimlar - afsona va ertaklardagi onomastik makonning har qanday sohasining xayoliy obyektining nomi: “Pandora” club-restoran, “Yoqub” o‘quv markazi, “Humo” savdo markazi, “Iqro” o‘quv markazi Pandora (qadimgi yunoncha Pandora - «hamma narsa bilan in’om etilgan» yoki hamma narsaga qodir), qadimgi yunon mifologiyasida - Zevs buyrug‘i bilan Prometey tomonidan odamlarga olov o‘g‘irlaganliklari uchun jazo sifatida yaratilgan birinchi ayol. Qiziq, u Zevsdan olingan idishni (Pandora qutisi) ochdi, undan barcha baxtsizliklar va ofatlar darhol dunyo bo‘ylab tarqalib ketdi va faqat Umid yopilgan qopqoq ostida qoldi. Bunday nomlar tasodifiy emas, ular hissiy va ruhiy ta’sirga, mijozning ijobiy assotsiatsiyasini uyg‘otishga qaratilgan.

Tirik mavjudotlar va tirik deb qabul qilingan mavjudotlarni deb ataydigan ergonimlarni alohida guruh sifatida ajratib ko‘rsatish kerak: *“Delfin”* mashina ehtiyot qismlari do‘koni, *“Gul”* do‘koni, *“Burgut”* kafesi, *“Lochin”* kafesi;

Noaniq semantikaga ega bo‘lgan eng kichik nomlar guruhi *“Apil-Tapil”* tez oshxonasi, *“C-3”* kiyim-kechak do‘koni.

Tahlil qilingan barcha inshoot nomlarini ham haqiqiy, ham ramziy nomlarga bo‘lish mumkin:[4, B. 667]

1. Haqiqiy ergonimlar, o‘z nomi bilan ushbu obyektning mazmuni yoki faoliyatini bildiradi: “Kabob” choyxonasi, “Tirik baliq” oshxonasi, “Do‘mboq jo‘ja” grill markazi, milliy taomlar oshxonasi, oziq-ovqat do‘koni, “Tabiiy gullar” do‘koni, telefon ustaxonasi, “Kitoblar olami” kitob do‘koni tez ovqatlanish restorani. Haqiqiy motivatsiyaga ega ergonimlar kompaniya xizmatlari haqida eng keng tasavvurni beradi, aniq, iste’molchilarni chalg‘itmaydi, lekin ular ramziy bilan solishtirganda kichik guruh bilan ifodalanadi. Haqiqiy motivatsiyadan farqli o‘laroq,

ramziy motivatsiya korxona faoliyati va mazmuni haqidagi savolga javob bermaydi. U masofadan turib bu haqda tasavvur berishi mumkin:[5, B. 54] “Samo” to‘yxonasi, “Qadam” restorani va boshqalar. U mavhum tabiatga ega bo‘lib, tadbirkorning faoliyat doirasini ko‘rsatmaydi: “Mo‘jiza” tikuv-sexi, “Inmart” ko‘ngilochar savdo markazi, “Afsona” kelin liboslari uyi, “Daymart” savdo majmuasi va boshqalar. Zamonaviy davr xilma-xillik bilan ajralib turadi. Har bir nom alohida ajralib turishini va mijozni hayratda qoldirishini xohlaydi. Zamonaviy til modasining ajoyib xususiyatlaridan biri bu til o‘yinidan foydalanishdir.

Grafikani qasddan buzish, ismning istalgan qismini ajratib ko‘rsatish unga yangi ma’no tus beradi va ko‘pincha qabul qiluvchini qiziqtiradi: “BiroN” oshxonasi, “Chinor” shashliklari.

Shunday qilib, biz tomonidan tahlil qilingan ergonimlarning aksariyati pragmatik xususiyatga ega bo‘lib, bu Surxondaryo viloyati ergonimlariga xos bo‘lgan, mijozni nomga nisbatan ijobiy his-tuyg‘ularni uyg‘otishga qaratilgan umumiy ovqatlanish, o‘quv markazlari, savdo do‘konlari nomlari o‘z-o‘zini reklama qilish elementini ko‘rsatadi. Ergonimlar nominativ(nomlash, atash) funksiyani bajarish uchun tan olingan, ular tarkibiy jihatdan ajralmas birliklardir. Tijorat tashkilotlarining ergonimlarining aksariyati bir so‘zli oddiy shakllar bilan ifodalanadi, ikki so‘zli konstruktsiyalar kamroq tarqalgan. ergonimlarning strukturaviy turlari:

1. Bir asosli (uning asosi bir ildizga to‘g‘ri keladi): “Mo‘jiza” tikuv sexi, “Lochin” kafesi, “Burgut” kafesi, Grand” kafesi “Diydor”kafesi, “Orzu” to‘yxonasi, “Afruz” restorani, «Versal» kafesi, «Qadam” restorani, «Samo» to‘yxonasi, «Fonon» tilla do‘koni, «Malika» maishiy-texnika do‘koni, «Orient» soat do‘koni, «Barbie» to‘y saloni, «Iqro» o‘quv markazi, “Cnos” kafesi, “Marjona” to‘y liboslari, “Ostona” eshiklar do‘koni.

2. Yasamalar (asos va qo‘shimcha mavjud): “O‘qituvchi” kitob do‘koni, “Sevimli” kafe-bar, “Shukrona” kafesi kabilar.

3. Kompleks (asosida ikki yoki undan ortiq ildiz morfemalari mavjud): Technoshop, “Shoxsaroy” to‘yxonasi, kabilar.

4. Tegishli ismlar: “Dildora” to‘y saloni, “Nora” to‘y saloni, “So‘g‘diyona” kafesi, “Bek” kafe-bari, “Ozoda” kafe-bari.

5. Ikki-uch so‘zdan iborat ergonimlar: “Yettinchi osmon” kafe-bari, “MEaT BURGERS” gastrobari, “TEMIR GULLAR” mashina ehtiyot qismlari do‘koni, “Ilm nuri” o‘quv markazi, “Story grand” qurilish mollari do‘koni, “Oltin shaharcha” maktabgacha ta’lim muassasasi.

O‘rganilayotgan ergonimlar, qoida tariqasida, ikki usulda shakllanadi: semantik va to‘g‘ri so‘z yasalishi. Semantik usul ergonim-so‘zlar bilan ifodalanadi:

a) Transonimizatsiya. Bu jarayonida boshlang‘ich onamastik birliklarning faoliyati har xil bo‘ladi. Bunda antroponimlar va toponimlar ko‘proq ishlatiladi, boshqa sinflarning birliklari esa kamroq ishlatiladi. Transonimizatsiya paytida asl nomning shakli o‘zgarishsiz qoladi. Umumiy ovqatlanish muassasalari, asosan, antroponimlar bilan ifodalanadi: “Nikol” restorani, “So‘g‘diyona” kafesi, “Dinara” go‘zallik saloni kabilar. Familiyalar alohida holatlarda qo‘llaniladi: “Xojiyev” restorani, “Mirzayev” oziq-ovqat do‘koni;

b) Transonimizatsiyaning dastlabki nomlari sifatida toponimlar ko‘p emas: “Sho‘rchi” restoran majmuasi, “Denov zakas somsasi”, “Jarqo‘rg‘on tandir go‘shti”.

Surxondaryo viloyatidagi ergonimlarning strukturaviy xususiyatlarini hisobga olib, nominatsiyalarni mulkchilik nuqtai nazaridan quyidagi tarzda tasnifladik:

1) ergonim - ot: “CHinor” restorani, “Diydor” kafesi, “Bek” kafe-bar, “Zefir” va boshqalar;

2) ergonim - sifat: “Pokiza” restorani, “Tanho” restorani;

3) ot + ot: restoran “Dengiz marvaridi”, bar “Rustami Doston” restoran “Sayohatchilar klubi”;

4) sifat + ot: “Shirin hayot” mebellar uyi.

Termiz shahrining ergonimlarini nomlar tomonidan bajariladigan asosiy funksiyalar nuqtai nazaridan tahlil qilaylik:

1) obyektni aniqlashning nominativ funksiyasi (har qanday to‘g‘ri nomning asosiy funksiyasi);

2) informatsion;

3) reklama (tovar va xizmatlarni sotish sohasi uchun mo‘ljallangan nomlarning o‘ziga xos funksiyalari).

Agar ushbu uchta funksiyani amalga oshirish nuqtai nazaridan ergonimlarning yig‘indisini ko‘rib chiqsak, unda ergonomik sohada har xil turdagi nominatsiyalar quyidagicha taqsimlanadi: ergonomik maydonning periferiyasi informatsion bo‘lmagan va bo‘lganlardan iborat. Reklama nomlari: “Chinor” restoran majmuasi, “Yettinchi osmon” kafe-bari, “Lady” kafe-bari, “Fayz” savdo majmuasi, “Siz uchun” restorani, “Bek” kafe-bari, “Dengiz marvaridlari” restorani, “Kushon” qahvaxonasi, “Ali Murtazo” savdo majmuasi va boshqalar. Tegishli nomlardagi ma’lumotlar va ularni taqdim etish usuli tufayli ergonimlar o‘zlarining asosiy maqsadiga - mijozlar e’tiborini jalb qilishga erishadilar. Termiz shahridagi kafelar, barlar, restoran, savdo inshootlari, to‘y salonlarining nomlari ikki turdagi: reklama va informativ ma’lumotlarni o‘z ichiga oladi.

Iste’molchining e’tiborini jalb qilish uchun ergonimdagi oqilona ma’lumotlar eng dolzarb hisoblanadi. Odatda, bu so‘zning keng ma’nosida taqdim etilayotgan xizmat haqida ma’lumot: “Abdulla bobo” osh markazi, “Brilliant” tilla taqinchoqlar do‘koni, “Baraka” oziq-ovqat do‘koni va boshqalarni misol tariqasida ko‘rsatish mumkin. Reklama ma’lumotlari - bu farqlovchi va baholovchi xususiyatning boshqa turi: “Kushon” restoran tipidagi kafe, “Inmart” va “Daymart” savdo markazi kabilar.

Shunday qilib, go‘zallik salonlari nomlarining ishlash xususiyatlarini tahlil qilib, muvaffaqiyatli ergonim yaratish uchun bir nechta muhim shartlarni aniqlash mumkin deb hisoblaymiz:

1. Ergonimlarni hosil qilishda ism qolipli, ma’nosiz va omonim bo‘lmasligi uchun nom yaratish imkoniyatlaridan to‘liq foydalanish kerak. Siz individual nom yaratishga harakat qilishingiz kerak. Omonimiyaning mavjudligi ergonimiyada jiddiy kamchiliklardan biridir.

2. Nom yaratishda adresant o’zbek tili imlo qoidalariga e’tibor berishga majburdir.

3. Nomlarda korxona faoliyati turiga zid bo‘lgan soxta uyushmalar bo‘lmasligi kerak.

4. Chet el lug‘atini haddan tashqari ko‘p ishlatishga berilmaslik kerak.

5. Iloji bo‘lsa, o‘rganilayotgan hududning o‘ziga xos xususiyatlarini hisobga oling.

Shunday qilib, ergonimning samarali ishlashining muhim shartlaridan biri bu nominatorning maqsadi va nomning turli xil psixologik va motivatsion xususiyatlarga ega bo‘lgan potentsiali adresat tomonidan idrok etilishining tengligidir. Iste’molchi uchun ergonimning informativligi ustuvor komponent hisoblanadi. Bu tovar sotishda birinchi navbatda katta rol o‘ynaydigan oqilona ma’lumotdir. Shu sababli, reklama jozibadorligidan ko‘ra ma’lumotlilikni tanlagan nomzodlar muqarrar ravishda g‘alaba qozonishadi. Hozirgi vaqtda informatsion ergonimlar sonining ko‘payishi tendentsiyasi kuzatilayotgani bejiz emas.

Foydalanilgan adabiyotlar ro‘yxati:

1. Мезенин С. М. Образность как лингвистическая категория // Вопросы языкознания. – 1983. – №6. – С. 49
2. Kultayeva, Fazilat. "Ergonimlarning jahon va o ‘zbek tilshunosligida o ‘rganilishi." E Conference Zone. 2022.
3. Kultaeva, Fazilat. "Study of ergonyms in uzbeki linguistics." World Bulletin of Social Sciences 16 (2022): 11-14.
4. Bakhtiyorovna, Kultayeva Fazilat. "Production Characteristics of Ergonomics." American Journal of Language, Literacy and Learning in STEM Education (2993-2769) 1.10 (2023): 666-670.
5. KULTAEVA, Fazilat. "Analysis of ergonomic units at the onomastic level." (2022).

MORFEM LUG‘ATLAR KLASSIFIKATSIYASI VA TABIATI

Boysariyeva Saodat
Termiz davlat universiteti o‘qituvchisi.

Annatatsiya: Maqolada so‘zlarning morfo-sintaktik, morfo-fonomik, morfo-orfepik morfo-orfografik tabiatini o‘zida mujassam etishi bilan boshqa lug‘atlardan farq qiluvchi lug‘at turi—morfem lug‘atlarningo‘ziga xos xarakter xususyiatini haqida so‘z boradi. Shuningdek, morfem lug‘atlar leksikografik xususiyatiga dooir tadqiqottlar tahlilga tortilgan.

Kalit so‘zlar: Morfemika, morfema, o‘zak morfema, affiks, morfem lug‘at, so‘z yasalish uyasi, morfem analiz.

Annotation: The article talks about the unique feature of morpheme dictionaries, a type of dictionary that differs from other dictionaries by incorporating morpho-syntactic, morpho-phonemic, morpho-orphepic and morpho-orthographic features of words. will go. Also, studies on lexicographic features of morpheme dictionaries are analyzed.

Keywords: Morphemics, morpheme, stem morpheme, affix, morpheme dictionary, word formation nest, morpheme analysis.

Аннотация: В статье говорится о специфической характеристике морфемных словарей - типа словаря, который отличается от других словарей воплощением морфосинтаксической, морфонематической, морфо-орфепической и морфо-орфографической природы слов. Также анализируются исследования лексикографических особенностей морфемных словарей.

Ключевые слова: Морфемика, морфема, основа морфемы, аффикс, морфемный словарь, словообразовательное гнездо, морфемный анализ.

Lug‘atshunoslik va morfemika tilning mustaqil sohalari hisoblanadi. Filologik xarakter kasb etgan lug‘atlarning metama’lumotlari tilning qaysi aspektiga doir ma’lumotni taqdim etishini hisobga olgan holda o‘sha sohaning tabiatidan kelib chiqib yaratiladi. Ya’ni sinonim lug‘atlar ma’nodoshlik qatori va ular orasidagi ma’noviy farqlanish va uslubiy xoslanishlarni ochib berishi bilan xarakterlansa, omonimlar lug‘atida shakldosh so‘zlarning ma’noviy, grammatik tafovutlarini ochib berishi bilan ajralib turadi. Xuddi shuning kabi jahon va o‘zbek lug‘atchiligi morfem lug‘at yaratish tajribalariga ega. So‘zning morfem tarkibi haqida ma’lumot beruvchi lug‘atlarning o‘ziga xos tabiati, leksikografik xususiyatlari mavjud.

Til vakilida morfem ma’lumotlar, tilning morfemik modellariga ega ekanlik meta-kognitiv jarayon hisoblanadi. Tilga doir bilimlarni sun’iy intellektga moslash o‘zbek tili kompyuter lingvistikasi oldida turgan muhim masalalardan biri hisoblanadi. Morfem bilimlarni lug‘atlashtirish masalasi shuning uchun ham dolzarb. Morfem tahlil so‘zning lug‘aviy ma’nolari va grammatik kategoriyalari haqida ma’lumot beradi. Lug‘atshunoslar morfem lug‘atlar so‘zlarning morfo-sintaktik, morfo-fonomik, morfo-orfepik morfo-orfografik tabiatini o‘zida mujassam etishi bilan boshqa lug‘atlardan farq qilishini e’tirof etishadi[Abdurakhmonova, N.2022].

Morfemika bo‘limida so‘z yasalishi, o‘zak va affikslarining o‘ziga xos xususiyatlari, tasnifi, vazifalari, joylashish tartibi, bir-birlariga munosabatlari tadqiq etiladi. Morfema til sistemasida tovushdan keyin turuvchi birlik bo‘lib, so‘zning eng kichik ma’noli qismidir. Morfema tushunchasi hamma vaqt tilning bir semantik-morfologik birligini uning ikkinchi bir semantik-

morfologik birligi bilan nisbatlanganda mavjud bo‘ladi. Masalan, gulchi so‘zi ikki ma’noli qismdan iborat bo‘lib, ular *gul va -chi* qismlaridir. Uning birinchi qismi alohida olinganda so‘zdir. Biroq bu so‘zga *-chi* qo‘shimchasi qo‘shilar va undan yangi so‘z hosil qilar ekan (gulchi), birinchi qism ikkinchi qismga nisbatan alohida emas, so‘zning qismi deb baholanadi. Shunga ko‘ra -chi elementi ham so‘zning qismi sifatida olinadi va u oldingi qismga nisbatan baholanadi – biri ikkinchisiga nisbatan alohida morfema hisoblanadi: *o‘zak morfema; affiks morfema*. Morfemikaga shu nuqtayi nazardan yondashiladigan bo‘lsa, u morfemalar (so‘zning ma’noli qismlari) haqidagi ta’limotdir. Morfemaning aynan sematik-morfologik tabiati mashinani o‘qittirish (machine learning) uchun muhim hisoblanadi. So‘zlarning morfem tahlili, odatda, tahlil qilinayotgan so‘zdagi bo‘linuvchi har bir qismning o‘zbek tilidagi boshqa so‘zlarda uchraydigan tilda tayyor holda mavjud bo‘lgan xuddi shunday elementlar (qismlar) bilan qiyoslanishi orqali amalga oshirilishi lozim. O‘zak tushunchasi morfemalarga ajraluvchi so‘zlardagi morfemalarning xarakteriga ko‘ra belgilanadi. So‘zlarda nechta morfema bo‘lishidan qat’i nazar ularning mustaqil holda leksik ma’no ifodalay olishi yoki ifodalay olmasligi bir-biriga zid qo‘yiladi. So‘zdagi ma’lum bir morfema mustaqil holda leksik ma’no ifodalaydi, boshqa morfemalar ana shu xususiyatga ega emas. Ana shu zidlik o‘zak morfema (o‘zak) va affiks morfema tushunchalarining vujudga kelishi uchun asos bo‘ladi:

a) semantik jihatdan ular bir-biriga zid bo‘ladi;

b) grammatik jihatdan affiks morfemalar odatda (turkiy tillarda, jumladan o‘zbek tilida ham) o‘zakdan keyin joylashadi, unga tobe bo‘ladi.

Demak, o‘zak morfema va affiks tushunchalari faqat morfem analiz — so‘zlarni morfemalarga ajratish jarayoni uchungina xosdir. Bunda affiks morfemalarning turlari, boshqa xususiyatlari hisobga olinmaydi. Affiks morfemaning qanday bo‘lishi (so‘z yasovchi yoki forma yasovchi ekanligi) ahamiyatsizdir. Bu o‘rinda ahamiyat kasb etadigan narsa u yoki bu so‘zning ma’noli qismlardan iborat ekanligi, o‘sha ma’noli qismlarni to‘g‘ri belgilash, turlarini to‘g‘ri baholash bilan cheklanadi.

Lug‘atning asosiy maqsadi ma’lumot saqlash bo‘lganidek, so‘zlarning morfem xarakterini aks ettiruvchi lug‘atlarga ehtiyoj mavjud. Odatda muayyan turdagi lug‘atlar quyidagilardan iborat bo‘lishi mumkin: 1) bosh so‘z qismi; 2) metamatn qismi; 3) illyustratsiya[Abdurakhmonova, N.2022].

Morfem lug‘atlarda illyustratsiya, tavsif va izohlar bo‘lmaydi. Har bir lug‘atning o‘ziga xos strukturasi mavjud. Foydalanish shartlarida lug‘atdan foydalanish yo‘riqnomasi keltiriladi. So‘z yasalish lug‘atlari derivatsion uya munosabatiga asoslanib tuzilgan bo‘ladi.

Rus leksikografiyasida morfem lug‘atlarning turli tipik ko‘rinishlari mavjud: morfemalarning sintagmatik lug‘ati, morfemalarning izohli lug‘ati, morfemalarning chastotali lug‘ati, rus tili so‘z yasalish lug‘ati, rus tilining derivativ-semantik lug‘ati kabi namunalarini uchratishimiz mumkin. A.N.Tixinov so‘z yasalish lug‘atlarini tuzish tamoyillari hamda so‘z yasalish uyalarining nazariy asoslarini ishlab chiqishda katta hissa qo‘shgan. Olim tomonidan yaratilgan quyidagi lug‘atlar rus tili morfem lug‘atlari ichida o‘z o‘rniga ega [Тихонов, 2007; 700].

1. Katta so‘z yasalish lug‘ati (Большой словообразовательный словарь) (1990)
2. Morfem-imlolug‘ati (Морфемно-орфографический словарь) (2002)
3. Rustiliningyangiso‘ztarkibilug‘ati (Новый словообразовательный словарь русского языка) (2014)

A.Kuznetsova, T.Efrimovalarning “Словарь морфем русского языка” so‘z yasovchi lug‘atdagi maqolalar izohli, etimologik va boshqa turdagi lug‘atlardagi maqolalardan farq qiladi. So‘zni tarkibiga ko‘ra tahlil qilish lingvistik tahlil turlaridan biri bo‘lib, uning maqsadi so‘zning tarkibi yoki tuzilishini aniqlashdan iborat. Lug‘at rus tilining 3000 dan ortiq keng tarqalgan so‘zlarini o‘z ichiga oladi, so‘zlar chastotali lug‘atlarga muvofiq tanlanganini guvohi bo‘lishimiz mumkin. Lug‘at maqolasini berishda quyidagi to‘rt jihat hisobga olingan [Кузнецова, Ефремова, 1986;1136]:

1. So‘z urg‘usi (aksentologik jihati);
2. So‘z yasalishi (hosilaviy jihati);
3. So‘z tarkibi (barcha morfemalar);
4. Orfoepik-orfografik jihati.

Morfologik-aksentologik komponent har qanday so‘zning to‘liq paradigmasini (barcha mumkin bo‘lgan so‘z shakllarini) yaratadi, har bir so‘z shakliga urg‘u qo‘yiladi, asosiy morfologik xususiyatlar keltiriladi hamda fleksiya xususiyatlari sharh bilan birga beriladi. Shuningdek, so‘z shakllarining maqbul variantlari taklif qilinadi.

Rus derivativ leksikografiyasi boy tarixga ega. Tilning morfemik sathini lug‘atlashtirish bo‘yicha E.Zaykova, N.Golyev, Doronina monografik planda bajarilgan ishlarni hamda A.Tixinov, A.Potixa, A.Kuznetsva, T.Efromova kabi mualliflar tomonidan nashr etilgan bir qancha lug‘atlarni ko‘rishimiz mumkin.

V.D.Tolmacheva, S.I.Kokorina muallifligidagi “Учебный словарь глаголных форм русского языка” deb nomlangan kitobda 5 mingga yaqin rus tilidagi fe’llar shakllari ma’lum bir model asosida alifbo tartibida shakllantirilgan [Толмачева, Кокорина.,1988; 333]

Rus tiliga doir lug‘atlarda morfem tahlilning berilishi orfografiyaga oid ko‘nikmalarni ham shakllantirishda muhum hisoblanadi. Shuning uchun ham rus tilida shu vaqtga qadar 59 ta morfem hamda derivativ lug‘atlar nashr etilgan. Shular orasida eng mashhurlari A.Kuznestva va T.Efrimovalarning “Словарьморфемрусскогоязыка” lug‘atidir. O‘zida 52 000 so‘zni jamlagan lug‘atda morfemik tahlil tamoyillari hamda lug‘at tuzish qonun-qoidalari keltirib o‘tilgan[Кузнецова, Ефремова, 1986;1136]. Lug‘atda quyidagicha morfem tahlilni ko‘rish mumkin:

АВТОР
а'втор
а'втор-***ств***-о
а'втор-***ск***(ий)
а'вторск-***ие***,*сущ.*
со-а'втор
соа'втор-***ств***-о
соа'втор-***ск***-ий
автор-***изова'***-ть
авториз-***а'ци́й (й)*** -а,1[авторизация]
авторизо'ва-***нн***-ый,*прич.-прил.*
автор-***изи'рова***-ть
авторизи'ровать-***ся***
авториз-***а'ци́й (й)*** -а,2
авторизи'рова-***нн***-ый,*прич.-прил.*

A.Shirshovning “Rus tilining so‘z yasalish izohli lug‘ati”da rus tilidagi 37 mingga yaqin so‘z kiritilgan. Ushbu so‘zlardan 2 mingdan ortiq so‘z yasalish uyalari shakllantirilgan [Ширшов, 2004; 1024] . Rus tilidagi so‘z tarkibi lug‘atlarida alifbo tartibi urfdan qolganiligining deyarli barchasi uyalashtirish tamoyiliga asoslanganini guvohi bo‘lishimiz mumkin. So‘zlarning o‘z

tarkibiy morfemalariga bo‘linishini, so‘zning so‘z yasalish tarkibini, shuningdek, berilgan morfema – ildiz yoki affiks bilan so‘zlar to‘plamini (so‘z yasalish uyasini) ko‘rsatuvchi lug‘atlar hosila lug‘atlardir. So‘z yasovchi lug‘atlardagi so‘zlar morfemalarga bo‘linib, urg‘u bilan beriladi. Shunday qilib mavjud morfem lug‘atlarni tahlil qilib bunday turdagi lug‘atlarning asosiy to‘rt turi mavjudligi ta’kidlamoqchimiz:

1. *O‘zak lug‘atlar* (bunday lug‘atlarning birliklari o‘zak morfemalardir; so‘zlar alifbo tartibida turkum so‘zlarning so‘z yasalish munosabatlari ko‘rsatilmagan holda keltirilgan);
2. *So‘zlarning morfem jihatdan segmentatsiyalangan lug‘atlar* (bunday lug'atlarning vazifasi nafaqat har bir so‘zning morfemik tarkibini ko‘rsatish, balki uning so‘z yasalish tarkibini ham ochib berishdir);
3. *Affiks morfemalarining izohli lug‘atlari* (bunday lug‘atlarda affikslarning ma’nosi, amal qilish xususiyatlari ochib beriladi);
4. *Chastotali so‘z yasovchi lug‘atlar* (morfemalar kamayish chastotasiga ko‘ra tartibga solinadi).

Foydalanilgan adabiyotlar:

1. А.Н. Тихонов Морфемно-орфографический словарь русского языка: около 100 000 слов /. – М. [АСТ]: Астрель, 2007. - 700 с.
2. Кузнецова А.И., Ефремова Т.Ф. Словарь морфем русского языка: Ок. 52000 слов. - М.: Рус. яз., 1986.- 1136 с. (5000 морфем (корней, префиксов, суффиксов))
3. Зайкова Е.С. Русская деривационная лексикография: история, теория, практика: автореф. дис. … канд. филол. наук. – Екатеринбург, 2008. – 23 с.
4. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In *2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)* (pp. 1790-1793). IEEE.
5. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
6. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
7. Голев Н.Д., Доронина Н.И. Деривационно-ассоциативный словарь русского языка // Проблемы коммуникаци и номинаци в концепции общегуманитарного знания: сб. науч. тр. в честь профессора Л.А. Шкатовой. – Челябинск, 1999. – С. 276–282
8. Потиха З.А. Школьный словообразовательный словарь / 1964
9. Кузнецова А.И., Ефремова Т.Ф. Словарь морфем русского языка: Ок. 52000 слов. - М.: Рус. яз., 1986.- 1136 с.
10. Ефремова Т.Ф. Толковый словарь словообразовательных единиц русского языка. - М.: Русский язык, 1996.
11. Толмачева В. Д., Кокорина С. И. Учебный словарь глагольных форм русского языка. – 1988 — 333 с.

O‘ZBEK TILIDA REKLAMALARNING SINTAKTIK STRUKTURASI

Botirova Maxliyo Baxodir qizi
Termiz davlat universiteti o‘qituvchisi.
E-mail: mbotirova@tersu.uz

Annotatsiya: Maqolada reklama matnlarining sintaktik qurilishi haqida so‘z yuritiladi. Reklama matnining asosiy vazifasi nisbatan qisqa xabar orqali ma’lum bir tovar, mahsulot, narsa yoki hodisaga e’tiborni jalb qilishdir. Bu vazifani bajarishga atov gaplarning qisqaligi va ma’nodorligi mos keladi, bu esa ularning reklama matnlarida tez-tez ishlatilishiga olib keladi.

Kalit so‘zlar: reklama matnlari, sintaktik sturktura, reklamalarning gap qurilishi, atov gaplar.

Аннотация: В статье рассматривается синтаксическое построение рекламных текстов. Основная задача рекламного текста – привлечь внимание к конкретному товару, товару, вещи или событию посредством сравнительно короткого сообщения. Краткость и содержательность именных словосочетаний подходят для выполнения этой задачи, что приводит к их частому использованию в рекламных текстах.

Ключевые слова: рекламные тексты, синтаксическая структура, структура предложений рекламных объявлений, именные словосочетания.

Annotation: The article deals with the syntactic construction of advertising texts. The main task of the advertising text is to attract attention to a specific product, product, thing or event through a relatively short message. The brevity and meaningfulness of noun phrases are suitable for fulfilling this task, which leads to their frequent use in advertising texts.

Keywords: advertising texts, syntactic structure, sentence structure of advertisements, noun phrases.

Reklama matnlarining leksik-semantik, axborotni jamlash xususiyatlari, pragmatik va ijtimoiy hayotdagi o‘rni bilan bog‘liq jihatlari nafaqat tilshunoslarni balki, tarixchi, madaniyatshunos va jamiyatshunoslarning ham diqqat markazidagi masalalardan biri hisoblanadi. Til esa o‘z kommunikativ vazifasini gap vositasida yuzaga chiqaradi. Tildagi barcha fonetik, leksik, morfologik hodisalar, qolaversa, nolisoniy vositalar gapning to‘g‘ri shakllanishi uchun xizmat qiladi. Ammo sanalganlar orasida leksikologiya va morfologiyaning gapni grammatik shakllantirishdagi ishtiroki o‘ziga xos. Gap barcha tillarda, jumladan, o‘zbek tilida ham grammatikaning asosiy tarkibiy qismlaridan biri bo‘lgan sintaksis bo‘limida tadqiq qilinadi. Gap turlari va ularning sintaktik qurilishi, matndagi ahamiyatini o‘rganish har doim tilshunoslarning diqqat markazida bo‘lib kelgan. Jumladan, reklama matnlaridagi gap qurilishi ham ayrim o‘ziga xosliklarini bilan ajralib turadi. Iste’molchini ishontirish, uni hakatga keltirish uchun reklama beruvchi nafaqat tanlangan so‘zlarning leksik-semantik, balki ularning matndagi xususiyatlarini, sintaktik qurilmadagi o‘zgarishlarini, demakki, sintaktik o‘ziga xosliklarini ham bilishi talab etiladi.[1, B. 108] Shu ma’noda reklama matnidagi sintaktik vositalarni tavsiflash, ularning lingvistik-uslubiy xususiyatlari va bu xususiyatlarning iste’molchiga ta’sirini tadqiq qilish reclama matnlarini boyitish va ularning turli-tumanligini ta’minlashda muhim ahamiyatga ega. Reklama

beruvchi tomonidan tanlangan har bir gap, uning qurilishi muayyan bir maqsadga: iste'molchining diqqatini jalb qilish, uni reklama qilininayotgan tovarni sotib olish lozimligiga ishontirish, reclama obyektiga munosabat va qiziqishni uyg'otish kabilarga xizmat qiladi.[2, B. 3] Reklama nazariyasi bilan shug'ullangan tadqiqotchilarning ta'kidlashicha: "Reklama matni sintaksisi qisqalik, ko'tarinkilik, his-hayajonlilik, asosiysi, huquqiyligi bilan ajralib turadi".[3, B. 62] Reklama matnlarida asosan sodda gaplardan, ularning egasiz turlaridan, kiritmalardan, undovlardan, his-hayajonli gaplardan, buyruq, so'roq va darak gaplardan foydalaniladi.

Reklama matnining sintaktik mukammalligi undan ko'zlangan maqsadga erishishda katta ahamiyatga ega. Reklama matnlari sintaksisi uchun soda gaplarni qo'llash foydaliroq. Chunki aynan sodda gaplar reklama matnidan ko'zlangan maqsad – tinglovchining tuyg'ulari, dunyoqarashi va tanloviga ta'sir qilishga erishishda muhim ahamiyatga ega. Atov gaplar reklama matnini kichik qismlarga bo'lishda qo'l kelsa-da, katta hajmdagi ma'lumotlarni ifodlash xususiyatiga ega ekanligi bilan ham ajralib turadi. Atov gaplar mahsulot yoki uni ishlab chiqaruvchi korxonaning nomini ifodalash bilan birga ichki dinamikaga egaligi, alohida ohangni talab qilishi kabi xususiyatlari bilan matnning tez o'zlashtirilishiga xizmat qiladi. Masalan: Oqtepa lavash. Mazali gaplar joyi. To'liqsiz, murakkablashgan sodda gaplar reklama matnlarida nafaqat ma'lumot yetkazish, balki foydalanuvchi diqqatini jalb qilish va shu orqali ularni harakatga kelishga undash vositasi ham. Qo'shma gap reklama matnida batafsil ma'lumot talab qilinadigan o'rinda, shuningdek, boshqa uslubiy vositalar talabi bilan qo'llaniladi.

Reklamalarda ekspressivlikni oshiruvchi, diqqatni jalb qilishda ko'maklashuvchi muhim vosita sifatida undalmalardan ham foydalaniladi. "Undalmalar so'zlovchning nutqi qaratilgan shaxs, jonli, jonsiz narsa-buyumlarni ifodalashiga asosan muhim uslubiy vosita sanaladi. Undalmalar so'zlovchining tinglovchiga munosabatini ifodalab baho xarakteristikasini ottenkasiga ega bo'ladi".[4, B. 250]Nestle decoration, mayinlikning shirin ta'mi; Tomoq og'rig'i, tomoshang tugadi. Septolete total.

Reklama matnlaridagi gap turlari haqida so'z borganda har xil gap, turli sintaktik konstruksiya nutqda turli xil uslubiy vazifalarni bajarishini unutmaslik kerak. O'zbek tilidagi reklama matnlarida sodda gaplar keng qo'llanilishi bilan birga oz bo'sa-da, qo'shma gaplardan ham o'rni bilan foydalaniladi. Qo'shma gaplarning kam qo'llanilishi o'zbek tilida nutqiy jarayonning lakonizm (yun. lakonismos – qisqalik – fikrni bayon qilishdagi qisqalik va aniqlik [5, B. 101-104]) ga asoslanganligi bilan izohlanadi. Masalan, 100 m³ beton sotib oling, 3 soatlik bepul avtobetonnasos xizmatiga ega bo'ling! Fanta iching, quvonchga to'ling!Reklama matnlari o'zbek tilidagi kam uchraydigan qo'shma gap shaklidagi reklamalardir. Ma'lumki, o'zbek tili gap qurilishida eganing ishtirok etmasligi me'yor hisoblanadi. Chunki egaga ishora qilib turuvchi morfologik vositalar (shaxs-son shakllari) kesimda mavjud bo'lishi bunga imkoniyat yaratadi.

Reklama matnining amaliy ahamiyati shundaki, uning tarkibida qo'llangan har bir so'z reklama qilinayotgan mahsulotga yoki xizmat turiga xaridorlarda qiziqish uyg'otishdek yukni o'z zimmasiga oladi. Bu adresatga qiziqayotgan obyekti bo'yicha aniq konseptual tushuncha olishiga imkon yaratadi.[5, B. 20] Shu munosabat bilan har xil leksik-sintaktik va grafik vositalardan foydalanishga asoslangan reklama imidjini shakllantirish muhim hisoblanadi. Reklama o'zining tashqi shakli bilan uni qabul qiluvchi (adresat)ning ongida taqdim etilayotgan mahsulot yoki xizmat turi to'g'risida ma'lum bir g'oyani shakllantiradi va harakatga javob qaytaradi. Reklama matniga qo'yiladigan asosiy talab ixchamlik va aniqlikdir, shu bilan bir qatorda uning matni ekspressivlik bo'yog'iga ega so'zlardan tuzilishi kerak. Bu ekspressivlik reklama matnidagi

adresatga yetkazilishi ko‘zlangan ifoda maqsadi turlicha shakldagi gaplardan tashkil topadi. Masalan, xabar mazmunidagi darak gap shaklida: Viferon-Idaferon va vitaminlar kompleksi –virusni bloklashga va immunitetni tiklashga yordam beradi. Viferon –shamollash va gripga qarshi komleks himoya. (Dori reklamasi)

So‘roq mazmunda: Toshkent markazidan uy olmoqchimisiz? Buning uchun nima qilish zarur? (Turar-joy reklamasi)

Buyruq mazmunda: Ta’mlar jangida ishtirok et! O‘z ta’mingni tanla! Pepsini tanla!!! (Pepsi reklamasi). Mazkur ichimlik reklamasida buyruq gapning kesimi ikkinchi shaxs ko‘plikda emas, balki birlikda qo‘llanishi reklama yaratuvchilarining va mahsulot ishlab chiqaruvchilarning o‘zlarini iste’molchilarga juda yaqin tutayotganligini ifodalashga xizmat qilgan.

Istak mazmunda: Katta bo‘lganimda oshpaz bo‘laman. Oyimlardan ham mazaliroq pishiraman. («Pikovit» vitamini reklamasi).

Xulosa qilib aytadigan bo‘lsak, o‘zbek tilidagi reklama matnlarining asosiy qismini sodda gaplar tashkil etadi. Sodda gaplar ma’lumotlarning tez yetkazilishi, diqqatni oson jalb qilishi bilan ajralib turadi. Darak gaplarga ko‘proq murojaat qilinsa-da, o‘rni bilan so‘roq gaplardan ham foydalaniladi. So‘roq gaplarning har ikki turi ham faol qo‘llaniladi. Ritorik so‘roq gaplar jarangdorlikni oshirishda qo‘l kelsa, sof so‘roq gaplar tafsilotlarni aniqlashtirish lozim bo‘lgan o‘rinlarda muhim.

Foydalanilgan adabiyotlar ro‘yxati:

1. Чиниқулов Н.Ж. Пешлавҳаларнинг лисоний хусусиятлари. Монографий. – Тошкент: Навоий университетинашриёт матбаа уйи, 2019. –108 б.
2. Ботирова М. O ‘ZBEK TILSHUNOSLIGIDA REKLAMA MATNLARI TADQIQIGA DOIR ISHLAR TAHLILI //МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА. – 2023. – Т. 6. – №. 3.
3. Chabanyuk T.A. Teoriya I praktika reklami. Komsomolsk-na-Amure: KiAGTU. 2013. – S. 62
4. Olyanich A.V. Reklamnix diskurs I ego konstituvniye priznaki // Reklamnix diskurs I reklamnix tekst kollektivnaya monografiya / Nauch. Red. T.N.Kolokoltseva. – M Flinga, 2011. – S.250
5. Matyoqubova D. Undalmalar ruhiyatni ifodalovchi vosita sifatida // cyberleninka.ru, researchfocus, volume 1. issue 1, 2022. – B. 101-104
6. Мамирова Д. Ўзбек реклама матнларининг социолингвистик тадқиқи: Автореф.дис. ... докт. филос. филол. наук. – Самарканд, 2021.

TARIXIY DIGNITONIM VA FALERONIMLARNING LISONIY TADQIQI

Xusanova Maksuda Nishonovna
Renessans ta'lim universiteti tadqiqotchisi.

Annotatsiya: Maqolada tarixiy merosimizning bir bo'lagi bo'lgan dignitonim va faleronimlar o'z davrda faol qo'llanib, bugungi kunda arxaizmga aylanganini hamda ularning lisoniy tadqiqi o'rganilgan**.**

Kalit so'zlar: orden, medal, unvon, madaniy boylik, imtiyoz, rag'batlantirish, bitiktosh**.**

Аннотация: В статье дигнитонимы и фалеронимы, являющиеся частью нашего исторического наследия, активно использовались в свое время и сегодня стали архаизмами, а также изучается их лингвистическое исследование.

Ключевые слова: орден, медаль, звание, культурные ценности, льгота, поощрение, надпись.

Annotation: In the article, dignitonyms and faleronyms, which are a part of our heritage, were actively used in their time, archaism today, and document linguistic research.

Keywords: order, medal, title, cultural property, benefit, rag treatment, epitaph.

Milliy tilimizning bebaho boyligi bo'lgan onomastik so'zlarni, xususan, dignitonim (davlat va xalqaro tashkilotlar tomonidan ta'sis qilingan unvonlar)lar va faleronim (orden,medal) nomlarining shakllanishi va ma'nolarining talqini hamda ilmiy onomastik tahliliga jiddiy e'tibor qaratish lozimki, ular ham tarixiy merosimizning bir bo'lagidir. Zero, "Tarixiy merosni asrab-avaylash, o'rganish va avlodlarga qoldirish, davlatimiz siyosatining eng muhim ustuvor yo'nalishlaridan biridir".[1.Мирзиёев Ш.М. 2017.29]

Har qanday xalq tarixida muayan tarixiy qimmatga ega bo'lgan unvon (faxriy, diniy va ilmiy unvon), orden, medal va mukofotlarning nomlari)lar ta'sis etilgan va ular ma'lum davrda ijtimoiy ehtiyoj tufayli faol qo'llangan hamda bugungi kunda ham ishlatilmoqda. Unvon (faxriy, diniy va ilmiy unvon), orden, medal va mukofotlarning nomlariga mavjud voqelik, ijtimoiy talablar doirasida narsalar, voqea-hodisalarning mazmunidan kelib chiqib maxsus nomlar beriladi. Mana shu nomlar ya'ni unvon , mansablar onomastikada dignitonim(davlat va xalqaro tashkilotlar tomonidan ta'sis qilingan unvonlar) va medal ,orden, mukofot, ko'krak nishonlari faleronimlar deb yuritiladi. Demak, insonning aqliy, g'oyaviy, badiiy madaniy, ma'rifiy, ijtimoiy-siyosiy faoliyatini rag'batlantirish maqsadida shakllangan hamda ta'sis etilgan unvon, orden, medal va mukofotlarning nomlari atoqli otlardir. Mana shuning uchun ham unvon(faxriy, diniy va ilmiy unvon), orden, medal va mukofotlarning nomlari sanaladigan dignitonim va faleronimlarning tarixiy-qiyosiy tadqiqiga zaruriyat sezilmoqda.

Tarixiy va zamonaviy mazmundagi unvon, orden, medal hamda mukofotlarning nomlarida milliy va umumiy insoniy qadriyatu an'analar va urf-odatlar o'z ifodasini topgan. Shu qadimiy an'analar, urf-odatlar dignitonim va faleronimlarning mazmun-mohiyatida bo'lgani uchun nomi xalqning ma'naviy meros, madaniy boylik, tilning o'ziga xosligini ma'lum darajada saqlagan qadimiy xususiyatlarini saqlab qolgan tarixiy yodgorliklardir .

Biz badiiy asarlarimiz orqali xalq tomonidan e'tirof etilgan quyidagi tarixiy dignitonimlarni ko'p uchratish mumkin: aqli, dono va bilimli kishi, shoirlarning odam otasi bo'lgan Rudakiy "*Odam ush-shuaro*", XII asrning Yusuf degan so'z san'atkori "*Xos Hojib*" (eshik og'asi), Amir Temur "*Sohibqiron*", Ahmad Yassaviy "*Shayx ul mashoyix*", "*Sulton ul orifin*" (oriflar sultoni), "*Karomatlar xazinasi*", Husayn Bayqaro esa "*Karim ul tarafayin*", Alisher Navoiy esa "*Nizomiddin Mir Alisher*" va "G'azal mulkining sultoni", "*Manzuri shohonshohi Xuroson*" kabi unvonlari.

Shad — dignitonimlar yaratilishining ilk bosqichiga V-VII asrlarda bitilgan O'rxun — Enasoy obidalari borib taqaladi. Bu yodgorliklar qadimgi turk tili va yozuvi bilan tarixiydir.

To'nyuquq bitiktoshda ,sarkarda , maslahatchi To'nyuquq "shad" unvoni olganini ta'kidlab o'tadi. Yetti yuz kishini birlashtirgani boshliqqa "shad" unvonini beradi. Shuningdek,ushbu bitigtoshlarda bag- bek, at- unvon, yuz basi-yuz boshi, bas –boshliq, rahbar, qan- xoqon, apatarqan –bosh qo'mondon kabi ma'nolarda unvon va mansablar qo'llangan .

Xotun -zodagon ayol. Mazkur so'z qadimgi turkiy tillarda, xususan, Enasoy bitiktoshlarida ham qo'llanib, "xonim", " obro'li kishining xotini ma'nosini anglatadi [2.Малов С.Е.1952.442]." At-tufha" da bu so'z "xonim", "xonim afandi" tarzda berilgan[3. O'rozboyev A.2021, 73] .Tilning taraqqiyoti natijasida bu so'z ayol kishi , uy bekasi , erkak kishining (jufti) ma'nolarini bildirgan. Qotin so'zi fonetik o'zgarishlar oqibatida qotin so'zi "xotun" tarzda aytilgan va yozilgan. Qatun/xatun so'zlari "malika","ayol, xotin kishi " ma'nolarida keng qo'llangan. Mirzo Ulug'bekning "To'rt ulus tarixi "asarida "malika " ma'nosida ishlatilgan. Turkon xotun,vazir Nasiriddin va bu yerda bo'lgan har bir kishi pastga tushishga majbur bo'ldi.[4.Мирзо Улуғбек.1994.154]

„Shayx ar-Rais" — Abu Ali ibn Sino buyuk qomusiy olim, tabib va faylasuf bo'lib, 980-yili Buxoro yaqinida Afshona qishlog'ida tavallud topgan. Xalq tomonidan „Shayx ar-Rais" („donishmandlar sardori, allomalar raisi") unvonini beriladi. Islomiy davlatlarda olimlar, diniy faqihlarga berilgan unvon. Keyinchalik taniqli so'fiylar , pirlarga nisbatan ham ishlatilgan. G'arbda esa Avitsenna nomi bilan mashhur.

Inson borki, qilgan kashfiyotlari, mehnatlarini, ixtirolarini omma oldiga rag'batlantirishini istaydi. U bir zamonni, davrni o'z faleronimlari mavjud. Davr o'zgarishi bilan eskirib, yangi faleronimlar paydo bo'ladi. Inqilobdan keyin insonlarni turli jabhalardagi faoliyatini rag'batlantirish maqsadida quyidagi orden va medallar tasis etilgan: "Mehnat Qizil bayroq " ordeni , "Xalqlar do'stligi" ordeni , "Qahramon ona" , "Qizil yulduz" ordeni, "Sharafli mehnatlari uchun" , "Onalik shuhrati" medali, "Besh yillik zarbdori " medali kabilar. "Qahramon ona" ordeni sobiq SSSRda 1944-yilda, demografik inqirozning oldini olish maqsadida ta'sis etilgan bo'lib, 10 yoki undan ko'proq bolani dunyoga keltirgan onalar o'ninchi farzandi 1 yoshga to'lgan va qolgan bolalari hayot bo'lgan taqdirda mazkur orden bilan taqdirlangan. Mukofot egalariga qator imtiyozlar joriy etilgan.

Xulosa qilib aytganda, har bir davrda insonlarni rag'batlantirish ehtiyoj sezilgan.Dignitonim va faleronimlar esa davr o'zgarishi bilan o'zgarib borgan, ya'ni eskisini o'rniga yangisi yaratilgan.

Foydalangan adabiyotlar ro'yxati:

1. Мирзиёев Ш.М. Миллий тараққиёт йўлимизни қатъият билан давом эттириб, янги босқичга кўтарамиз. IIжилд. – Тошкент: Ўзбекистон, 2017. – Б.29

2. Малов С.Е.Энисейская писменност тюрков.-М.-Л., 1952.-С.442
3. O‘rozboyev A.Ogahiy tarixiy asarlari leksikasi.-Toshkent.2021 - 73 b
4. Мирзо Улуғбек.Тўрт улус тарихи.-Тошкент.1994.Б-15.

FARIDA AFRO‘Z SHE’RIYATINING LINGVOPOETIK TADQIQI VA O‘ZIGA XOS JIHATLARI

Topildiyeva Feruza Raximjonovna
O‘zMU, katta o‘qituvchi.

Annotatsiya: Ushbu maqolada lingvopoetika, lingvistik stilistika, stilistikaning badiiy uslubi va shoira Farida Afro‘z she’riyatining lingvopoetik tadqiqi, o‘ziga xos tomonlari va jihatlari, shoiraning so‘z tanlash va so‘z qo‘llash mahorati,misralarda inju marjondek tizilgan so‘zlarning o‘rni, sodda so‘zlardan, zid ma’noli so‘zlarning antiteza hosil qilishi, ma’nodosh so‘zlarning va o‘ziga xos she’riy tashbehlarida leksik birliklarning uslubiy obrazli vazifalari xususida fikr yuritilgan va unda lirik hislar erkin mushohadasi tahlili berilgan, uning she’rlaridagi so‘zlar vositasida poetik tasvir yaratilganligi, unda har bir so‘z boshqa so‘zlarning poetik faollashuviga xizmat qiluvchi fidokorlikni his etishga chorlovchi vosita ekanligi namunalar misolida izohlangan. Unda hech eskirmas mavzuning Afro‘zona go‘zal bir talqini berilganligi ta’kidlangan.

Kalit so‘zlar: lingvopoetik tadqiq, stilistika, antiteza, obrazli ifodalar, realiya, til birliklari.

Abstract: In this article, linguopoetics, linguistic stylistics, artistic style of stylistics and linguopoetic research of poetess Farida Afro‘z’s poetry, its specific aspects and aspects, the poetess's skill in choosing words and using words, words arranged like a pearl necklace, the role of words with opposite meaning in the formation of antithesis of words from simple words, the stylistic and figurative functions of lexical units in specific poetic allusions and the free observation of lyrical feelings are discussed and the analysis is given, the poetic image is created by means of the words in his poems, in which each word is a tool that invites to feel self-sacrifice that serves the poetic activation of other words, and it is explained with the examples. It was noted that a beautiful Afro‘z’s interpretation of the timeless theme was given.

Keywords: linguopoetic research, stylistics, antithesis, figurative expressions, reality, language units.

Аннотация: В данной статье рассмотрены лингвопоэтика, лингвистическая стилистика, художественный стиль, стилистика и лингвопоэтическое исследование поэзии поэтессы Фариды Афруз, ее специфические аспекты и стороны, умение поэтессы выбирать и употреблять слова, слова, расположенные как жемчужное ожерелье в стихах. Обсуждается роль слов с противоположным значением в образовании антитезы слов от простых слов, стилистические и образные функции лексических единиц в конкретных поэтических аллюзиях, свободное наблюдение лирических чувств, дается анализ, поэтический образ создается средствами слов в его стихотворениях, в которых каждое слово является орудием, призывающим к чувству самопожертвования, служащим поэтической активизации других слов, и поясняется на примерах примеров. Было отмечено, что дана прекрасная интерпретация вечной темы, присущей стилю Афруз.

Ключевые слова: лингвопоэтическое исследование, стилистика, антитеза, образные выражения, реалия, языковые единицы.

Bizga ma'lumki, lingvistik poetika XX asrning 70-yillarida filologiya fanining mustaqil bir yo'nalishi sifatida vujudga kelgan. lingvistik poetika, o'zining hozirgi kungacha bo'lgan taraqqiyotida bir qancha o'zgarishlarni boshidan kechirdi, shuningdek,badiiy matnning lingvistik xususiyatlarini o'rganishning ko'plab turli usullarini tadqiq etdi. Bilamizki , lingvopoetika avvalo matnlarning lingvistik xususiyatlariga e'tibor qaratadi, faqat ular bilan chegaralanib qolmaydi. Lingvopoetikaning mohiyati va vazifalariga qarashning o'zgarishi tilshunoslikdagi hukmron tarkibiy –tizimli paradigmaning antroposentrik yondashuv bilan almashtirilishi tufayli o'rganish mumkin bo'ldi. Lingvopoetikaning maqsad va vazifalari konkkretlashtirildi, bu esa uning lingvistik stilistikadan farqlarini shakllantirish imkonini berdi. Ya'ni lingvopoetika til darajasining obrazli vositalarini matn yasovchi jihatida, ularning ma'no va obraz hosil qilishga ta'sirini hisobga olgan holda tadqiq qiladi. Ikkinchi tomondan, figura va uslubni ularning mazmunini hisobga olgan holda tavsiflaydi. [Keldiyorova, 2021, 58]. Hozirgi vaqtda funksional til faoliyatining barcha jabhalarini qamrab oluvchi lingvistik stilistikaning mustaqil sohasi sifatida shakllandi, xorijiy tillarni o'rganayotganlar uchun nutq mazmuni, muloqot muhiti va bayon maqsadiga qarab lingvistik vositalardan foydalanish usullari va shakllarini o'rganadigan va tavsiya etadigan amaliy stilistikani egallash alohida ahamiyatga egadir.

Odatda, lingvopoetik tahlilning mahsuldorligi, ahamiyati nafaqat matnda, balki ishlatilgan stilistik birliklar soniga, ulardan foydalanish sifatiga ham bevosita bog'liq bo'ladi. Agar lingvistik birliklardan foydalanish usullari yetarlicha aniq bo'lsa, lingvopoetik tadqiq qilish,agar matnning o'zi ham mazmunan, ham lingvistik jihatdan ancha murakkab bo'lsa chinakam qiziqish uyg'otadi. Shu, lingvopoetik tahlilning asosi olimlarning ma'lum bir she'riy matnning bevosita taassurotini tushuntirishga intilishi bo'lganligi sababli lingvopoetik tadqiqot usullari individual she'riy va yorqin stilistik belgilar asosida ishlab chiqila boshlandi.

Dastaval rus tilshunosi A.A.Lipgart lingvopoetikaga quyidagicha ta'rif beradi. "Lingvopoetika –bu filologiyaning bir bo'limi bo'lib unda badiiy matnda qo'llaniladigan stilistik jihatdan belgilangan til birliklarining vazifalari va ma'lum bir g'oyaviy –badiiy mazmunni yetkazish va qiyosiy ahamiyati masalasi bilan bog'liq holda ko'rib chiqiladi".[Липгард, 2016, 18-19] Aynan shu ta'rif stilistik belgilangan lingvistik birliklarning roli va funksiyasiga estetik jarayon yaratishga qaratilgan. Stilistik belgilangan birliklarning matnning umumiy mavzusi bilan semantik bog'liqligiga qarab , bu birliklar oddiy kuchaytirgich vazifasini bajarishi mumkin. Badiiy adabiyot tilidadagi antiteza usuli bir- biriga nisbatan qarama –qarshi ma'nolarni beradigan so'zlarni, iboralarni bayonida yonma-yon qo'llash orqali obrazlilikni keltirib chiqarishga asoslangandir. She'riyatda antiteza badiiy nutqning ifodaliligi, ta'sirchanligini ta'minlashda qo'l keluvchi ifoda imkoniyati sanaladi. Yana shunday ifodalarni qiyoslash, o'xshatish, o'ziga xos realiya yoki ifodalar bilan obrazlilik she'riy matnda yaratiladi.

Bunday misollarni shoira Farida Afro'z she'riyatida ham juda ko'plab uchratamiz. Dunyo o'zi qarama qarshiliklardan iborat ekanligi aniq,.Demak,inson umri ham kurashlardan iborat ekanligini his qilamiz. Gohida kimlardandir dilimiz og'rib qoladi, gohida iroda, kuch topib kechirib yuboramiz.

Dilimni og'ritdi dilimdagilar,
Tunu kun nomlari tilimdagilar,
Oftobim shamolim selimdagilar
Men ularni kechiray bir xilqat bergin,
Allohim, mehringdan bir xilvat bergin.

Ushbu misrada sutkaning qorong‘i va yorug‘ qismini ifodalovchi tunu kun qarshilantirilishi bilan birga, antiteza hosil qilingan, antonimik juftlikni yonma-yon qo‘llash orqali umumiylik semasiga ega bo‘lgan so‘z bilan shoira fikri obrazli tarzda ifoda etilgan.

Dushmanim ayyordir, paytlar poylaydi,
Do‘stim esa nodon, suvni loylaydi,
Ikkisi bir bo‘lib ko‘zim boylaydi,
Men ularni kechiray, bir qudrat bergin,
Allohim mehringdan bir xilvat bergin.

Ushbu misralarda dushmanim, do‘stim so‘zlari orqali antiteza hosil qilingan, shuningdek bu she’rni o‘qigan har qanday insonda kechirimli bo‘lish hissi uyg‘onadi, chunki yaratganning o‘zi mehrli va kechirimlidir.

Bugungi kunda, o‘zbek she’riyatida o‘ziga xos uslubi, xalq qalbidan munosib joy olishi, so‘z tanlash va so‘z qo‘llash mahorati, nutq go‘zalligi tarovati hamda so‘zlarni inju marjondek terish borasida Farida Afro‘zning o‘rni beqiyosdir. Uning ijodi o‘ziga xosligi bilan, tasavvufiy ruhga, xalqona ifodaga, hayotiy mushohadaga boyligi bilan, zamonga hamnafasligi bilan ajralib turadi. Uning “O‘zbek”, “Ayol” ,“Ayolni sev…”, “Yomg‘ir”, “Bo‘shliq”, “Xatolar qurboni“, “Men ayolman”, “O‘sha kun bugundir” va tashbehlari kabi she’rlari yuqoridagi fikrlarimizning isbotidir. Shuningdek, bir qator muxammaslarida ham o‘zining bugungi zamonga xos qarashlari, misralardagi hamohanglik, uyg‘unlik shoiraning betakror uslubi bilan ajralib turadi. Shoira voqelikka munosabat bildirishda so‘zlarning jimjimadorligiga emas, aksincha fikrlarga jon bera oladigan oddiy xalqona so‘zlarni izlab topadi, ularni bir-biriga bog‘laydi, ularni taqqoslagan holda, ulkan poetik olam manzarasi tasvirini gavdalantiradi. Shoira talqinida Vatan mavzusini yoritishda, uning istiqbolini kuylashda lirik qahramonning cheksiz faxru-iftixorini, shukr qilishi uni yuksaltirishga da’vat etish hissi ustuvor tamoyil sifatida ifodalanganini anglaymiz. Farida Afro‘z vatanning yantog‘idan tortib azim tog‘lariga qadar sevgan soddalikni tarannum etgan shoiradir. Buni biz shoiraning tashbehlarida ham ko‘rishimiz mumkin.

Sendan go‘zal gul yo‘q dunyoda
Sendan xushbo‘y, sendayin mayin,
Otam qabrin yashnatgan, yantoq.

Ushbu tashbehdan ma’lumki, barcha uchun ota siymosi juda ham aziz, hamisha e’zozda bo‘lgan , buyuk bir tirgak misoli tog‘dir. Ota bu dunyodan o‘tib, jasadlari tuproqqa qorishib ketsa ham qabridan chiqqan yantoq tikanli bo‘lsa ham, dunyodagi eng go‘zal, xushbo‘y guldan ham iforli maftunkordir, azizdir, xush bo‘ydir. Go‘yo yantoq yolg‘izlikda qolgan ko‘ngil iztirobini eshitguvchi bir gul. Ko‘ngil sahrosida uningsiz sabrli gul bo‘lmasa kerak. Ota-bobolarimiz qabrlarini ziyorat etganimizda shoiraning dilidan o‘tgan lahzalarni guvohi bo‘lamiz va his qilamiz. Go‘yo yantoq ildizlari bilan o‘tganlarga qo‘l cho‘zib turganday. Tirik inson bajarolmaydigan holatni obrazli tarzda yantoq misolida ifoda etilgan.Haqiqiy inson uchun hayot haqiqati, ota-bobolar xotirasi,yantoq guli kabi sodda, tikanlari kabi og‘riqli jarayondir. Bu tashbeh har bir insonga ota qadrini yana bir bor anglatib qo‘yadi, qadr tushunchasini his qildiradi.

Yaxshilik duyoning ustuni,
Arazlar ko‘ngilning tuguni,
Bugun-ku dunyoning buguni,
Yomonlik yiroq ket bizlardan.

Ushbu tashbehda esa shoira **yaxshilik** dunyoning ustuni, e'zozi bo'lgani kabi, araz, gina, qudurat inson qalbidagi og'riq, **tugun** ekanini ta'kidlab**, yomonlik** yiroq ket deya viqorla aytadi. Ushbu antiteza orqali she'riyat ta'sirchanligini oshirib, qahramon ruhiyatidagi kontrastlik va muallif his-tuyg'ulari bo'rttirib tasvirlanganini ko'rishimiz mumkin.

Yana shoira Farida Afro'z "O'zimniki" deb nomlangan she'rida Ona -vatanining jonli manzarasini, urf-odatini shoirona talqinini go'zal ifodalar bilan suratini chizgandek ko'rinadi:

Tutmi, jiyda meva zoti larzon-larzon,
Har qadamda bug'doy, paxta, sholi xirmon.
Oltinga yo'q himmat sizda, bizda arzon,
Mo'lmi, ko'lmi, terim endi o'zimniki.

Ushbu misralarda mevalardan: tut, jiydalarning lapanglab pishib yetilishishi, har qadamda **paxta, bug'doy, sholi** xirmonlari uyumi, oltindan qimmat bo'lgan **himmat**, xalqimizning faxr tuyg'usi odamgarchilikni o'zbekda borligi, millat g'ururi borasida fikr yuritib mo'lmi, ko'lmi terim endi **o'zimniki** deydi.

Uyat qilmas hech kim, o'sma qo'yib yursam,
Sochlarimga pilik solib to'yga borsam.
Kaftlarimning bo'z yeriga xino eksam,
Ko'zmunchog'u-irim endi o'zimniki.

Millatning millat ekanligini ko'rsatuvchi va shu millat vakillarigagina xos bo'lgan o**'sma, pilik, to'y, xina ko'zmunchoq, irim** kabi so'zlar tashbehi misolida she'r mazmunini boyitgan. She'rdagi xalqona ohang esa milliy jozibadorlikni yuzaga keltirgan va Farida Afro'z ijodining yana bir qirrasini, o'ziga xos jihatlarini ko'rsatgan. Kishini o'ylantiradigan, ensasini qotiradigan fikrlardan yiroq shoira ,samimiy tuyg'ulari orqali "dilda borini sodda so'zlar bilan ifoda etgan." O'zbek ayollariga xos bo'lgan jihatlarni marjonday satrlarga joylay olgan "o'zbek ayoli" timsolining kamtarona qiyofasini ham yarata olgan deya baralla ayta olishimiz mumkin. An'anaviy uslubda bitilgan bu she'r zamirida vatan yoshlariga, shuningdek, shoiraning o'ziga qarata aytgan da'vat ,ta'kid ohangini sezish qiyin emas. Ifoda etilgan bu fikrlar shoira hislarini so'zlarda ranglantiradi. Eng muhimi bunday mavzudagi she'rlar go'zal insoniy tuyg'ularga zamin yaratadi. Farida Afro'z ijodining yana bir o'ziga xos jihatini belgilashda uning alohida-alohida to'rtliklar shaklida yaratilgan she'r va tashbehlarilari muhim o'rin tutadi.

Xulosa qilib aytadigan bo'lsak, o'zbek xalqining adabiy an'analarini yuksaltirgan va zamonaviy she'riyat maydonida o'z o'rnini topib ulgurgan Farida Afro'z ijodi bugungi kun she'riyatining ulkan yutug'idir. Uni atroflicha tahlil qiiish, lingvopoetik tadqiqi va o'ziga xos jihatlarini o'rganish masalasi tilshunoslik fani rivoji uchun ham ahamiyatlidir.

Foydalanilgan adabiyotlar:

1. Farida Afro'z. Ko'zim manim. – T.: G'afur G'ulom nomidagi Adabiyot va san'at nashriyoti. 1999.
2. Farida Afro'z. O'sha kun bugundir, 1-jid, 2013.
3. Karimov S. O'zbek tilining badiiy uslubi: Filol. Fanlari d-ri…dis. – Samarqand, 1993.
4. Keldiyorova G. Badiiy matnning lingvopoetik tadqiqi. O'zbek tili tadqiqi va ta'limi muammolari. Xalqaro ilmiy-amaliy anjuman. Toshkent.: "Universitet". 2021.
5. Липгард Ф.Ф. Основы лингвопоэтики. Либроком. 2016.

TIL, USLUBIYAT, NUTQ MADANIYATI VA ADABIY TAHRIR

Taxirov Zoir Taxirovich
O‘zbekiston Milliy universiteti professori.

Annotatsiya: Mazkur maqolada funksiyaviy uslublar hamda adabiy tahrir aloqadorligi haqida so‘z boradi. Unda filologik fanlarning tilshunoslik, amaliy tilshunoslik, uslubiyat, nutq madaniyati, adabiy tahrir xususida so‘z boradi va ular turli matn tuzishda bir-biri bilan jips bog‘liqligi yoritaladi.

Tayanch so‘zlar: Tasviriy tilshunoslik, kategoriya, til tizimi til birliklari, uslub, nutq, nutqiy muloqat, nutq shakli, funksiyaviy-kommunikativ tur, matn, funksiyaviy tur, tahrir, matn uslubi.

Annotation: This article deals with such philological sciences as linguistics, applied linguistics, stylistics, speech culture and literary editing. Their interrelationships with each other in the compilation of various text outputs are discussed.

Keywords: Visual linguistics, category, language system, language units, style, speech, speech communication, form of speech, functional-communicative type, text, functional type, editing, text style.

Umumxalq tili tilshunoslik (lingvistika)ning tadqiq etish obyekti hisoblanadi. Tilni o‘rganish jihatlari nuqtai nazaridan esa tilshunoslik shartli ravishda ichki va tashqi tilshunoslikka ajratiladi. Umumiy tilshunoslik (tasviriy tilshunoslik) ichki lingvistikaga kiradi umumtil kategoriyalari, turli darajalarini o‘rganadi. Tashqi lingvistika (paralingvistika, etnolingvistika, psixolingvistika, sotsiolingvistika, kompyuter lingvistikasi, neyr-olingvistika va b.) so‘zlovchining jamiyatdagi faoliyati bilan tilning bevosita bog‘liqligini o‘rganadi.

Tasviriy tilshunoslik “til tuzilishi” [1] haqidagi fandir. Bu fan til ichida nima ***bor*** ekanligini, fikr, his-hayajon (emotsiya), ma’lumot (xabar) etishni o‘rganadi va o‘rgatadi.Tasviriy tilshunoslik bilan uslubiyatning o‘rganish obyekti bitta ayni bir tildir, ammo predmeti boshqa-boshqa. Aytish joizki til tizimidagi birliklar va kategoriyalar tasviriy tilshunoslikning tadqiqot predmeti hisoblanadi. “Tildan, ayniqsa adabiy nutqdan foydalanishning barcha usullari uslubiyatning tadqiqot predmeti” [2] bo‘ladi. Uslubiyatning nutqiy muloqatda tildan foydalanish haqidagi fan ekanligini mohiyatini ochib berish uchun adabiy nutq tushunchasini aniqlashtirib olish kerak. Turli manbalarda *adabiy nutq* terminiga quyidagicha tariflar beriladi: “... umumxalq tilining ishlangan, sayqal berilgan, ma’lum normaga solingan, xalqning turli madaniy ehtiyojlariga xizmat qiluvchi firmasi”[3]. “... o‘zbek tilining dialektlari, oddiy so‘zlashuv tili va jargon-laridan tamoman farqlanadi, shu xalq vakillari uchun umumiylik bilan xarakterlanadi”[4]. “... muayyan umumxalq tilining qayta ishlangan va meyorlashtirilgan mazkur tilda so‘zlashuvchi xalqning madaniy ehtiyojlariga xizmat qiluvchi shakli”[5].

“Umummilliy tilning meyorlashtirilgan shakli, yozma va og‘zaki ko‘rinishga ega ijtimoiy sohalar barchasida amalda bo‘ladi hamda milliy madaniy hayotga xizmat qiladi”[6].

Adabiy nutq *ijtimoiy lingvistika*ning asosiy tushunchasi. U milliy til egalari tomonidan namuna sifatida qabul qilinadi. Ijtimoiy (ijtimoiy-madaniy) mavjud (amalda bo‘lish) shakli. Uning asosiy vazifasi milliy tilda so‘zlashuvchilarning tarixan shakillangan jamoasi barcha a’zolari

faoliyatining barcha sohasida to‘laqonli aloqa-aralashuvi (muloqat)ni ta’minlashdan iborat. Adabiy nutq bajarishi lozim bo‘lgan “... ana shu asosiy vazifa uni mohiyatan adabiy nutqga aylantiradi, ya’ni umumqabul qilingan, umumtushunarli bo‘lishini ta’minlaydi”[7]. Yuqorida aytilganlarning barchasiga bitta qo‘shimcha qilish mumkin: *umummajburiy.*

Nutq jarayonida adabiy nutqdan, umuman tildan foydalanish umumiy nomda uslubiyat ataladi. U – nutqiy (yozma, og‘zaki) muloqat vaqtida til (birliklari, kategoriyalari)dan foydalanish haqidagi mustaqil fan. Uslubiyatda, asosan, *til birliklaridan foydalanish yoki “til birliklarining muayyan vazifani bajarishi”* haqida so‘z boradi. Ana shu vazifa bajarishligiga ko‘ra u funksiyaviy (vazifaviy) uslubiyat deb ataladi.

Til vositalari muayyan vazifa bajarar ekan, nutq egasining maqsadiga ko‘ra ma’lum o‘zgarishlarga uchraydi. Bu o‘zgarishlar ayrim til birliklari, kategoriyalari, so‘zlar guruhi, grammatik shakllar, konstruksiyalar (tuzilmalar) va fonemalarda yuz beradi.Uslubiyat tilshunoslik fanlari orasida alohida o‘rinni egallaydi. Boshqa tilshunoslik fanlari tilda *nima* borligini o‘rgansa, uslubiyat til birliklari va kategoriyalardan *qanday* foydalanishni tadqiq etadi. Nutq (og‘zaki, yozma)da fikr, his-hayajon, fikr almashish, informatsiya yetkazish qanday kechadi, mana shular uslubiyatning mohiyatini tashkil etadi. Bir so‘z bilan aytganda, *harakatdagi* tilni o‘rganadi. Tilda shakllangan birliklar va kategoriyalar nutqiy muloqot jarayonida quyidagilarga bog‘liq ravishda muayyan funksiya (vazifa)ni bajaradi:

– xabarlar mazmuniga;
– matn (umumiy nutqiy muloqot) ga;
– nutqiy muloqot amalga oshirilayotgan sharoitlarga.

Tilning qanday vositalari birikmaga kirishishi, bir-birlari bilan bog‘lanishi, bir-biriga moslashishi, muloqot sohasi, nutq shakli (og‘zaki, yozma), matn mavzusi, muhokamaning mohiyati, “gaplashish” va h. ga ko‘ra uslubiyat quyidagilarni aniqlashtiradi:

– tildan kishilarning birgalikda faoliyat olib borishlari, ish yuritishlari, umuman, turmush kechirishlari kabi jamiyatning muhim ehtiyojini qondirishda qanday foydalanilishini;
– har bir til birligi qanday vazifani bajarishligini;
– muayyan matn tarkibida (yoki muayyan matn turida), u yoki bu muloqiy sharoitda til birliklari mutanosiblashishi yoki bir-biriga ta’sir etishligini.

Uslubiyat tildan foydalanish masalasini aniqlashtiradi, tadqiq etadi. Bu uning predmetini tashkil etadi. Uning tilshunoslikning boshqa fan sifatidagi mazmun va mohiyatini tashkil etadi. Ana shunga ko‘ra uslubiyat tilshunoslikning funksiyaviy-kommunikativ (vazifaviy-muloqot) turi hisoblanadi va nutq madaniyatini, ortologiya (to‘g‘ri nutq haqidagi fan)ni, til me’yorlari to‘g‘risidagi ma’lumotni, ijtimoiy lingvistikani, adabiy nutq nazariyasini o‘z ichiga oladi. Uslubiyatga oid mashg‘ulotlarda quyidagilarni a’lo darajada bilish (o‘rgatish) nazarda tutiladi.

1. Til “material”ini, kategoriyalarini, birliklarini – uning tizimini tashkil etuvchilarni;
2. O‘zbek tili fonetikasi, leksikasi, grammatikasi (morfologiya sintaksis), frazeologiyasi, so‘z yasalishi va b. ni;
3.Uslubga ko‘ra ega va kesim moslashuvini;
4. Sodda va qo‘shma gaplar tuzilishini;
5. Tildagi obektiv va subektiv munosabatni;
6. O‘timli va o‘timsiz fe’llarni;
7. Qaratqich va tushum kelishiklarining farqini;
8. Mavhum otlar, shakllar omonimiyasi va h.k. ni.

Uslubiyatda, avval, mikromatn, shuningdek, muallifning umumiy fikri, kontekst ta'siriga ko'ra tahlil etilayotgan matnda til birliklaridan foydalanishdagi me'yordan chetlanish – sintaktik konstruksiyalardagi kamchiliklar, frazeologizmlarning xato qo'llanganligi, so'zlarning birikmaga noto'g'ri kiritilganligi, so'zning o'z ma'nosida qo'llanilmaganligi, ma'no bo'yog'idan noto'g'ri foydalanilganligi e'tiborga olinadi. *Havo bulut edi. Oq bulutlar bo'lganlikdan quyoshning nuri uncha kesilmagan, quyosh to'r ro'mol ichida o'tirgan kelinlar kabi yer yuziga suzilib qarar edi.* (A.Qodiriy). Tabiat manzarasini chizar ekan, adib *quyosh kelinlar kabi suzilib qarar edi* o'xshatish konstruksiyasini tuzadi. Uslubiy nuqtayi nazardan bu konstruksiya e'tiborni o'ziga tortadi. Grammatika bo'yicha mutaxassislarning fikricha, bu konstruksiya grammatik jihatdan to'g'ri hisoblanadi. Leksikolog esa, *suzilmoq* so'zini qo'llashda me'yordan chekinish bor, deydi. Uslubiyat bo'yicha mutaxassis esa mazkur konstruksiyadagi me'yordan chekinishlikni aniqlay boshlaydi: nima uchun, nega? Matn uchun bu nima beradi, muayyan aniq ifodaga yana qanday yangi ifoda (ottenka) bag'ishlaydi.

Mazkur konstruksiya tasviriy tilshunoslik nuqtayi nazaridan qaraladigan bo'lsa, aslida, *quyosh kelinlar kabi qaraydi.* Konstruksiya to'rt komponentdan tashkil topadi: *quyosh, kelin, kabi, qaraydi.* Bu konstruksiyada *quyosh, kelin* – ot, *kabi* – ko'makchi, *qaradi* – fe'l (o'tgan zamon). Ulug' adib quyoshning nur taratishini inson qarashiga o'xshatar ekan, qarashga xos ko'plab belgilar orasidan faqat bittasini tanlaydi: *suzilib.* Ana endi tasviriy tilshunoslikdan uslubiyatga o'tiladi: nega? Inson qarashiga turli belgilar xos: *kulib qarash, xo'mrayib qarash, bezrayib qarash, tirjayib qarash, jilmayib qarash, tumshayib qarash, tersayib qarash, ola qarash, irjayib qarash, sirli qarash, shubhali qarash, ishonqiramay qarash, ishonib qarash, chaqchayib qarash, baqrayib qarash, yovqarash, do'stona qarash,bo'zrayib qarash, iljayib qarash, yotsirab qarash, yig'lamsirab qarash* va h. Nega yozuvchi qa-rashga xos ko'pgina belgilar ichidan faqat *suzilib*ni tanladi. Ana shu tanlash ayni uslubni anglatadi. Agar *suzilib* so'zi – ravishdoshning nominativ (atash) ma'nosiga e'tibor bersak, u insonga xos qarash belgisini anglatadi, ma'no bo'yog'i, tabiiyki, ijobiy. Bunday qarashga insonlar yoshi, jinsidan qat'i nazar, barchasi ham ega bo'lishi mumkin. Lekin ular barchasi birdek ijobiy emas. Adib yuqorida ayol zotiga mansub tur – *kelin qarashi*ga quyoshning *nursochish*ini o'xshatadi. Nega kelinning, nega qizlarning, nega katta yoshdagilar (kampirlar)ning? Vaholanki, ular ham ayol! Chunki qizlarning *suzilib qarashi – odobsizlik,* yoshi kattalarniki – *yarashmagan qiliq*. Kelinlarning bu xil qarashlari ham ma'lum davrda (yangi kelinlik davrida), shunda ham o'z jufti halollariga nisbatan bo'lsa, *ijobiy* sanaladi. Yuqoridagi parcha badiiy matn ekanligini hisobga olsak, unda muallif quyosh nurlarini oq bulut to'sib qo'yganligi va uning ham yorug'ligi xira tortgan, hovuri pasayganligi va bu kishiga yoqishlik darajasida ekanligini ta'kid etmoqchi bo'ladi va buni kelinlarning suzilib boqishligi yoqimliligiga o'xshatadi. Uslubiyat tildagi bu hodisani tadqiq etarkan, tasviriy tilshunoslik bilan o'zaro jips bog'liqdir, chunki u mohiyatan tilni tadqiq etishni nihoyasiga yetkazadi. Uslubiyat tilni tizimli-tuzilmaviy tahlil qilish vazifasi (tasviriy tahlil) dan o'tib, lingvistika qismlariga asoslangan holda quyidagilarni tadqiq etadi.

1. Fikrni ifoda etishning yondosh (parallel) usullaridan, til vositalari yordamida yuzaga chiqadigan his-hayajonlarni;

2. Turli maqsadlarda va nutqiy muloqot sharoitlarida til vositalarining turli yo'sinda birikuvi, qo'shiluvi usullari, tartibini;

3. Nutqiy muloqotda, yozma va og'zaki nutqda til unsurlari, kategoriyalarining amalda qo'llanilish qonuniyatlari va o'ziga xosliklarini. *U darvozaga yaqinlashayotgan edi, nimadandir*

cho‘chib chetga burildi. Chunki darvozadan chiqib qolgan ikki kishini ko‘rib, bularning birini qayin otasi bo‘lib chiqmagidan hurkkan edi. (Abdulla Qodiriy).

Mazkur parchadagi jumlalar grammatika nuqtayi nazaridan to‘g‘ri ekanligini hech bir tilshunos inkor etmaydi. Faqat leksikolog so‘z qo‘llash xususida me’yordan ma’lum darajada chetlashish borligini qayd etadi va shu bilan cheklanadi. Faqat uslubiyatchi mutaxassisgina birinchi jumlada *cho‘chimoq,* ikkinchi jumlada *hurkmoq* so‘zining qo‘llanilganiga *nega* deya e’tibor beradi va uning tahliliga kirishadi, bunday ikki xil qo‘llash mohiyatini ochib beradi. Birinchi jumladagi *cho‘chimoq* fe’lining genetik ma’nosi *kutilmagan holatdan xavotirga tushmoq* ifodasini beradi, ya’ni holat, harakat xavotir bilan bog‘liq va bu holat ayni insonlarga xos. Ikkinchi jumladagi *hurkmoq* so‘zining genetik ma’nosi jonivorlarga, ularning ham muayyan turi – ot, eshak, qora mol, tuyaga xos holat, harakat ifodasini beradi va bu holat, harakat *qo‘rquv* bilan bog‘liq bo‘ladi. Yuqoridagi har ikki jumla grammatika qoidalariga muvofiq tuzilgani, ya’ni ega, kesim, ikkinchi darajali bo‘laklar o‘z o‘rnida, so‘zlarning birikmaga kirishuvi ham me’yorida ekanligi tasviriy tilshunoslik nuqtayi nazaridan to‘g‘ri, deya tan olinadi va bular uning o‘rganish obekti hamda predmeti hisoblanadi. Leksikologlar *cho‘chimoq, hurkmoq* so‘zlarining bir-biriga sinonim ekanligi, faqat *cho‘chimoq* so‘zining ma’no ifodasi betaraf, *hurkmoq* so‘zi ma’nosi (insonga nisbatan qo‘llanganda) bo‘yoqqa egaligini ta’kidlash bilan kifoyalanadilar.

Uslubiyat bo‘yicha mutaxassislar, boshqacha aytganda, semik tahlil bilan shug‘ullanuvchi tilshunoslar esa, har ikki so‘z, garchi sinonim bo‘lsa-da, nega qo‘llanildi, degan savolni qo‘yadi va leksik birliklar grammatik kategoriyalarni semik tahlil qilishga o‘tadi. Tilshunoslikdagi semik tahlil “Adabiy tahrir” fanida mantiqiy tahlil deyiladi – har ikki ibora teng ma’noga ega. Tajribasiz muharrir (uslubiyat, semik tahlil, mantiqiy tahlildan xabarsiz) nega endi insonga xos holat, harakat belgisini jonivorlarga xos holat, harakat belgisini ifoda etuvchi so‘z bilan atash kerak, deya *hurkmoq* so‘zini *qo‘rqmoq* so‘zi bilan almashtirsa yoki konstruksiyani o‘zgartirsa, tasviriy tilshunoslik nuqtayi nazaridan qoidadan tashqari hisoblanmaydi. Mikrotahrir nuqtayi nazaridan ham shunday. Ammo uslubiyat nuqtayi nazaridan qaralsa, boshqa holat ayon bo‘ladi. Adib badiiy matn yaratishdan avval, albatta, uning kompozitsiyasi, o‘z maqsadi va bu maqsadni amalga oshirishda tildagi vositalardan qanday foydalanish va boshqalar haqida chuqur mulohaza yuritadi. Natijada muallifning **individual** uslubi yuzaga keladi.

Abdulla Qodiriy birinchi jumlada *cho‘chib* ravishdoshini qo‘llaganining boisi, uni cho‘chitgan narsa, hodisa hali ma’lum emasdi. Ana shu nogahoniy ta’sirdan Otabek cho‘chib tushadi. Albatta, cho‘chishlikka qo‘rquv sabab bo‘ladi. Keyingi jumlada esa *hurkib* ravishdoshi – *cho‘chib*ning sinonimidan foydalanadi. Nega? Uslubiyat ana shu negaligini tadqiq etadi va quyidagi xulosaga keladi:

– *cho‘chib* so‘zi ma’no ifodasida hali *qo‘rqish* hissi yo‘q, u *xavotirlanish, sergaklanish* ma’no ifodasini beradi;

– *hurkib* so‘zi ma’no ifodasida esa qo‘rqish hissi ma’no ifodasi ayon seziladi;

– birinchi holatda xavotirga solgan, sergaklantirgan obyekt noma’lum;

– ikkinchi holatda qo‘rquvga sabab bo‘lgan obyekt – ikki kishi ma’lum;

– ikki kishidan biri “qayin otasi bo‘lib chiqmagi”.

Otabek, agar har ikki kishi begona ekanini avvaldan bilganda edi qo‘rqib o‘tirmasdi. Chunki ular Otabekka tanish bo‘lmagan bir yo‘lovchi sifatida qarar va ahamiyat bermas, to‘qnash kelganda ham salom-alikdan nariga o‘tmasdi. Lekin, ikki kishidan biri qayin otasi bo‘lsa-chi, ...

Ot, tuya ... lar uncha-munchaga hurkimaydi. Ularni hurkituvchi ta'sir, ya'ni qo'rqinch kuchli bo'lishi kerak. Qayin otasi bilan uchrashib qolish Otabek uchun kuchli qo'rqinch edi. Undagi ana shu qo'rqinchni ifoda etish uchun Abdulla Qodiriy *hurkib* ravishdoshidan foydalanadi. Aslida *hurkib* so'zini kishilar tushib qolgan holat va ularga xos harakat ifodasi uchun qo'llash faqat salbiy ma'noda bo'lishi mumkin. Lekin Abdulla Qodiriy o'z ijobiy qahramoniga nisbatan qo'llaydi va qayin otasidan dashnom olib, qattiq izza bo'lgani uchun bunday uchrashuv qay darajada og'ir ekanini ifodalash maqsadida *hurkib* so'zini – ijobiy ma'no bo'yog'iga ega bo'lishga "majbur" etadi. Uslubiyat ana shuni tadqiq etadi, adib individual uslubini ochib beradi. Biz yuqorida uslubiyat nutq madaniyati bilan jips aloqada degan edik. Zero, uslubiyat funksiyaviy-kommunikativ fanlar sirasiga kiradi.Bu fanlarni tilga bo'lgan umumiy asosiy yondashuv birlashtiradi. Mavzkur fanlar til egalarining nutqiy muloqot jarayonida til vositalaridan foydalanishlaridagi o'ziga xoslik va qonuniyatni o'rganadi. Demak, uslubiyat, nutq madaniyati, ortologiya, ritorika til unsurlari va kategoriyalaridan adabiy nutq doirasida oqilona, me'yorida foydalanishni tahlil etadi.

Nutq madaniyati haqidagi ta'limda til birliklaridan kundalik nutqiy muloqotda, yozma va og'zaki tarzda adabiy jihatlari doirasida foydalanish ko'rib chiqiladi.Nutq madaniyati deyilganda alohida olingan bir kishi va aholi nutq madaniyatining umumiy darajasi ham nazarda tutiladi. Nutq madaniyati bo'yicha mutaxassislar amaldagi adabiy nutq me'yorlari tizimi nuqtayi nazaridangina emas, balki, individual hamda umumiy nutqiy madaniyatga uning qanchalik mantiqiy bayonidagi aniq-tiniqlik va ifodalilik (estetik) ekanligini ham nazarda tutadilar.

Adabiy me'yorlar nuqtayi nazaridan nutq madaniyatiga quyidagi talablar qo'yiladi:

– til vositalaridan foydalanishning asoslangan bo'lishi;

– matn tuzilishi mantiqan izchillikka ega bo'lishi;

– bayon aniq-tiniq, tushunarli bo'lishi.

Shunday qilib, nutq madaniyati sohasidagi tadqiqot o'z ichiga faqat adabiy nutqnigina emas, balki xalq so'zlashuv tilini ham to'liq qamrab oladi, demak, oddiy og'zaki so'zlashuv unsurlari, xalq shevalari, jargonlar – jinoyatchilar argosidan tortib, to kasb-hunarga oid nutqdagi bilgan-bilmagan holda og'zaki, yozma nutqga, shu jumladan adabiy nutqga o'tadigan barcha-barcha unsurlar tahlil etiladi. Ayni mana shu masalalar bilan shug'ullanishiga ko'ra nutq madaniyati uslubiyatga yaqinlashadi. Uslubiyat esa nutq madaniyatidan farqli o'laroq asosiy e'tiborni adabiy nutqga qaratadi. Adabiy nutqga xos bo'lmagan hodisalarga ular tufayli erishilgan uslubiy samara nuqtayi nazaridan qaraydi va unga qanchalik erishilgani (yoki erishilmagani)ni, adabiy nutq me'yorlariga asoslangan matnga qancha o'rinli yoki o'rinsiz kiritilganini baholaydi. Nutq madaniyati tadqiq etadigan adabiy nutqga xos bo'lmagan hodisalar doirasi uslubiyatdagidan kengdir. Buni kundalik nutqda me'yorlashmagan hodisalar cheklanmagan miqdorda uchrashi bilan izohlash mumkin. Hatto, shaxsiy yozishmalar ham bundan holi emas. Agar kundalik nutqiy muloqotda me'yorlashmagan nutqdan bevosita turmush tashvishlari xususidagi suhbatlarda, nutqiy muloqot vaziyatga ko'ra bo'lganda foydalanilsa, adabiy matnga ular muayyan uslubiy maqsadga ko'ra kiritiladi.

Individual nutq ham, umumiy nutqiy madaniyat ham til egasining saviyasiga, ijtimoiy ongga, adabiy nutq rivoji va uning hozirgi darajasiga bog'liq. Abdulla Qodiriyning xoh katta, xoh kichik asarlaridagi obrazlarning, albatta, hayotda o'z prototiplari bo'lgan. Masalan, Abdulla Qodiriyning o'g'li Habibulla Qodiriyning yozishicha, Toshpo'lat tajangning hayotda prototipi

bo‘lgan, u ham gapni xuddi Toshpo‘lat kabi uzuq-uzuq, chertib-chertib so‘zlagan. Polvon ismli bu shaxsning Qodiriy bilan suhbatdagi nutqiga e’tibor beraylik.

– *Shu, degin, Julqun uka, – hammayoqni qimorga boy berib, juda abgor bo‘lib yurgan chog‘im edi. Bir kun boshim qotib ko‘chada borayotsam, yonimdan boyvachcha kiyingan yosh bir yigit o‘tib qoldi. “To‘xta”, dedim, “san boyvachchani yoningdagi pulingni bir sanab ko‘raychi, qancha ekan”, dedim ichimda.*[8].

Abdulla Qodiriy Polvonning nutqi bilan Toshpo‘lat tajangni so‘zlatishdan maqsadi qahramoni tabiatini, qiyofasini o‘quvchi ko‘zi o‘ngida yaqqol gavdalantirish maqsadida badiiy uslub talabiga ko‘ra hajviy obraz yaratishdir. O‘tgan asrning 20–30-yillarida, jamiyatda ana shunday nutq egalariga ko‘plab duch kelingan va ular zamonasiga mos edi.Bugungi kun “toshpo‘lat”larining nutqi esa butunlay o‘zgacha:

– *Shu, desang, bordim ... nima, gudid qilurasan, etgan gapingga atvechat qilasanmi, yo‘q, atvechat qilasanmi? Korochchisi, boshqa gudid qima ... sanminan voobshe gaplashmiman, xo‘, tushundimi? Konchay ... Yo‘q, san, o‘zing kimminan gaplashvotganini bilasanmi? Lyuboyini sindirib qo‘yaman ...*

Har ikki nutqdan uning egasi kim, ongi, saviyasi, madaniyati qay darajada ekanligi jamiyatda qanday o‘rin tutishligi ayon bo‘ladi.Demak, nutq madaniyati, aytish mumkinki, baholovchi fan hisoblanadi. U nutq asoslarini baholaydi, me’yoriylashganlik darajasini aniqlaydi. Muayyan nutqiy vaziyatda mazmunan o‘rinli, o‘rinsiz ekanligini tahlil etadi.

Uslubiyat nutq madaniyati sohasidagi tadqiqotlarga nazariy asos bo‘lib hisoblanadi. O‘rganilayotgan nutq unsurlari uslubiyat nuqtayi nazaridan aniq tavsifga ega bo‘ladi. “Til madaniyati yoki nutq madaniyati haqidagi fan ham nazariy, ham amaliy fan (yoki tadqiqot sohasi), til uslubiyati va nutq uslubiyati bilan turdosh hisoblanadi”[9].

Demak, nutq madaniyati fani bilan uslubiyat fanining tadqiqot obyekti yozma va og‘zaki nutq hisoblanadi.

Adabiy tahrir – filologiyaning amaliy fanlararo tarmoqlaridan biri. Bu fan uslubiyatni, uning asosiy tuzilmaviy qismlarini, shuningdek nutq madaniyatini, asosiy tasviriy lingvistik fanlarni chuqur bilishni talab etadi. Adabiy tahrir matn, qo‘lyozma maqsadga muvofiqligi, aniq maqsadga qaratilganligi masalalari bilan ular intellektual,ekspressiv (ham estetik, agar matn badiiy bo‘lsa), mazmuni, muallifning o‘ylagani, muayyan matn, qo‘lyozma nimaga mo‘ljallanganligi, hammasi to‘g‘ri bo‘lishligi masalalari bilan shug‘ullanadi. Mustaqil fan sifatida o‘tgan asrning 50-yillarida shakllangan adabiy tahrir muharrirlik ishlovi berilishi lozim bo‘lgan qo‘lyozmalarga nisbatan o‘z yondashuvini belgilab olgan. Shulardan asosiysi:

– tahrir – yaxlit ijodiy jarayon, matnni tekshirish va qayta ishlashni o‘z ichiga oladi;

– har bir matnni adabiy jihatda yaxlit matn sifatida tekshirish va unga tuzatishlarni kiritish.

Tahrirga tortilgan matn, qo‘lyozmaga bir qator talablar qo‘yiladi:

– keltirilayotgan ma’lumotlar, berilayotgan xabarlar ham ilmiy, ham asosli bo‘lishi kerak;

– bayon etilayotgan nazariy fikrlar, farazlar yoki talqinlar, hodisalar, voqea va b.lar ilmiy yoki aqlan sog‘lom bo‘lishi lozim;

– muallif bayoni mantiqan ketma-ketlik (uzviylik)ka ega bo‘lishi shart.

Qo‘lyozmaga muharrir tomonidan ishlov berish deganda, matnga uslubiy jihatdan ishlov berib, uning adabiyligini ta’minlash nazarda tutiladi. Bundan tashqari, uslubiy jihatdan ishlov berish bilan birga, qo‘lyozma tarkibiy tuzilishi (kompozitsiyasi) ham aniqlanadi (to‘g‘rilanadi). Adabiy tahrir fanida “matn, qo‘lyozma tili ustida ishlash” degan ibora borligi sababsiz emas.

Tahrir qilinishi kerak bo‘lgan qo‘lyozmaga uslubiy ishlov berish faqat matndagi imloviy hamda tinish belgilari bo‘yicha xatolarni tuzatishdan iborat emas, balki:

– sodda va qo‘shma gaplarning ichki tuzilishi grammatik me’yorlarga mosligi tekshiriladi:

– funksiyaviy uslubiyat qonuniyatlariga to‘liq amal qilinganligi ko‘rib chiqiladi.

Xullas, matn, qo‘lyozmani tahrir qilish jarayonida uslubiy talablarning barchasi so‘zsiz amalga oshirilishi shart.

Birinchi talab matnlar bayonidagi adabiy nutqning muayyan funksiyaviy turga xos izchilligi, ifodaviyligi, ikkinchisi – g‘ayriuslubiy unsurlarni qo‘llashning asosli ekanligi, uchinchisi – fikrni ifodalash shaklining mazkur uslubga mosligi, to‘rtinchisi – funksiyaviy taalluqligiga ko‘ra so‘zlar va iboralardan matn uslubi (yoki janr-nutq)ga mos holda foydalanilganligi, beshinchidan – rasmiy-ish yuritish, ilmiy uslublarda, so‘zlashuv nutqida, badiiy matnda matn shu tur asarlarga xos tuzilganligi.

Demak, matn, qo‘lyozmani adabiy tahrirdan chiqarish turdosh fanlar – tasviriy tilshunoslik (leksikologiya, frazeologiya, grammatika, punktuatsiya va b.), funksiyaviy uslublar (umuman uslubiyat), semantologiya, ortologiya va nutq madaniyati kabi fanlar bilan uzviy bog‘liq ekan. Adabiy tahrir muvaffaqiyati tahrirchidan chuqur bilimli bo‘lishni taqozo etadi.

Foydalanilgan adabiyotlar:

1. Vinokur G.O. Filologicheskiye issledovaniya. –Moskva, 1989.–221-b.
2. Vinogradov V.V. Problemы russkoy stilistiki. –Moskva, 1981. –173-b.
3. Hojiyev A. Lingvistik terminlarning izohli lug‘ati. –T.: O‘qituvchi, 1985.
4. Shoabdurahmonov Sh., Asqarova M., Xojiyev A. va b. Hozirgi o‘zbek adabiy tili. –T.: O‘qituvchi, 1980. –7-b.
5. O‘zME, 1-j. –127-b.
6. Ensiklopedicheskiy slovar. –Moskva: Sovetskaya ensiklopediya, 1984.–716-b.
7. Щerba L.V. Izbrannыe rabotы po russkomu yazыku.–Moskva.1978.–117-b.
8. Qodiriy H. Otam haqida. –T.: G‘afur G‘ulom nashriyoti , 1983. –46-b.
9. Vinogradov V.V. Problemы russkoy stilistiki. –Moskva, 1981. –175-b.

BOSHLANG‘ICH SINFLARDA IMLO QOIDALARI ASOSIDA TOG’RI YOZISHNI SHAKLLANTIRISHGA O‘RGATISH

Jumayeva Dilnoza Baxshulloyevna
Navoiy davlat pedagogika 581rotocol581
Maktabgacha va boshlang‘ich ta’lim fakulteti
Boshlang‘ich ta’lim yo‘nalishi katta o‘qituvchisi f.f.f.d(PhD).

Norova Nastarin Otabek qizi
Navoiy davlat pedagogika 581rotocol581
Maktabgacha va boshlang‘ich ta’lim fakulteti
Boshlang‘ich ta’lim yo‘nalishi 3-kurs talabasi.

Annotatsiya: Boshlang‘ich sinf o‘quvchilariga imloviy xatolarsiz to‘g‘ri yozishni o‘rgatish va bosh harflar imlosini boshlang‘ich sinf o‘quvchilariga 1- sinfdan boshlab o‘rgatib 581rotoco borasida fikr yuritilgan.

Kalit so‘zlar:imlo, qo‘shimcha,so‘z,tamoyil,qoida,bosh harf, atoqli otlar, 581rot, 581rotoc, metodika, ot, sifat, olmosh, tovush, ona tili, nomlar, talaffuz, grammatika, harf.

Аннотация: Задумано научить младших школьников правильному решению задачи без орфографических ошибок и научить правописанию букв младших школьников после 1-го класса.

Ключевые слова: орфография, наречие, слово, принцип, правило, заглавная буква, имена собственные, ошибка, методика, существительное, уточнение, местоимение, звуки, родной язык, имена, произношение, грамматика, буква.

Abstract: It was thought about teaching elementary school students to write correctly without spelling mistakes and teaching the spelling of capital letters to elementary school students from the 1st grade.

Keywords: spelling, adverb, word, principle, rule, capital letter, proper nouns, error, qualification, methodology, noun, adjective, pronoun, sound, native language, names, pronunciation, grammar, letter.

Ona tilimiz-milliy ma’naviyatimizning bitmas tuganmas bulog‘idir. Shunday ekan,unga munosib hurmat va ehtirom ko‘rsatish barchamizning nafaqat vazifamiz, balki muqaddas insoniy burchimizdir.

Sh.M. Mirziyoyev.

O‘zbekiston kelajagi bo‘lmish yoshlarning bilimli,aqlli va salohiyatli bo‘lib yetishishlari uchun maktab davri muhim rol o‘ynaydi. Ayniqsa, fanlarni yaxshi o‘rganishlari har bir mavzuni o‘zlashtirishlari zarur. O‘quvchilar ilk bora maktabga borar ekan, ularga birinchi navbat yozish o‘rgatiladi. Ona tilida chiroyli va imloviy xatolarsiz yozish katta ahamiyatga ega. Shuning uchun savodxonlikka o‘rgatish davrida harflarni, so‘zlarni to‘g‘ri yozishga ko‘p e’tibor qaratiladi.

O‘qituvchi imloviy malakaning psixologik tabiatidan kelib chiqib, kichik yoshdagi o‘quvchilarda imloga oid malakani shakllantirish ustida ishlash metodikasini belgilaydi.

O‘quvchilarda imloviy malakalar grammatikaga bog‘lab o‘rgatiladi. Imlo qoidalari bir so‘znigina emas, balki umumiylik mavjud bo‘lgan butun so‘zlar guruhining yozilishini tartibga soldi. Bu xususiyati bilan u qoida xat yozuvchini har bir so‘zni yodda saqlash, xotiralashdan qutqaradi va qoidaga amal qilib, belgilangan me’yorga muvofiq butun so‘zlar guruhini yozish imkonini yaratadi.

Imlo qoidalarini grammatik,fonetik,so‘z yasalishga oid materiallarni ma’lum darajada bo‘lmasdan turib o‘zlashtirish mumkin emas. Grammatik nazariya imlo qoidalari uchun poydevor hisoblanadi. Shuning uchun boshlang‘ich sinflarda imlo qoidasi shu qoidaga asos bo‘ladigan Grammatik nazariyaga bog‘liq holda o‘rganiladi. Masalan,shakl yasovchi qo‘shimchalarning yozilishi haqidagi qoidalar ot,sifat,son.kishilik olmoshlari,fe’l mavzulari ichiga kiritilgan.

Fonetik tamoyil asosidagi asos va qo‘shimchalar imlosidagi ayrim qoidalar 3-4-sinfda fe’l yasovchi qo‘shimchalar o‘rganiladi,son so‘ziga fe’l yasovchi ***-a*** qo‘shimchasi qo‘shish bilan yasalgan so‘zning sona emas sana tarzida talaffuz qilinishi va shunday yozilishi, -a,-la fe’l yasovchi qo‘shimchasi o‘rganilganda, yuqoridagi so‘zlar kabi yosh, ot so‘zlarining talaffuziga muvofiq 582roto, ata tarzida yozilishi o‘rgatiladi. Ong so‘ziga fe’l yasovchi –la qo‘shimchasini qo‘shish bilan yasalgan so‘z ongla emas, angla tarzida talaffuz qilinadi va yoziladi.

4-sinfda egalik qo‘shimchalari bilan qo‘llangan otlarning imlosi o‘rganilganda og‘iz,qorin, o‘rin,singil,ko‘ngil,burun,582rotoc kabi so‘zlar o‘zagidagi I, u, a unlilari talaffuzda tushib qoladi va shunday yozilishi o‘rgatilganda yuqoridagi fonetik tamoyil asosida o‘zak negiz va qo‘shimchalar imlosiga asoslanadi(og‘iz+im=og‘zim, qorin+im=qornim, o‘rin+im=o‘rnim, singil+im=singlim, ko‘ngil+im=ko‘nglim, burun+im=burnim, 582rotoc+im=shahrim).

Imlo qoidalari juda ham muhim hisoblanganligi sababli bu to‘g‘rida qonun ham qabul qilingan.

O‘zbekiston Respublikasi Vazirlar Mahkamasining 1995-yil 24-avgustdagi 339- qaroriga muvofiq O‘zbek tilining asosiy imlo qoidalari to‘g‘risidagi qaror qabul qilindi.

Kishining ismi, ota ismi, familiyasi, taxallusi, ramziy atoqli oti bosh harf bilan boshlanadi: Dilbar, O‘rinova, Muhabbat Majidovna, Azamat Shuhrat o‘g‘li, Hamza Hakimzoda, Muhammadsharif So‘fizoda, Mannon Otaboy, Navoiy, Furqat; Yelpig‘ichxon, Salomjon Alikov kabi.

Bu qoidani o‘quvchilarga tushuntirish uchun o‘zlarining ismlaridan foydalanishimiz mumkin.

Joy nomlari bosh harf bilan boshlanadi: Andijon, Yangiyo‘l (shaharlar), Naymancha, Buloqboshi (qishloqlar), Bodomzor, Chig‘atoy (mahallalar), Zavraq (dara), Yorqoq (yaylov), Qoratog‘, Pomir (tog‘lar), Oqtepa, Uchtepa (tepalar), Zarafshon, Sirdaryo (daryolar), Yoyilma (kanal); Turkiya, Hindiston (mamlakatlar) kabi. Bunday atoqli ot tarkibidagi sifatlovchi ham bosh harf bilan yoziladi: Shimoliy Kavkaz, Markaziy Qizilqum kabi.

O‘quvchilarning yashash joylarini aytib, o‘zlariga doskada yozdirilsa, yaxshi eslarida qoladi va qiziqarli bo‘ladi.

Yulduz va sayyoralar, boshqa xil osmon jismlarining atoqli oti bosh harf bilan boshlanadi: Hulkar, Qavs, Mirrix (yulduz va sayyoralar nomi), Tinchlik dengizi (Oydagi 582rotoc nomi) kabi. Yer, quyosh, oy turdosh otlari sayyora nomi bo‘lib kelgandagina bosh harf bilan yoziladi: Yer Quyosh atrofida, Oy Yer atrofida aylanadi.

Madaniy-maishiy va savdo korxonalariga, adabiyot 583rotoc'at asarlariga, sanoat va oziq-ovqat mahsulotlariga, shuningdek, transport vositalari, sport inshootlariga qo‘yilgan nomlar bosh harf bilan boshlanadi: “Tong” (mehmonxona), “Saodat” (firma), “Navro‘z” (xayriya jamg‘armasi), “Kamalak” (matbaa birlashmasi), “G‘uncha” (bog‘cha), “Botanika” (sanatoriy), “Paxtakor” (stadion), “Qutlug‘ qon” (roman), “Dilorom”(opera), “Tanovar” (kuy), “Ozodlik” (583rotoc), “Jasorat” (yodgorlik), “S/лo” (sovutgich) kabi.

Muhim tarixiy sana va bayramlarning nomlari tarkibidagi birinchi so‘z bosh harf bilan boshlanadi: Mustaqillik kuni, Xotira kuni, Ramazon hayiti, Navro‘z bayrami kabi.

Davlatlarning, davlat oliy tashkilotlari va mansablarining, xalqaro tashkilotlarning nomidagi har bir so‘z bosh harf bilan boshlanadi: O‘zbekiston Respublikasi, Rossiya Federatsiyasi, Misr Arab Respublikasi, 16 O‘zbekiston Respublikasi Prezidenti, O‘zbekiston Respublikasi Oliy Majlisining Raisi, O‘zbekiston Respublikasi Oliy Sudining Raisi, Birlashgan Millatlar Tashkiloti, Jahon Tinchlik Kengashi kabi. Boshqa tarkibli nomlarda oliy mansabni bildiruvchi birinchi so‘zgina bosh harf bilan boshlanadi: Bosh vazirning o‘rinbosari, Mudofaa vaziri, Yozuvchilar uyushmasi, O‘zbekiston Milliy tiklanish demokratik partiyasi kabi. Vazirliklar va idoralar, korxonalar va tashkilotlar nomi tarkibidagi birinchi so‘z bosh harf bilan boshlanadi: Sog‘liqni saqlash vazirligi, Fan va texnika davlat qo‘mitasi, Fanlar akademiyasi, Tilshunoslik 583rotocol583 kabi.

Davlatning oliy darajali mukofoti nomi tarkibidagi har bir so‘z bosh harf bilan boshlanadi: “O‘zbekiston Qahramoni” (unvon), “Oltin Yulduz” (medal). Boshqa mukofotlar, faxriy unvonlar, nishonlar nomidagi birinchi so‘zgina bosh harf bilan boshlanadi: “Sog‘lom avlod uchun” (orden), “O‘zbekistonda xizmat ko‘rsatgan fan arbobi” (faxriy unvon), “Matbaa a’lochisi” (nishon) kabi.

Gapning birinchi so‘zi bosh harf bilan boshlanadi: Yer tagidan Muqaddasga bir qarab oldim (O.Yoqubov). Eslatma: 1) ko‘chirma gapdan keyin kelgan muallif gapining birinchi so‘zi (agar u atoqli ot bo‘lmasa) kichik harf bilan yoziladi: «Bu men», - qo‘rqibgina javob berdi ko‘laga (O.Yoqubov); 2) xatboshiga gapning sanaluvchi qismlari chiqarilganda, bunday qismlar oldidan chiziq qo‘yiladi va ular kichik harf bilan yoziladi: Ma’muriy huquqbuzarlik to‘g‘risidagi ishni ko‘rishga tayyorlash vaqtida tegishli organ (mansabdor shaxs) quyidagi masalalarni: - mazkur ishni ko‘rib chiqish uning huquq doirasiga kirish-kirmasligini; - ma ‘muriy huquqbuzarlik to‘g‘risidagi 583rotocol va ishga oid boshqa materiallar to‘g‘ri tuzilgan-tuzilmaganligini hal qiladi; 3) gapning qismlari qavsli raqam yoki qavsli harf qo‘yib sanalsa, bunday qismlar ham kichik harf bilan yoziladi: Hozirgi o‘zbek adabiy tilining lug‘at boyligi, asosan, besh manba negizida tarkib topgan: 1) umumturkiy so‘zlar; 2) o‘zbekcha so‘zlar; 3) tojik tilidan kirgan so‘zlar; 4) arab tilidan kirgan so‘zlar; 5) rus tilidan kirgan so‘zlar (“O‘zbek tili” darsligidan).

Tarkibli nomlarning bosh harfidan iborat qisqartmalar, atoqli ot bo‘lmagan ba’zi birikmalarning qisqartmalari bosh harf bilan yoziladi: AQSh (Amerika Qo‘shma Shtatlari), BMT (Birlashgan Millatlar Tashkiloti), AES (atom elektr stansiyasi) kabi. Qisqartma tarkibida bo‘g‘inga teng qism bo‘lsa, uning birinchi harfigina bosh harf bilan yoziladi: ToshDTU (Toshkent davlat texnika universiteti) kabi.

Xulosa qilib shuni aytish mumkunki, bosh harflar bilan yozish ko‘proq kishilarning ism-falimiyasi, joy nomlari, bayram va tarixiy sanalarga xos bo‘lib, ot so‘z turkumiga kiradi. Boshlang‘ich sinflarda ot so‘z turkumi 4-sinfda tushuntirilib, shu o‘rinda bosh harf bilan yoziladigan so‘zlar ham o‘rgatiladi.

Boshlang‘ich sinf o‘quvchilari bosh harf bilan yoziladigan so‘zlarni farqlashni o‘rganishlari qiyinroq bo‘ladi, chunki ular faqat nuqtadan so‘ng hamda kishilarning ismi kelganda bosh harflar bilan yoziladi, deb o‘ylashadi. Ularga o‘zbek tilining asosiy imlo qoidalaridagi bosh harflarga tegishli bo‘lgan boshqa qoidalarni ham o‘rganish lozim. O‘quvchilarga bularni o‘rgatayotganda hayotiy misollar va rasmlar orqali tushuntirsak, ularning o‘zlashtirishlari oson bo‘ladi.

Foydalanilgan adabiyotlar:

1. S.Matchonov, H.Baniyeva, X.G‘ulomova, Sh.Yo‘ldosheva, G.Xolboyeva. Boshlang‘ich sinflarda ona tili o‘qitish metodikasi darslik. Toshkent.2021.
2. M.E.Toirova. Ona tili 4-sinf 1-qism. Tosh.2023
3. Jumaeva D.B. Analysis of antroponyms on the basis of ethnolinguistic factors // Czech Republic. Volume 12. 2021. May. – P. 102-107 (Impact Factor – 8,758).
4. Жумаева Д.Б. Антропонимлар ва антропонимларнинг ўзига хос хусусиятлари // Namangan davlat universiteti ilmiy axboroti. – Namangan, 2022. – № 6. – B. 622-626 (10.00.00. № 6).
5. O‘zbek tilining imlo lug‘ati. Toshkent 2015y.

O‘ZBEK TILIDAGI HARAKAT MAZMUNLI SLENGLAR

Rasulov Namoz
O‘zbekiston Milliy universiteti o‘qituvchisi.

Akbarova Aziza
O‘zbekiston Milliy universiteti talabasi.

Annotatsiya: ushbu maqolada sotsiolektlardan biri slenglar, ularning jahon va o‘zbek tilshunosligidagi o‘rni haqida fikr yuritiladi. O‘zbek yoshlari orasida ommabop qo‘llanilayotgan ayrim fe’lli slenglar mazmuni va qo‘llanish doirasi haqida daliliy fikrlar keltiriladi. Slenglar standart bir tilda so‘zlashuvchi aholining yirik qismini tashkil etuvchi yoshlar nutqi hisoblanib, ularning tilda o‘rganilishi jamiyatdagi til taraqqiyotini kuzatishning samarali omili hisoblanadi. Slenglarni ilmiy tadqiq qilish vaqt o‘tishi bilan yoshlar ongida til va til unsurlarining o‘zgarishini kuzatishga yordam beradi. Tilning zamonaviylashuv jarayoni va boshqa tillar bilan aralashuvi yoshlar nutqida namoyon bo‘lishini hisobga olsak, ushbu sotsiolektning jamiyat va til hayotini bevosita bog‘liqlikda o‘rganishda naqadar muhim omil ekanligini his qilish mumkin.

Kalit so‘zlar: til, jamiyat, sotsiolingvistika, dialekt, sotisolekt, adabiy til, argo, jargon, sleng, lug‘at.

Аннотация: в данной статье рассматривается один из социолектов-сленгов, их место в мировом и узбекском языкознании. Представлены доказательные мнения о содержании и сфере применения некоторых глагольных сленгов, широко используемых среди узбекской молодежи. Сленги — это речь молодых людей, составляющих значительную часть стандартного одноязычного населения, и их изучение является эффективным фактором контроля языкового развития в обществе. Научное изучение сленга помогает наблюдать за изменением языка и языковых элементов в сознании молодых людей с течением времени. Учитывая тот факт, что в речи молодежи проявляется процесс языковой модернизации и смешения с другими языками, можно почувствовать, насколько важен этот социолект при изучении жизни общества и языка в непосредственной связи.

Ключевые слова: язык, общество, социолингвистика, диалект, социолект, литературный язык, сленг, словарь.

Abstract: this article examines one of the sociolects-slangs, their place in world and Uzbek linguistics. Evidence-based opinions on the content and scope of some verbal slangs widely used among Uzbek youth are presented. Slang is the speech of young people who make up a significant part of the standard monolingual population, and their study is an effective factor in controlling language development in society. The scientific study of slang helps to observe the changes in language and linguistic elements in the minds of young people over time. Considering the fact that the speech of young people shows the process of linguistic modernization and mixing with other languages, one can feel how important this sociolect is when studying the life of society and language in direct connection.

Keywords: language, society, sociolinguistics, dialect, sociolect, literary language, slang, dictionary.

Til va jamiyat tushunchalari ma'no anglatish nuqtayi nazaridan farqli ma'nolarni ifodalasada, bir-biri bilan uzviy bog'langandir. Tilni xalq yaratadi va o'z navbatida, u jamiyat taraqqiyoti bilan hamnafas ohangda takomil topib boradi. Xalqning madaniy va ijtimoiy-tarixiy taraqqiyoti o'z izlarini, birinchi galda, tilda qoldiradi. Shu bois, jamiyatdagi har qanday o'zgarish, birinchi navbatda, tilda aks etadi. [I.Yo'ldoshev va boshq. 2007. 192 b.]

Tilda adabiy til va dialekt farqlanganidek, o'z navbatida dialekt atamasi ham ikki turda: hududiy va ijtimoiy farqlanadi. Ijtimoiy dialekt sotsiolingvistik termin sifatida sotsiolekt atamasi bilan yuritiladi. Sotsiolekt atamasi XX asrning ikkinchi yarmida paydo bo'lgan. Umumiy ijtimoiy xususiyatlariga ko'ra birlashgan ayrim guruhlarning lisoniy ehtiyojlarini qondirish uchun nutq amaliyotida ishlab chiqilgan til shakli sotsiolekt hisoblanadi. [Sh.Usmanova va boshq. 2014. 85 b.]

Sotsiolekt termini ostida "argo", "jargon", "sleng" tushunchalari birlashtiriladi. E'tiborli jihati shundaki, ushbu atamalar qator manbalarda bir-biriga sinonim qo'yiladi. Ayniqsa, sleng atamasi va tushunchasi paydo bo'lganidan keyin jargon tarkibida bir narsa yoxud argoga qiyosan tahlil qilina boshlandi. Keyingi yillarda slenglarning o'rganilishiga e'tibor kuchaya boshlaganini ham shu o'rinda ta'kidlab o'tish joiz.

Sleng inglizcha "slang" so'zidan olingan. Unga haligacha muayyan bir ta'rif mavjud emas. Ingliz tilshunosligida 1756-yilda ilk bor qo'llangan bu atama omi va savodsiz xalq leksikoni sifatida qaralgan. XIX asr boshlariga kelib endi sleng termini quyi tabaqa vakillari nutqi sifatida e'tirof etilsa-da, standart madaniyatli tabaqa nutqidan past leksika sifatida e'tiborda bo'ladi. Turli lug'atlar, xususan, ingliz tilidagi lug'atlarda slengga "jargon"ga yaqin yoki aynan jargon atmasining ta'rifi beriladi. Fransuz tilida bu atamaning aynan muqobili argo bo'lsa, turk tilida argo, jargo va slenglar bir atama – argo ostida birlashtiriladi. Ba'zi lug'atlarda aynan quyidagi ta'rifni uchratish mumkin: juda norasmiy hisoblangan so'z va iboralardan iborat bo'lgan, yozma nutqdan ko'ra og'zaki nutqda keng tarqalgan va odatda ma'lum bir kontekst yoki odamlar guruhi bilan chegaralangan til turi. [https://dictionary.cambridge.org/dictionary/english/slang]

Rus tilshunoslaridan biri V.Belikov: "Sleng g'arb lingvistik madaniyati uchun xarakterlidir. Mazmunan u jargon atamasiga yaqin turadi", – deydi. [В. И. Беликов, 2001: 317 ст.] Ko'p tillarda bo'lgani kabi rus tilida ham sleng uchun lingvistik qoida jargon atamasining izohi bilan bir xil.

Jahon tilshunosligida qayd etilgan ta'rif va tavsiflarni qiyoslagan holda o'zbek tilshunosligidagi sleng tushunchasini argo va jargondan farqlashimiz mumkin:

- slenglar umumtil va dialektal xarakter kasb qila oladi, argo va jargonlarda bu xususiyat mavjud emas. Masalan, *dimka bermoq* slengi Buxoro hududida keng qo'llanishiga guvoh bo'lishimiz mumkin. (Dimka bermoq – dimedrol bermoq *uxlatmoq slengi* bilan sinonimlik hosil qilib, "aldab ketmoq" ma'nosida qo'llanadi);

- slenglar, asosan, disfemik xarakterda bo'ladi, argo va jargonlarda esa bu toifalanish kuzatilmaydi.

- slenglar arxaik va neologik xususiyatlarga ega. Ba'zi slenglarning oralarida o'n yillik farq bo'lgan yoshlar o'rtasida ham qo'llanmasligining guvohi bo'lishimiz mumkin. Masalan, *strelka (uchrashuv)* slengini bugungi kunda arxaiklashayotgan deb ko'rsatishimiz mumkin.

- bir slengning valentlik imkoniyati uning turli ma'nolarni anglatishiga sabab bo'lishi mumkin. Masalan, *tepmoq* slengi *Endi tepdimi?* Gapida *tushunmoq* ma'nosida qo'llansa, *quloqqa tepmoq* birikuvida esa *aldamoq* ma'nosida ishlatilgan.

- slenglar o'zaro sinonimlik hosil qilishi ham mumkin. Masalan, *sochib tashlamoq* slengi *yormoq* slengi bilan ma'nodosh; *uxlatib ketmoq, stalbani ushlatib ketmoq, kiydirib ketmoq* slenglari o'zaro sinonim ekanligini kuzatish mumkin.

- slenglar aslida til tarkibidagi leksikaning ko'chma ma'no hosilasi hisoblanadi. Shuningdek, ma'lum tilning izohli lug'atida so'z tarkibidagi metoforik usul orqali ko'chgan ma'nolar ham sanab o'tiladi. E'tiborli jihati shundaki, O'zbek tilining izohli lug'atida so'zlarning ko'chma ma'nolari qatorida slenglarni uchratmaymiz. Masalan, *uxlamoq* fe'li O'zbek tilining izohli lug'atida quyidagicha izohlanadi:

1) uyquga ketmoq, uyqu holatida bo'lmoq. *Bola... goh u ishga, goh bu ishga yugurar, kechalari esa holdan toyib, horib, dong qotib uxlar edi.*

2) ko'chma: tinch holatda bo'lmoq. *Og'ir sukunatga cho'mib uxlardi qishloq.*

Bizningcha, slenglarning yuqorida sanab o'tilgan o'ziga xos jihatlari uni yondosh hodisalardan, xususan, jargon va argodan farqlaydi.

O'zbek tilshunosligidagi sotsiolingvistikaga oid darslik va qo'llanmalarda sleng atamasi quyidagicha izohlanadi: Sleng – muayyan kasb egalari yoki ijtimoiy guruhlarning og'zaki nutqida qo'llanadigan, emotsional-ekspressiv bo'yoqqa ega bo'lgan so'z va iboralar. [Sh.Usmanova va boshq. 2014. 85 b.]

Sleng adabiy tildan farqli ravishda til talablariga javob bermaydi, u, asosan, o'smirlarning ko'cha tili hisoblanadi. Lekin bu ta'rif bilan uni faqat yoshlargina tushunadi deyish xato. Ba'zi slenglarni jamiyatda ham yoshlar, ham kattalar qatlami tushunishi mumkin. Bu slenglarning ommalashuvi bilan bog'liq jarayon. Statistik ma'lumotlarga qaraganda, bugungi kunda 36 millionli O'zbekiston aholisining 65 foizini 30 yoshgacha bo'lgan yoshlar tashkil qiladi. Ularning nutqida qo'llaniladigan so'z va iboralarning turli vositalar: ijtimoiy tarmoqlar, ko'cha va oila suhbatlari orasida ishlatilishi asnosida tarqalish imkoni ham yuqori. Ba'zi slenglarning ommalashmasligiga asosiy sabab esa ularning ma'nosi anglashilmay qolishidir, chunki slenglar ko'chma ma'noda qo'llanadi. Ya'ni qo'llanilayotgan biror so'z lug'atdagi izohidan tubdan yoki qisman farqlanadi.

Yoshlarning nutqidagi bu slenglar grammatik jihatdan turli so'z turkumlariga oid. Ularning ko'pchiligini mana shu so'z turkumidagi leksika tashkil qiladi, deb qat'iy hukm berib bo'lmaydi. Quyida esa fe'l turkumiga oid ayrim slenglar va ularning mohiyatini tahlil qilib chiqamiz. Aytib o'tish joizki, harakat mazmunidagi slenglar ot va sifat xarakteridagi slenglardan miqdoran ko'p va ular, asosan, tarkibli tarzda qo'llanadi.

Chopmoq: darsni chopmoq kabi tarkibli yo mustaqil tarzda qo'llanadi. Darsni chopmoq tarzida kelganda undan voz kechib, boshqa mashg'ulot bilan shug'ullanmoq ma'nosida qo'llanadi. Mustaqil tarzda qo'llanganda esa "yomon qiladi", "dabdalamizni chiqaradi" (masalan, Domla endi hammamizni chopadi) ma'nosi anglashiladi.

Kalla qo'ymoq: "ishni buzib qo'ymoq", "vaziyatni ostin-ustun qilmoq" ma'nolarida qo'llanadi.

Surmoq "ketmoq, yo'qolmoq" ma'nolarida ishlatiladi va bu slengda, asosan, buyruq ohangi ustunlik qiladi.

Chizmoq tarkibli tarzda quyonni chizmoq sifatida ishlatiladi va “qochib qolish” ma’nosida qo‘llanadi.

Yemoq/yemaslik voqelikka mos gap yoxud narsani ishlatish va qo‘llashni anglatadi.

Yedirmoq “muomalasiga loyiq javob bermoq” ma’nosida ishlatiladi.

Yoqmoq: 01ni/ jinnini so‘zlari bilan tarkibli qo‘llanadi. 01 ni yoqma – toshkentlik bo‘lib olma; jinnini yoqmoq – tushunmaganga olmoq ma’nolarida ishlatiladi. Shu sleng tarkibida dialektal xarakterdagi sleng borligini yana bir bor ko‘rishimiz mumkin.

*Uxlatmoq “*o‘g‘irlab o‘zlashtirmoq”; *uxlatib ketmoq* esa “aldamoq” ma’nosida ishlatiladi. Faqat bu “o‘g‘irlik” chinakamiga bo‘lmaydi. Yangi kiyim olgan odamga ham yaqin o‘z doirasidagilar “qayerdan uxlatding?” tarzida ishlataveradi.

Sochib tashlamoq slengi “biror ishni qoyilmaqom qilib bajarish” ma’nosida keladi.

Kiydirmoq “aldamoq” ma’nosida ishlatiladi. Masalan, kiydirib ketdi – aldab ketdi.

Bugungi kunda o‘zbek tilida qo‘llanadigan bir qator fe’l mazmunidagi slenglarga yana misollar keltirish mumkin: ramkada bo‘lmoq, xaltaga kirmoq, yax bo‘lmoq, bechorani yasamoq, qurimoq, dam bermoq, stolbani ushlatib ketmoq, dimka bermoq, shamol bo‘lmoq, sirpanmoq, xit qilmoq, hit qilmoq//bo‘lmoq, ekib qo‘ymoq, erimoq, pishmoq, yuvmoq, pat(i)ni tashlamoq, tepmoq, chirimoq, bezzvukda gaplashmoq, gromka bermoq, omin bo‘lmoq, hoji ona//hoji akani yoqmoq, lox qilmoq, sindirmoq, quritmoq, bloklamoq, gazini bermoq, trendga chiqmoq, qusmoq, tug‘moq, o‘ramoq, chorniyga tiqmoq//kirmoq, yormoq, shmon qilmoq, yeb qo‘ymoq, qovun tushirmoq, tarvuz yutmoq, sho‘xlik qilmoq, suv quymoq, paxta qo‘ymoq

Foydalanilgan adabiyotlar:

1. Sayfullayeva, R.R. va boshq. Hozirgi o‘zbek adabiy tili [darslik] Toshkent: Fan va texnologiya, 2010. – B. 267
2. Usmanova Sh., Bekmuxamedova N., Iskandarova G., Sotsiolingvistika (o‘quv qo‘llanma), Toshkent: Universitet, 2014. – B. 85
3. Yo‘ldoshev I., Sharipova O‘., Tilshunoslik asoslari (o‘quv qo‘llanma), Toshkent: Iqtisod-Moliya, 2007. – B. 192
4. В. И. Беликов, Л. П. Крысин. Социолингвистика. Учебник. Москва, 2001. 317 ст.
5. O‘zbek tilining izohli lug‘ati (gov.uz)
6. Oxford English Dictionary. Oxford University Press. Retrieved March 4, 2010.
7. https://dictionary.cambridge.org/dictionary/english/slang

IZOGLOSSA

Po‘latova Sevara
Filologiya fanlari nomzodi (PhD),
O‘zbekiston milliy universiteti.
E-mail: sevaramahmudqizi@gmail.ru

Annotatsiya. Ushbu maqolada izoglossa tushunchasi va uning qo‘llanishi haqida ma’lumot berilgan. Izoglossalar yaxlid (agar hodisa ma’lum bir mintaqada qo‘llanilib undan tashqarida uchramasa), uzuq (bir-biridan uzilgan turli mintaqalarda uchrasa), yopiq (ma’lum bir chegaralangan arealdan boshqa joyda uchramasa), ochiq (turli mintaqalarda uchrasa) bo‘lishi mumkinligi ko‘rsatilgan.

Kalit so‘zlar. Tavsifiy dialektologiya, taksonomiya, landshaft, izoglossa, izoliniya, izomorfema, izofonema, izoleksema

Abstract. This article provides information about the concept of isoglossa and its application. Isoglosses are complete (if the phenomenon is used in a certain region and does not occur outside of it), ring (if it occurs in different regions that are disconnected from each other), closed (if it does not occur anywhere but a certain limited area), open (if it occurs in different regions) it is shown that it can be

Keywords. Descriptive dialectology, taxonomy, landscape, isoglossa, isoline, isomorpheme, isophoneme, isolexeme.

Areal lingvistika lisoniy hodisalarning ma’lum hududlarda tarqalishi va shu asosda tillararo aloqalarni o‘rganuvchi yo‘nalishdir. Areal lingvistika asosiy tushunchalari va ish usullaridan biri **izoglossa** (gr. isos o‘xshash, bir xil va gr. glossa til hodisasi) ayni bir lisoniy hodisaning tarqalish chegaralarini belgilovchi chiziq, hudud chegaralaridir.

Izoglossalar bilan uzviy bog‘liq xaritalash va lisoniy (lingvistik) atlaslar tuzish areal lingvistikani lingvistik geografiya bilan yaqin aloqada ekanligini ko‘rsatadi.

Izoglossa arealogik tadqiqotning asosiy tushunchalaridan biri bo‘lib, o‘zaro o‘xshash hodisa (qonuniyat)ning tarqalish (qo‘llanish) nuqtalarini va bu nuqtalarni tutashtiruvchi chiziqni hamda hodisaning tarqalish chegaralarini xaritada belgilovchi chiziq yoki belgi.

Izoglossani nima asosida o‘tkazish (izoglossa mazmuni, tarqalishi, o‘rganilayotgan hodisa yoki qonuniyat) arel tilshunoslikda juda katta ahamiyatga ega. Tavsifiy (dialektolografik) ishlarda, ya’ni ma’lum bir hududdagi shevalar tavsiflanganda, odatda, shu shevaning o‘lchov mezoni sifatida qabul qilingan etalondan yoki boshqa bir shevadan farqlovchi katta-kichik har qanday farq qayd etib boriladi. Tavsifiy ishlarda, ko‘p hollarda, bu hodisalardan qay biri qay darajada ahamiyatli ekanligi haqida tasavvurga ega emas. Shuning uchun ko‘p hollarda masalan: ma’lum bir shevaning o‘ziga xos xususiyatlari, sifati dialektologik 200-300 turdagi o‘ziga xoslikni qayd etsa, arealog shu hodisalardan eng ahamiyatli ma’lum bir shevaning o‘ziga xos antologik xususiyatda aks ettiruvchi 2-3 belgini (hodisa, qonuniyatni) izoglossa mazmuni sifatida tanlaydi va shu hodisaning tarqalish makonini (arealini) aniq belgilashga harakat qiladi. Lisoniy hodisaning (qonuniyatning) tabiatiga muvofiq ravishda izoglossa har xil, jumladan, fonetik fonologik, morfemik, morfologik, leksik, sintaktik bo‘lishi mumkin.

Odatda izoglossa bitta hodisa asosida belgilanadi. Agar ikki hodisa bir-biri bilan (sabab-oqibat munosabati) o‘zaro bog‘langan bo‘lsa, bunda ikki hodisa bitta izoglossa bilan berilishi mumkin (masalan: umuturkiy “a” ning yarim lablangan “o”ga o‘tishi va tilda “o” fonemasining mavjudligi). Agar hodisalar orasida sabab-oqibat munosabatlari mavjud bo‘lmasa, bunday hodisalar alohida-alohida izoglossalar bilan berilishi lozim (masalan: tilda yarim lablangan “o” fonemasining mavjudligi va umuturkiy “a”ning “o”ga o‘tishi. Bunday holatda ushbu tildagi “o”ning umuturkiy “a”dan kelib chiqishi shart emas. Vaholanki, birinchi holatda “o” faqat umuturkiy “a”dan kelib chiqadi. Shuning uchun arealogik tadqiqotlar tavsifiy dialektologik tadqiqotlar nihoyasiga yetar davridagi bosqich sanaladi. Aytish mumkinki, arealogik tadqiqotlar tavsifiy dialektologik ishlarning navbatdagi bosqichidir. Ya’ni tavsifiy dialektologiya ma’lum bir shevaning chegaralarini va o‘ziga xos hodisalarni aniqlashni o‘z oldiga maqsad qilib qo‘ysa, arealogik tadqiqotlar shevalar orasidagi hududiy chegaralarni olib tashlaydi. Chunki turli mintaqada amalda bo‘lgan turli shevalarda qo‘llanilgan bir xil hodisa (qonuniyat) arealogik tadqiqotlarda mintaqaviy-shevaviy mansublikdan qat’iy nazar bir arealga birlashtiriladi. Chunonchi: so‘z boshida jarangsiz “t” tovushining jarangli “d” fonemasiga o‘tishi. Umuman olganda, jaranglilashish turkiy tillarning o‘g‘uz guruhiga mansub tarmoqlari uchun xos hodisa sanalsada, umuturkiy *t~d* hodisasi, birinchidan, o‘g‘uz guruhiga kiritiladigan *turk, ozorbayjon, turkman,* shevashunoslari bu tillarning juda ko‘plarida qadimgi turkiy “t”ni “d”ga o‘tmasligini qayd etadi. Ikkinchidan, umuturkiy “t”ni “d”ga o‘tish hodisasini o‘g‘uzlardan ancha uzoq bo‘lgan *totor, boshqird, qirg‘iz, o‘zbek* tilining amalda bo‘lgan mintaqalaridagi shevalarida ham uchratish mumkin. Shuning uchun dialektologlarning bergan oddiy tavsiflari ichida o‘ta ahamiyatlilarini til yoki shevaning eng xarakterli xususiyatlarini izoglossa asosida tanlash va o‘rganish talab qiladi. Shuning uchun A.Boltaev, G.Blagovalarning ishlarida o‘zbek dialektologik tadqiqotlarida shevaning xususiyatlari sifatida ajratilgan juda ko‘p sonli belgilar aksar holatlarda shevaning mohiyatini belgilash uchun ikkinchi darajali belgilar bo‘lib, izoglossa vazifasini o‘tay oladigan eng asosiy belgilar bu hodisalar ichida yo‘qolib ketishini qattiq tanqid qilishadi.

Shevalar orasidagi chegaralarni yo‘qotib, ma’lum bir hodisaning tarqalish mintaqalarini aks ettiruvchi izoglossa mazmuni (belgisi) til tizimida ahamiyatli hodisa bo‘lishi ma’lum bir guruh (urug‘, elat, lahja, til…) yoki bir necha xususiyatlarni belgilovchi hodisa (qonuniyat) bo‘lmog‘i lozim. Hozirgi o‘zbek tilshunosligi oldidagi eng muhim vazifalaridan biri ham mana shu bilan belgilanadi; o‘zbek shevalari, turkiy tillar orasidagi o‘xshashlik va farqlar behad ko‘p. Lekin ularning barchasi ham izoglossa mazmuni (asosi, belgisi) vazifasini o‘tay olmaydi. O‘zbek shevalaridagi jo‘qichilar, galdirlar, yo‘qchilar deb qipchoq, o‘g‘uz va qarluq-uyg‘ur lahjalarini atigi bitta xususiyati − so‘z boshida *y~j, k~g* almashinuvi bilan ajrata olish ham yaqqol ko‘rsatib turibdi. Ana shu xususiyat (qonuniyat) (*“k”*ning *“g”*ga o‘tishi kabilar ham) qipchoq, o‘g‘uz va qarluq-uyg‘ur shevalari orasidagi muhim belgi sanalishi mumkin. Chunki bu xususiyatlar leksik farqlar bilan ham uzviy bog‘liqdir. Shuning uchun arealogik tadqiqot o‘tkazishdan oldin har bir tadqiqotchi har bir izoglossa mazmunini (asos, belgisi) til tizimidagi taksonomik (tasnif etish uchun ahamiyatli bo‘lgan) ahamiyatini asoslab bermog‘i lozim. Tabiiyki, arealogik tadqiqot o‘rganilayotgan arealning to‘liq tasnifini ham o‘z ichiga olishi kerak.

Izoglossalar yaxlid (agar hodisa ma’lum bir mintaqada qo‘llanilib undan tashqarida uchramasa), uzuq (bir-biridan uzilgan turli mintaqalarda uchrasa), yopiq (ma’lum bir chegaralangan arealdan boshqa joyda uchramasa), ochiq (turli mintaqalarda uchrasa) bo‘lishi mumkin.

Areal va izoglossalarni xarita shaklida berish shart emas. Izoglossa va areal toponimlarini sanab o‘tish, ro‘yxatda berish usuli bilan ham ifodalash mumkin. Lekin xaritalashtirilgan izoglossa va areal haqiqiy lisoniy manzarani − lingvistik lanshaftni beradi. Mohiyatan atama bo‘lgan lanshaft (*nems*. landchaft − manzara, ko‘rinish) atamasi arealogiyada juda keng qo‘llaniladi.

Har bir kartani tuzishda dastlab kartalashtirilgan til hodisasining tarqalish chegarasi — izoliniyasi aniqlanib chiqilishi kerak. So‘ngra, har xil sheva va dialektlarda mavjud bo‘lgan bir xil til hodisalari izoglossa sifatida kartalashtiriladi. Ba’zi hududlarda izoglossalarning tarqoq holda uchrashi qachonlardir aholining bir joydan ikkinchi joyga ko‘chishining natijasi ekanligidan dalolat beradi. Ayrim dialektlarda leksik izoglossalar yig‘indisida noaniqlik uchrasa, ba’zi dialektlarda esa, fonetik izoglossalar ayniqsa, zich joylashgan bo‘ladi. Morfologik izoglossalar zichligi leksik va fonetik izoglossalar zichligi orasida bo‘lib, keyingisiga yaqinroq turadi.

Bu xil izoglossalarning ko‘proq massivga tarqalishida qadimda aloqa yo‘llari va daryo oqimi katta rol o‘ynagan bo‘lishi mumkin. Kartalashtirilgan hudud turli tomonga tarqalgan izoglossalarni o‘z ichiga olganda har xil lingvistik hodisalarning tarqalishi ba’zan bir-biriga mos kelmay qolishi mumkin. U holda o‘xshash bo‘lmagan ayrim gruppa izoglossalar bir-biriga juda yaqin yo‘l bilan o‘tib, izo- glossalar to‘dasini, yig‘indisini tashkil etgan joylarni aniqlash lozim bo‘ladi.

Umuman, xulosa qilib aytish mumkinki, mos hodisalarning u yoki bu a’zolari, bo‘lagi tarqalgan eng chet nuqtalarni tutashtiruvchi atlas kartasidagi chiziq izoglossadir.

Foydalanilgan adabiyotlar:

1. Akhmedov A. R., Sevara P. Theoretical and practical issues of creating uzbek areal linguistics //Middle European Scientific Bulletin. – 2021. – T. 12. – C.142-146.
2. Pulatova S.M. Unas of modern usbek literary norms (on the example of Vobkent district). T., international journal of word art volume 4, issue 1. 2021.
3. Pulatova S.M. Areal linguistics. Globe. 2020.
4. Rahmatovich A. A. Po‘latova S.M. Vobkent tumani shevasida so‘zning morfemik tarkibi //scientific reports of bukhara state university. – c. 123.
5. Akhmedov A. R., qizi Po‘latova S. M. O ‘zbek shevalarini lingvo-geografik xaritalashtirish //Conferences. – 2021.

T. JUMAMURATOV ASARLARI TILIDAGI PRAGMATIK BELGILAR

Kazaxbaeva Anargul Tolıbaevna
Filologiya fanlari nomzodi, docent
Berdoq nomidagi qoraqalpoq davlat universiteti.

Annotation: Due to the wide range of social experience, the communication processes are changing, and the pragmatic structure, which was originally structured in a natural way, is changing and becoming more complex. This, in turn, leads to the continuous expansion of the semantic shell of pragmatics. Therefore, it is necessary to start linguistic analysis from pragmatics.
Keywords: pragmatics, hidden meaning, spoken language, composition (work), poet.

Аннотация: В связи с широким спектром социального опыта коммуникативные процессы изменяются, а прагматическая структура, изначально структурированная естественным образом, меняется и усложняется. Это, в свою очередь, приводит к постоянному расширению смысловой оболочки прагматики. Поэтому лингвистический анализ необходимо начинать с прагматики.
Ключевые слова: прагматика, скрытый смысл, разговорная речь, произведение, поэт.

Insonning nutqi uning ishonchini belgilaydi. Inson o‘z nutqi orqali obektiv dunyoga munosabatini bildiradi, shu bilan birga milliy o‘zligini ham gavdalandiradi. Insonning dunyo qarashi, ongi, tuyg‘ulari, qabul qilishi, bilimi, hayotiy tajribasi, iymon-e’tiqodi, umumiy insonning jamiyat faoliyatida ko‘rinadigan siyosiy, falsafiy, aqliy, diniy va estetik ko‘z-qarashlari jamlangan nutqining tekshirilishi va pragmatik analizi insonlarning bugungi mustaqillik davrida qizǵın mayishiy sohalardagi o‘rnini belgilashda ahamiyatli hisoblanadi. Shu manoda, insonnin jamiyat faoliyatini o‘zida to‘liq gavdalantirilgan “harakat”dagi, “jonli” nutqning inson ongi bilan bog‘liq siyosiy va til birligini o‘rganishda katta ahamiyatga ega.

Odatda, insonning nutq faoliyati uning ijtimoiy hayotdagi barcha faoliyatlar kabi, muayyan maqsadga yo‘naltirilgan bo‘ladi. U bayon qilinayotgan fikr muayyan maqsadni nazarda tutgan, o‘z tuyg‘ulari, munosabatini tinglovchiga yetkazish, shuningdek suhbat muayyan kommunikativ harakatda qo‘zg‘ash, unga ta‘sir qilish natijasida uni navbatdagi nutq jarayoniga tortish va h.

Haqiqattan da, qandayda bir fikr yuritishi uchun xizmat qiluvchi gapning asosida so‘zlovchining ko‘zlagan maqsadi, niyati yashiringan bo‘ladi. Demak, so‘zlovchi har bir aytgan gapini muayyan bir maqsad bilan bog‘laydi va u kommunikatsiyada shu gapning pragmatik mazmuni sifatida shakllanadi.

Til biliminig so‘ngi yillardagi muvaffaqiyatlarining biri tilga butun bir tizim sifatida qarash va uni nutqga qarama-qarshi qo‘yib o‘rganish til birligining nutq faoliyatida matn va sharoit bilan bog‘liq holda yuzaga kelishi pragmatik halotlarga aniqlik kiritdi. Til birligini pragmatik yo‘nalishda o‘rganish esa, asar yoki matn mazmunini o‘quvchiga tog‘ridan-tog‘ri ko‘zga tashlanmaydigan, faqat nutqda yuzaga chiqadigan ishki imkoniyatlarni aniqlashga yo‘l ochdi. Natijada, til birligining nutq jarayoni bilan bog‘liq holda real pragmatik vazifalarini aniqlash vazifalarni aniqlash muammosi yuzaga keldi.

Bizning maqolada ana shu yo‘nalishdagi tadqiqot ishi bo‘lib, unda shoir T.Jumamuratov asarlari tilining pragmatik belgilarni analiz qilish bilan ahamiyatli hisoblanadi.

«Til egasi» tushunchasi hozirgi kunda til bilimida quyidagi ma’nolarda qo‘llanilmoqda: 1) muayyan tilda nutq faoliyatini amalga oshirish, yani nutq va uni qabul qilish qobiliyatiga ega bo‘lgan shaxs; 2) tildan muloqat vositasi sifatida foydalanuvchi shaxs, kommunikant; 3) o‘z millatining milliy-madaniy, ma’naviy qadriyatlarni gavdalantiruvchi lug‘aviy tarkibini egallagan, uni yuzaga chiqaruvchi shaxs, muayyan til vakili. Badiiy asarni pragmalingvistik tahlil qilishda ana shu «til egasi» asosiy faktor hisoblanadi. Badiiy matn yaratishda ijodkor faktorini tadqiq etish muammosini chuqur oshib berishga tog‘ri keladi.

Qaysidir yozuvchi yoki shoir stilini, uning so‘z qo‘llash yoki gap tuzish mahorati nuqtai-nazaridan matn tadqiq etish usuli bugungi kun talabiga javob bermaydi. Shu sababdan, badiiy matn yaratish masalasi shaxs faktori bilan birga o‘rganish badiiy matnning lingvistik tomonlari teran nazar bilan qarashga tog‘ri keladi.

Bunda yozuvchi va o‘quvchi munosabati ijtimoiy xususiyatga ega bo‘lib, bunda ma‘lumot uzatish «yuki» til birliklari «yelkasiga» tushadi. Biroq, shaxslararo munosabat normasi, milliy madaniy qadiriyatlash tizimini egallamasdan, to‘liq aloqaga kirishish imkoni yo‘q. Yani, kommunikativ akt turli bilimlarga ega bo‘lishni talab qiladi, masalan, til bilimini til tuzilishi bilan bog‘liqli bo‘lsa, ensiklopedik bilim bo‘lsa, bo‘lmishni gavdalantiruvchi interaktiv bilim esa o‘zaro munosabat asosida turgan, ijtimoiy guruh uchun umumiy mavjud bo‘lgan qoidalar to‘plamini bilishni talab qiladi.

Sulıwlıq ta jarasıqlı, sırttağı teri...
Biraq júrek say kelmese, kim payda tabar?
Aqılsızdıń sol boladı, soǵılar jeri,
Sonnan ájel tapqanınday qundız hám qamar.

Berilgan parchadagi so‘z etilgan qunduz (qundız) va qamar (qamar) terisi juda chiroyli bo‘lganligi uchun ov qilinadigan va qimmatbaho ekanligi haqida o‘quvchi oldindan ma‘lumotga ega bo‘lmasa, she’r mazmunini to‘liq tushuna olmaydi

Badiiy matnning umumiy tuzulishi, mazmuni til birliklari kommunikativ ma’no shakllantiruvchi tajriba maydoni hisoblanadi. Butun kommunikativtil tuzilma bo‘lgan matnning semantik tarkibi denotative hám signifikativ bo‘laklardan iborat. Bundan birinchisi matn mazmunining bolayotgan hodisalar bilan bog‘liq taraflariga iqtibos bersa, ikkinchisi esa matn ijodkorning nutq mahorati bilan bog‘liq. Masalan,

Ótedi kárwan, mashina tınbay,
Asfaltlı jollar – mısalı ayna.
Kósheler dúziw, tartılǵan sımday,
Jarasqan gileń qalpaqlı jayǵa [3.59].

Haqiqatdan ham, uyning tepasidagi shifer ko‘rinishi qalpog‘ni (qalpaq) eslatadi. Yurtimizda uy-joylarning tom qismini shifer qilish XX asrning 50-yillaridan boshlab urfga kirdi. Mana shu yangiliklarni bildirish maqsadi da to‘g‘ri ko‘chalarning mazmuni ostida yashirin turda uylarning shifer qilinganligini tushuntirmoqchi bo‘ladi. Yani, «Uylar shifer qilingan» deb yozilmasa ham gap bilan o‘z-o‘zidan uylarning shiferli ekanligi tushuniladi.

Til birliklari kontekst, nutq sharoiti, til egalarining tashqi dunyo haqidagi umumiy bilim va konlikmalari kabi faktorlar bilan bog‘liq holda qo‘llanilganda ularning funksiyaviy imkoniyatlari yanada kengayadi. Shunday qilib, til birliklarining ma‘noviy xususiyatlari nutqda

gavdalanadi, aniq yo‘nalish oladi. Har bir gapiruvchi shaxs real hayotda bo‘layotgan hodisalarni o‘zicha qabul qiladi, uni belgili joy, vaqt oralig‘ida tushunib, hosil bo‘lgan tushunchaga libos beradi. Lug‘aviy ko‘rinish va mazmun shakllanishi nutq muallifi tanlagan hodisani vaqt-joy maydonida joylashtirish bilan bog‘liq.

Masalan, suhbatdoshga yoki manzilga bildirilayotgan fikr yoki yashirin yoki oshiq, ijobiy yoki salbiy, to‘g‘ridan-to‘g‘ri yoki majoziy va boshqa ko‘rinishda bo‘lishi mumkin.

Hámel menen jazıqsızdı tırnaysań,
Sestiń bálent, ishiń gewek sırnaysań.
Jaramsaqlar jalǵan maqtar kózińshe,
Shının aytsam, bes tıyınǵa turmaysań [3.37].

Bunda manzilga bo‘lgan so‘zlavchining fikri ochiq-oydin bayon qilingan. Biroq, fikr hardoim binday ochiq turda bildirila bermaydi

Shóbin shapqan alańlıqta gúl óser,
Sonlıqtan jurt otaq penen gúreser,
Adam qurtsa adamdaǵı sheńgeldi,
Sonda ǵana ǵalawıttı túweser [3.114].

T.Jumamuratov XX asr qoraqalpoq adabiyotining yirik vakillaridan biri. Shoirning ta’sirshan xarakterdagi, ashshiq haqiqatlarni teran oshib beruvchi, didaktik mazmundagi she’rlari kitob iqlosmandlarida katta ta’sir qoldiradi.

T.Jumamuratov asarlari tili oddiyligi, badiiyligi, obrazliligi bilan o’ziga tortadi.

Kók teńizde suwdıń betin jaynatıp,
Tolqınlardı jarǵanda kór tamasha,
«Motor» degen arǵımaqtı oynatıp,
Tereńge aw salǵanda kór tamasha [3.17].

Shoir bunda dengizda baliqchilarining baliq ovlash, kemani suzish harakatlarini badiiy jihatdan yetkazib beradi. She’rda kema so‘zi qo‘llanilmagan bo‘lsa ham, kemaning xarakterli belgilarini (suv yuzasini yashnatish, to‘lqinlarni parchalash, motorni o‘ynatish) berish orqali suvda kemaning suzip ketayotganligi bayonlanmoqda.

«Dengiz bilan o‘ynamang» («Teńiz benen kóp oynamań») she’rida bazibir baliqchilarning o‘z vazifalariga javobgarchilik bilan befarvo qarashi, o‘z manfaatlari yo‘lida, jamiyat ishlarining bajarilmasligi tanqid qilinadi.

Olardıń hesh xabarı joq,
Bul júristen tabarı joq,
Maylı úyrek jegen menen,
Planına tamarı joq [3.21].

Bunda planning bajarilmasligi ularni yeb yurgan taomlari ma’nosida yuklab tushuntirmoqda.

Danıshpan dep júr ekenbiz mańbastı,
Hayran bolıp qıyallarım aljastı,
Keshegi kún jurttıń bári maqtaǵan,
«Perishteler» búgin kórseń albaslı,

Deb jamiyat turmishidagi hodisalarni, odamlar xarakterini farishta va alvasti so‘zlarini qo‘llash orqali berilgan. Bunda farishta so‘ziga alohida ma’no yuklatilgan. Yani, farishtaga begunoh, pok, halol odamlar tenglashtiriladi. Lekin, bunda shu kunga qadar yaxshi deb yurgan

odamlarning bugungi kunda haqiqiy ko‘rinishini bildik degan maqsadda aytilgan. Kommunikativ maqsad ushbu farishta va alvasti so‘zlari orqali ifodalangan.

Balıq penen qusı boldı kólinde,
Gúnde menen ketpen turdı jerinde.
Jasıl giya – mallarınıń azıǵı,
Oraq turdı qıstırılǵan belinde [3.50].

Bu she’r parchasida shoir xalqimizning chorvachilik va dehqonchilik bilan shug‘illanganligini to‘g‘ridan-to‘g‘ri aytmasdan, badiiy ko‘rinishda ularning belgilari orqali tasvirlangan. Bu shoirning fikrni yetkazish, so‘z qo‘llash mahoratidan darak beradi.

Túyek basqan ógizlerdiń tuyaǵı,
Bárin oylar shayırdıń da qıyalı.
Gáwhar tuwǵan qozısınıń ómirin,
Qoy biyshara adam ushın qıyadı [3.59].

Bu she‘rda T.Jumamuratov qora ko‘l qoylarini nazarda tutib, qo‘zichoq gavharga tehglashtiriladi. Demak, shoirning maqsadi qamali (qamalı) terili qo‘zichoq bahosini bildirish. Qora ko‘l qoylarining qo‘zichoqlarining terisi qimmatbaho hisoblanib, qo‘zichoq tug‘ilgandan keyin ikki kun ichida terisini shilib olinishini, odamlar uchun ushbu gavhar qo‘y yaratilishidan bexabar odam bu qatorlarning ma’nosini tushunmaydi. Demak, gapda pragmatik ma’no yuklash uchun xabarbi qabul qilivchining pragmatik qobilyati bo‘lishi kerak ekan.

Sonday-sonday urlıqlardıń túri kóp,
Ayta bersek, jırlaytuǵın jırı kóp.
Jańa tuwǵan buzawlardıń ornına,
Uzatılıp ketetuǵın «iri» kóp [3.62].

Bunda shoir o‘sha vaqtdagi jamiyatda, ijtimoiy hayotida, ayrim rahbarlar befarvolik qilishi, vazifalarni o‘z vaqtida bajarmaslik oqibatida planlarning bajarilmay qolishi va natijada ularning davlatga yolg‘on hisobat berishini, qo‘shib yozishlarni qattiq tanqidga oladi. Bu oshkoralik fikrlarni to‘g‘ridan-to‘g‘ri aytmasdan, badiiy til bilan, ixcham so‘zlar asosida ochib beradi.

Galileo Galiley Italiyaning Piza shahrida yozgan «Dialogo» asarida inson o‘zining «eng yashirin sirlarni ham boshqa birovga yigirma to‘rt belgining har turli birikishi orqali» yetkiza olishidan hayron qolganligini yozgan edi. Tildagi bunday cheksiz imkoniyatni Por-Royal filosof-grammatistlar «Galileyning ajoyib kashf etishi» deb nomlagandi.

Júregim soqtı dúrsildep,
-Oyınıń seniń qurısın,- dep,
Kózin maǵan qadadı:
Raxmet saǵan, qırshın,- dep [3.78].

She‘rda qirchin, qarg‘ish (qırshın ǵarǵıs) ma‘nosini bildiruvchi so‘z orqali qizning yigitga tuyg‘ularini bildirgan. Bu o‘quvchiga bir o‘qishning o‘zida seziladi.

Haqiqatdan ham shoir ajoyib o‘xshatish orqali jinoiy ish oshib bermoqda.

Didaktemalar - naql-maqollar, aforizm, ma‘noli gaplar, rivoyatlar, turli hayotiy voqealar asosidagi hikoyalarni didaktemalar sifatida umumlashtirish mumkin. Sababi, ularning mazmunida tarbiya, targ‘ibot, didaktika yotadi. Gapda didaktemalarni berish orqali ham muallif qo‘shimcha ma‘no bera olishi mumkin

Badiiy matndagi muayyan maqsadda qo‘llaniladi. Didaktemalarning matn ichida singdirilish sifati ham xar xil bo‘ladi. Ayrim holatlarda muallif didaktemalarni matnga to‘g‘ridan-to‘g‘ri olib kiradi va uning qanday janrga tegishli ekanligini aytib beradi. Masalan:

Eshki sawıp kúnin kórgen paqırlar,
«Sawın sıyır – bulaq» degen naqıl bar.
Tuw kóterip dushpan dúbep kelgende,
Tulpar minip jaw qaytarǵan batırlar [3.75].
Ayırım jaǵdayda, tikkeley kirgizip jibere beredi.
Jalataydan saq bol, márt bolsań eger,
Aqmaqqa jantassań, abıroydı tóger,
Jamanǵa jaqsılıq – basıńa bále,
Eshekler semirse, iyesin teber [3.89].

Xulosa qilib aytganda, T.Jumamuratov asarlarida fikrni yetkazib berish, qanday da bir hodisa yoki holatni tariflashda pragmatik belgilarni ko‘plab uchratish mumkin.

Jamiyat tajribasining keng ko‘lamliligi sababli, kommunikatsiya jarayonlari o‘zgarib turadi hamda boshlang‘ish holatda oddiy tuzilishda bo‘lgan pragmatik struktura o‘zgarib, qiyinlashib boradi. Bu esa, o‘z navbatida pragmatikaning to‘xtamasdan semantik qobig‘i kengayib borishiga olib keladi. Demak, til analizni pragmatikadan boshlash zarur.

Foydalanilgan adabiyotlar:

1. Karimov I.A. Yuksak ma‘naviyat – yengilmas kuch. – Toshkent, «Ma’naviyat». 2008.
2. Азнаурова Э.С. Прагматика художественного слова. – Тошкент, Фан, 1988.
3. Jumamuratov T. Ómirińniń ózi filosofiya. Nókis, 2005.
4. Mamajonov A. Abdupattoev M. Matn sintaksisi. Farg'ona, 2002.
5. Урисбоева Г. Производные слова в синтаксической и прагматической функции в современном русском языке: Автореф. дисс… канд. фил. наук. – Тошкент, 1988.
6. Safarov Sh. Pragmalingvistika. Toshkent, 2008.
7. Тоҳиров З. Ўзбек тилида прагматик семаларнинг ҳосил бўлиши ва уларда субъектив муносабатларнинг юзага чиқиши: Филол. фан. номз… дисс. автореф. Тошкент, 1994.
8. Yo'ldoshev Marufjon. Badiiy matn lingvopoetikasi. Toshkent, 2008.

International scientific-practical conference: Vol.2 / No. 22.04(2024).
«CONTEMPORARY TECHNOLOGIES OF COMPUTATIONAL LINGUISTICS - CTCL.2024».
Proceedings homepage: https://www.myscience.uz/index.php/linguistics

QO‘SHMA GAPLARDA TRANSFORMATSIYA: NOMINATIV BIRLIK (CHEGARALI QURILMA) LARDA POG‘ONAVIYLIK

Sulaymanova Dilobar
tayanch doktorant,
Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti.
E-mail: dilobar7796@gmail.com

Annotatsiya: Ushbu maqolada markaz va qurshov munosabatining sintaktik sathda, xususan qo‘shma gaplarda namoyon bo‘lishi, gap bo‘laklari iyerarxiyasi haqida so‘z yuritiladi. Turkiy tillarda yuzaga keladigan sintaktik-struktur jarayon-nutqiy gaplarning bitta nominativ birlikka transformatsiya etilishi va boshqa bir gap tarkibida bitta konstruktiv bo‘lak yoki mazkur konstruktiv bo‘lakning kengaytiruvchisi mavqeyida kelishi tahlil qilinadi. O‘zbek tili grammatikasining muhim sintaktik-struktur xususiyatlardan biri- transformatsiya jarayonida gap tarkibidagi komponent (aktant)larning pog‘onali joylashuvi masalasi hamda ushbu iyerarxik tahlilda gap va so‘z kengaytiruvchisining maqomi oydinlashtiriladi. Gapning LSQida so‘zlar bog‘lashuvini tahlil qilish orqali aktantlarning gap qurilish sathidagi vaziyati aniqlashtiriladi, gapning til bosqichidagi lisoniy gap bo‘laklari (tarkibiy qismlari) o‘rtasidagi semantik, grammatik bog‘lanish masalasi oydinlashtiriladi.

Kalit so‘zlar: qo‘shma gap, transformatsiya, nominativ birlik, aktant, gap va so‘z kengaytiruvchisi

Аннотация: В данной статье рассматривается проявление связи центр-окружение на синтаксическом уровне, в частности в сложных предложениях, и иерархии предложений. Анализируется трансформация синтаксически-структурных процессуально-речевых предложений в турецких языках в одну номинативную единицу и наличие одного конструктивного предложения или расширителя этого конструктивного предложения в другом предложении. В процессе трансформации, которая является одной из важных синтаксически-структурных особенностей грамматики узбекского языка, уточняются вопросы ступенчатого расположения компонентов (актантов) в предложении и статус расширителя предложения и слова в этом иерархическом анализе. Положение актантов на уровне построения предложения выясняется путем анализа связи слов в ЛСО предложени, анализируется вопрос семантической и грамматической связи между частями (компонентами) языкового предложения на языковой стадии предложения.

Ключевые слова: сложное предложение, трансформация, номинативная единица, актант, расширитель предложения и слова.

Abstract: This article highlights the manifestation of the center and subordinate relationship at the syntactic level and the hierarchy of components, especially in compound sentences. The phenomenon in Turkish languages - the transformation of a sentence into one nominative unit and the occurrence of one constructive clause or an expander of one constructive clause in another sentence is analyzed. One of the syntactic-structural features characteristic of the grammar of the Uzbek language- the issue of the transformation process in the hierarchy of sentence fragments and the status of the sentence and word expander in the hierarchy are clarified

in stepwise syntactic analysis. The position of the actants at the level of sentence construction is clarified by analyzing the connection of words in the LSQ of the sentence, the issue of semantic and grammatical connection between the parts (components) of the linguistic sentence at the language stage of the sentence is clarified.

Keywords: compound sentence, transformation, nominative unit, actant, sentence and word expander.

Kirish

Ma'lumki, shakl va mazmun (semantika) yorqin namoyon bo'ladigan sintaksisdagi muhim masalalaridan biri bu-gapning sintaktik strukturasi, undagi komponentlarning ma'lum izchillik asosida pog'onali joylashuvi hamda shu asosda markaz qurshovlilik munosabatining yuzaga kelishi hisoblanadi. Aynan shu semantik-struktur izchillik va iyerarxiya munosabati tufayli borliqning inson tafakkuridagi in'ikosi yuzaga keladi. R.A.Budagov: «Sintaktik munosabatlarning o'zi to'laligicha ma'noga asoslanadi», — deb ta'kidlagan edi [Budagov, 1974; 93]. Darhaqiqat, gap so'zlovchining muayyan fikrini tinglovchiga yetkazish maqsadi bilan tuziladigan sintaktik butunlik sanaladi.

Har qanday lisoniy birlik shakl va ma'no birligidan iborat. Yuqori sath birligi bo'lgan sintaktik birliklar ham bundan mustasno emas. Shunday ekan, barcha lisoniy birliklarni bu ikki tomonning o'zaro qanday munosabatda ekanini o'rganish, ular o'rtasidagi muvofiq va nomuvofiq jihatlarni yoritish katta ahamiyatga ega (Asqarova, 1960: 66).

1. Qo'shma gaplarda iyerarxiya: markaz va qurshov munosabati

Qo'shma gap- sodda gap bilan chambarchas aloqador bo'lsa-da, undan grammatik asoslar soni, fikriy mundarija (asosiy axborot)ning sifat va miqdori nuqtayi nazaridan farqlanuvchi sintaktik butunlik. Struktur jihatdan sodda va qo'shma gap o'rtasidagi tafovut aniqki, kesimlik belgisi ularni farqlovchi asosiy mezon sanaladi.

Qo'shma gap lisoniy birlik sifatida murakkab tabiatli bo'lib, bu murakkablik uning lisoniy sathda tutgan o'rni, tarkibidagi sodda gapning grammatik shakli va bog'lovchi vositalari hamda mazmuniy munosabatining rang-barangligi bilan belgilanadi (Sayfullayeva, Mengliyev, Qurbonova, Abuzalova, Boqiyeva, Yunusova, 2009: 395).

Qo'shma gapning sintaktik qurilishi juda chuqur tadqiq qilingani, shuningdek, sintaktik qurilish aniq ko'rish mumkin bo'lgan sath ekanligi bois qo'shma gap qurilishini belgilash qiyinchilik tug'dirmaydi. Bunda qo'shma gapni shakllantiruvchi komponentlar o'rtasidagi munosabat va uning sintaktik ifodalanishini belgilashda mazkur munosabatni shakllantiruvchi mantiqiy-mazmuniy mexanizm va bu mexanizmning sintaktik tuzilishga nisbatlanishiga alohida e'tibor qaratiladi.

Ma'lumki, qo'shma gapning ma'noviy va struktur tasnifi ko'p holda o'zaro farqlanadi. Formal sintaksis qo'shma gap tasnifida ko'p hollarda ularning nutqiy belgi-xususiyati, zohiriy alomatiga tayanadi. Substantsial sintaksis tasnifi esa qo'shma gapning lisoniy mohiyatidan kelib chiqqan holda amalga oshiriladi.

Keyingi yillarda qo'shma gaplar tarkibidagi sodda gaplarning qurilish qolipiga ko'ra tasnif keng qo'llanilmoqda. Qo'shma gaplar tasnifida ularning tarkibidagi sintaktik aloqa va munosabatlarni alohida hisobga olish lozim bo'ladi.

Sodda, uyushgan va qo‘shma gaplarni farqlovchi asosiy belgi-[P_m] ning markaz va qurshov munosabatining tashkil topishida o‘rni beqiyos ekanligini ta’kidlab o‘tish lozim. Chunki sodda gapning eng kichik qolipi sanalgan [WP_m]da o‘z atrofiga gap kengaytiruvchilaridan iborat yagona markazni shakllantirsa, qo‘shma gapda mos ravishda bunday markazlar (iyerarxiya daraxti) soni [P_m] soniga ekvivalent ravishda ikkidan boshlab cheksiz davom etishi mumkin:

$$[WP_m^1 + WP_m^2]$$

$$[WP_m^1 + WP_m^2 + WP_m^3]$$

$$[WP_m^1 + WP_m^2 + WP_m^3 + \ldots + WP_m^n]$$

Qo‘shma gapning lisoniy-sintaktik qolipini tahlil qilish orqali uning voqelanish usullari turli ko‘rsatkichlarga asoslanishini ko‘rish mumkin:

Jumladan, gapning fikriy mundarijasini ifodalovchi grammatik asoslar miqdoriga (ikki bosh bo‘lakli yoki bir bosh bo‘lakli ekanligiga)

a) bir subyekt qo‘shma gapning har ikki qismi uchun umumiy bo‘lgan qolip:

$$S$$
$$[WP_m^1 + WP_m^2]$$

b) ikki subyektli qo‘shma gapning lisoniy-sintaktik qolipi:

$$S \quad S$$
$$[WP_m^1 + WP_m^2]$$

c) subyekt (ega) ifodalanmagan yoki noma’lum bo‘lgan qismlardan tashkil topgan qo‘shma gap qolipi:

$$[WP_m^1 + WP_m^2]$$

Har uchala holatda ham subyektning ifodalanish-ifodalanmasligi, bir yoki bir nechta ekanligi aktantlari gap kengaytiruvchilardan iborat bo‘lgan markaz miqdoriga ta’sir qilmaydi, chunki ushbu markazlar soni subyekt emas, balki [P_m] lar miqdoriga ko‘ra shakllanadi.

2. Qo‘shma gaplarda transformatsiya: nominativ birlik (chegarali qurilma)larda pog‘onaviylik

Ta’kidlash joizki, LSQ nutqiy hodisa sifatidagi gapning muhim grammatik vastruktur (qurilish) xususiyatini o‘zida mujassamlashtirgan mavhum qurilma sifatida yashaydi. Bu gapning struktur elementi orasidagi munosabatni ifodalab, uning umumiy grammatik ma’nosini umumlashtiradi [Sayfullayeva, Mengliyev, Qurbonova, Abuzalova, Boqiyeva, Yunusova, 2009: 331].

1. O‘zbek tilida gap qurilishidagi eng kichik qolip [WPm] sanaladi, shu bois nutqiy voqelikda W-leksik ma’no anglatuvchi qismi vazifasini lisoniy sintaktik qolipning voqelanish jarayonida sintaktik tabiati turli xil bo‘lgan, biroq [WPm] ni yuzaga keltira oladigan istalgan atov birliklar: nominativ birlik-leksemalar, erkin va turg‘un bog‘lanishli birikmalar, kengaygan birikmalar (harakat nomi, ravishdosh, sifatdoshli oborotlar) bajarishi mumkin (Mahmudov, 1984: 41).

O‘zbek tilining grammatikasiga xos bo‘lgan sintaktik-struktur xususiyatlardan biri gap bo‘laklari ierarxiyasidagi *transformatsiya* jarayonidir. Ta’kidlab o‘tish joizki, turkiy tillarda istalgan kattalikdagi nutqiy gap sifatdosh, ravishdosh va harakat nomlari orasida bitta nominativ birlikka transformatsiya etilishi va boshqa bir gap tarkibida bitta konstruktiv bo‘lak yoki bitta

konstruktiv bo‘lakning kengaytiruvchisi mavqeyida kiritilishi mumkin (Qurbonova, 2001: 167). Turkiy tillarning ushbu tipologik xususiyati ilmiy tadqiqi turkolog olima N. Z. Gadjiyeva, o‘zbek tilshunosligida J. Omonturdiyev, A. Mamatov, A. Nurmonov, N. Mahmudov, A. Berdialiyevlar nomlari bilan bog‘liq bo‘lib, ushbu ilmiy tadqiqotlarda istalgan turdagi sodda va havola bo‘lakli qo‘shma gaplarda grammatik asoslar, xususan E va [Pm] qismlarini shakliy o‘zgartirish orqali yakka nominativ birlikka transformatsiya qilinishi, shuningdek, ushbu kengaygan birikmaning gapning konstruktiv bo‘lagi yoki boshqa konstruktiv bo‘lak kengaytiruvchisi (aktanti) bo‘lib kelishi mumkinligi haqidagi asosiy ma’lumotlar berib o‘tiladi. Buni quyidagi havola bo‘lakli qo‘shma va uning kengaygan birikmali (chegarali qurilma, ramochnaya konstruksiya) sodda gap varianti talqini orqali tahlil qilishimiz mumkin:

1.	To‘ldiruvchi ergash gapli havola bo‘lakli qo‘shma gap	*Xalqimiz shuni ta’kidlaydiki, mehnat insonni yuksaltiradi, dangasalik halokatga boshlaydi. (Oybek)*
2.	kengaygan birikmali (chegarali qurilmali) sodda gap	*Xalqimiz mehnatning insonni yuksaltirishi, dangasalikning halokatga boshlashini ta’kidlaydi.*

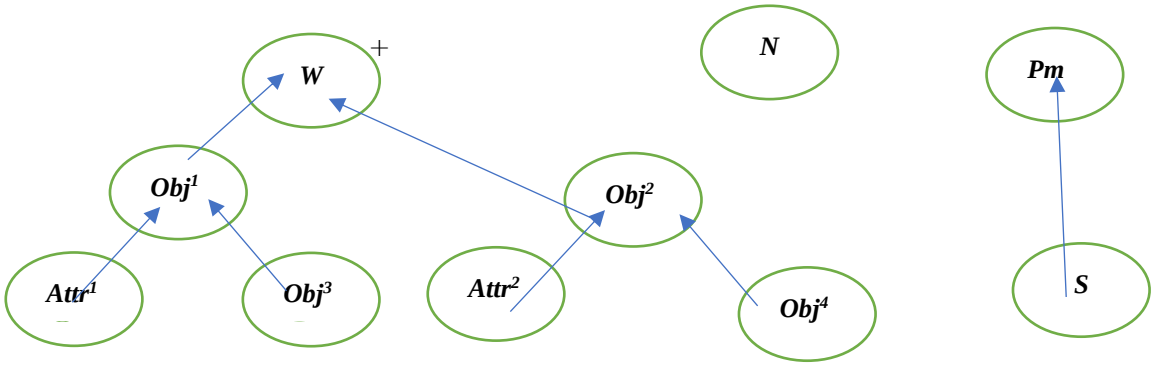
Yuqoridagi havola bo‘lakli qo‘shma gapdan bilishimiz mumkinki, ergash gap butunligicha bosh gap tarkibidagi ko‘rsatish olmoshiga havola qilingani bois sodda gapdagi “*mehnatning insonni yuksaltirishi, dangasalikning halokatga boshlashini*” kengaygan harakat nomi bilan ifodalangan oborotni “*shuni*” ko‘rsatish olmoshiga almashtirish mumkin:

1.	kengaygan birikmali (chegarali qurilmali) sodda gap	[S]	[Obj]	[W]	[N]	[M]	[T]	[P]
		Xalqimiz	***mehnatning insonni yuksaltirishi, dangasalikning halokatga boshlashini***	*ta’kidla*	-	*-y, -di*	*-y*	*-di*
2.	Ko‘rsatish olmoshili sodda gap	*Xalqimiz*	***shuni***	*ta’kidla*	-	*-y, -di*	*-y*	*-di*

Har ikkala sodda gap tahliliga e’tibor bersak, gapning konstruktiv bo‘laklari tahlilida farqli jihatlarning mavjud emasligiga guvoh bo‘lamiz. Biroq gaplarning yuza struktur elementlarining pog‘onaviy bog‘lanishdagi farq quyidagicha:

1-gapda gap kesimining leksik ma’no anglatuvchi qismiga bog‘langan “yuksaltirishi” va “boshlashi” so‘z kengaytiruvchilari o‘z navbatida “mehnatning”, “insonni”; “dangasalikning”, “halokatga” aktantlariga markaz vazifasini bajargan va daraxtsimon pog‘onaviy bog‘lanishni yuzaga keltirgan.

Xalqimiz	*Mehnatning*	*insonni*	*yuksaltirishi*	*dangasalikning*	*halokatga*	*boshlashini*	*ta'kidla*	*-y*	*-di*
S	***Attr¹***	***Obj³***	***Obj¹***	***Attr²***	***Obj⁴***	***Obj²***	***W***	***T***	***Pm***



Ta'kidlash jozki, gapning LSQ yuzaga kelishida konstruktiv birliklarning qanday nominativ yoki struktur birliklar bilan ifodalanishi ahamiyatsiz. R. Sayfullayeva , M. Qurbonova, M. Abuzalova kabi formal-funksional yo'nalish tadqiqotchilarining ilmiy ishlarida keltirilganidek, [WPm] uchun relevant belgi bitta, u ham bo'lsa, atov birligi vazifasini bajara olishidir (Qurbonova, 2001: 169).

Xulosa

Ushbu maqola nutqiy gaplarning bitta nominativ birlikka transformatsiya etilishi va boshqa bir gap tarkibida bitta konstruktiv bo'lak yoki bitta konstruktiv bo'lakning kengaytiruvchisi mavqeyida kelishi hodisasini yoritishga yo'naltirilgan. Tilning lisoniy tizimidagi markaz qurshovlilik munosabatining sintaktik sathda voqelanishi, qo'shma gap tarkibiy qismlarida markaziy bo'lak- kesim (atov birlik[W]+kesimlik ko'rsatkichi [Pm])ni to'ldirishdagi o'rni; mutlaq hokim bo'lak-konstuent (kesim)ga bilvosita yoki bevosita bog'langan aktant (tobe bo'laklar)larning nisbiy pog'onali-iyerxik o'rinlashuvi yoritilgan. Tilshunoslikdagi pog'onaviylik tamoyiliga asosan dastlabki bosqichda qurshov vazifasini bajargan bo'lak (jumladan, [Pm] bilan valentlik hosil qilgan gap kengaytiruvchisi yoki [W] atov birligiga tobelangan so'z kengaytiruvchisi) ning keyingi bosqichda mos ravishda so'z kengaytiruvchilari bilan kengayib markaz mavqeyida kelishini ochib berishga qaratilgan. O'zbek tili grammatikasiga xos bo'lgan sintaktik-struktur xususiyatlardan biri-gap bo'laklari ierarxiyasidagi transformatsiya jarayoni masalasi hamda qo'shma gapning pog'onali sintaktik tahlilida gap va so'z kengaytiruvchisining iyerxiyadagi maqomi oydinlashtirilgan.

Foydalanilgan adabiyotlar:

1. Asqarova, M. *Hozirgi zamon o'zbek tilida qo'shma gaplar*. Toshkent: O'zFA nashriyoti, 1960, -B. 66
2. Будагов, Р. *Человек и его язык.* — М.: ИМУ, 1974, — С.93
3. Mahmudov, N. *O'zbek tilidagi sodda gaplarda semantik-sintaktik asimmetriya*. -T.: O'qituvchi, 1984, -B. 41

4. Qurbonova, M. *O‘zbek tilshunosligida formal-funksional yo‘nalish va sodda gap qurilishining talqini*: Filologiya fanlari doktori… dis.-Toshkent, 2001, -B. 167, 169
5. Sayfullayeva, R., Mengliyev, B., Qurbonova, M., Abuzalova, M., Boqiyeva, G., Yunusova, Z. *Hozirgi o‘zbek adabiy tili*. – T.: Fan va texnologiya , 2009. -B. 331, 395

ZAMONAVIYLASHGAN VA “TIRILGAN” ESKI SO‘ZLAR

Abjalova Manzura Abdurashetovna
Filologiya fanlari doktori (DSc), dotsent
Toshkent davlat o‘zbek tili va adabiyoti universiteti.
E-mail: abjalova.manzura@gmail.com
ORCID: 0000-0002-1927-2669

Safarova Gulshoda Shuhrat qizi
ToshDO‘TAU O‘zbek filologiyasi fakulteti
Amaliy filologiya yo‘nalishi 3-bosqich talabasi.
E-mail:gulshodasafarova4@gmail.com

Ravshanova Nozima
ToshDO‘TAU O‘zbek filologiyasi fakulteti
Amaliy filologiya yo‘nalishi 3-bosqich talabasi.

Annotatsiya. Bugungi kunda vaqt o‘tishi bilan eskirgan so‘zlar qatlamidan joy olgan muayyan so‘zlarning iste’molga qaytishi, hatto ayrimlarining qo‘llanishi ommalashayotganini ham kuzatish mumkin. Ijtimoiy, siyosiy, shaxsiy va diniy omillar bilan bog‘liq mazkur xolat tilda eskirgan so‘zlar semantikasidagi o‘zgarishlarni tadqiq etishni taqazo etadi. Ushbu maqolada aynan eskirgan so‘zlar qatlami va hozirda iste’molga qaytgan ba’zi so‘zlar xususida so‘z yuritilgan.
Kalit so‘zlar: eski so‘zlar, eskirgan qatlam, “tirilgan” so‘zlar, zamonaviylashgan so‘zlar.

Abstract. Today, over time, one can observe the return of individual words from the layer of obsolete words to use, and even the use of some of them is becoming popular. This situation, associated with social, political, personal and religious factors, requires the study of changes in the semantics of obsolete words in the language. This article is about a layer of obsolete words and some words that are now back in use.
Keywords: old words, outdated layer, modernized words.

Аннотация. Сегодня с течением времени можно наблюдать возвращение отдельных слов из слоя устаревших слов в употребление, и даже употребление некоторых из них становится популярным. Такая ситуация, связанная с социальными, политическими, личностными и религиозными факторами, требует изучения изменений семантики устаревших слов в языке. Эта статья о слое устаревших слов и некоторых словах, которые сейчас снова в употреблении.
Ключевые слова: старые слова, устаревший слой, модернизированные слова.

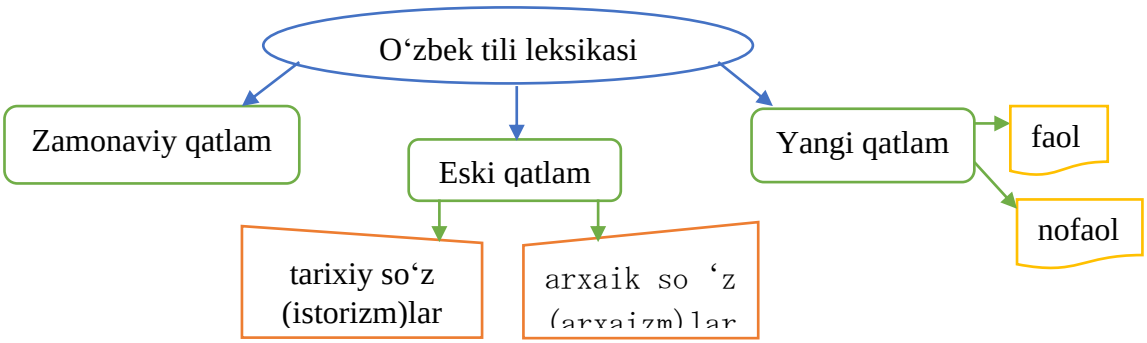
Kirish. Tilning leksikologiya – leksik qatlami tilshunoslikning boshqa sohalariga qaraganda ko‘proq o‘zgarishlarga duch keladigan qismidir. Bugungi kunda raqamli texnologiyaga moslashtirish uchun amalga oshirilayotgan ishlan qatorida tilning leksik qatlami ham zamonaviylik xususiyatlari bilan tadqiq etilishini talab qiladi. Tabiiy tilning qayta ishlanishi (NLP)da barcha sathlar muhim ahamiyat kasb etadi [Abjalova, 2020; 2023]. So‘zning leksik-

semantik xususiyatlarining aniq o‘rganilishi semantik tasnif mukammalligi va to‘liqliligini ta’minlaydi. Til korpuslarida kontekstual semantik tahlilning to‘g‘ri amalga oshirilishiga erishiladi [Abjalova, 2022].

O‘rganish sohasini so‘z va uning lug‘aviy ma’nosi bilan bog‘liq hodisalar, ya’ni o‘z va ko‘chma ma’no, bir ma’nolilik va ko‘p ma’nolilik, so‘zning shakl va ma’no munosabatiga ko‘ra turlari (sinonim, omonim, antonim, paronim, giper-giponim, xolo-meronim) tashkil etgan leksikologiya o‘zbek tili leksikasidagi eskirish va yangilanish jarayonlarini ham tahlil etadi. Vizual tarzda tilni inson tanasiga qiyoslanganda, fonetika tanadagi qon, grammatika suyaklari, leksikologiya esa uning ustuxoni, ya’ni etidir. So‘z boyligining ko‘pligini tananing etdor, bo‘liq bo‘lishiga mengzash mumkin. Et eng ko‘p o‘zgarishga uchraydi. Tana ozg‘in yo semizligini ham et belgilaydi. Demak, tilning boyligi leksik qatlam qalinligi bilan aniqlanadi. Soddalashuv, murakkablashuv, ixtisoslashuv, so‘zlarning yangilanishi, eskirishi, yangi ma’no ottenkalarini kasb etishi, neologizmlar va so‘z o‘zlashtirish orqali til leksikoni doimiy o‘zgarish jarayonida bo‘ladi.

S.I.Ojegov ta’kidlaganidek, leksika tilning boshqa sathlariga qaraganda murakkab va ko‘p qirralidir. Uning vazifasi, bir tomondan, jamiyat nafas olayotgan borliqni barcha murakkabliklari bilan in’ikos etish bo‘lsa, ikkinchi tomondan, leksik, semantik, uslubiy-sinonimik so‘z yasash va shuning kabi bir qator hodisalarning murakkab lisoniy munosabatlarini ifoda etishdan iboratdir. [Suvonova, 2006:48].

Manbalar tahlili. Leksikadagi yangilanish va eskirish ham muhim jarayon sanaladi. Bu jarayonlar tilshunos olimlar tomonidan ham o‘rganilgan, tadqiq etilgan, tahlil qilingan. Olimlar o‘zlarining tilshunoslikka bag‘ishlangan asarlarida bu haqida ma’lumotlar keltirib o‘tishgan [Shoabdurahimov va boshq.; Tursunov va boshq.; Nurmonov va boshq., Mahmudov; Xolmonova]. Leksik qatlamning yangilanish va eskirish jarayonlarini quyidagicha tushuntiriladi: tarixiylik jihatdan o‘zbek tili leksikasi 3 xil: 1) zamonaviy qatlam; 2) eski qatlam; 3) yangi qatlam. Zamonaviy qatlam deya yangilik va eskilik buyog‘iga ega bo‘lmagan so‘zlar hisobga olinadi, ya’ni tilda faol qo‘llanuvchi so‘zlar. Unda sof o‘zbek va o‘zlashgan so‘zlar ham bo‘ladi. Zamonaviy qatlam so‘zlari lug‘at boyligimizning asosini tashkil etadi. Eskilik bo‘yog‘iga ega bo‘lgan so‘zlar esa eski qatlamni tashkil etib, ular ikki xil bo‘ladi: 1) tarixiy so‘z (istorizm)lar; 2) arxaik so‘z (arxaizm)lar.



1-chizma. O‘zbek tili leksikasi tarkibi

1. Istorizm (tarixiy so‘z)lar tarixga oid (hozirgi kunda foydalanilmaydigan) narsa-hodisalarni ifodalaygan so‘zlardir. Til evolyutsiyasi natijasida ba’zi narsalar yo‘qoladi, shuningdek, ularniifodalagan so‘zlar ham birgalikda iste’moldan chiqib ketadi. Eskirish (arxaiklashish) shu tarzda sodir bo‘ladi. Istorizm hisoblangan so‘z ham, shu so‘z ifodalagan narsa-hodisa ham hozirgi davrda mavjud bo‘lmaydi. Shunday so‘zlarga misollar: *yalak, podsho,*

qo‘rboshi, ellikboshi, uloba, dakma, cho‘ri, kanizak, jallod, zindon, sovut,shirdog‘, sodoq, paranji, poruchik, kaftan, uyezd va boshqalar. Shunday tarixiy so‘zlar borki, ular bugungi kunda iste’molga qaytgan, “tirilgan”. Masalan: *vazir, misvok* (tish tozalagich / zubochistka), *ellikboshi* (haj, umra ziyoratlariga boruvchi guruhlarga boshchilik qiluvchi), *madrasa, imom, mufti, fitr, vaqf* kabi so‘zlar. Misollardan yaqqol ayon bo‘ladiki, tilda so‘zlar o‘lmaydi, balki davriylikka mansub bo‘ladi. Eskiradi, yangilanadi, qo‘llanilish ko‘lami o‘zgaradi va hkz. Bu hodisani so‘zning lug‘aviy ma’nosi ifodalagan narsa, hodisa yo holatlar eskirmasdan, uning avvalgi semantikasi eskirib, hozirda yangi semantik tabiatga ega bo‘lishi bilan tushuntirish mumkin, ya’ni so‘zning oldingi ma’nosi eskirib, semantik istorizmga aylanadi. Demak, istorizmning leksik istorizm – so‘zning tubdan iste’moldan chiqishi va semantik istorizm – so‘zning avvalgi ma’nosi iste’moldan chiqishi (ammo hozirda so‘zning o‘zi yangi ma’no yo vazifada qo‘llanadi) turlarini farqlash joiz. Semantik istorizmga yana misol tariqasida *rais* so‘zini keltirish mumkin: umuman, *rahbar* lug‘aviy ma’nosiga ega mazkur leksik birlik ilgari shayxlar rahnamosi (1) O‘rta Osiyo xonliklarida shariat qonun-qoidalari, diniy marosimlar, urf-odatlar, ro‘za tutish, namoz o‘qish, xayr-u ehson ishlarining o‘z vaqtida aniq bajarilishini nazorat qiluvchi, ota-onalarning bolalarini maktabga yuborishlarini, savdo ahlining xaridorlar haqiga xiyonat qilmasligi uchun tarozi va toshlarning to‘g‘ri bo‘lishini tekshirib turuvchi mansabdor. Shuningdek, shahar orastaligi va xavfsizligi uchun ham mas’ul hisoblangan shaxs.) semantikasiga ega bo‘lgan bo‘lsa, hozirda 1) muayyan saylov organlari va tashkilotlar rahbari (Mas, O‘zbekiston Respublikasi Oliy Majlis raisi, O‘zbekiston Respublikasi Vazirlar Mahkamasi huzuridagi Oliy attestatsiya komissiyasi raisi va h.k.); 2) anjuman, majlis yoki yig‘inga rahbarlik qiluvchi shaxs [O‘zME] sememalariga ega. Shuningdek, *saroy* – podshoh qarorgohi, *saroy* – muhtasham bino (san’at saroyi).

2. Arxaizmlar (eskirgan so‘zlar) hozirda mavjud bo‘lgan narsa-hodisalarning eski nomlaridir. Arxaik so‘zlar ifodalagan narsa-buyumlar mavjud bo‘lib, nomlari yangi, ya’ni boshqa ekvivalentlari bilan ifodalanadi. Masalan: *dirham* (tanga), *jayb* (yoqa, cho‘ntak), *tovochi* (xabarchi), *ulus* (xalq), *jumhuriyat* (respublika), *lang* (cho‘loq), *firqa* (partiya), *dorilfunun* (universitet), *tilmoch* (tarjimon), *budun* (xalq), *bitik* (kitob), *irn* (lab), *dudoq* (lab), *meng* (xol), *ilik* (qo‘l) va boshq. Arxaik so‘zlarni iste’mol jarayonida urfdan chiqqan, vaqtinchalik so‘zlar deyish o‘rinli.

Shuni ham alohida ta’kidlash joizki, ayrim ma’lum ya’ni yangilik bo‘yog‘iga ega bo‘lmagan so‘zlar boshqa ma’no kasb etgan holatlar mavjud. *Reaksiya* so‘zining ma’nosi izohli lug‘atda quyidagicha berilgan. [O‘TIL, 2006-2008. 3:365]

REAKSIYA I [re + lot. – aks, teskari; actio – harakat] **1.** *psxl. tib.* Tashqi va ichki qo‘zg‘atuvchilarga javoban inson va uning tana a’zolarida paydo bo‘ladigan harakat, jarayon yoki holat; aks ta’sir

2. *kim.* Ikki yoki bir necha moddalar o‘rtasida kechadigan va yangi moddalar hosil qiladigan o‘zaro kimyoviy ta’sir. Kimyoviy reaksiya.

REAKSIYA II *siyos.* Eski, umri tugagan ijtimoiy tartiblarni saqlab qolish, qayta tiklash va mustahkamlash maqsadida ijtimoiy taraqqiyotga faol qarshilik ko‘rsatish.

Hozirgi kunda raqamli texnologiyalarning ommalashishi natijasida uning yana bir ma’nosi yuzaga kelganini ko‘rishimiz mumkin. Mohiyatan *ta’sir* mazmuni bo‘lsa-da, uni anglatish yo‘li o‘zgachadir. Reaksiya so‘zi hozirda ijtimoiy tarmoqlarda o‘z hissiyoti va munosabatini turli shakldagi stikerlar orqali ifodalashdir. Asosan, ma’lumot va yangiliklar so‘ngida *reaksiya*

bildiring birikmasi tarzida keladi. Shuningdek, *to‘plangan reaksiyalar, bildirilgan reaksiyalar* tarzida ham qo‘llaniladi.

Xulosa qilib shuni aytish joizki, leksikaning yangilanish va eskirish jarayonlari o‘z ichida bo‘lmay, ekstralingvistik omillar, asosan, yangi innovatsiyalarning paydo bo‘lishi hamda siyosiy tuzumlarning o‘zgarishi va bular ortidan yuzaga keluvchi yangidan yangi faoliyat-jarayonlar, ularning mahsulotlari esa yangi nomlar bilan atalgani sababli bu turdagi leksemalar paydo bo‘ladi. So‘zlarning eskirishi va iste’moldan chiqishi esa mavjud narsalarning iste’moldan chiqishi va uni ifodalagan so‘zning ham u bilan birga eskirishi natijasida, qolaversa, so‘zlarning talaffuz iqtisodiga mosi yoxud talaffuzga qulayligi ham sinonim so‘zlarning birini chiqib ketishini ta’minlashi mumkin [Dadaboyev va boshq]. Ammo davr o‘tishi bilan eski so‘zlar yangi ma’no kasb etib “tirilishi” yoxud o‘z ma’nosida zamonaviy hodisa yoki narsa-buyum nomini ifodalashga xizmat qilishi mumkin. Bu holat tilning ijtimoiyligini yana bir bora dalillab beradi va tilda ham “atom hodisalar” borligini asoslaydi.

Foydalanilgan adabiyotlar:

1. Abjalova M.A. Korpus lingvistikasi: uslubiy qo‘llanma. – Toshkent: Bookmany print, 2022. – 110 b.
2. Abjalova M.A. Tahrir va tahlil dasturlarining lingvistik modullari. [Matn]: monografiya / M.A. Abjalova. – Toshkent: Nodirabegim, 2020. – 176 b. ISBN 978-9943-6939-0-6
3. Abjalova M., Gulomova N. Author’s Corpus of Alisher Navoi and its Semantic Database. // IEEE – UBMK – 2022: 7th International Conference on Computer Science and Engineering. 24-26 September 2022. – Diyarbakir, Turkey. – pp. 182-187. Impakt Factor 5.5. http://doi.org/10.1109/UBMK55850.2022.9919546
4. Abjalova M. Tagging and Annotation of Corpus Units. // International Journal of Language Learning and Applied LinguisticsISSN: 2835-1924 Volume 2 | No 12 | Dec -2023. – PP. 103-107. https://inter-publishing.com/index.php/IJLLAL/article/view/3228/2733
5. Dadaboyev H., Abjalova M., Rashidova U. O‘zbek tili sinonimlarining o‘quv lug‘ati. – Toshkent: BookPrint, 2023. – 810 b.
6. Suvonova R. So‘z o‘zlashtirish va metonimiya / O‘zbek tili va adabiyoti, – Toshkent. 2006/6. 48-bet.
7. Usmonova U. Leksik arxaizm va uning semantizatsiyasi. O‘zbek tili milliy va ta’limiy korpusining nazariy va amaliy masalalari. compling.navoiy-uni.uz.
8. O‘zbek tilining izohli lug‘ati: 80 mingdan ortiq so‘z va so‘z birikmasi. / A. Madvaliyev tahriri ostida. – Toshkent: O‘zbekiston Milliy ensiklopediyasi. 5 jildli. 2006.
9. O‘zbekiston Milliy Ensiklopediyasi. Birinchi jild. Toshkent, 2000-yil.
10. Маҳмудов Н. Ўзбек тилшунослигининг тараққиёти // Ўзбек тили ва адабиёти. 2001.4-сон. 5–7-бетлар.
11. Нурмонов А., Шаҳобиддинова Ш., Искандарова Ш., Набиева Д. Ўзбек тилининг назарий грамматикаси. Морфология. Т.: Шарқ, 2001. 164 б.
12. Турсунов У., Мухторов Ж., Раҳматуллаев Ш. Ҳозирги ўзбек адабий тили. Олий ўқув юртлари филология факультетлари талабалари учун дарслик. – Т.: Ўзбекистон, 1992. – 339 б.
13. Шоабдураҳимов Ш., Асқарова М., Ҳожиев А., Расулов И., Дониёров Х. Ҳозирги ўзбек адабий тили. – Тошкент: Ўқитувчи, 1980.

THE SIGNIFICANCE OF TRANSLATION TEXTS IN TEACHING THE UZBEK LANGUAGE

Halmukhamedova Umida Rasulovna,
National University of Uzbekistan
named after Mirzo Ulugbek, Teacher of the Department
of Computer linguistics and applied Linguistics.

Abstract: Translation texts are of urgent importance in teaching the Uzbek language. Topics related to the issue of equivalence, the role of levels in the emergence of equivalence, and the issue of the method of application are among the current problems in translation studies. The purpose: to identify the problem raised in the issue of equivalence, to determine the significance of the levels of equivalence, to compare the semantic proximity of equivalence between the original and the translated text.

Keywords. Equivalence, adequacy, reference, communication, denotative, expressive, phatic, metalinguistic function.

Аннотация: Переводные тексты имеют актуальное значение в преподавании узбекского языка. К числу актуальных проблем переводоведения относятся темы, связанные с проблемой эквивалентности, ролью уровней в возникновении эквивалентности, вопросом о способе применения. Цель: выявить проблему, возникающую в вопросе эквивалентности, определить значимость уровней эквивалентности, сравнить семантическую близость эквивалентности между оригиналом и переведенным текстом.

Ключевые слова: Эквивалентность, адекватность, референция, коммуникативная, денотативная, экспрессивная, фатическая, метаязыковая функции.

Annotatsiya: O‘zbek tilini o‘qitishda tarjima matnlari dolzarb ahamiyatga ega. Ekvivalentlik masalasi, ekvivalentlikning paydo bo‘lishida darajalarning o‘rni, qo‘llash usuli masalasiga oid mavzular tarjimashunoslikning dolzarb muammolaridandir. Maqsad: ekvivalentlik masalasida ko‘tarilgan muammoni aniqlash, ekvivalentlik darajalarining ahamiyatini aniqlash, asl va tarjima matni o‘rtasidagi ekvivalentlikning semantik yaqinligini solishtirish.

Kalit so'zlar: Ekvivalentlik, adekvatlik, murojaat, aloqa, denotativ, ekspressiv, fatik, metallingvistik funksiya.

Introduction

The main factor of culture that connects different nations is translation studies. Translation is the most important form of inter-ethnic communication, and through translation, the works of scientists and poets are translated into different languages, and ties between peoples are connected to each other. The most laborious type of translation for any language is the translation by finding the equivalent version.

The translated text is equal to the original text and should create an impression of unity with it. Accordingly, one of the tasks of the translation is that the translated text takes the full communicative place of the original and presents a harmonious balance of artistic pleasure, content

and structure equal to the possibility of the reader of the original. Such a balance should cover all aspects of translation.

Literature Review

Analysis of linguistic and methodological literature on translation issues shows that there is still no About this phenomenon and its components. Thus, L.K. Latishev distinguishes the main and pragmatic parts of translation comretention.

The main part unites the elements of this comretence that are always involved in the translation and takes into account concertal and technological components [1]. Concertual competence refers to the translator's general knowledge about the essence of translation, its specific features, what distinguishes it from other types of language mediation, the goals and tasks of translation, etc. A technological consultant employs a set of basic translation skills that help the translator overcome the usual "technical" challenges encountered in the translation process and solve various problems. The translator is not in all cases, but the knowledge, skills, skills necessary for translating a certain topic, a certain stylistic genre, etc., constitute a pragmatic component of the translation competence, including specific and special components. does. Specific knowledge includes the acquisition of certain translation methods, and special knowledge is the ability of the translator to understand the necessary topic and stylistic genre of translation as a special ability to fully understand the text in the original language. will be Roger Bell describes the translation process as a complex process, which includes the following components: knowledge of the language in which the translation is being performed; knowing the type of text; knowledge of the language in which the translation is carried out; knowledge of the field of translation; contrasting knowledge and skills in decoding and encoding, usually described as "communicative competence" [3]. According to the scientist, translation competency consists of 4 types of skills: grammatical, sociolinguistic, discursive and strategic. Unlike Roger Bell, some researchers are trying to group the components of translation comprehension, for example, M. Rresas divided the main language comprehension (understanding the original, creating a draft of the translation text, creating the final version of the translation) and a. the number of reriferic contentions (use of dictionaries, documents, knowledge in the field of science and use of technological tools). K. Nord, in turn, emphasizes linguistic and cultural knowledge, as well as the ability to understand the text, search and transfer information, create a text, evaluate the quality of translation as components of translation competence [5].

The main work in the development of the components of the professional competence of the translator was carried out by N.N. Gavrilenko, based on his analysis, all components of the professional competence of a translator of professionally oriented texts are related to three main aspects of his activity, which allowed his research to distinguish several components [6].

Results and discussion

Linguists and translators have different opinions on the issue of equivalence. According to some translators, giving the main content while strictly adhering to the norms of the language being translated is an equivalent translation, and it is emphasized that this is an acceptable situation. In general, "equivalent" is equality, firstly, when equivalence implies no similarity at all, and secondly, equivalent is substitution. For example in Hindi (bachche man ke raja hote heyn) "Children will be the king of hearts" or we can translate as “childhood- royalty". In this case, it is

possible to achieve equivalence in translation by giving its main meaning without compromising the level of translation. Therefore, in the process of working on a piece of the original text, several variants of translation arise. It is important that the translator can choose the best alternative. In this case, the translator selects individual elements of the text according to the level of importance, which are grammatically and semantically correct. Choosing the translation option with the least loss is the most important factor that determines the level of translation.

It is important to recognize the levels of equivalence in translation. This has been interpreted differently by linguists and translators.

Translation, as a special type of interlingual communication, is a creative process that requires the main attention to be paid to the meaning aspect of two language systems. Because the basis of completeness of information is the spiritual harmony of the texts of different languages. One of the tasks of the goal of communication is to fully ensure the aspects of the semantic alternative of any image means in translation. For example, in India (eyk haath se taali nahi bajati) i.e "No clapping with one hand" idiom in uzbek "clapping out of two hands". The general structure and content of the text, which is the main unit of communication, is the field of experience that forms the communicative meaning of linguistic units. The semantic (spiritual) content of the text, which is a complete communicative linguistic structure, consists of denotative and significant parts. The first of these refers to the aspects of the content of the text related to the events taking place in reality, and the second is related to the activity of speech thinking of the creator of the text or speech.

We mentioned above that the content of communicative units requires the combination of all spiritual elements, the expression of a single content, the goal of communication. Accordingly, in the field of communicative semantics, we are interested not in the relations outside the sign, but in the manifestation of these relations in the speech structure through the sign.

J.Lyonz, the author of the book "Semantics", which is popular among semasiologists, says that the speaker chooses a linguistic phrase purposefully while performing an action and "expresses the reference in the content of this phrase".[1,40- b] ra "interprets the reference-not as the result of a one-way action performed by the speaker or writer, but as a type of linguistic activity formed on the basis of a two-way, i.e., "interpersonal action" involving the speaker and the listener".[2 , 8-15b]

We also like this approach, because any language activity requires a communication environment, and language communication is a manifestation of interpersonal relationships, regardless of what form it takes. In addition to describing a situation or together with it, the purpose of communication may be to express an emotion, an emotional reaction to the information provided or the situation being described. The language has special tools to express this attitude: a specific tone, signs and kh.z. We are not talking about a specific emotional component of this sign, but about special signs that provide information about the assessment of the entire situation being described. By interacting with the content of other levels, this information is often an important aspect of the communicative act.

Conclusion

In conclusion, it can be said that in the analysis of the above examples, we considered whether expressions and similes in Bhisham Sahni's novel "Basanti" are given their equivalent directly, that is, adequately or indirectly.

The original and translation units can be equivalent at all levels or at some of them. Fully or partially equivalent units and potentially equivalent definitions objectively exist in the original language and in the translated language, but their correct evaluation, selection and application depends on the knowledge, skills and creativity of the translator, his language and depends on the skill of considering and comparing extralinguistic factors. When studying the translation of phrases, the translator's skill is clearly visible mainly in finding equivalent and alternative options, because the words given in simple language in Hindi are expressed by phrases in the translation. It was observed that the meaning shifts in the equivalent phrases in the original and translated text, and when the original phrases move to the translation field, their meaning narrows and they take on a new meaning as an addition. The translator translated the similes given in the original in a way specific to the culture of Uzbek readers and managed to preserve the meaning in both languages.

References:

1. Ланчиков В.К., Семантические и стилистические аспекты перевода., –М., 1977. – С.40.
2. Лайонз Джон., Лингвистическая семантика //Языки славянской культуры. – У.,2003. – С.8-15.
1. Швейцер А.Д. Теория перевода. //N. Neubert A. Text and translation. Leipzig, 1985.- R. 55. //– М., 1988.
2. Сирожиддинов Ш., Одилова Г. Бадиий таржимаасослари. – Тошкент: Mumtoz so„z, 2011.
3. Басантий романи. Амир Файзулла. – Т: “Turon-Iqbol”, 2010.
4. Латышев Л. К. Курс перевода: Эквивалентность перевода и способы ее достижения., –М.: Международные отношения, 1981.
5. Швейцер А.Д. Теория перевода. Статус, проблемы, аспекты. – М., 1988.
6. Рахматов, З. Н., & Рашидов, Д. Н. (2023). Пути совершенствования механизма разработки маркетинговой стратегии АО «Ўзтемирйўлйўловчи». Innovative achievements in science 2022, 2(17), 55-60.
7. Комиссаров, В. Н. Лингвистика перевода. – М.: Международные отношения, 1980.

TILSHUNOSLIKDA MEDIAMATNNING O‘RNI

Abdullayeva Madinabonu Qurbonali qizi
O‘zMU 1-bosqich magistranti.

Annotatsiya: Ma’lumki, hozirgi kunda tilshunoslikda yangi yo‘nalishlar va sohalar o‘rganilmoqda. Ushbu maqolada ham mediamatn haqida umumiy ma’lumotlarni berilgan. Shu bilan birga , mediamatn- tilshunoslik bilan ko‘plab fanlarning qiziqishlari doirasiga kirishi, bu asosan ommaviy kommunikatsiya mohiyati bilan izohlanadi va mediamatnlarning tilshunoslikda o‘rganilishi, ularning turlari haqida ma’lumot beradi. Bundan tashqari mediamatnlarning funksional guruhlanishi, ularning o’ziga xoslik jihatlari izohlangan. Shuni qo’shimcha qilish mumkinki, mediamatnlardagi bo’linishlar tilshunoslar tomonidan turlicha bo’lagani va ularning farqlari tushuntirib berilgan.

Kalit so‘zlar: Medialingvistika, mediamatn, diskurs, kommunikatsiya, reklama, matbuot, siyosiy nutq, oilaviy nutq, sport diskursi, internet diskursi, telematn, radiomatn, ilmiy matn, badiiy matn.

Abstract: It is known that new directions and fields are being studied in linguistics today. This article also provides general information about media text. At the same time, mediatext-linguistics is the interest of many disciplines, which is mainly explained by the essence of mass communication, and the study of mediatexts in linguistics provides information about their types. In addition, the functional grouping of media texts and their characteristics are explained. It can be added that the divisions in media texts are different by linguists and their differences are explained.

Keywords: Media linguistics, media text, discourse, communication, advertising, press, political discourse, family discourse, sports discourse, internet discourse, teletext, radio text, scientific text, artistic text.

Аннотация: Известно, что сегодня в лингвистике изучаются новые направления и области. В этой статье также представлена общая информация о медиатексте. В то же время медиатекст-лингвистика вызывает интерес многих дисциплин, что в основном объясняется сущностью массовой коммуникации, а изучение медиатекстов в лингвистике дает информацию об их видах. Кроме того, поясняется функциональная группировка медиатекстов и их характеристики. Можно добавить, что лингвисты различают подразделения медиатекстов и объясняют их различия.

Ключевые слова. Медиалингвистика, медиатекст, дискурс, коммуникация, реклама, пресса, политический дискурс, семейный дискурс, спортивный дискурс, интернет-дискурс, телетекст, радиотекст, научный текст, художественный текст.

So‘nggi yillarda ijtimoiy fanlarga qiziqish ortdi, bu bir necha omillar bilan izohlanadi: tilshunoslik fanining ichki ehtiyojlari; ijtimoiy sohani kuzatib boorish uchun siyosiy matnlarni va ommaviy axboroy vositalarining matnlarini tahlil qilish zarurati. Bizning asosiy muloqot vositamiz til hisoblanadi va u o‘zini matnlar orqali namoyon qiladi. Matnlar og’zaki va yozma shaklda axborot beradi.

Rus olimi M.V. Lyapon matnga quyidagicha izoh beradi:

1. Matn - til sistemasining kommunikativ jarayonda muhim funksiya bajaruvchi eng yuqori pog’onasi.

2. Matn – bu jumlalarning o‘zaro munosabatidan kelib chiquvchi xabardir. Matn tadqiqotchi uchun til hodisalarini o‘rganishda birdan- bir manbadir.

3. Matn bu so‘zlovchining nutqiy qobilyatini ro‘yobga chiqaruvchi vositadir. Bu jarayonda matn tilning nutqda real qo‘llanishini ta’minlovchi makon vazifasini ham bajaradi.

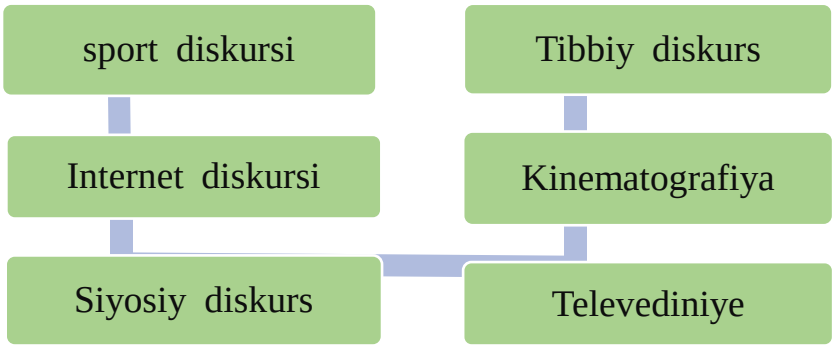
4. Matn bu so‘zlovchining nutqiy faoliyati mahsulidir.[68]

Bundan tashqari matn haqida V.Matezius, F.Danesh, V.Dressler, P.Hartman, I.Galperin, T.Silman, I.Kovtunova, N.Pospelov, L.Loseva kabi xorijlik tilshunoslar o‘z fikrlarini keltirgan, o‘zbek tilshunosligida esa M.To‘xtasinov, A.Mamajonov, M.Hakimov, A.Nurmonov, M.Yo‘ldoshev, E.Qilichev o‘z ilmiy tadqiqotlarida ma’lumotlar beradi.[69]

Demak, matnni asosiy axborot beruvchi til birligi sifatida belgilaymiz. Hozirgi kunda siyosiy va ijtimoiy sohadagi ma’lumotlarni ommaviy axborot vositalari orqali tez bilib olishning imkoni katta. Bu o‘rinda tilshunoslik mediamatnlarga tayanadi.

Tilshunoslikning zamonaviy sohalaridan biri medialingvistika hisoblanadi. Medialingvistikaning asosi mediamatndir. Bu soha 2000- yilda paydo bo‘lgan bo‘lib, ilk marta rus tilshunosligida T.G.Dobrosklonskayaning “Medialingvistika nazariyasi va usullari” nomli dissertatsiyasida qo‘llangan. Ba’zi ma’lumotlarda bu atama britaniyalik olim J.Kornerning “Medialingvistika doirasida” nomli maqolasida uchraganligi ta’kidlanadi.

Medialingvsitika- bu ommaviy axborot vositalari tilidan foydalanishni lingvistik o‘rganishdir. Sohaning asosiy jihati shundaki, media sohadagi tilning amaliy yoki bosma, raqamli va tarmoqli ommaviy axborot vositalari orqali taqdim etilgan zamonaviy ommaviy kommunikatsiyalarni o‘rganadi. Britaniyalik tadqiqotchilar mediamatnlar va ularning komponentlari og‘zaki va yozma matnlardan farqli deya ta’kidlaydilar. Mediamatnlarning ma’nosi oddiy matn anglatgan ma’nodan kengroq hisoblanadi. Media lingvistika nazariyasining muhim tarkibiy qismi mediamatnlarning barcha mumkin bo‘lgan turlarini to‘liq va izchil tavsiflash uchun parametrlar to‘plamdan iborat. Media diskurslar bugungi kunda reklama, PR, matbuot, televideniye, radio, internet aloqalari (ijtimoiy tarmoqlar, onlayn nashrlar, bir paytlar media xarakteriga ega bo‘lmagan nutqlarning (siyosiy nutq, kundalik nutq, oilaviy nutq, ko‘ngilochar nutq) tarjimoni bo‘lib xizmat qiluvchi ommaviy kommunikatsiya kanallari va shakllarining zamonaviy ko‘pligi gumanitar fanlarning o‘zaro munosabatlarini rivojlantirishga turtki beradi.Media diskurslarning obyektini belgilashda tilshunoslar, asosan, ommaviy axborot vositalariga urg‘u berishadi. Biroq media diskurslarning ichki janriy bo‘linishini quyidagicha belgilash mumkin:



[68] Добросклонская Т.Г. Медиалингвистика. Системный подход к изучению языка СМИ. – Москва: Флинта: Наука, 2008. 264 с.
[69] Kurbanova M., Yo’ldoshev M. Matn tilshunosligi. – T., 2014.

Boshqa manbalarda, asosan, rus tilshunosligida mediamatnlarning diskursiga ko‘ra turlari quyidagicha guruhlanadi:[70]

1.Ommaviy- siyosiy matn
2. Ijtimoiy- muammoli matn
3. Tarixiy matn
4. Badiiy matn
5. Madaniyatshunoslik matni
6. Falsafiy matn
7. Diniy matn

Shubhasiz, matnlarning bunday tasniflanishi shartli hisoblanadi va bu ularni asosiy toifalarga bo‘linishini ta’minlaydi. Mediamatnlar o‘zining xarakterli xususiyatlariga ham ega. Masalan, ommaviy bo‘lishi, tematik mazmun jihatdan ochiqlikka ega bo‘lishi, og’zaki shakllarda amalga oshiriladigan intertekstuallik mavjudligi, uyg’unlik va boshqalarni sanashimiz mumkin.

Nihoyat, funksional- uslubiy vazifasiga ko‘ra olimlar mediamatnlarni guruhlarga ajratadi.

Ammo bu bo‘linishda ham ikki xil variantini ko‘rishimiz mumkin. Birinchi guruhda mediamatnlar besh guruhga ajratilgan:

1.Tahliliy matnlar
2.Tarbiyaviy matnlar
3.Ikki markazli matnlar
4.Aloqa matnlar
5.Axborot matnlar

Ikkinchi guruhlanish rus tilshunosi O.V.Krasnoyarovaning “Matn va mediamatn muammolari” nomli tadqiqotida keltirilgan. Unda quyidagicha guruhlangan:

1. Hujjatli- publistik matn
2. Ilmiy matn
3. Badiiy matn
4. Ommaviy- ko‘ngilochar matn

Bundan ko‘rinadiki, mediamatnlarning ichki guruhlanishi va o‘rganilishida ham turli xil qarashlar mavjud bo‘lgan ekan. Mediamatn hozirgi vaqtda uning ishtirokchilarining nutq xatti-harakatlari strategiyasi va taktikasidagi sezilarli o‘zgarishlar bilan ajralib turadi, bu esa uni kommunikativ- pragmatik jihatdan o‘rganish uchun asos bo‘lib xizmat qiladi. Bunday matnlarni tashkil etish tamoyillari va tarkibiy va semantik tashkil etilishidagi o‘zgarishlarni o‘rganish alohida o‘rin tutadi. Misol uchun, gazeta matnining nutq tuzilishini tavsiflash uchun, gazeta nutqining elementi sifatida, matn sathining minimal tarkibiy va semantik birligi- nutq shakli ajratilgan.

Ma’lum bir ma’noda mediamatn tushunchasi turli ko‘rinishlari va belgilari bilan farqlanadi. Masalan, jurnalistik matn, reklama matni, internet matni tur va ko‘rinishiga, janr va shakliga ko‘ra tasniflanadi, ammo ularning barchasi mediamatn sifatida ifodalanadi. Shunga ko‘ra mediamatnlar tur va ko‘rinish jihatdan quyidagucha tasniflaniladi:[71]

I. OAV turlari bo‘yicha mediamatnlarning tuzilishi

[70] Красноярова О.В. Текст и медиатекст: проблема дифференциации понятий. // Вопросы теории и практики журналистики, 2013. – Т. 4. – №1.
[71] Добросклонская Т.Г. Медиалингвистика. Системный подход к изучению языка СМИ. – Москва: Флинта: Наука, 2008. 264 с.

1. Bosma nashrlar matni (yangilik, reportaj, maqola, sharx);
2. Telematn (tok- shou);
3. Radiomatn (janr turlari bo‘yicha);
4. Kinomahsulot (janr turlari bo‘yicha);
5. Ijtimoiy tarmoqlar matni (sms, chat);

II. Qabul qilish usuliga ko‘ra:

Bosma- qog’oz matni

Audiomatn

Telematn

Veb- nashr

Mediamatnlarni o‘rganishda ularni guruhlarga ajratish yaxshi samara beradi. Mediamatnning asosiy toifalari mediaviylik, ommaviylik, interativlik hisoblanadi. Demak, mediamatn yaratishda ma’lum natijaga erishish uchun nazariy bilimlarga ega bo‘lish, ularning turlarini bilish va ularing talablarini mukammal bilish kerak. Chunki, hozirgi kunda axborot maydoni juda ham katta. Unda namoyon bo‘lish uchun esa bunday matnlarni bilish talab qilinadi. Mediamatnlarda asosiy ta’sir kuchi intertekstual birliklarda namoyon bo‘ladi. Misol, uchun birgina sarlavha yoki epigraf insonni jalb qiluvchi vositalardan hisoblanadi. Negaki, medimatnlar bir shaxs uchun emas omma uchun yaratiladi va ularni birdek jalb qilishda intertekstual birlik alohida ahamiyatga ega bo‘ladi.

Xulosa, mediamatn media nutqni tushunishda muhim rol o‘ynaydi. Hozirgi kunda ommaviy axborot vositalarining nutqi eng globallashgan va dolzarbdir. Media- nutq juda xilma-xil bo‘lib, bu holat uning turlarini o‘rganishni talab qiladi. Mediamatn o‘rganishda eng muhim ilmiy vazifa nutqning o‘zaro ta’siri va ta’sir modellarini aniqlashdir. Bundan tashqari, mediamatn turlarini janr jihatdan o‘rganish, ularning til birliklarini izohlash ham muhim ahamiyatga ega.

Foydalangan adabiyotlar:

1. Бакиева Г.Х., Тешабаева Д.М. Оммавий ахборот воситалари тили. – Тошкент: O‘zkitobsavdonashriyot, 2020.
2. Блокдва О. Медиатекст и его виды. Особенности стилистики и риторики. Proza.ru/2013/04/
3. Добросклонская Т.Г. Медиалингвистика. Системный подход к изучению языка СМИ. Москва: Флинта: Наука, 2008.
4. Красноярова О.В. Текст и медиатекст: проблема дифференциации понятий. // Вопросы теории и практики журналистики, 2013. – Т.
5. Kurbanova M, Yo‘ldoshev M. Matn tilshunosligi. - T., 2014.
6. Тошалиев И., Абдусатторов Р. Оммавий ахборотнинг тили ва услуби. – Тошкент: Zar-qalam, 2006.
7. Xomidova, M. F. (2019). INTERTEXTUALITY AND LITERARY TRANSLATION. Scientific Bulletin of Namangan State University, 1(6), 309-311.
8. Xomidova M. F. The problem of intertextuality in modern linguistics //Central Asian Problems of Modern Science and Education. – 2019. – Т. 3. – №. 4. – С. 259-263.