

**FARG‘ONA DAVLAT UNIVERSITETI
HUZURIDAGI ILMIY DARAJALAR BERUVCHI
DSc.03/30.12.2019.Fil.05.02 RAQAMLI ILMIY KENGASH**

BUXORO DAVLAT UNIVERSITETI

AVEZOV SUXROB SOBIROVICH

**O‘ZBEK-RUS VA RUS-O‘ZBEK PARALLEL KORPUSLARI ONLINE-
PLATFORMASINI YARATISHNING ASOSIY TAMOYILLARI
(A.QODIRIYNING «O‘TKAN KUNLAR» ROMANI MISOLIDA)**

10.00.11 – Til nazariyasi. Amaliy va kompyuter lingvistikasi

**FILOLOGIYA fanlari bo‘yicha falsafa doktori (PhD) dissertatsiyasi
AVTOREFERATI**

Farg‘ona – 2024

**Filologiya fanlari bo‘yicha falsafa doktori (PhD) dissertatsiyasi
avtoreferati mundarijasi**

**Оглавление автореферата диссертации доктора философии (PhD) по
филологическим наукам**

**Contents of dissertation abstract of Doctor of Philosophy (PhD) on
Philological sciences**

Avezov Suxrob Sobirovich

О‘zbek-rus va rus-o‘zbek parallel korpuslari online-platformasini yaratishning asosiy tamoyillari (A.Qodiriyning «O‘tkan kunlar» romani misolida).....3

Авезов Сухроб Собирович

Основные принципы создания узбекско-русской и русско-узбекской online-платформы параллельных корпусов (на материале романа А.Кадыри «Минувшие дни»).....25

Avezov Sukhrob Sobirovich

Basic principles of creating uzbek-russian and russian-uzbek online platforms for parallel corpora (based on the material of A.Kadyri’s novel “Bygone days”).....49

E‘lon qilingan ishlar ro‘uxati

Список опубликованных работ

List of published works.....53

**FARG‘ONA DAVLAT UNIVERSITETI
HUZURIDAGI ILMIY DARAJALAR BERUVCHI
DSc.03/30.12.2019.Fil.05.02 RAQAMLI ILMIY KENGASH**

BUXORO DAVLAT UNIVERSITETI

AVEZOV SUXROB SOBIROVICH

**O‘ZBEK-RUS VA RUS-O‘ZBEK PARALLEL KORPUSLARI ONLINE-
PLATFORMASINI YARATISHNING ASOSIY TAMOYILLARI
(A.QODIRIYNING «O‘TKAN KUNLAR» ROMANI MISOLIDA)**

10.00.11 – Til nazariyasi. Amaliy va kompyuter lingvistikasi

**FILOLOGIYA fanlari bo‘yicha falsafa doktori (PhD) dissertatsiyasi
AVTOREFERATI**

Farg‘ona – 2024

Falsafa doktori (PhD) dissertatsiyasi mavzusi Oliy attestatsiya komissiyasida №B2022.3.PhD/Fil2798 raqam bilan ro'yxatga olingan.

Dissertatsiyasi Buxoro davlat universitetida bajarilgan.

Dissertatsiya avtoreferati uch tilda (o'zbek, rus, ingliz (rezyume) Farg'ona davlat universiteti veb-sahifasining www.fdu.uz hamda "ZiyoNet" axborot-ta'lim portalining www.ziynet.uz manzillariga joylashtirilgan.

Ilmiy rahbar:

Nigmatova Lolaxon Xamidovna
filologiya fanlari doktori (DSc), professor

Rasmiy opponentlar:

Rahimov Usmonjon Ergashevich
filologiya fanlari nomzodi, dotsent

Muxiddinov Anvar G'afurovich
filologiya fanlari doktori (DSc), professor

Yetakchi ilmiy tashkilot:

Sharof Rashidov nomidagi Samarqand davlat universiteti

Dissertatsiya himoyasi Farg'ona davlat universiteti huzuridagi ilmiy darajalar beruvchi DSc.03/30.12.2019.Fil.05.02 raqamli Ilmiy kengashning «27» iyul 2024-yil soat 10.00 dagi majlisida bo'lib o'tadi (Manzil: 100151, Farg'ona shahri, B. Marg'inoniy ko'chasi, 105-uy. Tel.: (993873) 244-57-82; e-mail: info@fdu.uz).

Dissertatsiya bilan Farg'ona davlat universiteti Axborot-resurs markazida tanishish mumkin (356 raqami bilan ro'yxatga olingan). (Manzil: 100151, Farg'ona shahri, Murabbiylar ko'chasi, 19-uy. Tel.: (99873) 244-71-28

Dissertatsiya avtoreferati 2024-yil «6» iyul kuni tarqatildi.
(2024-yil «6» iyul dagi 18 raqamli reyestr bayonnomasi)



S.M. Muminov
Ilmiy darajalar beruvchi Ilmiy kengash
raisi, f.f.d., professor.

M.T. Zokirov
Ilmiy darajalar beruvchi Ilmiy kengash
kotibi, f.f.n., professor.

A. Mamajonov
Ilmiy darajalar beruvchi Ilmiy kengash
qoshidagi Ilmiy seminar raisi,
f.f.d., professor.

KIRISH (falsafa doktori (PhD)dissertatsiyasi annotatsiyasi)

Dissertatsiya mavzusining dolzarbligi va zarurati. Jahon tilshunosligida XX asrning ikkinchi yarmida kompyuter va korpus tilshunosligining jadal rivojlanadi. Bu rivojlanish o‘z navbatida mashina tarjimasini, tabiiy tilni qayta ishlash, matematik tilshunoslik dasturlarini yaratish va ishlab chiqish uchun zamin yaratdi. Tilshunoslikda axborot texnologiyalaridan foydalanish bilan bog‘liq yuzlab til korpuslari mavjud bo‘lib, ular til muhandisligi, avtomatik tarjima, kompyuter tahlili, tahrirlash, tezaurus, elektron lug‘atchlik kabi sohalar imkoniyatlarni kengaytirishga imkon beradi. Nomukammal korpus texnologiyalari o‘rnini yanada zamonaviylashtirilgan va foydalanuvchilarga yo‘naltirilgan dasturlar egallab bormoqda, ularning imkoniyatlari va vositalari lingvistik tadqiqotlarni yuqori ilmiy darajada amalga oshirish imkonini beradi.

Bugungi kunda amaliy tilshunoslikning asosiy vazifalaridan biri milliy korpuslarni yaratish va rivojlantirishdir. Korpus tilshunosligining rivojlanish tamoyillari, uning tarkibiy mexanizmlari, parallel korpusni yaratish muammolari haqida ma‘lumotlarning kamligi tilning milliy korpusini yanada takomillashtirishning asosiy sababi bo‘ldi. Ruscha-o‘zbekcha va o‘zbekcha-ruscha parallel korpusni yaratilishi va har tomonlama tahlil qilinishi tarjimashunoslikni takomillashtirishga, asl nusxa va tarjima matnlari o‘rtasida funksional mutanosiblik va proporsionallikka erishishga yordam beradi.

Mamlakatimizda axborot texnologiyalarining jadal rivojlanishining hozirgi bosqichida “... davlat tilining mukammalligini saqlash, uni boyitish va aholining nutq madaniyatini oshirish; davlat tilining zamonaviy axborot texnologiyalari va kommunikatsiyalariga faol integratsiyasini ta‘minlash”¹ ehtiyeji olim tilshunoslar oldiga o‘zbek kompyuter tilshunosligi sohasiga doir yangi vazifalarni belgilab berdi. Respublikamizda til siyosati sohasida amalga oshirilgan keng ko‘lamli islohotlar O‘zbekistonda korpus tilshunosligining rivojida nafaqat yangi davr boshlanishiga, balki tilshunoslikda, inson omilini hisobga olgan holda, tadqiqotlarning yangi bosqichiga o‘tishga, amaliy tadqiqotlar olib borishga, til o‘rganuvchilarga e‘tiborning kuchayishiga sabab bo‘ldi. Bunday korpusning mavjudligi ilmiy tadqiqotlar, lingvistik tahlil, til modellarini ishlab chiqish va boshqa tegishli sohalar uchun juda muhimdir.

O‘zbekiston Respublikasi Prezidentining 2019-yil 21-oktyabrdagi PF-5850-son «O‘zbek tilining davlat tili sifatidagi nufuzi va mavqei tubdan oshirish chora-tadbirlari to‘g‘risida»gi farmoni, 2020-yil 20-oktyabrdagi PF-6084-son “Mamlakatimizda o‘zbek tilini yanada rivojlanish va til siyosatini takomillashtirish chora-tadbirlari to‘g‘risida”, 2020-yil 6-noyabrdagi PF-6108-son “O‘zbekistonning yangi taraqqiyot davrida ta‘lim-tarbiya va ilm-fan sohalarini rivojlantirish chora-tadbirlari to‘g‘risida” 2022-yil 28-yanvardagi PF-60-son “2022-2026-yillarga mo‘ljallangan yangi O‘zbekistonning taraqqiyot stradejiyasi to‘g‘risida”gi Farmonlari; 2017-yil 17-fevraldagi PQ-2789-son «Fanlar akademiyasi faoliyati,

¹ O‘zbekiston Respublikasi Prezidenti Shavkat Mirziyoyevning 2020-yil 20-oktyabrdagi «Mamlakatimizda o‘zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to‘g‘risida»gi PF-6084-son farmoni // <https://lex.uz/docs/5058351>

ilmiy-tadqiqot ishlarini tashkil etish, boshqarish va moliyalashtirishni yanada takomillashtirish chora-tadbirlari to‘g‘risida», 2019-yil 4-oktabrdagi PQ-4479-son “O‘zbekiston Respublikasining “Davlat tili haqida”gi qonuni qabul qilinganining o‘ttiz yilligini keng nishonlash to‘g‘risida”gi qarorlari, 2022-yil 20-dekabrda O‘zbekiston Respublikasi Prezidentining Oliy Majlisga Murojaatnomasi hamda mazkur faoliyatga tegishli boshqa me‘yoriy-huquqiy hujjatlarda belgilangan ustuvor vazifalarni amalga oshirishda ushbu tadqiqot muayyan darajada xizmat qiladi.

Tadqiqotning respublika fan va texnologiyalari rivojlanishining ustuvor yo‘nalishlariga mosligi. Dissertatsiya respublika fan va texnologiyalari rivojlanishining I. “Axborotlashgan jamiyat va demokratik davlatni ijtimoiy, huquqiy, iqtisodiy, madaniy, ma‘naviy-ma‘rifiy rivojlantirish, innovatsion iqtisodiyotni shakllantirish va ularni amalga oshirish yo‘llari” ustuvor yo‘nalishiga muvofiq bajarilgan.

Muammoning o‘rganilganlik darajasi. To‘plangan ma‘lumotlarni tartibga keltirish va tizimlashtirish zarurati kompyuterlar va kompyuter korpuslarni yaratishidan ancha oldin paydo bo‘lgan. Olimlar ma‘lumotnoma yig‘gan, kartochka va kartotekalar yaratgan, konkordans, lug‘atlar va ensiklopediyalar to‘plagan. Ma‘lumot to‘plash va qayta ishlash ko‘p vaqtni olar edi. XX asrning o‘rtalarigacha kompyuter texnologiyalari aniq fanlarning o‘ziga xos “quroli” edi. Gumanitar fanlar kompyuter va axborot texnologiyalaridan uzoq vaqt davomida mustasno edi, ammo yuqoridagi texnologiyalarning rivojlanishi hamda tilshunoslik sohasiga kirib kelishi oqibatida kompyuter tilshunosligi fani, keyinchalik uning negizida korpus tilshunosligi shakllandi. Korpus tilshunosligi XX asrning 90 yillarida to‘liq shakllandi. Albatta, uni mutlaqo yangi va yosh deb atash mumkin ham emas, chunki XIII asrning boshlarida “Muqaddas Bitik” uchun konkordans yaratilgan («Concordantiae morales sacrae scripturae», “Muqaddas Bitikning axloqiy konkordansi”). Shunga qaramay, lingvistik tadqiqotlar uchun kompyuter texnologiyalaridan foydalanadigan korpus lingvistikasi bir paytning o‘zida axborotni qayta ishlash, saqlash va to‘liq foydalanishga imkon beradigan jahon tilshunosligida yangi, noan‘anaviy tendensiya deb hisoblanadi.

Jahon tilshunosligida lingvistik ta‘minotning nazariy va uslubiy muammolarini o‘rganish o‘tgan asrning 60-yillarida boshlangan: axborot-qidiruvsiz kompyuter bu oddiy hisoblash mashinasi hisoblanadi, lekin axborot-qidiruv tizimi bilan jixozlangan kompyuter bu “mexanik” bosqichdan “mantiqiy-lingvistik” bosqichga o‘tgan mashina deb hisoblanadi, ya‘ni, axborot-qidiruv tizimlari bosqichiga o‘tish tan olindi². Dunyo tilshunosligida tabiiy tilni qayta ishlashga A.Keller, K.V.Linden, N.Varda³, M.A.Mohri⁴ larining asarlari bag‘ishlangan. 1960-yilda V.Frensis va G.Kuchera⁵ tomonidan ishlab chiqilgan Braun lingvistik korpusini parallel korpus, uning rivojlanishi va dasturlashiga bag‘ishlangan birinchi ish deb hisoblash

² Соколов А.В. Автоматизация библиографического поиска / А.В.Соколов. – М.: Книга, 1981. – С.91.

³ Daniel S. Jurafsky and James H. Martin. Speech and Language Processing. Contributing writers: Nigel Ward, –2000. pages: 950

⁴ Mohri M.A. Finite-state transducers in language and speech processing. Computational Linguistics. – P. 269-312.

⁵ Френсис Н., Кучера Г. Вычислительный анализ современного американского варианта английского языка. – М., 1967.

mumkin. D.Sinklarning ishi chet tilini o‘qitish va o‘rganishda korpus texnologiyalaridan foydalanish muammosiga bag‘ishlangan⁶.

Rus tilining parallel korpusi, uning tuzilishi, dasturiy ta‘minoti va bajaradigan asosiy funksiyalari D.O.Dobrovolskiy⁷, Sichinava⁸, Yu.Tao⁹, V.Zaxarov¹⁰, S.A.Mannik¹¹ va A.A.Kokoreva¹² asarlarida o‘z aksini topgan.

Shuningdek, parallel korpusni rivojlantirish muammosini A.G.Gilemshina¹³ va A.H.Xudoyberdiyeva¹⁴ asarlarida ham qayd etilgan. O‘zbek tilshunosligi allaqachon lingvistik va kompyuter korpusini o‘rganish ustida faol ishlamoqda. Til korpusi va parallel korpusning dolzarb masalalari N.Z.Abduraxmonova¹⁵, Sh.M. Xamrayeva¹⁶, M.A.Abjalova¹⁷, A.A.Eshmuminov¹⁸, G‘.G‘.Abdavaxobov¹⁹, N.A.Ataboyev²⁰,

⁶ Френсис Н., Кучера Г. Вычислительный анализ современного американского варианта английского языка. – М., 1967.

⁷ Добровольский Д.О. Какие задачи решают параллельные корпуса? (Электрон ресурс): <https://postnauka.ru/video/>; Добровольский Д.О. Корпус параллельных текстов в исследовании лексической семантики. (Электронресурс): <http://www.dialog-21.ru/media/2503/dobrovolskij.pdf>

⁸ Д.В.Сичинава. Параллельные тексты в составе национального корпуса русского языка: новые направления развития и результаты. https://ruslang.ru/doc/sitchinava/sitchinava-2015-parallel_corpus.pdf; Д.В.Сичинава. Параллельные корпуса в составе национального корпуса русского языка: новые языки и новые задачи. Труды Института Русского Языка Им. В.В.Виноградова. –2019. –С. 41-61 <https://elibrary.ru/item.asp?id=42567226>

⁹ Тао Ю., Захаров В.П. Разработка и использование параллельного корпуса русского и китайского языков // Информ. процессы и системы. – 2015 №4 <https://www.researchgate.net/publication/306911853>

¹⁰ Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург. – 2005

¹¹ Маник С.А. Параллельный корпус в практике перевода общественно-политических текстов (с английского на русский). // Modern Studies of Social Issues – 2019, Volume 11, Number 4. <http://soc-journal.ru>

¹² Кокорева А.А. Корпус параллельных текстов в обучении иностранному языку. – Вестник ТГУ (Тамбовский государственный университет), выпуск 2 (118), – 2013. – С. 57-62.

¹³ Гилемшина А.Г. Разработка параллельного корпуса арабского, русского, татарского языков: проблемы и перспективы. Казанская наука. №1, – 2020. – С. 24-29.

¹⁴ Худойбердиев Х.А., Солиев О.М., Солиев П.А. Разработка параллельного корпуса таджикского и русского языков // Новые информационные технологии в автоматизированных системах. – 2019. №22. URL: <https://cyberleninka.ru/article/n/razrabotka-parallelnogo-korpusa-tadzhikskogo-i-russkogo-yazykov> (дата обращения: 29.05.2023). 2. URL: <https://cyberleninka.ru/article/n/razrabotka-parallelnogo-korpusa-ta29.05.2023>).

¹⁵ Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (Содда гаплар мисолида). Филол. фан. бўйича фалсафа доктори (PhD)...дис. афтореф. – Тошкент, – 2018. – 52 б.;

¹⁶ Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол.фан. бўйича фалсафа доктори (PhD)...дис. – Бухоро, – 2018.;

¹⁷ Абжалова М.А. Ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи дастурининг лингвистик модуллари (Расмий ва илмий услубдаги матнлар таҳрири дастури учун): Филол.фан. бўйича фалсафа доктори (PhD)...диссер. – Фарғона, – 2019. – 164 б.;

¹⁸ Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол.фан. бўйича фалсафа доктори (PhD)...диссер. – Қарши, – 2019. – 140 б.;

¹⁹ Абдувахобов Ғ.Ғ. Янги авлод омонимлар ўқув лўғати мобил иловасини яратишнинг лингвистик асослари. Автореферат фил. фанлари бўйича фалсафа доктори (PhD). – Фарғона. – 2023;

²⁰ Атабоев Н. Б. Инглиз тили корпусининг функционал хусусиятлари (соса мисолида). Автореферат фил. фанлари бўйича фалсафа доктори (PhD). – Тошкент. – 2020;

R.A.Karimov²¹, A.N.Axmedova²², O.Xidirov²³, U.Xoliyarov²⁴, G.Begmatova²⁵, G.Toirova²⁶ tadqiqotlarida ko‘rib chiqilgan.

Parallel korpus yaratishning asosi matn va uning tarkibiy qismlari hisoblanadi. Korpusni shakllantirish jarayonida matn tanlash va tahlil qilish muhimdir. Shu sababdan o‘zbek tili “matn lingvistikasi” yo‘nalishini yangi bosqichga olib chiqishga ulkan hissa qo‘shgan olimlar A.Mamajonov²⁷, N.M.Turniyozov²⁸, E.Qilichev²⁹, M.Hakimov³⁰, M.Yuldoshev, S.Muhamedova, M.Saparniyazovalar³¹ ishlariga alohida e’tibor qaratildi.

Dissertatsiya tadqiqotining dissertatsiya bajarilgan oliy ta’lim muassasasining ilmiy-tadqiqot ishlari rejalari bilan bog‘liqligi. Dissertatsiya Buxoro davlat universitetining ilmiy-tadqiqot ishlari rejasining “O‘zbek tilshunosligida til, shaxs va jamiyat munosabatlari tadqiqi muammolari” (2023-2024-yy.) yo‘nalishi doirasida bajarilgan.

Tadqiqotning maqsadi o‘zbek-rus parallel korpusini yaratishning nazariy asoslarini o‘rganish, A.Qodiriyning “O‘tkan kunlar” romani misolida parallel korpuslarning o‘zbek-rus va rus-o‘zbek onlayn platformasini ishlab chiqish hamda mashina tarjimasini takomillashtirish, o‘zbek va rus tillari orasidagi jumlar, iboralar va matn birliklari darajasida lingvistik tahlil va madaniyatlararo almashinuvni o‘rganish uchun qulay vositani yaratishdan iborat.

Tadqiqotning vazifalari quyidagilardan iborat:

parallel korpusning umumiy xususiyatlarini tavsiflash, parallel korpus va boshqa korpuslar orasidagi farqlarni ko‘rsatish;

A.Qodiriyning “O‘tkan kunlar” romani misolida o‘zbek va rus tillarining parallel matnlarini to‘plash va tayyorlash;

o‘zbek va rus tillarining parallel korpuslari bilan amallar bajaruvchi onlayn platformani ishlab chiqish va joriy etish;

parallel matnlar bilan ishlash, shu jumladan tarjimalarni qidirish, taxrirlash, taqqoslash va tahlil qilish uchun funksional imkoniyatlarni ishlab chiqish;

foydalanuvchi tajribasi asosida ishlab chiqilgan platformaning sifati va samaradorligini baholash;

yaratilgan o‘zbek-rus va rus-o‘zbek parallel korpusdan foydalangan holda

²¹ Каримов Р.А. Ўзбек ва инглиз тили параллел корпусини тузишнинг лингвистик ва дастурий масалалари. Автореферат фил. фанлари бўйича фалсафа доктори (PhD). – Тошкент. – 2021;

²² Ахмедова А.Н. Параллел корпусда ўхшатишларнинг лексик-семантик муносабатлари (инглизча-ўзбекча, ўзбекча-инглизча параллел матнлар мисолида). Автореферат фил. фанлари бўйича фалсафа доктори (PhD). – Тошкент. – 2021.

²³ Хидиров О. Миллий корпус учун парсинг дастури яратишнинг лингвистик асослари: Филол.фан.бўйича фалсафа доктори (PhD)...дис. автореф. – Жиззах, – 2021. – 55 б.;

²⁴ Холиёров Ў. Ўзбек тили таълимий корпусини тузишнинг лингвистик асослари: Филол.фан.бўйича фалсафа докт. (PhD) ...дисс. – Термез, – 2021. – 181 б.;

²⁵ Бегматова Г. Ўзбек миллий корпусида идиомалар базасини яратиш: Филол.фан.бўйича фалсафа докт. (PhD) ...дисс. – Термез, – 2021. – 143 б.;

²⁶ Тоирова Г. Ўзбек тили миллий корпусини яратишнинг назарий ва амалий масалалари: Филол.фан.докт. (DSc) ...дисс. – Бухоро, – 2021. – 208 б.

²⁷ Мамажонов А. Текст лингвистикаси. – Тошкент, – 1989.

²⁸ Турниёзов Н., Турниёзова К., Хайруллаев Х. Структур синтаксис асослари. – Тошкент: Фан, – 2019.

²⁹ Қиличев Э. Ўзбек тилининг практик стилистикаси. – Тошкент: Ўқитувчи, – 1985.

³⁰ Ҳакимов М. Ўзбек прагмалингвистикаси асослари. – Тошкент: Akademnashr, – 2013. – 172 б.

³¹ Yo‘ldoshev M., Muhamedova S., Saparniyazova M. Matn lingvistikasi. – Toshkent, – 2021.

lingvistik tahlillar va tadqiqotlar, shu jumladan frazeologik birliklar va lingvistik voqeliklar tahlili va ularning tarjimalarini amalga oshirishdan iborat.

Tadqiqot obyekti o‘zbek-rus tillari parallel korpusi hisoblanadi.

Tadqiqot predmeti o‘zbek-rus va rus-o‘zbek parallel korpuslarida frazeologik iboralar va til voqeliklarini tarjima qilishning o‘ziga xos xususiyatlari, shuningdek, A.Qodiriyning “O‘tkan kunlar” romani misolida ushbu til hodisalarini samarali tarjima qilish metod va usullarini ishlab chiqishdan iborat.

Tadqiqotning usullari. Belgilangan maqsadlarga erishish va vazifalarni amalga oshirish uchun dissertatsiya ishida quyidagi ilmiy uslub va usullardan foydalanildi: tasniflash, tavsiflash, qiyosiy, distributiv tahlil, tarjima, parallel matnlar korpusining qiyosiy tahlili, statistika, korpus matni tahlil qilish usuli, kontrastiv, ekspert bahosh va ma’lumotlarni tahlil qilish.

Tadqiqotning ilmiy yangiligi quyidagilardan iborat:

A.Qodiriyning “O‘tkan kunlar” romanidagi til realiyalari matnlar parallel korpusi orqali tahlil qilinib, tarjimaning asosiy jihatlari va murakkabliklari, ikki tilli korpuslar metodologiyasini ishlab chiqishda leksik, semantik, stilistik va madaniy xususiyatlarni hisobga olish muhimligi asoslangan;

o‘zbek-rus parallel matnlar bilan ishlashga mo‘ljallangan, o‘xshashi bo‘lmagan, real vaqt rejimida ma’lumotlarni tahlil qilish va tahrirlash uchun imkon va vositalarni beradigan, ushbu tillar uchun parallel korpuslar bilan ishlashdagi mavjud bo‘shliqni to‘ldiradigan onlayn platforma ishlab chiqilgan;

lingvistik tadqiqotlar uchun parallel korpusning asosiy jihatlari, ularning o‘ziga xos xususiyatlari, xususan, tarjima strategiyalarini o‘rganish va madaniyatlararo xususiyatlarni tushunishga yordam beradigan turli tillardagi matnlarni solishtirish qobiliyati dalillangan;

o‘zbek-rus parallel korpuslarida A.Qodiriyning “O‘tkan kunlar” romanidagi frazeologik birliklarning batafsil tahlili o‘tkazilib, ularni tarjima qilish usullari hamda madaniy moslashuvi, semantik va kontekstual xususiyatlari aniqlangan.

Tadqiqotning amaliy natijalari quyidagilardan iborat:

rus-o‘zbek va o‘zbek-rus parallel korpusining yaratilgan bazasiga oid materiallar va xulosalar o‘zbek tilining milliy korpusini yaratish, uning ilmiy tajribasini boyitishning eng muhim ilmiy-nazariy asosi ekanligi isbotlangan;

ilgari mavjud bo‘lmagan va hozirda ishlab chiqilgan va yaratilgan rus-o‘zbek va o‘zbek-rus parallel korpusi tarjima tadqiqotlari uchun qimmatli manba ekanligi asoslab berilgan;

parallel korpusning umumiy xususiyatlari tavsiflangan, parallel korpusning boshqa korpuslardan ajralib turadigan jihatlari ko‘rsatilingan hamda chet tilini o‘rganishda parallel korpusning o‘ziga xos o‘rni belgilab berilgan;

parallel korpuslarda frazeologik iboralar va lingvistik voqeliklarning batafsil tahlili, shu jumladan ularning tarjima xususiyatlari, semantik o‘zgarishlari va kontekstual xususiyatlarini aniqlash amalga oshirildi. Ushbu tahlil tillararo farqlar va o‘xshashliklarni chuqurroq tushunish va tadqiq qilishga yordam beradi.

Tadqiqot natijalarining ishonchliligi ilmiy baza, tadqiqot metodikasi va metodologiyasi, shuningdek, yangi muammolarni shakllantirish, yangi vazifalarni

bajarish va tadqiqot natijalarining dissertatsiya ishi nazariy qoidalariga muvofiqligi; O‘zbekiston Respublikasi Oliy attestatsiya komissiyasi tomonidan tavsiya etilgan O‘zbekiston va xorijdagi jurnallarda chop etilgan ilmiy maqolalar; xalqaro va respublika miqyosidagi ilmiy-amaliy anjumanlardagi maqolalar to‘plamlaridagi tezislar bilan ta’minlanadi.

Tadqiqot natijalarining ilmiy va amaliy ahamiyati.

Tadqiqot natijalari korpus lingvistikasi sohasidagi muammolarni o‘rganish, ulardan ham rus, ham o‘zbek tillarining kompyuter tilshunosligi va korpus lingvistikasini o‘rganishda manba sifatida ahamiyatga ega.

Tadqiqot natijalarining amaliy ahamiyati shundan iboratki, dissertatsiya tadqiqotlari materiallari va natijalaridan ma’ruza kurslari va amaliy mashg‘ulotlarni, amaliy filologiya bo‘yicha maxsus kurslarni tayyorlashda, darsliklar, o‘quv qo‘llanmalari, uslubiy tavsiyalar, mashqlar to‘plamlarini yaratish jarayonida foydalanish mumkin.

Tadqiqot natijalarining joriy qilinishi. Parallel korpuslarni yaratishning nazariy va amaliy jihatlarini o‘rganish va tabiiy tillarni qayta ishlash sohasida texnologiyalarni qo‘llashda olingan ilmiy natijalar asosida bir qator tadbirlar amalga oshirildi:

o‘zbek-rus parallel korpuslarida A.Qodiriyning “O‘tkan kunlar” romanidagi frazeologik birliklarning batafsil tahlili o‘tkazilib, ularni tarjima qilish usullari hamda madaniy moslashuvi, semantik va kontekstual xususiyatlariga oid natijalaridan Toshkent davlat o‘zbek tili va adabiyoti universitetida amalga oshirilgan I-OT-2019-42 “O‘zbek va ingliz tillarining elektron she’riy lug‘atini yaratish (inson qiyofasi, xarakteri, tabiati va milliy ramzlari tasvirlari)” fundamental loyihasida foydalanilgan (Toshkent davlat o‘zbek tili va adabiyoti universiteti 2023-yil 9-iyundagi 01/10-1150-son ma’lumotnomasi). Natijada ko‘p tillilik tushunchasi, tarjima algoritmlari va tadqiqotdan matn korpusini yaratishga yondashuv tarjima jarayonini osonlashtirdi, tarjimaning aniqligini oshirdi, hamda lug‘atda qidirish va tahlil qilish uchun keng ma’lumotlar bazasini yaratilgan;

A.Qodiriyning “O‘tkan kunlar” romanidagi til realiyalari matnlar parallel korpusi orqali tahlil qilinib, tarjimaning asosiy jihatlari va murakkabliklari, ikki tilli korpuslar metodologiyasini ishlab chiqishda leksik, semantik, stilistik va madaniy xususiyatlarni hisobga olish muhimligi xususidagi natijalar “Umumiy tilshunoslik” darsligining 13-paragrafiga kiritilgan (Oliy ta’lim, fan va innovatsiyalar vazirligi Buxoro davlat universitetining 2023-yil 12-oktyabrdagi 01-04/3562-son ma’lumotnomasi, GRIF №.112-13). Natijada o‘quv jarayoni osonlashtirdi va tinglovchilarga til tuzilmalarini tahlil qilish va amaliy qo‘llash uchun kengroq vositalar taqdim etilgan;

lingvistik tadqiqotlar uchun parallel korpusning asosiy jihatlari, ularning o‘ziga xos xususiyatlari, xususan, tarjima strategiyalarini o‘rganish va madaniyatlararo xususiyatlarni tushunishga yordam beradigan turli tillardagi matnlarni solishtirish qobiliyati bo‘yicha olingan nazariy va uslubiy bazadan “Madaniyat va ma’rifat” telekanalida “Zamondosh” va “Milliy ta’lim” ko‘rsatuvlariga ssenariy yozishda foydalanilgan (O‘zbekiston milliy teleradiokompaniyasining «Madaniyat va

ma'rifat» telekanali 2023-yil 1-iyundagi 02-02/125-sonli ma'lumotnomasi). Natijada mazkur teledastirlarning ilmiyligi va ommabobligi oshgan.

Tadqiqot natijalarining aprobatsiyasi. Mazkur tadqiqot natijalari 5 ta xalqaro va 2 ta respublika ilmiy-amaliy anjumanlarida muhokamadan o'tkazilgan.

Tadqiqot natijalarining e'lon qilinganligi. Dissertatsiya mavzusi yuzasidan jami 15 ta ilmiy ish chop etilgan, jumladan, O'zbekiston Respublikasi Oliy attestatsiyasi komissiyasining doktorlik dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan ilmiy nashrlarda 8 ta maqola, ulardan 3 tasi respublika va 5 tasi xorijiy jurnallarda e'lon qilingan.

Dissertatsiyaning tuzilishi va hajmi. Dissertatsiya kirish, uch bob, umumiy xulosa, foydalanilgan adabiyotlar ro'yxati va ilovadan tashkil topgan bo'lib, ishning umumiy hajmi 152 sahifadan iborat.

DISSERTATSIYANING ASOSIY MAZMUNI

Dissertatsiyaning **“Kirish”** qismida mavzuning dolzarbligi va zarurati asoslanib, tadqiqotning maqsadi va vazifalari, obykti va predmeti, ilmiy va amaliy ahamiyati aniqlangan. Tadqiqotning O'zbekiston Respublikasi fan va texnologiyalari taraqqiyotining ustuvor yo'nalishlariga mosligi ko'rsatilib, uning ilmiy yangiligi va amaliy natijalari bayon etilgan. Tadqiqot natijalarining amaliyotga joriy qilinishi, ishning aprobatsiyasi, natijalari, e'lon qilingan ishlar va dissertatsiyaning tuzilishi bo'yicha ma'lumotlar keltirilgan.

“Kompyuter va korpus tilshunosligi: muammolar va istiqbollar” deb nomlangan dissertatsiyasining birinchi bobi uch bo'limdan iborat.

“Kompyuter tilshunosligining rivojlanishi va hozirgi holati” nomli birinchi bo'limda ushbu ilmiy sohaning rivojlanishi va hozirgi holatining asosiy jihatlari ko'rib chiqilgan. Unda turli jihatlar, jumladan, kompyuter tilshunosligi tarixi, maqsad va vazifalari, yondashuv va usullari, shuningdek, uning hozirgi vaqtdagi ahamiyati va qo'llanilishi ko'rib chiqiladi.

Kompyuter tilshunosligining rivojlanishi 1957-yilda Noam Xomskiy g'oyalarining paydo bo'lishi bilan boshlandi. Olim til grammatikasini cheksiz to'g'ri jummalarni yaratishga qodir mexanizm sifatida taqdim etishni taklif qildi. Bu XX asr oxiri – XXI asr boshlarida matematika va tilshunoslik kesishmasida, kompyuter tilshunosligining mustaqil fan sifatida paydo bo'lishiga olib keldi.

Vaqt o'tishi bilan olimlar kompyuterlarning so'zlarni qayta ishlashni avtomatlashtirish imkoniyatlarini angladilar. 1959-yilda italyan olimi R.Buza kompyuterdan foydalangan holda qadimiy qo'lyozmalar uchun lug'atlar yaratdi. 1969-yilda Bezansonda (Fransiya) lingvistik tadqiqot Markazi tashkil etildi, u yerda katta lug'atlar uchun kataloglar yaratildi.

Rossiyada kompyuter tilshunosligi rus kibernetikasining asoschisi hisoblangan taniqli matematik va kibernetik A.Lyapunov nomi bilan chambarchas bog'liq. Ammo kompyuter tilshunosligi tarixining boshlanishi XX asr boshlariga to'g'ri keladi va semiotikaga – belgilar va ramzlar haqidagi fanga qiziqishning ortishi bilan bog'liq deb hisoblanadi.

Kompyuter tilshunosligining ta'rifi har xil bo'lishi mumkin, ammo tor ma'noda

u tilning faoliyatini modellashtirish uchun kompyuter vositalaridan foydalanadigan fanlararo amaliy soha hisoblanadi.

Bo‘limda turli darajadagi til tuzilishini tahlil qilish va modellashtirish, matn yaratish va lingvistik statistikani o‘z ichiga olgan tilni avtomatik qayta ishlashga alohida e‘tibor qaratilgan. Kompyuter tilshunosligining turli sohalari, jumladan, tabiiy tillarni qayta ishlash, elektron korpuslar va lug‘atlarni yaratish, matnlarni avtomatik tarjima qilish, matndan ma‘lumot olish, avtomatik havola qilish, bilimlarni boshqarish tizimlari va boshqalar ko‘rib chiqilgan.

“Korpus va uning tarkibiy qismlari” deb nomlangan ikkinchi bo‘limda XX-XXI asrlarda korpus tilshunosligining rivojlanishi va uning asosiy tarkibiy qismlari haqida ma‘lumot keltirilgan.

1990-yillarning boshlarida TEI – matnni kodlash tashabbusi degan dastur (Text Encoding Initiative) ishlab chiqildi, bu matnni belgilash uchun asosiy standartga aylanib, turli korpuslarning mosligini ta‘minladi. Ushbu tashabbus ixtisoslashtirilgan dasturlardan foydalangan holda raqamli matnlarni yaratish bilan shug‘ullanadigan ishlab chiqaruvchilarni birlashtirdi.

Korpus tilshunosligi rivojlanishidagi muhim qadam — standartlar va tabiiy tillarni qayta ishlashning yangi usullarini (NLP) ishlab chiqilishi bo‘ldi. Bu matnli ma‘lumotlar bilan ishlash uchun yangi imkoniyatlarni ochdi, matnni tahlil qilish va avtomatik qayta ishlashda yutuqlarni ta‘minladi.

Yevropa mamlakatlarida “korpus bumi”, ya‘ni turli xil ingliz tilidagi korpuslarni yaratishi boshlandi. Shuningdek, Rossiyada matnlarni raqamlashtirish va rus tilining milliy korpusini yaratish jarayoni boshlandi. Ushbu rivojlanish doirasida Y.V.Bogoyavlenskaya, M.I.Solniskina, G.M.Gatiyatullina, N.V.Kozlova, V.P.Zaxarov, D.Sichinava kabi olimlar tomonidan korpuslarning turli tasniflari taklif qilindi.

Zamonaviy korpus texnologiyalari birinchi o‘ringa foydalanuvchi qulayligini qo‘yadi. “Sketch Engine” kabi platformalar tadqiqotchilar uchun korpuslar bilan ishlashni yanada qulay va oson qilish uchun turli xil sozlanish variantlari, interfeyslar, maslahatlar va texnik yordamni taklif etadi.

Uchinchi bo‘lim **“Parallel korpusning umumiy xususiyatlari, tuzilishi va tarkibi. Parallel korpusning umumiy tavsifi”** deb nomlanadi. Bu bo‘limda parallel korpuslarning umumiy xususiyatlari va tuzilishi va tarkibini haqida gap ketadi.

Parallel korpuslar – bu tegishli tarjimalarga ega bo‘lgan ikki yoki undan ortiq tillardagi matnlar to‘plami. Ular tilshunoslik, mashina tarjimasini, kompyuter tilshunosligi va boshqa sohalarda keng qo‘llaniladi.

Parallel korpusning tarkibi uning maqsadi va ko‘lamiga bog‘liq. U badiiy adabiyot, ilmiy maqolalar, rasmiy hujjatlar va boshqalar kabi turli janrdagi matnlarni o‘z ichiga olishi mumkin. Bundan tashqari, korpusda til o‘zgarishi va rivojlanishini o‘rganish uchun turli davr va uslubdagi matnlar bo‘lishi mumkin.

Rus tilining milliy korpusida ingliz, nemis, fransuz, ispan, italyan, polyak, ukrain va belorus tillarini o‘z ichiga olgan 8 ta ikki tilli parallel korpuslar mavjud. Korpusning umumiy hajmi 37 milliondan ortiq so‘zni tashkil etadi va rus tili bilan bog‘langan har bir til uchun morfologik belgilash keltirilgan. Asosan slavyan

tillaridan iborat, 25 tilda 9 ta parallel matnlarni o‘z ichiga olgan, 5 millionga yaqin so‘zlardan iborat ko‘p tilli korpus ham mavjud.

Bo‘limda parallel korpuslar tarjima usullarini o‘rganish, tarjima lug‘atlarini takomillashtirish va tarjimani yangi bosqichga olib chiqish uchun keng imkoniyatlarga ega ekanligi ta’kidlanadi. Shuningdek, ular asl matn va tarjima o‘rtasidagi grammatik va stilistik farqlarni bartaraf etish uchun tarjimonlar tomonidan qo‘llaniladigan turli strategiyalarni o‘rganish uchun manba matnlari va ularning tarjimalarini solishtirish imkonini beradi.

“O‘zbek-rus va rus-o‘zbek parallel korpusini ishlab chiqish va joriy etish” deb nomlangan dissertatsiyasining ikkinchi bobi uchta paragrafdan iborat bo‘lib, ular o‘zbek-rus va rus-o‘zbek parallel korpusini ishlab chiqish va joriy etish masalalariga bag‘ishlangan.

Ushbu bob parallel korpusni yaratishning asosiy vazifalarini, korpusga kirish uchun veb-dasturni ishlab chiqish texnologiyasini, shuningdek parallel korpus ma’lumotlar bazasini yaratishning asosiy bosqichi sifatida matnni tenglashni o‘z ichiga oladi.

“O‘zbek-rus rus-o‘zbek parallel korpusini yaratishning asosiy vazifalari” deb nomlangan 2.1-bo‘limida ushbu korpusni yaratish bilan bog‘liq muhim jihatlar va maqsadlar muhokama qilinadi. O‘zbek va rus tillari va madaniy merosini asrab-avaylash va ommalashtirish mazkur bo‘limda asosiy vazifalardan biri sifatida keltirilgan. Parallel korpus yordamida ikkala tilda ham matnlarni tuzish va tizimlashtirish mumkinligi, bu esa ushbu xalqlarning tarixi, adabiyoti va madaniy an’analarini saqlash va o‘rganish imkonini berishi ko‘rsatib o‘tilgan.

Parallel korpusning yana bir vazifalaridan biri o‘zbek va rus tillarini o‘qitish va o‘rganishni takomillashtirishdan iborat. Korpusda mavjud bo‘lgan haqiqiy va xilma-xil matnlar talabalarga o‘qish, yozish, tinglash va gapirish qobiliyatlarini rivojlantirishga yordam beradi. Bu talabalarga til muhitiga chuqurroq kirib borish va ikkala tilning madaniy xususiyatlarini o‘rganish imkonini beradi.

O‘zbek-rus va rus-o‘zbek parallel korpusini yaratishdagi qiyinchiliklardan biri o‘zbek tilining milliy korpusining yo‘qligi hisoblanadi. Biroq, o‘zbek tilining morfoanalizatorini yaratish bo‘yicha ishlar olib borilgan va hozirda tadqiqotchilar oldida qolgan korpuslarni ishlab chiqish vazifasi turibdi.

Shuning uchun o‘zbek-rus va rus-o‘zbek parallel korpusi lingvistik va madaniy merosni saqlash va rivojlantirish, tillarni o‘qitish va o‘rganishni takomillashtirish, til texnologiyalarini rivojlantirish uchun muhim ahamiyatga ega bo‘lib, O‘zbekiston, Rossiya va boshqa mamlakatlar o‘rtasidagi aloqa va hamkorlikni rivojlantirishga yordam beradi.

“O‘zbek-rus va rus-o‘zbek parallel korpusli veb-illovalarni ishlab chiqish texnologiyasi” 2.2-bo‘limda o‘zbek va rus tillarida parallel korpusni yaratish uchun veb-ilovani ishlab chiqish texnologiyasi va usullari tavsiflangan. Loyiha zamonaviy texnologiyalar va vositalar yordamida ishlab chiqilgan, shuningdek, bir necha bosqich va funktsionallikni o‘z ichiga qamrab oladi.

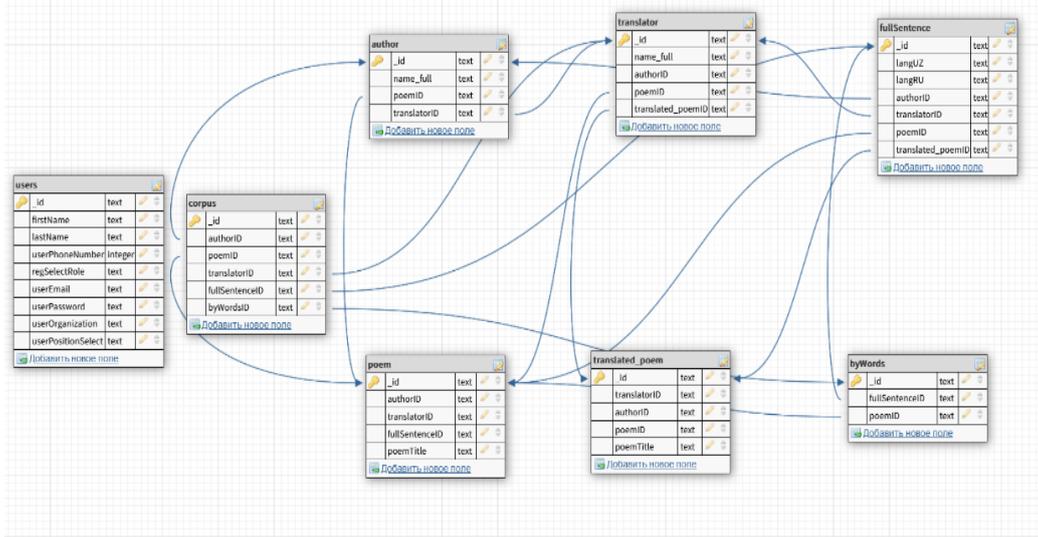
Loyiha quyidagi texnologiyalar va dasturlash tillaridan foydalanishga asosida yaratilgan: React, Node.js, JavaScript, MongoDB va boshqalar. React -

foydalanuvchi interfeysini yaratish, Node.js – server qismi va MongoDB ma'lumotlar bazasi bilan o'zaro ishlash uchun, JavaScript dastur mantig'iyiligini yozish uchun qo'llanilgan.

Dasturni yaratish turli xil matn muharrirlarini tahlil qilgandan so'ng boshlandi va natijada VS Code muharriri tanlandi. Ushbu muharrir foydalanuvchilarga qulay interfeysga, ko'plab funksiyalar va plaginlarga ega, shuningdek, u krossplatformalidir.

MongoDB, ma'lumotlarni saqlash uchun eng mashhur ma'lumotlar bazasini boshqarish tizimlaridan biri (MBBT) tanlangan. MongoDB foydalanish qulayligi, krossplatformalik va masshtablikka ega. Uchinchi tomon dasturlarini o'rnatmasdan MBBT bilan ishlash uchun o'rnatilgan Node.js serveri ham tanlandi.

Ma'lumotlar bazasi dizayni 8 ta obyektни aniqlash va bog'lashni o'z ichiga oladi: user, corpus, author, translator, fullSentence, poem, byWords va translated_poem.



1- Rasm. MB tuzilishi

1-rasmda DbDesigner web-servisi yordamida yaratilgan ma'lumotlar bazasi tuzilishi ko'rsatilgan. Bundan tashqari, rasimga kiritilmagan ba'zi obyektlar (1-rasm) rasmlar va foydalanuvchi ma'lumotlarini saqlash uchun ishlatiladi.

Foydalanuvchilarni avtorizatsiya qilish uchun tegishli kod tuzilmalari ishlab chiqilgan, jumladan server tomonida ro'yxatdan o'tishni tekshirish va foydalanuvchi ma'lumotlarini tekshirish. Muvaffaqiyatli avtorizatsiyadan so'ng, foydalanuvchi ma'lumotlari va token brauzerning lokal xotirasida saqlanadi.

Asosiy fayllardan biri bu usersReg.js. UsersReg.js fayli foydalanuvchi ro'yxatdan o'tganligini tekshirish uchun ishlatiladi. Ushbu tekshirish natijasiga qarab, foydalanuvchi tegishli sahifaga yo'naltiriladi. Ikkita rol mavjud: administrator va oddiy foydalanuvchi. Faqat root huquqiga ega foydalanuvchilar yangi foydalanuvchilarni ro'yxatdan o'tkazish imkoniyatiga ega. Agar foydalanuvchi administrator sifatida tizimga kirgan bo'lsa, u qo'shimcha interfeys sahifalariga kirish huquqiga ega bo'ladi, u yerda siz parallel korpus uchun materiallarni yuklashingiz, ularni boshqarishingiz, yangi foydalanuvchilar va dastur mualliflarini ro'yxatdan o'tkazishingiz mumkin. Oddiy foydalanuvchilar bunday

imtiyozlarga ega emaslar, ular faqat parallel holatda materiallar bilan ishlashlari va o'z ishlarining natijalari to'g'risida hisobot olishlari mumkin.

Quyidagi struktura foydalanuvchi avtorizatsiya uchun mas'ul (2-rasm).

```
return (  
  <section className='login_section'>  
    <div className='login_box'>  
      <div className='login_form'>  
        <h2>Login</h2>  
        <div>  
          <input  
            type='email'  
            placeholder='Foydalanuvchi email'  
            required  
            value={logEmail}  
            onChange={(e) => setLogEmail(e.target.value)}  
          />  
          <span className='login_user_icon'>  
            </span>  
        </div>  
        <div className='login_inputBx'>  
          <input  
            type='password'  
            placeholder='Password'  
            required  
            value={logPassword}  
            onChange={(e) => setLogPassword(e.target.value)}  
          />  
          <span className='login_lock_icon'>  
            </span>  
        </div>  
      </div>  
    </div>  
  </div>  
)
```

2-rasm. Avtorizatsiya uchun kod tuzilmasi

```
UserModel.findOne({ userEmail })  
  .then(savedUser => {  
    // console.log(savedUser)  
    if (!savedUser) {  
      res.status(422).json({ error: "Bunday foydalanuvchi topilmadi" })  
    }  
    bcrypt.compare(userPassword, savedUser.userPassword)  
      .then(doMatch => {  
        if (doMatch) {  
          const token = jwt.sign({ _id: savedUser._id }, SECRET_KEY)  
          const  
            {  
              _id,  
              firstName,  
              lastName,  
              userEmail  
            } = savedUser  
          res.status(200).json({  
            token,  
            user: {  
              _id,  
              firstName,  
              lastName,  
              userEmail  
            }  
          })  
        }  
      })  
  })
```

3-rasm. Foydalanuvchining verifikatsiya (tekshirish) tuzilmasi

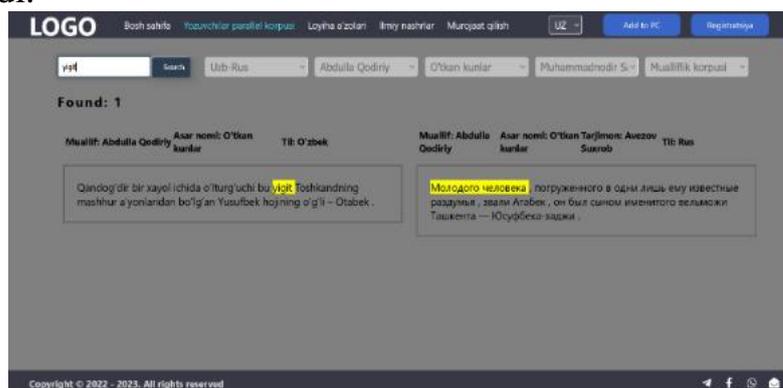
Foydalanuvchi o'z ma'lumotlarini matn va parol atributlari yordamida avtorizatsiya shakliga kiritadi. Ushbu formadagi yuborish tugmasi foydalanuvchiga ushbu ma'lumotlarni tekshirish uchun yuborish imkonini beradi. Ushbu tugmani bosgandan so'ng, usersReg.js funksiyasi serverdagi kiritilgan ma'lumotlarni ma'lumotlar bazasidagi ma'lumotlar bilan taqqoslab tekshiradi. Agar kiritilgan ma'lumotlar ma'lumotlar bazasidagi ma'lumotlarga mos keladigan bo'lsa, u holda foydalanuvchi uchun seans yaratiladi va u saytning asosiy sahifasiga - Homepage.js ga o'tadi. (3-rasm)

Tadqiqotchi-foydalanuvchi uchun asosiy sahifa "Parallel korpus" sahifasi bo'lib, unda parallel korpus matnlarining barcha materiallari ko'rsatiladi. Foydalanuvchi korpusni tili va turi bo'yicha filtrlash, shuningdek tegishli ma'lumotlarni olish uchun qidiruv maydonidan foydalanish imkoniyatiga ega.



4-rasm. Parallel korpus qidiruv tizimi sahifasi

Ushbu sahifada asosiy vosita qidiruv maydonidir. So‘rovni kiritganingizda, platforma qidiruv natijalarini o‘z ichiga olgan filtrlangan ma‘lumotlar bilan tegishli matnni taqdim etadi.

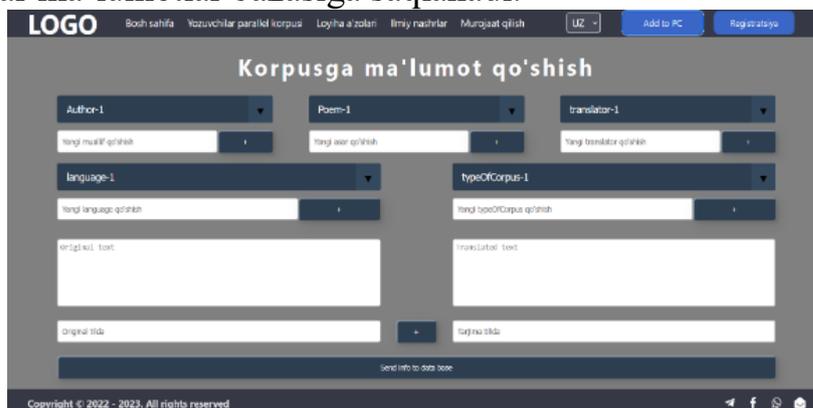


5-rasm. Parallel korpusdan qidiruv natijasi

Ushbu platformada kiritilgan ma‘lumotlar semantik ravishda teglanadi va qidiruv paytida foydalanuvchi teglangan natijani oladi. Ushbu platformaning afzalliklaridan biri shundaki, unda boshqa hech qanday platforma qila olmaydigan o‘zbekcha-ruscha matn bilan markerlangan holda ishlash mumkin.

Sketch Engine – bu matn korpuslari bilan ishlashning eng ilg‘or platformalaridan biri, ammo u o‘zbek tilida matn bilan ishlash imkoniyatini bermaydi. Shu munosabat bilan Sketch Engine o‘zbekcha matnni markerlash funksiyasiga ega emas.

Parallel korpusga ma‘lumot qo‘shish muallifning ismi va familiyasini kiritish, matn nomini kiritish, tarjimon haqidagi ma‘lumotlar, tillarning konfiguratsiyasi va korpus turini tanlash, matnni asl tilga va tarjima matnini yuklash kabi bir necha bosqichlarni o‘z ichiga oladi, shuningdek semantik teglashni ham bajaradi. Shundan so‘ng, ma‘lumotlar ma‘lumotlar bazasiga saqlanadi.



6-rasm. Parallel korpusga ma‘lumot qo‘shish sahifasi

Shunday qilib, ishlab chiqilgan veb-ilova tadqiqotchilarga o‘zbek va rus tillarining parallel korpusi bilan ishlash, yangi ma‘lumotlarni qo‘shish, matnlarni filtrlash va qidirish, ma‘lumotlarni qidirish va tahlil qilishni yaxshilash uchun semantik teglash o‘tkazish imkoniyatini beradi.

“Tenglashtirish – parallel korpus ma‘lumotlar bazasini yaratishning asosiy bosqichi sifatida” nomli 2.3. - bo‘limda parallel korpus ma‘lumotlar bazasini yaratish jarayonida tenglashtirishning roli va ahamiyatini muhokama qilinadi. Tenglashtirish – bu mos keladigan matn qismlarini turli tillarda bog‘lash va ular

o‘rtasida parallel aloqalarni yaratishga imkon beradigan muhim qadamdir.

Tenglashtirishning asosiy jihatlaridan biri bu turli tillardagi matnlar o‘rtasidagi o‘xshashlikni eng samarali va aniq usul yoki algoritmnin tanlashidir.

Abdulla Qodiriyning “O‘tkan kunlar” romanidagi frazeologik birliklarni o‘zbek tilidan rus tiliga tarjima qilish uchun ibora darajasida tenglashtirish usuli qo‘llanilgan. Bu usul jumalarga qaraganda kichikroq matn birliklarini hisobga olish imkonini beradi, bu esa turli tillardagi matnlar orasidagi o‘xshashliklarni to‘g‘ri aniqlash imkonini beradi.

Ibora darajasida tenglashtirish usulidan foydalanish o‘zbek va rus tillaridagi tegishli frazeologik birliklarni aniqroq bog‘lash va olingan ma‘lumotlardan frazeologik birliklarning tarjimasini tahlil qilish uchun imkon berdi.

Quyida parallel korpusdan misol keltirilgan:

Muallif: Абдулла Қодирий Asar nomi: Ўткан кунлар Til: O‘zbek	Абдулла Қодирий Ўткан кунлар Мухаммаднодир Til: Rus
Yusufbek hojining «O‘g‘lim yosh, sen dunyoning issiq-sovug‘ini tatig‘an va manim ishongan kishimsan.	Из головы не выходил наказ Юсуфбека хаджи «Сын мой еще молод, ты же прошел огонь и воду, видел и знаешь многое, предан мне.

7-rasm. Tarjima natijasi

Ushbu gapda “sen dunyoning issiq-sovug‘ini tatig‘an” o‘zbek maqoli ishlatilgan bo‘lib, so‘zma-so‘z “ТЫ ИСПЫТАЛ ТЕПЛО И ХОЛОД МИРА” degan ma‘noni anglatadi va uning ruscha tarjimasini romanda “ТЫ ЖЕ ПРОШЕЛ ОГОНЬ И ВОДУ” deb berilgan.

“Sen dunyoning issiq-sovug‘ini tatig‘an” maqoli ko‘p hayotiy sinovlarni boshidan kechirgan, tajriba, donolik orttirgan insonni tasvirlash uchun ishlatiladi. Shu nuqtayi nazardan, bu maqolda Yusufbek-hoji so‘zlari adresatning (Hasanali) tajribali va ishonchli ekanligi ta‘kidlangan.

“ТЫ ЖЕ ПРОШЕЛ ОГОНЬ И ВОДУ” ning ruschadan tarjimasini ham inson ko‘p qiyinchilik va sinovlarni boshdan kechirganini, tajribali va ishonchli bo‘lishini ko‘rsatadi. Bunday holda, tarjima umumiy o‘zbek maqolining semantik yukini va ruhini saqlab qoladi.

Ushbu misolda frazeologik birliklarni tarjima qilish texnikasi va usullarini tahlil qilib, biz “ekvivalentlik” tarjima usulini ajratib ko‘rsatishimiz mumkin: asl maqolni shunga o‘xshash ma‘noga ega bo‘lgan maqsadli tilda (rus tilida) o‘xshash maqol bilan almashtirish. Bunday holda, o‘zbek maqoli iboraning umumiy ma‘nosi va hissiy rangini saqlaydigan “ПРОШЕЛ ОГОНЬ И ВОДУ” rus maqoli bilan almashtirildi.

Shunday qilib, tarjima asl nusxaga to‘g‘ri keladi va maqolning semantik jihati saqlanib qoladi. Ekvivalentlik usulidan foydalanib, tarjimon rus iborasidan foydalangan holda o‘zbek maqolining ma‘nosini muvaffaqiyatli yetkazdi, matnini rus tilida so‘zlashuvchi auditoriya uchun qulay va tushunarli qildi.

Bo‘lim mazmuni taqdimotini davom etib, shuni ta‘kidlash joizki, so‘z darajasida tenglashtirish ham turli tillardagi matnlarni aniqroq bog‘lash imkonini beruvchi muhim usul sifatida ko‘rib chiqiladi.

Bundan tashqari, so‘z darajasidagi tenglashtirish mashina tarjimasini va tabiiy tilni qayta ishlash sohasida keng qo‘llanilishi ta‘kidlangan. Bunday usul avtomatik

tarjima, matnni tahlil qilish, nutqni aniqlash va boshqa vazifalarda qo'llaniladi. Moses, GIZA++, FastAlign va OpenNMT kabi o'z algoritmlarida so'z darajasida tenglashtirishni qo'llaydigan maxsus loyihalar va tizimlarga misollar keltirish mumkin.

Bo'lim xulosasida har bir tenglashtirish turining afzalliklariga qaramay (iboralar, so'zlar va belgilar darajasida), ulardan foydalanishda muayyan tillar va vazifalarning o'ziga xos xususiyatlarini hisobga olishi kerakligi ta'kidlab o'tilgan. O'zbek va rus tillaridagi matnlarni tabiiy tilni qayta ishlash sohasidagi xususiyatlari va ehtiyojlarini hisobga olgan holda, ularni samarali tenglashtirish bo'yicha qo'shimcha tadqiqotlar olib borish va yangi usul va algoritmlarni ishlab chiqish muhim ahamiyatga ega ekanligi aytib o'tilgan.

“Tarjima amaliyotida parallel korpuslar usullari, afzalliklari va istiqbollari: “O'tkan kunlar” romanining o'zbekcha-ruscha tarjimasi misolida” deb nomlangan dissertatsiyasining uchinchi bobi uch paragrafdan iborat.

“Abdulla Qodiriyning “O'tkan kunlar” romani misolida o'zbek va rus tillarining parallel korpusidagi frazeologik birliklarni tarjima qilish uslub va ustunliklarini tahlil qilish” nomli 3.1-bo'limda matnlarning parallel korpusi yordamida frazeologik birliklarni o'zbek tilidan rus tiliga tarjima qilish metodlari va usullarini tahlil qilish amalga oshirildi. Ushbu tadqiqotda Abdulla Qodiriyning “O'tkan kunlar” romani o'zbek tilidagi ko'plab frazeologik birliklar va idiomatik iboralarni o'z ichiga olgan matn namunasi sifatida ishlatilgan.

Tadqiqot frazeologik birliklarni va ularning rus tiliga tarjimalarini parallel korpusda taqqoslashga asoslangan bo'lib, bu tarjima usullarini batafsil tahlil qilishga imkon beradi. Tahlil aniq sanalar, familiyalar yoki ismlarga havola qilinmasdan amalga oshiriladi, chunki asosiy e'tibor frazeologik birliklarni tarjima qilish usullari va metodlariga qaratiladi.

Bo'lim doirasida frazeologik birliklarni tarjima qilishning ekvivalentlik, moslashish, kalka va so'zma-so'z tarjima kabi turli usullari va uslublari ko'rib chiqiladi. Frazeologik birliklarning tarjimasiga misollar keltirilgan va ularning asl nusxaga mosligi, ma'no va obrazning yetkazilishi tahlil qilingan. Tahlilda frazeologik birliklarning semantik xususiyatlari, shuningdek madaniy va kontekstual jihatlari hisobga olinadi.

Quyida parallel korpusdan misol keltirilgan:

Muallif: Abdulla Qodiriy	Asar nomi: O'tkan kunlar	Til: O'zbek	Muallif: Abdulla Qodiriy	Asar nomi: O'tkan kunlar	Til: Rus
Ammo sen bo'lsang uning bilan tenglashar eding. «Teng-tengi bilan, tezak qopi bilan».			А вот ты под стать ему. Недаром говорят: «Ровня к ровне, а кизяку место в мешке».		

8-rasm. Parallel korpusda frazeologizm tarjimasi

To'ybekaning munosib kuyov – “yosh mehmon” haqidagi yarim hazil so'zlari Kumushning g'azabini qo'zg'atadi, Kumush masxara bilan unga To'ybekaning o'zi turmushga chiqishini taklif qiladi, bunga xizmatkor maqol bilan javob beradi:

“Teng-tengi bilan, tezak qopi bilan”³². Bu maqolni M.Safarov so‘zma-so‘z tarjima qilgan: “Ровня к ровне, а кизяку место в мешке”³³. Bu yerda muvaffaqiyatli tarjima bordek ko‘rinishi mumkin: ma’no aniq yetkazilgan va aytilishi maqolga loyiqdir. Boshqa tomondan, maqolning o‘zbek xalqi hayoti bilan chambarchas bog‘liq bo‘lgan milliy rangga ega ekanligini hisobga olsak, tarjimaning muvaffaqiyatini yana bir bor tan olishimiz kerak bo‘ladi. Biroq, “кизяк” so‘zining satr osti sharhida tushuntirish berilgan – “навоз”, bu tarjima to‘g‘ri ekanligini inkor etadi. Gap shundaki, “кизяк” so‘zi yoqilg‘i sifatida ishlatiladigan g‘isht (yoki yadro) shaklida quritilgan go‘ngni anglatadi ³⁴, ya’ni, bu yerda biz go‘ngning o‘zi haqida emas, balki yoqilg‘i sifatida ishlatiladigan va ko‘pincha qoplarga solingan undan tayyorlangan mahsulot haqida gapiramiz. Noto‘g‘ri tushuntirish natijasida rusiyazabon o‘quvchi “навоз” so‘zini “qop” so‘zi bilan bog‘lay olmaydi.

Parallel korpusdagi frazeologik birliklarni tarjima qilish metodlari va usullarini o‘rganish muhim ahamiyatga ega, chunki u o‘zbek va rus tillari o‘rtasidagi frazeologik iboralarning ma’no va obrazlilikini yetkazishning tarjima jarayoni va samarali usullarini yaxshiroq tushunish imkonini beradi. Parallel korpus frazeologik birliklar tilning madaniy konteksti va xususiyatlarini yetkazishda muhim rol o‘ynaydigan adabiy, badiiy va boshqa matnlarni tarjima qilishda amaliy qo‘llaniladi.

“O‘zbek-rus tili misolida voqelikni o‘rganish va tarjima qilish uchun parallel korpusdan foydalanishning afzalliklari” nomli 3.2.- bo‘limida voqelikni o‘rganish, o‘zbek va rus tillariga tarjima qilish uchun parallel korpusdan foydalanish bilan bog‘liq muhim fikrlar keltirilgan.

Parallel korpusning asosiy afzalliklaridan biri bu - o‘zbek tilidagi har bir matn rus tilidagi matn bilan bog‘langanligidir. Bu real matnlar kontekstida lingvistik hodisalar va tarjima usullarini taqqoslash imkoniyatini beradi.

Parallel korpusdan foydalanish bizga o‘zbek va rus tillarida so‘zlashuvchilarning madaniyati, dini, tarixi va hayotining boshqa sohalarini aks ettiruvchi voqelikni yanada samarali o‘rganish imkonini beradi.

Realiylar - bu kontekst va madaniy xususiyatlarni hisobga olmagan holda boshqa tilga to‘g‘ri va adekvat yetkazish qiyin bo‘lgan so‘zlar va iboralardir. Parallel korpusdan foydalanish voqelikni ularning mazmuni va kontekstini hisobga olgan holda tushunish va tarjima qilishni osonlashtiradi.

Shunday qilib, ushbu bo‘limning asosiy maqsadi “O‘tkan kunlar” romanining asl matni va M.Safarovning ruscha tarjimasini taqqoslash asosida realiyning o‘ziga xos xususiyatlarini va ushbu voqelikni tarjima qilish usullarini o‘rganishdir. Tahlil jarayonida biz ushbu tadqiqot uchun ishlab chiqilgan o‘zbek va rus tillarining parallel korpusidan foydalanamiz.

³² Qodiriy A. O‘tkan kunlar //Sharq nashriyoti. – Toshkent. – 1994. – C, 16

³³ Кадыри А. Минувшие дни // Главная редакция издательско-полиграфической акционерной компаний // –Ташкент: SHARQ, –2009. –С. 31

³⁴ <https://dic.academic.ru/dic.nsf/ushakov/833562>

Saroyning to'rida boshqalarga qarag'anda ko'rkamrak bir hujra, anovi hujralarga kiygiz to'shalgani holda bu hujrada qip-qizil gilam, uttalarda bo'z ko'rpalar ko'rilgan bo'lsa, munda ipak va adras ko'rpalar, narigilarda qora charog' sasig'anda, bu hujrada sham' yonadir, o'zga hujralarda yengil tabi'atlik, serchaqchaq kishilar bo'lg'anida bu hujraning egasi boshqacha yaratilishda.

В глубине двора — отличающаяся изысканным убранством уютная комната. Если в других комнатах полы устланы кошками, то здесь — вместо них лежат ярко-красные ковры; если остальные снабжены бязевыми одеялами, здесь их заменяют курпачи из шелка и адраса; если везде коптят плошки, тут резво пляшет пламя свечи. Да и обитатель помещения, не в пример легкомысленным, заводным постояльцам других комнат, держится иначе.

9-rasm. Parallel korpusda realiyar tarjimasasi

Bu misolda tanlangan so'zlarni realiya deb tasniflash mumkin, chunki ular tilning milliy va tarixiy tomonlarini aks ettiradi. “Hujra”, “kurpalar”, “adras” so'zlarini ma'nosiga ko'ra maishiy voqelik deyishimiz mumkin.

“Hujra” so'zi arab tilidan kelib chiqqan bo'lib, ikki ma'noga ega. Bu so'zning birinchi ma'nosi “kichik xona”, ikkinchisi – “madrasadagi o'qish yoki yashash uchun mo'ljallangan kichik xonasi” ma'nosiga ega bo'lib, bu xonadan asosan Markaziy va G'arbiy Osiyodagi ta'lim muassasalari o'qituvchilari va talabalari foydalangan.

Asarda “hujra” so'zi Otabekning Marg'ilonda yashagan paytidagi karvonsaroyning kichkina xonasiga nisbatan qo'llanilgan. Hozirgi kunda bu so'z o'zbek tilining faol lug'atini tark etgan va faqat ikkinchi ma'nosini saqlab qolgan, ya'ni “madrasa, karvonsaroy, masjidlardagi kichkina xona”. Muallifning tasviri bu szdan foidalanishi burchakdagi kichki va korongu honaning tarixiy kiyofasini qilishga yordam berdi.

So'zning tarjimasida “perifraza” usuli qo'llanilgan: tarjimon “qulay xona” iborasidan foydalanib, asl so'zning semantik ma'nosini yetkazishga harakat qilgan. Biroq, “hujra” so'zining tarixiy va diniy mazmuni tarjimada saqlanmagan, chunki “xona” so'zining o'zi uchta lug'atviy ma'noga ega, ammo ularning hech biri “hujra” so'zining ma'nosiga to'g'ri kelmaydi. Shunga qaramay, tarjimon ma'noni aniqroq aks ettirish uchun “qulay” sifatidan ham foydalanganligini ta'kidlash joiz. Shu tafayli kitobxon vaqtinchalik yashash uchun kichkina, ammo qulay xonaning tasvirini ko'radi.

Ko'rib chiqilayotgan misoldagi ikkinchi “ko'rpa” so'zi ham maishiy voqelikka ishora qiladi va aniq milliy xususiyatga ega. Romanning rus tiliga tarjimasida “ko'rpa” so'zini o'xshash “курпача” so'zi bilan noto'g'ri almashtirish holati kuzatiladi.

Asarda “ko'rpa” so'zi semantik jihatdan tancha yoki sandalni yopish uchun maxsus yoping'ich ma'nosini anglatadi. Romanda kitobxonga bu uy-ro'zg'or buyumi qishning sovuq kunlarida issiqlikni saqlash uchun maxsus narsa sifatida ishlatilishi ayon bo'ladi. Biroq, rus tiliga tarjimada bu so'z “курпача”ga almashtirilgan va shu bilan boshqa ma'noga ega bo'ldi. “Курпача” so'zi o'zbek tilida “maxsus matras, o'tirish uchun tor paxtadan yoping'ich, paxta yostig'i” degan ma'noni anglatadi. Bundan kelib chiqadiki, “ko'rpa” so'zining o'zbek tilidan rus tiliga tarjimasida noaniqlik paydo bo'lgan, bu tarjima qilingan birlikning ma'nosini

buzadi, garchi madaniy va tarixiy xususiyatlar o'zgarishsiz qolsa ham. Bunga o'zbek xalqiga xos xususiyatlar va turmush tarzini aks ettirishning samarali usullaridan biri bo'lgan tarjima usuli – transliteratsiya sabab bo'ldi.

Shuning uchun quyidagicha xulosaga kelish mumkin: 1) o'zbek tilidagi “ko'rupa” so'zining ekvivalenti rus tilidagi “odeyalo, pokrivalo” so'zlari degan xulosaga kelish mumkin; 2) “ko'rpacha” so'zi rus tilida to'g'ridan to'g'ri ekvivalentga ega emas, ammo semantikasi bilan bog'liq “ko'rupa” so'zidan sezilarli farq qiladi va “o'tirish uchun mo'ljallangan paxta bilan to'ldirilgan tor ipak yoping'ichni” anglatadi.

Ushbu misolda oxirgi “adras” so'zi ham maishiy voqeliklarning namunasi. O'zbek tilidagi “adras” so'zi sharqona naqshli paxta matosini tasvirlaydi. Asarda muallif bu so'zni Otabek tunab qolgan “hujra” - xonaning jihozlari va nafosatini tariflash uchun ishlatgan. Muayyan milliy va madaniy tusga ega bo'lgan “adras” so'zining rus tilida to'g'ridan to'g'ri ekvivalenti yo'q. Shunga asoslanib, tarjimada transliteratsiya usuli ishlatilgan.

Shunday qilib, voqelikni o'rganish va tarjima qilishda o'zbek-rus tillarning parallel korpusidan foydalanish muhim afzalliklarni beradi, voqelik bilan bog'liq madaniy, diniy va tarixiy jihatlarni tahlil qilish va tushunishga yordam beradi. Ushbu yondashuv aniqroq va madaniy jihatdan yetarli darajada tarjima qilishga yordam beradi, turli tillarda so'zlashuvchilar o'rtasida madaniyatlararo muloqot va o'zaro tushunishni kuchaytiradi.

“Mashina tarjimasi va NLP: tarjima samaradorligini oshirish uchun parallel korpuslardan foydalanish” nomli 3.3 bo'limda parallel korpuslarning mashina tarjimasi tizimlari samaradorligini oshirish va tabiiy tillarni qayta ishlash sohasini (NLP) rivojlantirishdagi roli muhokama qilinadi. Ushbu yondashuv o'zbek va rus tillari o'rtasida mashina tarjimasi tizimlarini yaratishda katta ahamiyatga ega.

Parallel korpuslar – bu turli tillarda katta hajmdagi parallel matnlar bilan mashina tarjimasi modellarini ta'minlaydigan qimmatli manba. Yaratilgan o'zbek-rus parallel korpusi ushbu ikki til o'rtasida mashina tarjimasi tizimlarini o'rgatish va takomillashtirish uchun qo'llanilishi mumkin bo'lgan bunday manbaga misoldir.

O'zbek-rus va rus-o'zbek parallel korpusi NLP modellariga til xususiyatlarini o'rgatish va o'zbek va rus tillari o'rtasida o'xshashliklarni aniqlashda muhim ahamiyatga ega. Parallel korpus o'zbek tili va uning madaniy merosini saqlash va ommalashtirishda ham o'z hissasini qo'shadi. Ushbu parallel korpusning yaratilishi va ishlatilishi turli sohalarda, jumladan, ta'lim, biznes va xalqaro munosabatlar uchun yangi imkoniyatlar ochib beradi.

O'zbek tilini mashinaviy tarjima qilish uchun BERT va GPT kabi chuqur o'rganish modellaridan foydalanish mumkin. Buning uchun katta matnli korpuslar asosida modellarni oldindan tayyorlash, so'ngra ularni ma'lum bir tarjima vazifasi bo'yicha keyinchalik o'qitish talab etiladi. Seq2Seq, Transformer, T5 va Marian NMT kabi boshqa modellardan ham foydalanish mumkin. Modelni samarali o'qitish uchun katta va xilma-xil parallel korpusga ega bo'lish muhimdir. Google, Facebook, OpenAI va Microsoft kabi kompaniyalar o'zlarining tarjima tizimlari uchun turli xil modellar va yondashuvlardan foydalanadilar. O'zbek tili uchun mashinaviy tarjima

platformasini yaratish madaniyatlararo muloqot va madaniy merosni asrab-avaylashga hissa qo‘shishi mumkin bo‘lgan dolzarb vazifadir.

Mashinaviy tarjima va NLP rivojlanishi sharoitida parallel korpuslardan foydalanish modellarni o‘qitish va tarjima sifatini yaxshilash uchun muhim vositadir. Shu nuqtayi nazardan yaratilgan o‘zbek-rus va rus-o‘zbek parallel korpusi o‘zbek va rus tillari o‘rtasida mashinaviy tarjima tizimlarini rivojlantirish, shuningdek, o‘zbek tili va uning madaniy merosini saqlash va ommalashtirish uchun zamonaviy va istiqbolli manba hisoblanadi.

XULOSA

1. Kompyuter tilshunosligi – tilshunoslik va axborot texnologiyalari negizida hosil bo‘lgan fan, korpus lingvistikasi esa yirik matn to‘plamlaridagi tabiiy til ma’lumotlarini tahlil qilish va tizimlashtirish bilan shug‘ullanadigan fan hisoblanadi. Ikkala fan tillarni o‘rganish va takomillashtirishda muhim rol o‘ynaydi. Ushbu yo‘nalishlarning joriy etilishi o‘zbek tilining tarkibiy va funksional xususiyatlarini chuqur o‘rganishga yordam beradi, avtomatik tarjima qilish, matnlarni qayta ishlash va lingvistik tahlil qilish uchun samarali vositalar yaratilishini ta’minlaydi.

2. Parallel korpus – asl matnlar va ularning bir yoki bir nechta tillarga tarjimalaridan tashkil topgan lingvistik ma’lumotlar bazasi bo‘lib, turli til tizimlarida til tuzilmalari va ularning muvofiqqligini tahlil qilish uchun noyob imkoniyatni taqdim etadi. Bir tilga mo‘ljallangan yoki tillarni bir-biridan mustaqil ravishda tahlil qiladigan bir tilli va qiyosiy korpuslardan farqli o‘laroq, parallel korpuslar tarjima ekvivalentlarini chuqur tahlil qilishga imkon beradi, tillarning o‘zaro ta’sirining nozik tomonlarini ochib beradi va natijada tarjima va tilshunoslik fanini sezilarli darajada boyitadi. Parallel korpuslar tilning semantik, sintaktik va kontekstual jihatlarini o‘rganishda bebaho yordam beradi, bu jihat ularni nafaqat tarjimonlar va tilshunoslar uchun, balki tabiiy tilni qayta ishlash sohasida machine learning uchun ham ajralmas vositaga aylantiradi.

3. Hozirgi kunda “World Wide Web” (Veb) gipermatn tamoyillariga asoslangan va internet orqali kirish mumkin bo‘lgan global axborot tizimi sifatida ishlaydi va foydalanuvchilarga ma’lumotlar va resurslarning keng doirasini ko‘rish va ular bilan o‘zaro aloqada bo‘lish imkonini beradi. Veb matn, audiovizual va multimedia materiallarining keng va dinamik ravishda to‘ldiriladigan omborini tashkil etadi, bu xususiyat uni kompyuter lingvistikasi, mashina tarjimasi va madaniyatlararo aloqa sohasidagi tadqiqotlar uchun muhim bo‘lgan parallel korpuslarni bazasini yig‘ish va tahlil qilish uchun eng muhim manbaga aylantiradi. Internetning ushbu noyob xususiyati tadqiqotchilarga parallel matnlarni olish va qayta ishlash uchun misli ko‘rilmagan imkoniyatlarni taqdim etadi, til tuzilmalari va madaniy xususiyatlarni tushunishni osonlashtiradi va yuqori sifatli va kontekstga moslashtirilgan mashina tarjima tizimlarini yaratish uchun asos bo‘ladi.

4. O‘zbek-rus parallel korpusi bilan ishlash uchun onlayn platformani ishlab chiqishda interaktivlik, keng qamrovlik va qulaylik talablarini inobatga olgan holda React, Node.js, Express.js, MongoDB va Visual Studio Codedan iborat texnologik

stekni tanlash optimal hisoblanadi. Reactning tezligi va komponent yondashuvi orqali yuqori ishlash va dinamik o'zaro ta'sirni ta'minlashi, parallel matnlarni vizualizatsiya qilish uchun muhimdir.

4. Node.js va Express.js asinxron so'rovlar va ma'lumotlarni boshqarish uchun engil, kuchli server tomonidagi ramka bilan ta'minlaydi, bu API yaratish va matn ma'lumotlarini saqlash uchun mos bo'lgan hujjatga yo'naltirilgan ma'lumotlar bazasi MongoDB bilan integratsiyalashni osonlashtiradi. Node.js va Express.js asinxron so'rovlar yuborish va ma'lumotlarni boshqarish uchun yengil, kuchli server muhitini yaratadi, API yaratish va matnli ma'lumotlarni saqlash uchun qulay bo'lgan MongoDB hujjatiga asoslangan ma'lumotlar bazasi bilan integratsiyani osonlashtiradi. Visual Studio Code - ishlab chiqishni tezlashtirish orqali samarali kodlash imkonini beradi.

5. Matnni tenglashtirish - parallel korpus ma'lumotlar bazalarini yaratish, turli tillar bo'yicha matn segmentlarini to'g'ri taqqoslashni ta'minlashning asosiy bosqichi bo'lib, bu lingvistik tadqiqotlarning aniqligi va samaradorligi uchun muhim ahamiyatga ega. Matnni tenglashtirish parallel korpus ma'lumotlar bazalarini yaratishda muhim bosqich bo'lib, turli tillardagi matn segmentlarining aniq, mos kelishini ta'minlaydi. Bu lingvistik tadqiqotlarning aniqligi va samaradorligi uchun muhimdir. Tenglashning har xil turlarining mavjudligi (so'z-so'z, gap-gap, paragraf-paragraf) kontekstning turli darajalarida tahlilni osonlashtiradi va matnlar orasidagi tarkibiy va semantik aloqalarni tushunishni chuqurlashtiradi. Bundan tashqari, lingvistik munosabatlarni batafsil tahlillash va chuqur tushunish uchun bunda kichik, masalan, iboralar yoki tovushlar darajasida tenglashtirish mavjud.

6. Rus tilining lingvistik tipi flektiv, o'zbek tili esa agglyutinativ tillarga kiradi. Strukturaviy til xususiyatlaridagi farqlar parallel korpusning o'zbek-rus va rus-o'zbek onlayn platformasini yaratish jarayonida matnlarni tenglashtirish jarayonini murakkablashtiradi, bu esa universal tenglashtirish usullarini bevosita qo'llashni imkonsizligini ko'rsatadi. Ushbu muammoni hal qilish uchun alohida leksik elementlarni tarjima qilishda maksimal aniqlikni ta'minlash maqsadida maxsus so'z darajasida tekislash usuli ishlab chiqildi.

Frazeologik birliklar va idiomalarni tarjima qilishda ularning madaniy va kontekstual o'ziga xosligini inobatga olgan holda tarjimaning adekvatligi va tabiiyligini ta'minlash uchun ibora va gaplar darajasida hizalanish qo'llanilgan.

Frazeologik birliklar va idiomalarni tarjima qilishda, ularning madaniy va kontekstual xususiyatlarini hisobga olgan holda, tarjimaning etarliligi va tabiiyligini ta'minlash uchun iboralar va jumlar darajasida hizalama ishlatilgan.

7. Frazeologik birliklar va idiomalarni tarjima qilishda, ularning madaniy va kontekstual o'ziga xosligini inobatga olgan holda, tarjimaning adekvatligi va tabiiyligini ta'minlash uchun ibora va gaplar darajasida tenglashtirish qo'llanilgan. Bu iboralarning o'zbek tilidan rus tiliga va aksincha tarjimasini tahlil qilish usullari nafaqat lug'aviy tarkibni, balki madaniy va kontekstual to'yinganlikni ham aniq yetkazish muhimligini ta'kidlaydi. Bunga funktsional yukni saqlab qolish uchun tarkibiy o'zgartirish, shunga o'xshash effektini ko'paytirish uchun kompensatsiya va maqsadli auditoriya tomonidan yaxshiroq tushunish uchun madaniy kontekstni

o'zgartirish orqali erishiladi. Bunga funktsional yukni saqlab qolish uchun strukturaviy transfotmatsiyalash, o'xshash effektini takrorlash uchun kompensatsiya va maqsadli auditoriya tomonidan yaxshiroq tushunish uchun madaniy kontekstni o'zgartirish orqali erishiladi. Tushuntirish elementlaridan foydalanish va ahamiyatsiz tarkibiy qismlarni tushirib qoldirish ham tarjimani chuqurroq tushunish va moslashtirishga yordam beradi.

8. A.Qodiriyning "O'tgan kunlar" romani lingvistik realiyalari tarjimasini tahlil qilishda transliteratsiya usullari asl eshutilish va kontekstual adaptatsiyani saqlab qolish, madaniy xususiyatlarni hisobga olgan holda ma'noni tushunishni ta'minlashda asosiy o'rin tutadi. Neologizmlarni yaratish noyob hodisalar uchun yangi atamalarni kiritishga imkon beradi, perifrastik tarjima esa ma'noni batafsil bayon qiladi. Kalkalash va funktsional analoglarni izlash mazmunning to'g'ri va madaniy jihatdan adaptiv berilishiga yordam beradi, kontekstual tarjima esa adekvatlik va tabiiylikka erishish uchun lingvistik va madaniy kontekstni hisobga oladi. Ushbu usullar birgalikda asl matnning chuqurligini aniq aks ettiruvchi tarjima yaratish uchun asos yaratadi.

9. Mashina tarjimasini yaratishda tabiiy tilni qayta ishlash (NLP) usullarini qo'llash muhim jihat hisoblanadi, chunki bu usullar til tuzilmalari va semantikasini chuqur tushunish va tahlil qilishni ta'minlaydi. Bunday usullarga misol qilib til birliklari o'rtasidagi ehtimoliy muvofiqlikni aniqlash uchun parallel matnlarning katta hajmlarini tahlil qilishga asoslangan statistik mashina tarjimasini (SMT) va tarjima ketma-ketliklarini modellashtirish uchun chuqur neyron tarmoqlardan foydalanadigan neyron mashina tarjimasini (NMT) kiradi. Bundan tashqari, BERT (Bidirectional encoder representations from transformers) va GPT (Generative pre-trained transformer) kabi transformerlarga asoslangan usullar tilning nuanslari tushunishda va matn generatsiyasining sezilarli yutuqlarga erishishiga yordam beradi, kontekstga dinamik moslashishni ta'minlaydi va matndagi sintaktik va semantik munosabatlarni chuqurroq tahlil qilish orqali tarjima sifatini yaxshilaydi.

10. Milliy til korpusi-bu tizimli ravishda tashkil etilgan va raqamlashtirilgan matnlarning keng to'plami bo'lib, u millatning madaniy, tarixiy va ijtimoiy jihatlarini aks ettiruvchi lingvistik ifodalarning keng doirasini qamrab oladi va shu bilan tilshunoslik, leksikografiya va tabiiy tilni avtomatik qayta ishlash bo'yicha tadqiqotlar uchun keng qamrovli ma'lumotlar bazasini taqdim etadi. Shu nuqtai nazardan, ikki yoki undan ortiq tillarga tarjima qilingan va turli tillardagi matnning tegishli segmentlarini taqqoslash mumkin bo'lgan tarzda tenglangan matnlardan iborat parallel korpus milliy til korpusida asosiy rol o'ynaydi, chunki u nafaqat tillararo munosabatlar va madaniy nuanslarni chuqur tushunishga yordam beradi, balki mashina tarjimasini texnologiyalari va madaniyatlararo muloqotni rivojlantirish uchun ham bebaho manba hisoblanadi.

**НАУЧНЫЙ СОВЕТ DSc.03/30.12.2019.Fil.05.02
ПО ПРИСУЖДЕНИЮ УЧЕНЫХ СТЕПЕНЕЙ ПРИ
ФЕРГАНСКОМ ГОСУДАРСТВЕННОМ УНИВЕРСИТЕТЕ**

БУХАРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

АВЕЗОВ СУХРОБ СОБИРОВИЧ

**ОСНОВНЫЕ ПРИНЦИПЫ СОЗДАНИЯ УЗБЕКСКО-РУССКОЙ И
РУССКО-УЗБЕКСКОЙ ONLINE-ПЛАТФОРМЫ ПАРАЛЛЕЛЬНЫХ
КОРПУСОВ (НА МАТЕРИАЛЕ РОМАНА А. КАДЫРИ «МИНУВШИЕ
ДНИ»)**

10.00.11 – Теория языка. Прикладная и компьютерная лингвистика

**АВТОРЕФЕРАТ
диссертации доктора философии по ФИЛОЛОГИЧЕСКИМ наукам (PhD)**

Фергана – 2024

Тема диссертации доктора философии (PhD) зарегистрирована в Высшей аттестационной комиссии за № B2022.3.PhD/Fil2798.

Диссертация выполнена в Бухарском государственном университете.

Автореферат диссертации на трех языках (узбекский, русский, английский (резюме)) размещен на веб-странице Ферганского государственного университета по адресу: (www.fdu.uz) и информационно-образовательном портале «Ziynet» (www.ziynet.uz).

Научный руководитель:

Нигматова Лолахон Хамидовна
доктор филологических наук, профессор

Официальные оппоненты:

Рахимов Усмонжон Эргашевич
кандидат филологических наук, доцент

Мухиддинов Анвар Гафурович
доктор филологических наук, профессор

Ведущая организация:

Самаркандский государственный университет имени Шарофа Рашидова

Защита диссертации состоится « 27 07 2024 года в 10:00 часов на заседании Научного совета DSc.03/30.12.2019.Fil.05.02 при Ферганском государственном университете по адресу 100151, г. Фергана, улица Б.Маргиноний, 105 тел.:(99873) 244-57-82; e-mail : info@fdu.uz

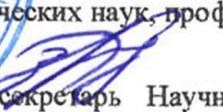
С диссертацией можно ознакомиться в Информационно-ресурсном центре Ферганского государственного университета (зарегистрирована под № 356 . Адрес : 100151, г. Фергана, улица Мураббийлар,19. Тел.:(+99873)244-71-28

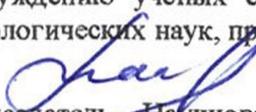
Автореферат диссертации разослан « 6 » 07 2024 года

(Протокол рассылки № 18 от « 6 » 07 2024 года)




С.М.Муминов
Председатель Научного совета по присуждению ученых степеней, доктор филологических наук, профессор


М.Т.Зокиров
Ученый секретарь Научного совета по присуждению ученых степеней кандидат филологических наук, профессор


А.Мамажонов
Председатель Научного семинара при Научном совете по присуждению ученых степеней, доктор филол. наук, профессор

ВВЕДЕНИЕ (аннотация к диссертации доктора философии (PhD))

Актуальность и востребованность темы диссертации. Мировое языкознание второй половины 20 столетия связано с интенсивным развитием компьютерной и корпусной лингвистики, явившейся основой к созданию и разработке программ для машинного перевода, обработки естественного языка, математической лингвистики. Появляются сотни языковых корпусов, связанных с применением информационных технологий в языкознании, позволяющих расширить возможности в области языковой инженерии, автоматического перевода, компьютерного анализа, редактирования, тезауруса, электронного словаря. На смену несовершенным корпусным технологиям приходят более модернизированные и ориентированные на потребности пользователя программы, возможности и инструменты которых позволяют проводить лингвистические исследования на более высоком научном уровне.

Одной из основных задач мировой лингвистики выступает создание и разработка национальных корпусов. Малоизученность принципов развития корпусной лингвистики, ее структурных механизмов, проблемы создания параллельного корпуса послужили основной причиной дальнейшего усовершенствования национального корпуса языка. Необходимость создания и комплексного анализа русско-узбекского и узбекско-русского параллельного корпуса способствует совершенствованию переводоведения, установлению достижения функциональной пропорциональности и соответствия между текстами оригинала и перевода.

На современном этапе интенсивного развития информационных технологий в государстве необходимость «... сохранения безупречности государственного языка, обогащение его и повышение речевой культуры населения; обеспечение активной интеграции государственного языка в современные информационные технологии и коммуникации»³⁵ поставила новые задачи перед языковедами в области узбекской компьютерной лингвистики. Проводимые в республике широкомасштабные реформы в сфере языковой политики обусловили не только начало новой эры в развитии корпусной лингвистики Узбекистана, но и переход на новый этап научно-исследовательских работ в языкознании, усилилось внимание к проведению прикладных исследований, изучающих язык с учетом человеческого фактора. Таким образом, наличие такого корпуса будет иметь существенное значение для научных исследований, лингвистического анализа, разработки языковых моделей и других смежных областей.

Настоящее диссертационное исследование в определенной степени служит реализации задач, обозначенных в УП-5850 от 21 октября 2019 года «О мерах по кардинальному повышению роли и авторитета узбекского языка в качестве государственного языка», УП-6084 от 20 октября 2020 года «О

³⁵ Ўзбекистон Республикаси Президенти Шавкат Мирзиёевнинг 2020 йил 20 октябрдаги «Мамлакатимизда ўзбек тилини янада ривожлантириш ва тил сиёсатини такомиллаштириш чора-тадбирлари тўғрисида»ги ПФ-6084-сон фармони // <https://lex.uz/docs/5058351>

мерах по дальнейшему развитию узбекского языка и совершенствованию языковой политики в стране», УП-6108 от 6 ноября 2020 года «О мерах по развитию сфер образования и воспитания в новый период развития Узбекистан», УП-60 от 28 января 2022 года «О стратегии развития нового Узбекистана на 2022 — 2026 годы»; УП-2789 от 17 февраля 2017 года «О мерах по дальнейшему совершенствованию деятельности Академии наук, организации, управления и финансирования научно-исследовательской деятельности», ПП- 4479 от 4 октября 2019 года «О широком праздновании тридцатилетия принятия закона Республики Узбекистан «о государственном языке», ПП-5117 от 19 мая 2021 года «О мерах по поднятию на качественно новый уровень деятельности по популяризации изучения иностранных языков в Республике Узбекистан», а также других нормативно-правовых документах, принятых в данной сфере.

Соответствие исследования приоритетным направлениям развития науки и технологий Республики. Диссертационное исследование выполнено в соответствии с приоритетным направлением развития науки и технологий Республики Узбекистан I. «Формирование системы инновационных идей и пути их реализации по социальному, правовому, экономическому, культурному, духовно-просветительскому развитию информационного общества и демократического государства».

Степень изученности проблемы. Необходимость в упорядоченности и систематизации всей накопленной информации появилась задолго до создания компьютеров и компьютерного корпуса. Учеными создавались карточки, картотеки, собирались печатные конкордансы, глоссарии, энциклопедии. Сбор и обработка информации занимала много времени. До середины XX столетия компьютерные технологии были привилегией точных наук. Компьютер и информационные технологии долго не являлись научным инструментом исследования гуманитариев, но на стыке наук образуется компьютерная лингвистика, а затем и корпусная. Корпусная лингвистика окончательно сложилась к 90 годам XX столетия. Конечно же нельзя ее назвать и совсем новой и юной, так как конкордансы были созданы еще в начале XIII века для Священного Писания (*«Concordantiae morales sacrae scripturae»*, «Нравственная конкорданция Священного Писания»). Несмотря на это корпусная лингвистика, использующая компьютерные технологии для лингвистического исследования, считается новым веянием в мировом языкознании, позволяющей в полной мере задействовать преимущества нового способа обработки и хранения информации.

Изучение теоретико-методологических проблем лингвистического обеспечения в мировой лингвистике началось в 60-х годах прошлого века: было признано, что компьютер без информационно-поискового языка остаётся просто вычислительной машиной, а компьютер оснащенный поисковой системой является машиной, перешедшей из «механического» этапа в «логико-лингвистический» этап, то есть машина переходит в этап

информационно-поисковых систем³⁶. Обработке естественного языка в мировой лингвистике посвящены труды А. Кейлера, К.В. Линдена, Н. Варда³⁷, М.А. Мохри³⁸. Первой работой, посвященной параллельному корпусу, его разработке и программированию можно считать Брауновский лингвистический корпус, спроектированный У. Френсис и Г. Кучера в 1960 году³⁹. Проблеме использования корпусных технологий в преподавании и изучении иностранного языка посвящена работа Д. Синклера⁴⁰.

Параллельному корпусу русского языка, его структуре, программному обеспечению и выполняемой основной функции посвящены работы Д.О. Добровольского⁴¹, Сичинавы⁴², Ю. Тао⁴³, В. Захарова⁴⁴, С.А. Маник⁴⁵, А.А. Кокорева⁴⁶.

Также проблеме разработки параллельного корпуса можно отметить работы А.Г. Гилемшиной⁴⁷, А.Х. Худойбердиева⁴⁸. В отечественном языкознании уже ведется активная работа по изучению лингвистического и компьютерного корпуса. По параллельному корпусу актуальные вопросы были рассмотрены в работах Н.З. Абдурахмановой⁴⁹, Ш.М. Хамроевой⁵⁰, М.А.

³⁶ Соколов А.В. Автоматизация библиографического поиска. – М.: Книга, – 1981. – С.91.

³⁷ Daniel S. Jurafsky and James H. Martin. Speech and Language Processing. Contributing writers: Nigel Ward, – 2000. pages: 950

³⁸ Mohri M.A. Finite-state transducers in language and speech processing. Computational Linguistics. – P. 269-312.

³⁹ Френсис Н., Кучера Г. Вычислительный анализ современного американского варианта английского языка. – М., – 1967.

⁴⁰ Синклер Д. Предисловие к книге “Как использовать корпуса в преподавании иностранного языка”/ Д. Синклер [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru/corpora-infro.html>, свободный.

⁴¹ Добровольский Д.О. Какие задачи решают параллельные корпуса? (Электрон ресурс): <https://postnauka.ru/video/>; Добровольский Д.О. Корпус параллельных текстов в исследовании лексической семантики. (Электронресурс): <http://www.dialog-21.ru/media/2503/dobrovolskij.pdf>

⁴² Сичинава Д.В. Параллельные тексты в составе национального корпуса русского языка: новые направления развития и результаты. https://ruslang.ru/doc/sitchinava/sitchinava-2015-parallel_corpus.pdf; Д.В.Сичинава. Параллельные корпуса в составе национального корпуса русского языка: новые языки и новые задачи. Труды Института Русского Языка Им. В.В. Виноградова. –2019. –С. 41-61 <https://elibrary.ru/item.asp?id=42567226>

⁴³ Тао Ю., Захаров В.П. Разработка и использование параллельного корпуса русского и китайского языков // Информ. процессы и системы. –2015. №4 <https://www.researchgate.net/publication/306911853>

⁴⁴ Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург. – 2005

⁴⁵ Маник С.А. Параллельный корпус в практике перевода общественно-политических текстов (с английского на русский). // Modern Studies of Social Issues – 2019, Volume 11, Number 4. <http://soc-journal.ru>

⁴⁶ Кокорева А.А. Корпус параллельных текстов в обучении иностранному языку. – Вестник ТГУ (Тамбовский государственный университет), выпуск 2 (118), – 2013. – С. 57-62.

⁴⁷ Гилемшина А.Г. Разработка параллельного корпуса арабского, русского, татарского языков: проблемы и перспективы. Казанская наука. №1, – 2020. – С. 24-29.

⁴⁸ Худойбердиев Х.А., Солиев О.М., Солиев П.А. Разработка параллельного корпуса таджикского и русского языков // Новые информационные технологии в автоматизированных системах. – 2019. №22. URL: <https://cyberleninka.ru/article/n/razrabotka-parallelnogo-korpusa-tadzhikskogo-i-russkogo-yazykov> (дата обращения: 29.05.2023). 2. URL: <https://cyberleninka.ru/article/n/razrabotka-parallelnogo-korpusa-ta29.05.2023>).

⁴⁹ Абдурахманова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (Содда гаплар мисолида). Филол. фан. бўйича фалсафа доктори (PhD)...дис. афтореф. – Тошкент, – 2018. – 52 б.

⁵⁰ Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол.фан. бўйича фалсафа доктори (PhD)...дис. – Бухоро, – 2018.

Абжаловой⁵¹, А.А.Эшмуминова⁵², Г.Г.Абдувахобова⁵³, Н.А. Атабоева⁵⁴, Р.А. Каримова⁵⁵, А.Н.Ахмедовой⁵⁶, О.Хидирова⁵⁷, У.Холиярова⁵⁸, Г.Бегматовой⁵⁹, Г.Тоировой⁶⁰.

Ключевым элементом в создании параллельных корпусов является текст и его компоненты. Выбор и анализ текста играют существенную роль в этом процессе. В связи с этим, мы обратили внимание на исследования выдающихся узбекских ученых, таких как А. Мамажонов⁶¹, Н. М. Турниёзов⁶², Э. Қиличев⁶³, М. Ҳакимов⁶⁴, М. Йулдошев, С.Мухамедова, М.Сапарниязова⁶⁵ и других. Их работы в области «Лингвистика текста» значительно продвинули данное направление, что имеет большое значение для сбора и обогащения данных в корпусах.

Связь диссертационного исследования с планами научно-исследовательских работ высшего образовательного учреждения, где выполнена диссертация. Диссертационное исследование выполнено в рамках плана научно-исследовательских работ Бухарского государственного университета на 2022-2024 годы по теме «Исследовательские проблемы отношений языка, личности и общества в узбекском языкознании».

Цель исследования состоит в изучении теоретических основ создания узбекско-русского параллельного корпуса, разработке онлайн-платформы для узбекско-русских и русско-узбекских параллельных корпусов на примере романа А.Кадыри «Минувшие дни», а также в совершенствовании машинного перевода, создании удобного инструмента для лингвистического анализа и межкультурного обмена на уровне предложений, фраз и текстовых единиц между узбекским и русским языками.

⁵¹ Абжалова М.А. Ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи дастурнинг лингвистик модуллари (Расмий ва илмий услубдаги матнлар таҳрири дастури учун): Филол.фан. бўйича фалсафа доктори (PhD)...диссер. – Фарғона, – 2019. – 164 б.

⁵² Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол.фан. бўйича фалсафа доктори (PhD)...диссер. – Қарши, – 2019. – 140 б.

⁵³ Абдувахобов Г.Г. Янги авлод омонимлар ўқув луғати мобил иловасини яратишнинг лингвистик асослари. Автореферат фил. фанлари бўйича фалсафа доктори (PhD). – Фарғона. – 2023.

⁵⁴ Атабоев Н.Б. Инглиз тили корпусининг функционал хусусиятлари (соса мисолида). Автореферат фил. фанлари бўйича фалсафа доктори (PhD). – Тошкент. – 2020.

⁵⁵ Каримов Р.А. Ўзбек ва инглиз тили параллел корпусини тузишнинг лингвистик ва дастурий масалалари. Автореферат фил. фанлари бўйича фалсафа доктори (PhD). – Тошкент. – 2021

⁵⁶ Ахмедова А.Н. Параллел корпусда ўхшатишларнинг лексик-семантик муносабатлари (инглизча-ўзбекча, ўзбекча-инглизча параллел матнлар мисолида). Автореферат фил. фанлари бўйича фалсафа доктори (PhD). – Тошкент. – 2021.

⁵⁷ Хидиров О. Миллий корпус учун парсинг дастури яратишнинг лингвистик асослари: Филол.фан. бўйича фалсафа доктори (PhD)...дис. автореф. – Жиззах, – 2021. – 55 б.;

⁵⁸ Холиёров Ў. Ўзбек тили таълимий корпусини тузишнинг лингвистик асослари: Филол.фан. бўйича фалсафа докт. (PhD) ...дисс. – Термез, – 2021. – 181 б.;

⁵⁹ Бегматова Г. Ўзбек миллий корпусида идиомалар базасини яратиш: Филол.фан. бўйича фалсафа докт. (PhD) ...дисс. – Термез, – 2021. – 143 б.;

⁶⁰ Тоирова Г. Ўзбек тили миллий корпусини яратишнинг назарий ва амалий масалалари: Филол.фан. докт. (DSc) ...дисс. – Бухоро, – 2021. – 208 б.

⁶¹ Мамажонов А. Текст лингвистикаси. – Тошкент, – 1989.

⁶² Турниёзов Н., Турниёзова К., Хайруллаев Х. Структур синтаксис асослари //Тошкент: Фан, – 2019.

⁶³ Қиличев Э. Ўзбек тилининг практик стилистикаси. – Тошкент: Ўқитувчи, – 1985.

⁶⁴ Ҳакимов М. Ўзбек прагмалингвистикаси асослари. Тошкент: Akademyashr, – 2013.

⁶⁵ Yo'ldoshev M., Muhamedova S., Saparniyazova M. Matn lingvistikasi. – Toshkent. – 2021.

Задачи исследования заключаются в следующем:

описать общие черты параллельного корпуса, указать отличия параллельного корпуса от других корпусов;

сбор и подготовка параллельных текстов узбекского и русского языков на примере романа А. Кадыри «Минувшие дни»;

разработка и реализация онлайн-платформы, обеспечивающей доступ к параллельным корпусам узбекского и русского языков;

разработка функционала для работы с параллельными текстами, включая поиск, выравнивание, сравнение и анализ переводов;

оценка качества и эффективности разработанной платформы на основе экспериментов и пользовательского опыта;

проведение лингвистических анализов и исследований, включая анализ фразеологизмов и языковых реалий, и их переводов, используя созданный узбекско-русский и русско-узбекский параллельный корпус.

Объектом исследования является узбекско-русский и русско-узбекский параллельный корпус.

Предметом исследования - особенности перевода фразеологических оборотов и языковых реалий в узбекско-русском и русско-узбекском параллельных корпусах, а также разработка и анализ методов и приемов для эффективного перевода данных языковых явлений на примере романа А.Кадыри «Минувшие дни».

Методы исследования. Для достижения выдвинутых целей и реализации поставленных задач в диссертации применялись следующие научные методы и приёмы: метод классификации, описательный метод, сравнительно-сопоставительный метод, дистрибутивный анализ, метод перевода, сопоставительный анализ корпуса параллельных текстов, статистика, метод корпусного анализа текста, контрастивный анализ, экспертные оценки и анализ данных.

Научная новизна исследования заключается в следующем:

анализ языковых реалий романа А.Кадыри «Минувшие дни» посредством параллельного корпуса текстов обосновал важность учета лексических, семантических, стилистических и культурных особенностей при разработке методологии двуязычных корпусов, а также выявил основные аспекты и сложности перевода;

разработана уникальная онлайн-платформа, предназначенная для работы с узбекско-русскими параллельными текстами, предоставляющая возможности и инструменты для анализа и редактирования данных в режиме реального времени и заполняющая существующий пробел в работе с параллельными корпусами этих языков;

обоснованы основные аспекты параллельного корпуса для лингвистических исследований, их специфические особенности, в частности, способность сравнивать тексты на разных языках, что помогает изучать стратегии перевода и понимать межкультурные особенности;

проведен подробный анализ фразеологических единиц в узбекско-русских параллельных корпусах на примере романа А.Кадыри «Минувшие дни», определены методы их перевода, а также культурная адаптация, семантические и контекстуальные особенности.

Практическая значимость результатов исследования состоит в следующем:

доказано, что материалы и выводы относительно созданной базы русско-узбекского и узбекско-русского параллельного корпуса являются важнейшей научно-теоретической базой для создания национального корпуса узбекского языка, обогащая его научный опыт;

выявлены общие черты параллельного корпуса, указаны отличительные особенности параллельного корпуса от других корпусов, а также обозначены основные особенности параллельного корпуса в изучении иностранного языка;

обосновано, что разработанный и созданный русско-узбекский и узбекско-русский параллельный корпус, который до этого не существовал представляет собой ценный ресурс для исследования перевода.

проведен детальный анализ фразеологических оборотов и языковых реалий в параллельных корпусах, включая выявление особенностей их перевода, семантических вариаций и контекстуальных особенностей. Этот анализ способствует более глубокому пониманию и исследованию межъязыковых различий и сходств.

Достоверность полученных результатов обеспечивается научной базой, методологией и методикой исследования, а также постановкой новых проблем, решением новых задач и соответствием результатов проведенного исследования теоретическим положениям диссертационной работы; научными статьями, опубликованные в журналах Узбекистана и за рубежом, рекомендованных ВАК РУз для публикации основных результатов диссертаций; тезисами в сборниках статей на международных и республиканских научно-практических конференциях.

Научная и практическая значимость результатов исследования.

Результаты исследования имеют теоретическое значение при изучении проблем в области корпусной лингвистики, возможностью их использования в качестве источника в изучении компьютерной лексики и корпусной лингвистики как русского, так и узбекского языков.

Практическая значимость результатов исследования заключается в том, что материалы и результаты диссертационного исследования могут быть использованы при подготовке лекционных курсов и практических занятий, спецкурсов по прикладной филологии, в процессе написания учебников, учебных пособий, учебно-методических пособий, методических рекомендаций, сборников упражнений.

Внедрение результатов исследования. На основе теоретических и практических аспектов создания параллельных корпусов и применения

технологий в области обработки естественного языка были проведены следующие мероприятия:

подробный анализ фразеологических единиц в романе А.Кадыри «Минувшие дни» в узбекско-русских параллельных корпусах, а также исследование методов их перевода, культурной адаптации, семантических и контекстуальных особенностей были использованы в фундаментальном проекте I-OT-2019-42 «Создание электронного поэтического словаря узбекского и английского языков (изображения человеческого облика, характера, природы и национальных символов)», реализованном в Ташкентском государственном университете узбекского языка и литературы (справка № 01/10-1150 от 9 июня 2023 года). В результате понятие мультязычия, алгоритмы перевода и подходы к созданию текстового корпуса облегчили процесс перевода, повысили точность перевода и создали обширную базу данных для поиска и анализа в словаре;

анализ языковых реалий в романе А.Кадыри «Минувшие дни» с помощью параллельного корпуса текстов, исследование основных аспектов и сложностей перевода, учет лексических, семантических, стилистических и культурных особенностей при разработке методологии двуязычных корпусов были включены в 13-й параграф учебника «Общее языкознание» (справка Министерства высшего образования, науки и инноваций Бухарского государственного университета № 01-04/3562 от 12 октября 2023 года, ГРИФ № 112-13). В результате учебный процесс был упрощен, а слушателям предоставлены более широкие инструменты для анализа и практического применения языковых структур;

теоретическая и методологическая база, полученная из исследований основных аспектов параллельного корпуса для лингвистических исследований, их уникальных характеристик, в частности, изучения стратегий перевода и способности сравнивать тексты на разных языках, помогающих понять межкультурные особенности, была использована для написания сценариев к передачам «Замондош» и «Миллий таълим» на телеканале «Маданият ва маърифат» (справка № 02-02/125 от 1 июня 2023 года). В результате повысилась научная значимость и популярность этих телепередач.

Апробация результатов исследования. Результаты данного исследования изложены в 7 докладах и выступлениях и прошли апробацию на 2 республиканских и 5 международных, научно-практических конференциях и семинарах.

Публикация результатов исследования. Результаты исследования отражены в 15 научных публикациях: 8 научных статей в изданиях, рекомендованных Высшей аттестационной комиссией Республики Узбекистан для опубликования результатов докторской диссертации (PhD), в том числе, 3 статей – в республиканских и 5 публикаций – в зарубежном издании.

Структура и объем диссертации. Диссертация состоит из введения, 3 глав, заключения, списка использованной литературы и приложения, общий объем составляет 162 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснованы актуальность и востребованность темы, указаны соответствия исследования приоритетным направлениям развития науки и технологий республики, охарактеризованы степень изученности проблемы и обусловленность диссертации планами исследования, определены объект, предмет, цель и задачи исследования, методы их решения, а также научная новизна и практические результаты работы, раскрыта научная и практическая значимость полученных результатов, представлены сведения о внедрении результатов, апробации работы, публикации результатов, структуре и объеме диссертации.

Первая глава диссертации **«Компьютерная и корпусная лингвистика: проблемы и перспективы»** состоит из трех параграфов.

Первый параграф **«Развитие и современное состояние компьютерной лингвистики»** представляет собой обзор ключевых аспектов развития и актуального состояния этой научной области. В нем рассматриваются различные аспекты, включая историю компьютерной лингвистики, цели и задачи, подходы и методы, а также ее важность и применение в современном мире.

Развитие компьютерной лингвистики началось с появления идей Ноама Хомского в 1957 году, когда он предложил представить грамматику языка как механизм, способный создавать бесконечное количество правильных предложений. Это привело к появлению компьютерной лингвистики на стыке математики и лингвистики.

С течением времени ученые осознали потенциал компьютеров в автоматизации обработки текстов. В 1959 году итальянский ученый Р. Буза создал словари для древних рукописей с использованием компьютеров. В 1969 году в Безансоне (Франция) был основан Центр лингвистических исследований, где создавались каталоги для больших словарей.

В России компьютерная лингвистика тесно связана с именем выдающегося математика и кибернетика А. Ляпунова, который считается основателем русской кибернетики. Но начало истории компьютерной лингвистики относится к началу XX века и связано с растущим интересом к семиотике – науке о знаках и символах.

Определение компьютерной лингвистики может варьироваться, но в узком смысле она связана с междисциплинарным прикладным направлением, использующим компьютерные инструменты для моделирования функционирования языка.

Особое внимание уделяется автоматической обработке языка, которая включает анализ и моделирование языковой структуры на различных уровнях, генерацию текстов и лингвостатистику. Рассматриваются различные

направления компьютерной лингвистики, включая обработку естественного языка, создание электронных корпусов и словарей, автоматический перевод текстов, извлечение информации из текста, автореферирование, системы управления знаниями и другие.

Во втором параграфе **«Корпус и его составляющие»** представлена информация о развитии корпусной лингвистики в XX-XXI веках и ее основных составляющих.

В начале 1990-х годов была разработана инициатива по кодированию текста TEI (Text Encoding Initiative), которая стала ключевым стандартом для разметки текстов и обеспечила совместимость различных корпусов. Эта инициатива объединила разработчиков, занимающихся созданием цифровых текстов с использованием специализированного программного обеспечения.

Важным шагом в развитии корпусной лингвистики стала разработка стандартов и новых методов обработки естественного языка (NLP). Это открыло новые возможности для работы с текстовыми данными и обеспечило прогресс в анализе и автоматической обработке текста.

В странах Европы начался «корпусный бум», сопровождаемый созданием различных корпусов английского языка. Также в России был начат процесс оцифровки текстов и создания национального корпуса русского языка. В рамках этого развития были предложены различные классификации корпусов учеными, такими как Ю.В. Богоявленская, М.И. Солнышкина, Г.М. Гатиятуллина, Н.В. Козлова, В.П. Захаров, Д. Сичинава и др.

Современные корпусные технологии ставят удобство пользователей на первое место. Платформы, такие как «Sketch Engine» предлагают различные настраиваемые опции, интерфейсы, подсказки и техническую поддержку, чтобы сделать работу с корпусами более комфортной и удобной для исследователей.

В третьем параграфе **«Общая характеристика параллельного корпуса, структура и состав. Общее описание параллельного корпуса»** освещается общая характеристика параллельных корпусов и представляется их структура и состав.

Параллельные корпуса являются собраниями текстов на двух или более языках, которые имеют соответствующие переводы. Они широко применяются в лингвистике, машинном переводе, компьютерной лингвистике и других областях.

Состав параллельного корпуса зависит от его цели и области применения. Он может включать тексты различных жанров, таких как художественная литература, научные статьи, официальные документы и другие. Кроме того, корпус может содержать тексты разных эпох и стилей для изучения языковых изменений и развития.

Национальный корпус русского языка содержит 8 двуязычных параллельных корпусов, включающих английский, немецкий, французский, испанский, итальянский, польский, украинский и белорусский языки. Общий объем корпуса составляет более 37 миллионов слов, и для каждого из языков

в паре с русским предусмотрена морфологическая разметка. Также существует многоязычный корпус на 25 языках, в основном славянских, содержащий 9 параллельных текстов и насчитывающий около 5 миллионов слов.

В параграфе отмечается, что параллельные корпуса открывают широкие возможности для исследования методов перевода, улучшения переводческих словарей и обучения переводу на новый уровень. Они также позволяют сравнивать исходные тексты и их переводы для изучения различных стратегий, используемых переводчиками для преодоления грамматических и стилистических различий между оригинальным текстом и переводом.

Вторая глава диссертации **«Разработка и реализация узбекско-русского и русско-узбекского параллельного корпуса»** состоит из трех параграфов, которые посвящены разработке и реализации узбекско-русского и русско-узбекского параллельного корпуса.

Данные параграфы включают в себя основные задачи создания параллельного корпуса, технологию разработки веб-приложения для доступа к корпусу, а также выравнивание текстов как основной этап в создании базы данных параллельного корпуса.

В параграфе 2.1. **«Основные задачи создания узбекско-русского русско-узбекского параллельного корпуса»** рассматриваются важные аспекты и цели, связанные с созданием данного корпуса. Одной из ключевых задач является сохранение и популяризация узбекского и русского языкового и культурного наследия. С помощью параллельного корпуса можно компилировать и систематизировать тексты на обоих языках, что позволяет сохранить и изучать историю, литературу и культурные традиции этих народов.

Кроме того, одной из задач параллельного корпуса является улучшение преподавания и изучения узбекского и русского языков. Аутентичные и разнообразные тексты, доступные в корпусе, помогают учащимся развивать навыки чтения, письма, аудирования и разговорной речи. Это позволяет студентам глубже погрузиться в языковую среду и изучать культурные особенности обоих языков.

Одной из сложностей при создании узбекско-русского и русско-узбекского параллельного корпуса является отсутствие Национального корпуса узбекского языка. Однако уже была проведена работа по созданию морфоанализатора узбекского языка, и в настоящее время перед исследователями стоит задача разработки остальных корпусов.

В параграфе 2.2. **«Технология разработки веб-приложения узбекско-русского и русско-узбекского параллельного корпуса»** описывается технология и методы разработки веб-приложения для создания параллельного корпуса на узбекском и русском языках. Проект разработан с использованием современных технологий и инструментов, а также включает в себя несколько этапов и функциональных возможностей.

Проект основан на использовании следующих технологий и языков программирования: React, Node.js, JavaScript, MongoDB и других. React был

выбран для создания пользовательского интерфейса, Node.js - для серверной части и взаимодействия с базой данных MongoDB. JavaScript использовался для написания логики приложения.

Разработка началась после проведения анализа различных текстовых редакторов, и в результате выбор был сделан в пользу VS Code. Этот редактор обладает удобным интерфейсом, множеством функций и плагинов, а также является кроссплатформенным.

Для хранения данных была выбрана MongoDB, одна из популярных систем управления базами данных (СУБД). MongoDB обладает простотой использования, кроссплатформенностью и масштабируемостью. Также был выбран встроенный сервер Node.js для работы с СУБД без необходимости установки стороннего программного обеспечения.

Проектирование базы данных включало определение и связывание 8 сущностей: user, corpus, author, translator, fullSentence, poem, byWords и translated_poem.

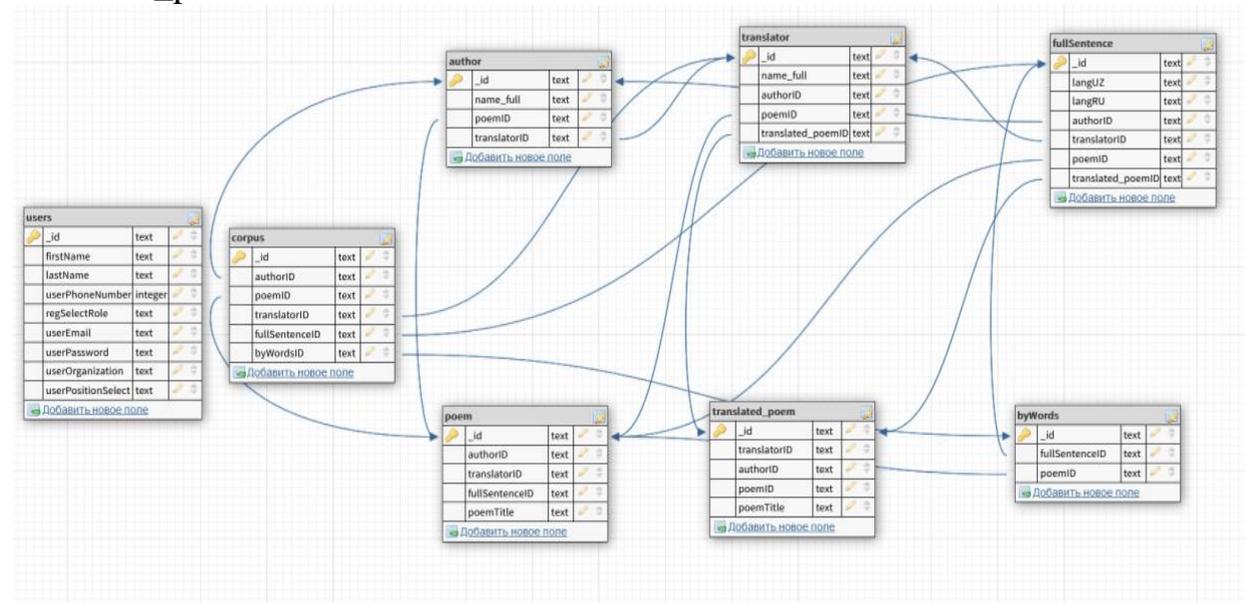


Рис. 1. Структура БД

На рисунке (Рис. 1) представлена структура базы данных, созданная с помощью web-сервиса DbDesigner. Кроме того, некоторые сущности, не включенные на рисунке (Рис. 1), будут использоваться для хранения изображений и пользовательских данных.

Для авторизации пользователей были разработаны соответствующие структуры кода, включающие проверку регистрации и проверку данных пользователя на серверной части. После успешной авторизации данные пользователя и токен сохраняются в локальном хранилище браузера.

Одним из основных файлов является usersReg.js. Файл usersReg.js служит для проверки регистрации пользователей. В зависимости от результата этой проверки, пользователь будет перенаправлен на соответствующую страницу. Существуют две роли: администратор и обычный пользователь. Только администраторы с root-правами имеют возможность регистрировать новых пользователей. Если пользователь вошел в систему как администратор,

то у него открывается доступ к дополнительным страницам интерфейса, где можно загружать материалы для параллельного корпуса, управлять ими, регистрировать новых пользователей и авторов приложения. Обычные пользователи не имеют таких привилегий, они могут только работать с материалами в параллельном корпусе и получать отчеты о результатах своей работы.

Для авторизации пользователя отвечает следующая структура (Рис. 2)

```
return (
  <section className='login_section'>
    <div className="login_box">
      <div className="login_form">
        <h2>Login</h2>
        <div>
          <div className="login_inputBx">
            <input
              type="email"
              placeholder='Foydalanuvchi emaili'
              required
              value={logEmail}
              onChange={(e) => setLogEmail(e.target.value)}
            />
            <span className='login_user_icon'>
              <FaUserAlt />
            </span>
          </div>
          <div className="login_inputBx">
            <input
              type="password"
              placeholder='Password'
              required
              value={logPassword}
              onChange={(e) => setLogPassword(e.target.value)}
            />
            <span className='login_lock_icon'>
              <FaLock />
            </span>
          </div>
          <label className='login_remember'>
```

Рис. 2. Структура кода для авторизации

```
UserModel.findOne({ userEmail })
  .then(savedUser => {
    // console.log(savedUser)
    if (!savedUser) {
      res.status(422).json({ error: "Bunday foydalanuvchi topilmadi" })
    }

    bcrypt.compare(userPassword, savedUser.userPassword)
      .then(doMatch => {
        if (doMatch) {
          const token = jwt.sign({ _id: savedUser._id }, SECRET_KEY)
          const
            {
              _id,
              firstName,
              lastName,
              userEmail
            } = savedUser
          res.status(200).json({
            token,
            user: {
              _id,
              firstName,
              lastName,
              userEmail
            }
          })
        }
      })
  })
```

Рис. 3. Структура верификации пользователя

Пользователь вводит свои учетные данные в форму авторизации с помощью атрибутов text и password. Кнопка submit на этой форме позволяет пользователю отправить эти данные на проверку. После нажатия на эту кнопку, функция usersReg.js проверяет введенные данные на сервере, сравнивая их с данными в базе данных. Если введенные данные совпадают с данными в базе, то для пользователя создается сессия и он переходит на основную страницу сайта - Noмерpage.js. (рис. 3)

Основной страницей для исследователя-пользователя является страница «Параллельный корпус», где отображается весь материал параллельного корпуса. Пользователь имеет возможность фильтровать корпус по языку и типу, а также использовать поле поиска для получения соответствующей информации.

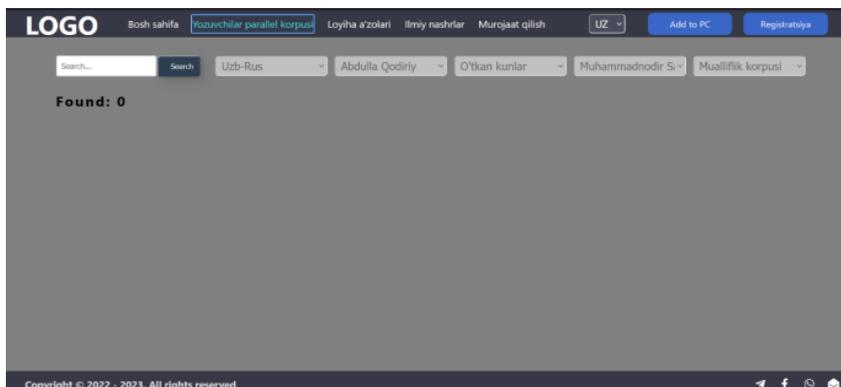


Рис. 4. Страница поиска информации

На этой странице основным инструментом является поле поиска. При введении запроса платформа предоставляет соответствующий текст с фильтрованной информацией, которая включает в себя результаты поиска

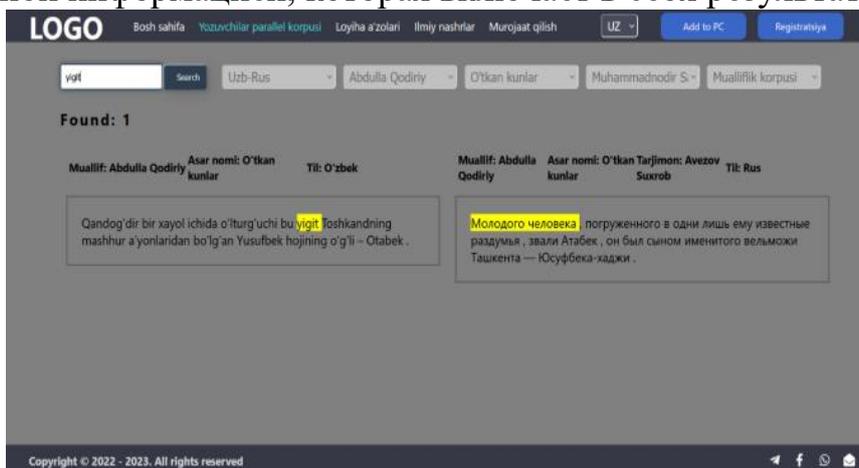


Рис. 5. Результат поиска

На этой платформе внесенная информация подвергается семантическому тегированию, и при поиске пользователь получает маркированный результат. Одним из преимуществ этой платформы является то, что на ней можно работать с узбекским текстом, чего не может сделать ни одна другая платформа.

Sketch Engine является одной из самых продвинутых платформ для работы с корпусами текста, но она не предоставляет возможности работы с текстом на узбекском языке. В связи с этим у Sketch Engine отсутствует функция маркирования узбекского текста.

Добавление информации в параллельный корпус включает несколько этапов, таких как ввод имени автора и фамилии, внесение названия текста, информации о переводчике, выбор конфигурации языков и типа корпуса, загрузка текста на языке-оригинале и текста-перевода, а также выполнение семантического тегирования. После этого информация сохраняется в базу данных.

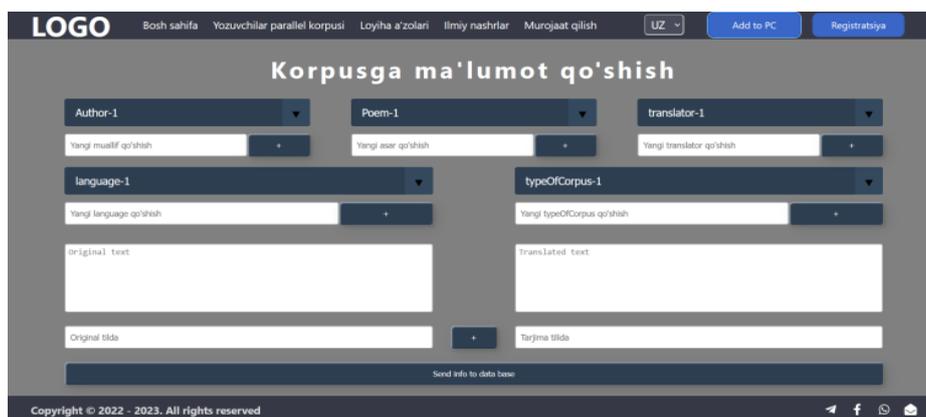


Рис. 6. Окно добавления информации в корпус

Таким образом, разработанное веб-приложение предоставляет исследователям возможность работать с параллельным корпусом узбекского и русского языков, добавлять новую информацию, фильтровать и искать тексты, а также проводить семантическое тегирование для улучшения поиска и анализа данных.

В параграфе 2.3. «**Выравнивание как основной этап в создании базы данных параллельного корпуса**» рассматривается роль и значение выравнивания в процессе создания базы данных параллельного корпуса. Выравнивание является важным этапом, который позволяет связать соответствующие фрагменты текста на разных языках и создать параллельные связи между ними.

Один из основных аспектов выравнивания – это выбор метода или алгоритма, который наиболее эффективно и точно устанавливает соответствие между текстами на разных языках. Для выравнивания фразеологизмов из романа «Минувшие дни» Абдуллы Кадыри на узбекском языке и их перевода на русский язык был использован метод выравнивания на уровне фраз. Данный метод позволяет учитывать более мелкие единицы текста, чем предложения, что позволяет более точно определять соответствие между текстами на разных языках.

Использование метода выравнивания на уровне фраз позволило более точно связать соответствующие фразеологизмы на узбекском и русском языках и использовать полученные данные для дальнейшего анализа перевода фразеологизмов.

Далее приведен пример из параллельного корпуса:

Muallif: Абдулла Қодирий	Asar nomi: Ўткан кунлар	Til: O'zbek	Абдулла Қодирий	Ўткан кунлар	Мухаммаднодир	Til: Rus
Yusufbek hojining «O'g'lim yosh, sen dunyoning issiq-sovug'ini tatig'an va manim ishongan kishimsan.			Из головы не выходил наказ Юсуфбека хаджи «Сын мой еще молод, ты же прошел огонь и воду, видел и знаешь многое, предан мне.			

Рис. 7. Результат перевода

В данном предложении используется узбекская пословица «sen dunyoning issiq-sovug'ini tatig'an», что дословно означает «ты испытал тепло и холод мира», и её русский перевод «ты же прошел огонь и воду».

Узбекская пословица «sen dunyoning issiq-sovug'ini tatig'an» используется для описания человека, который прошел через множество жизненных

испытаний и обрел опыт, мудрость. В данном контексте, эта пословица подчеркивает, что адресат слов Юсуфбека-хаджи (Хасанали) является опытным и заслуживает доверия.

Русский перевод «ты же прошел огонь и воду» также указывает на то, что человек пережил множество трудностей и испытаний, став опытным и надежным. В данном случае, перевод в целом сохраняет смысловую нагрузку и дух оригинальной узбекской пословицы.

Анализируя приёмы и методы перевода фразеологизмов в данном примере, можно выделить приём «Эквивалентность»: замена исходной пословицы на аналогичную пословицу на целевом языке (русском), которая имеет схожее значение. В данном случае, узбекская пословица заменена русской пословицей «прошел огонь и воду», что сохраняет общий смысл и эмоциональную окраску выражения.

Таким образом, перевод соответствует оригиналу, и смысловая нагрузка пословицы сохранена. Используя метод эквивалентности, переводчик успешно передал смысл узбекской пословицы с помощью русской идиомы, делая текст доступным и понятным для русскоязычной аудитории.

Продолжая изложение содержания параграфа 2.3., следует отметить, что выравнивание на уровне слов также рассматривается как важный метод, позволяющий более точно связать тексты на разных языках.

Далее подчеркивается, что выравнивание на уровне слов широко используется в области машинного перевода и обработки естественного языка. Оно применяется в задачах автоматического перевода, анализа текста, распознавания речи и других. Приводятся примеры конкретных проектов и систем, таких как Moses, GIZA++, FastAlign и OpenNMT, которые используют выравнивание на уровне слов в своих алгоритмах работы.

В заключении параграфа подчеркивается, что несмотря на преимущества каждого вида выравнивания (на уровне фраз, слов и символов), их использование должно учитывать специфику конкретных языков и задач. Важно проводить дополнительные исследования и разработку новых методов и алгоритмов для эффективного выравнивания текстов на узбекском и русском языках, учитывая их особенности и потребности в области обработки естественного языка.

Третья глава диссертации **«Параллельные корпуса в переводческой практике: методы, преимущества и перспективы на примере узбекско-русского перевода романа “Минувшие дни”»** состоит из трех параграфов.

В параграфе 3.1. «Анализ методов и приемов перевода фразеологизмов в параллельном корпусе узбекского и русского языков на примере романа Абдуллы Кадыри “Минувшие дни”» проведен анализ методов и приемов перевода фразеологических единиц из узбекского языка на русский язык с использованием параллельного корпуса текстов. В данном исследовании использовался роман Абдуллы Кадыри «Минувшие дни» как пример текста, содержащего множество фразеологизмов и идиоматических выражений на узбекском языке.

Исследование основывается на сопоставлении фразеологических единиц и их переводов на русский язык в параллельном корпусе, что позволяет провести детальный анализ методов перевода. Анализ проводится без привязки к конкретным датам, фамилиям или именам, поскольку основное внимание уделяется методам и приемам перевода фразеологических единиц.

В рамках параграфа рассмотрены различные методы и приемы перевода фразеологических единиц, такие как эквивалентность, адаптация, калькирование и дословный перевод. Приводятся примеры перевода фразеологизмов и анализируется их соответствие оригиналу, передача смысла и образности. При анализе учитываются как смысловые особенности фразеологических единиц, так и культурные и контекстуальные аспекты.

Далее приводится пример из параллельного корпуса:

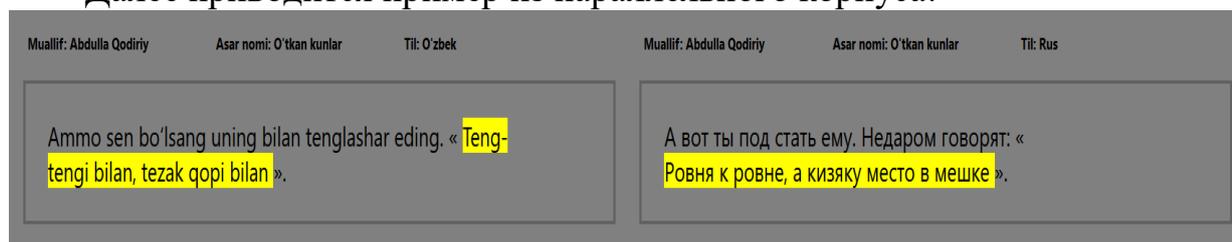


Рис. 8. Результат перевода фразеологизмов

Тойбекины слова, полушутливо описывающие подходящего жениха для Кумуш как «молодого гостя», вызывают у последней раздражение, на что она с иронией предлагает служанке вступить в брак с ним. В ответ служанка использует пословицу «Teng-tengi bilan, tezak qopi bilan»⁶⁶, которую М. Сафаров переводит как «Ровня к ровне, а кизяку место в мешке»⁶⁷. Этот перевод, кажущийся вначале удачным за счет точной передачи смысла и соответствия форме пословицы, при более глубоком анализе теряет свою значимость. Причина этого кроется в том, что пословица имеет узкоспецифическую связь с узбекским бытом, что добавляет к ней национальную специфику. Однако, в комментарии к слову «кизяк», объясняемому как «навоз», возникает несоответствие: в узбекской культуре «кизяк» относится к высушенному навозу, используемому как топливо, часто укладываемому в мешки, а не к самому навозу⁶⁸. Это неправильное толкование приводит к путанице у русскоязычного читателя, который не может сопоставить хранение навоза в мешке с его традиционным использованием, выходя за рамки общепринятых представлений.

Исследование методов и приемов перевода фразеологизмов в параллельном корпусе является значимым, поскольку позволяет лучше понять процесс перевода и эффективные способы передачи смысла и образности фразеологических выражений между узбекским и русским языками. Это имеет практическое применение при переводе литературных, художественных и других текстов, где фразеологизмы играют важную роль в передаче культурного контекста и особенностей языка.

⁶⁶ Qodiriy A. O'tkan kunlar. – Toshkent: Sharq, – 1994. – С. 16.

⁶⁷ Кадыри А. Минувшие дни. – Ташкент: Шарк, – 2009. – С. 31.

⁶⁸ <https://dic.academic.ru/dic.nsf/ushakov/833562>

В параграфе 3.2. «Преимущества использования параллельного корпуса для изучения и перевода реалий на примере узбекско-русского языка» подчеркиваются важные моменты, связанные с использованием параллельного корпуса для исследования и перевода реалий на узбекский и русский языки.

Одним из ключевых преимуществ параллельного корпуса является его сущность – это коллекция текстов, в которой каждый текст на узбекском языке соотносится с текстом на русском языке. Это предоставляет возможность сравнивать языковые явления и методы перевода в контексте реальных текстов.

Использование параллельного корпуса позволяет более эффективно исследовать реалии, которые отражают специфические аспекты культуры, религии, истории и других областей жизни носителей узбекского и русского языков. Реалии представляют собой слова и выражения, которые трудно передать точно и адекватно на другой язык без учета контекста и культурных особенностей. Использование параллельного корпуса облегчает понимание и перевод реалий, позволяя учесть их значение и контекст.

Таким образом, основной целью данного параграфа является исследование специфики слов-реалий и изучение методов перевода этих реалий на основе сопоставления оригинального текста романа «Минувшие дни» и его русского перевода, выполненного М. Сафаровым. В ходе анализа мы используем параллельный корпус узбекского и русского языков, разработанный нами для данного исследования.

Muallif: Абдулла Қодирий Asar nomi: Ўткан кунлар Til: O'zbek	Абдулла Қодирий Ўткан кунлар Мухаммаднодир Til: Rus
Saroyning to'rida boshqalarg'a qarag'anda ko'rkamrak bir hujra, anovi hujralarga kiygiz to'shalgani holda bu hujrada qip-qizil gilam, uttalarda bo'z ko'rpalar ko'rilgan bo'lsa, munda ipak va adras ko'rpalar, narigilarda qora charog' sasiq'anda, bu hujrada sham' yonadir, o'zga hujralarda yengil tabi'atlik, serchaqchaq kishilar bo'lg'anida bu hujraning egasi boshqacha yaratilishda.	В глубине двора — отличающаяся изысканным убранством уютная комната. Если в других комнатах полы устланы кошмами, то здесь — вместо них лежат ярко-красные ковры; если остальные снабжены бязевыми одеялами, здесь их заменяют курпачи из шелка и адраса; если везде коптят плошки, тут резво пляшет пламя свечи. Да и обитатель помещения, не в пример легкомысленным, заводным постояльцам других комнат, держится иначе.

Рис. 9. Результат перевода реалий

В данном примере выбранные слова можно классифицировать как реалии, поскольку они отображают национальные и исторические аспекты языка. Слова «хужра», «кўрпалар» и «адрас» имеют смысловую связь с бытовыми реалиями, исходя из их значения.

Слово «хужра» происходит от арабского языка и имеет два значения. Первое значение этого слова – «небольшая комната», в то время как второе – «келья» – обладает исторической окраской и описывает жилую комнату в медресе, которую в основном использовали преподаватели и студенты учебных заведений стран Центральной и Западной Азии.

В литературном произведении слово «хужра» относится к маленькой комнате караван-сарая, где Атабек проживал во время своих странствий в Маргилане. Ныне это слово вышло из активного словарного запаса узбекского

языка и сохранило только свое второе значение – «небольшая комната в медресе, караван-сараяе, мечетях». Использование данного слова автором способствовало воссозданию исторического образа небольшой и темной комнаты, скрытой в углу.

В переводе слова использовался перифрастический метод: переводчик, применив словосочетание «уютная комната», пытался передать смысловое значение оригинала. Однако при этом исторический и религиозный колорит слова «хужра» в переводе не сохранился, так как само слово «комната» имеет три словарных значения, но ни одно из них не соответствует значению слова «хужра». Тем не менее, стоит отметить, что для более точной передачи смысла переводчик использовал также прилагательное «уютный». Благодаря этому перед читателем возникает образ небольшой, но достаточно комфортной комнаты для временного проживания.

Второе слово «кўрпа» в рассматриваемом примере также относится к бытовым реалиям и обладает выраженным национальным характером. В переводе романа на русский язык произошла ошибочная замена слова «курпа» на его сходное слово «курпача».

В данном отрывке литературного произведения слово «курпа» семантически обозначает специальное покрывало для прикрытия танчи или сандала. В романе читателю становится ясно, что этот предмет быта используется в холодные зимние дни как специальное устройство для сохранения тепла. Однако в переводе на русский язык это слово было трансформировано в «курпача» и, таким образом, приобрело иное значение. Слово «курпача» на узбекском языке означает «специальный матрас, узкое ватное одеяло для сидения, ватная подстилка». Из этого следует, что в переводе слова «курпа» с узбекского на русский язык возникла неточность, которая искажает смысл переводимой единицы, хотя культурные и исторические особенности остаются неизменными. Причиной этого стал выбранный метод перевода – транслитерация, которая является одним из эффективных способов передачи характерных черт узбекского народа и быта.

Следовательно, можно сделать вывод, что 1) эквивалентом узбекского слова «курпа» на русском языке является «одеяло, покрывало»; 2) слово «курпача» не обладает прямым эквивалентом в русском языке, тем не менее, оно существенно отличается по семантике от связанного с ним слова «курпа» и означает «узкое шелковое одеяло, наполненное ватой, предназначенное для сидения».

В данном примере последнее слово «адрас» также является примером бытовых реалий. Слово «адрас» на узбекском языке описывает хлопчатобумажную ткань с восточным рисунком. В произведении автор использовал это слово для характеристики обстановки и изыскаемости комнаты – «хужра», где остановился Атабек. Обладая специфическим национальным и культурным колоритом, слово «адрас» не имеет прямого эквивалента на русском языке. Исходя из этого, в переводе был применен метод транслитерации.

Таким образом, использование параллельного корпуса узбекско-русского языка при изучении и переводе реалий предоставляет значительные преимущества, облегчая анализ и понимание культурных, религиозных и исторических аспектов, связанных с реалиями. Этот подход способствует более точному и культурно адекватному переводу, укрепляет межкультурный диалог и взаимопонимание между носителями разных языков.

В параграфе 3.3. «Машинный перевод и NLP: использование параллельных корпусов для повышения эффективности перевода» рассматривается роль параллельных корпусов в повышении эффективности систем машинного перевода и развитии области обработки естественного языка (NLP). Этот подход имеет большое значение для создания систем машинного перевода между узбекским и русским языками.

Параллельные корпуса являются ценным ресурсом, предоставляющим моделям машинного перевода большой объем параллельных текстов на различных языках. Созданный узбекско-русский параллельный корпус является примером такого ресурса, который может быть использован для обучения и улучшения систем машинного перевода между этими двумя языками.

Важность узбекско-русского параллельного корпуса заключается в его способности учить модели NLP языковым особенностям и устанавливать соответствия между узбекским и русским языками. Параллельный корпус также способствует сохранению и распространению узбекского языка и его культурного наследия. Создание и использование этого параллельного корпуса открывает новые возможности для межкультурного взаимодействия и сотрудничества в различных сферах, включая образование, бизнес и международные отношения.

Модели глубокого обучения, такие как BERT и GPT, могут быть использованы для машинного перевода узбекского языка. Для этого требуется предварительное обучение моделей на больших корпусах текста, а затем их дообучение на конкретной задаче перевода. Другие модели, такие как Seq2Seq, Transformer, T5 и Marian NMT, также могут быть использованы. Важно иметь большой и разнообразный параллельный корпус для эффективного обучения моделей. Компании, такие как Google, Facebook, OpenAI и Microsoft, используют различные модели и подходы для своих систем машинного перевода. Создание платформы машинного перевода для узбекского языка является актуальной задачей, которая может способствовать межкультурной коммуникации и сохранению культурного наследия.

В контексте развития машинного перевода и NLP, использование параллельных корпусов является важным инструментом для обучения моделей и улучшения качества перевода. Узбекско-русский параллельный корпус, созданный в данном контексте, представляет собой актуальный и перспективный ресурс для развития систем машинного перевода между узбекским и русским языками, а также для сохранения и распространения узбекского языка и его культурного наследия.

ЗАКЛЮЧЕНИЕ

1. Компьютерная лингвистика – дисциплина на стыке языкознания и информационных технологий, а корпусная лингвистика занимается анализом и систематизацией естественных языковых данных в больших текстовых коллекциях, обе науки играют ключевую роль в изучении и совершенствовании языков. Внедрение данных направлений способствует глубокому изучению структурных и функциональных особенностей узбекского языка, обеспечивая создание эффективных инструментов для автоматического перевода, обработки текстов и лингвистического анализа.

2. Параллельный корпус представляет собой лингвистическую базу данных, состоящую из оригинальных текстов и их переводов на один или несколько языков, что обеспечивает уникальную возможность для анализа языковых структур и их соответствий в различных языковых системах. В отличие от одноязычных и сопоставительных корпусов, которые либо фокусируются на одном языке, либо анализируют языки независимо друг от друга, параллельные корпуса позволяют проводить глубокий анализ переводческих эквивалентов, выявляя тонкости языкового взаимодействия, что значительно обогащает переводческую и лингвистическую науку. Они предоставляют неоценимую помощь в изучении семантических, синтаксических и контекстуальных аспектов языка, что делает их незаменимым инструментом не только для переводчиков и лингвистов, но и для машинного обучения в области обработки естественного языка.

3. В современном контексте, «Всемирная паутина» (Web) выступает в качестве глобальной информационной системы, основанной на принципах гипертекста и доступной через интернет, что позволяет пользователям просматривать и взаимодействовать с широким спектром данных и ресурсов. Она образует самый обширный и динамично пополняемый репозиторий текстовых, аудиовизуальных и мультимедийных материалов, что делает её наиболее значимым источником для сбора и анализа параллельных корпусов, важных для исследований в области компьютерной лингвистики, машинного перевода и межкультурной коммуникации. Эта уникальная особенность Web предоставляет исследователям беспрецедентные возможности для извлечения и обработки параллельных текстов, способствует углублению понимания языковых структур и культурных особенностей, закладывая тем самым фундамент для создания высококачественных и адаптированных к контексту систем машинного перевода.

4. В разработке онлайн-платформы для работы с узбекско-русскими параллельными корпусами выбор технологического стека, состоящего из React, Node.js, Express.js, MongoDB и Visual Studio Code, является оптимальным с учётом требований к интерактивности, масштабируемости и удобству разработки. React обеспечивает высокую производительность и динамичное взаимодействие за счёт реактивности и компонентного подхода, что важно для визуализации параллельных текстов. Node.js и Express.js создают легковесную, мощную серверную среду для асинхронных запросов и

управления данными, упрощая создание API и интеграцию с MongoDB — документо-ориентированной базой данных, подходящей для хранения текстовых данных. Visual Studio Code поддерживает эффективное кодирование и отладку, ускоряя разработку.

5. Текстовое выравнивание является ключевым этапом при создании баз данных параллельных корпусов, обеспечивая точное сопоставление текстовых сегментов в различных языках, что важно для точности и эффективности лингвистических исследований. Выравнивание может быть разного типа: от слова к слову, предложению к предложению, от абзаца к абзацу, обеспечивая анализ на различных уровнях контекста и углубляя понимание структурных и семантических связей между текстами. Существует также подуровневое выравнивание, например, на уровне фраз или синтаксических единиц, для детализированного сопоставления и глубокого понимания языковых отношений.

6. Языковой тип русского языка – флективный, узбекский же относится к агглютинативному языковому строю. Различия структурных языковых особенностей усложняют процесс выравнивания текстов, в процессе создания узбекско-русской и русско-узбекской онлайн-платформы параллельных корпусов, делая невозможным прямое применение универсальных методик выравнивания. Для решения этой проблемы был разработан специализированный метод выравнивания на уровне слов, для обеспечения максимальной точности перевода отдельных лексических единиц.

7. При переводе фразеологизмов и идиом, учитывая их культурную и контекстуальную специфику, использовалось выравнивание на уровне фраз и предложений для обеспечения адекватности и естественности перевода. Методы анализа перевода этих выражений из узбекского языка на русский и наоборот подчеркивают важность точной передачи не только лексического содержания, но и культурно-контекстуальной насыщенности. Это достигается через трансформацию структуры для сохранения функциональной нагрузки, компенсацию для воспроизведения аналогичного эффекта и модификацию культурного контекста для лучшего понимания целевой аудиторией. Использование пояснительных элементов и опускание нерелевантных компонентов также способствуют более глубокому пониманию и адаптации перевода.

8. В анализе перевода языковых реалий романа А.Кадыри «Минувшие дни» ключевую роль играют методы транслитерации для сохранения оригинального звучания и контекстуальной адаптации, обеспечивающей понимание смысла с учетом культурных особенностей. Создание неологизмов позволяет ввести новые термины для уникальных явлений, а перифрастический перевод детализирует смысл. Калькирование и поиск функциональных аналогов способствуют точной и культурно адаптированной передаче содержания, в то время как контекстуальный перевод учитывает лингвокультурный контекст для достижения адекватности и естественности.

Эти методы вместе формируют основу для создания перевода, точно отражающего глубину оригинала.

9. В контексте разработки машинного перевода, применение методов обработки естественного языка (NLP) является неотъемлемым аспектом, поскольку они обеспечивают глубокое понимание и анализ языковых структур и семантики. Среди примеров таких методов можно выделить статистический машинный перевод (SMT), который базируется на анализе больших объемов параллельных текстов для выявления вероятностных соответствий между языковыми единицами, и нейронный машинный перевод (NMT), использующий глубокие нейронные сети для моделирования последовательностей перевода на основе контекста и семантики, что позволяет достигать более высокой точности и естественности переводимого текста. Кроме того, методы, основанные на трансформерах, такие как BERT (Bidirectional encoder representations from transformers) и GPT (Generative pre-trained transformer), способствуют значительному прогрессу в понимании нюансов языка и генерации текста, обеспечивая динамическую адаптацию к контексту и улучшение качества перевода за счет более глубокого анализа синтаксических и семантических отношений в тексте.

10. Национальный корпус языка представляет собой обширное собрание текстов, систематически организованных и цифровизированных, которое охватывает широкий спектр лингвистических выражений, отражающих культурные, исторические и социальные аспекты нации, тем самым обеспечивая исчерпывающую базу данных для исследований в области лингвистики, лексикографии и автоматической обработки естественного языка. В этом контексте параллельный корпус, состоящий из текстов, переведенных на два или более языка и выровненных таким образом, что соответствующие сегменты текста на разных языках могут быть сопоставлены, играет ключевую роль в национальном корпусе языка, поскольку он не только способствует глубокому пониманию межъязыковых отношений и культурных нюансов, но и является неоценимым ресурсом для развития технологий машинного перевода и межкультурной коммуникации.

AVEZOV SUKHROB SOBIROVICH

**BASIC PRINCIPLES OF CREATING UZBEK-RUSSIAN AND RUSSIAN-
UZBEK ONLINE PLATFORMS FOR PARALLEL CORPORA (BASED ON
THE MATERIAL OF A. KADYRI'S NOVEL "BYGONE DAYS")**

10.00.11 – Language theory. Applied and computational linguistics

ABSTRACT
of the doctor of philosophy (PhD) dissertation in PHILOLOGY

Fergana – 2024

The theme of PhD dissertation was registered by the Higher Attestation Commission under the number B2022.3.PhD/Fil2798.

The doctoral thesis has completed at the chair of Bukhara State University.

The abstract of the dissertation is posted in three languages (uzbek, russian, english (resume)) on the website of the scientific council (www.fdu.uz) and on the "ziyonet" information and educational portal (www.ziyonet.uz)

Scientific supervisor:

Nigmatova Lolakhon Khamidovna
Doctor of Science in Philology, Associate Professor

Official opponents:

Rahimov Usmonjon Ergashevich
candidate of philological sciences,
associate professor

Mukhiddinov Anvar Gafurovich

doctor of philological sciences (DSc), professor

Leading organization:

Samarkand State University named after Sharof Rashidov

The Defense of the Dissertation will take place on 27 07 2024 at 10:00 at the meeting of the Scientific Council numbered DSc.03/30.12.2019.Phil.05.02 under Fergana State University (Address: 100151, Fergana city, 105 Burhoniddin Marginoniy street. Tel: (99873) 244-57-82; fax: (99873) 244-66-03; e-mail: info@fdu.uz)

The dissertation is available in the Information-Resource center of Fergana State University (registered under № 356) Address: 19, Murabbiy street, Ferghana, 100151, Tel: (+99873) 244-71-28;

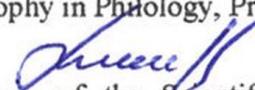
The abstract of the dissertation is distributed on « 6 » 07 2024.

(Mailing report No. 18 of « 6 » 07 2024)




S.M. Muminov
Chairman of the Scientific Council awarding scientific degrees, Doctor of Science in Philology, Professor


M.T. Zokirov
Scientific Secretary of the Scientific Council awarding scientific degrees, Doctor of Philosophy in Philology, Professor


A. Mamajonov
Chairman of the Scientific Seminar under Scientific Council awarding scientific degrees, Doctor of Science in Philology, Professor

INTRODUCTION (annotation of Doctor of Philosophy (PhD) dissertation`

The aim of the research is to study the theoretical foundations of creating ... Uzbek-Russian parallel corpus, to develop an online platform for Uzbek-Russian and Russian-Uzbek parallel corpora based on the example of A.Kadyri's novel "Bygone Days", as well as to improve machine translation, creating a convenient tool for linguistic analysis and intercultural exchange at the level of sentences, phrases, and text units between the Uzbek and Russian languages.

The object of the research work is the Uzbek-Russian and Russian-Uzbek parallel corpus.

The scientific novelty of the study is as follows:

the analysis of the linguistic realities of A.Kadyri's novel "Bygone Days" through a parallel text corpus has substantiated the importance of considering lexical, semantic, stylistic, and cultural features when developing a methodology for bilingual corpora, as well as identified the main aspects and challenges of translation;

a unique online platform has been developed for working with Uzbek-Russian parallel texts, providing opportunities and tools for real-time data analysis and editing, filling the existing gap in the work with parallel corpora of these languages;

the main aspects of the parallel corpus for linguistic research have been substantiated, their specific features, in particular, the ability to compare texts in different languages, which helps to study translation strategies and understand intercultural features;

a detailed analysis of phraseological units in Uzbek-Russian parallel corpora has been conducted using the example of A.Kadyri's novel "Bygone Days", methods of their translation, as well as cultural adaptation, semantic and contextual features have been identified.

Implementation of research results. Based on the theoretical and practical aspects of creating parallel corpora and applying technologies in the field of natural language processing, the following activities were conducted:

a detailed analysis of phraseological units in A.Kadiri's novel "Bygone Days" in Uzbek-Russian parallel corpora, as well as a study of methods for their translation, cultural adaptation, semantic, and contextual features, were used in the fundamental project I-OT-2019-42 "Creation of an electronic poetic dictionary of Uzbek and English languages (images of human appearance, character, nature, and national symbols)", implemented at the Tashkent State University of Uzbek Language and Literature (reference № 01/10-1150 dated June 9, 2023). As a result, the concept of multilingualism, translation algorithms, and approaches to creating a text corpus facilitated the translation process, improved translation accuracy, and created a vast database for search and analysis in the dictionary;

the analysis of linguistic realities in A.Kadiri's novel "Bygone Days" using a parallel text corpus, the study of key aspects and difficulties of translation, consideration of lexical, semantic, stylistic, and cultural features in the development

of bilingual corpora methodology were included in the 13th paragraph of the textbook “General linguistics” (reference from the Ministry of Higher Education, Science and Innovation of Bukhara State University № 01-04/3562 dated October 12, 2023, GRIF № 112-13). As a result, the educational process was simplified, and students were provided with broader tools for analysis and practical application of linguistic structures;

the theoretical and methodological basis obtained from the studies of the main aspects of the parallel corpus for linguistic research, their unique characteristics, in particular, the study of translation strategies and the ability to compare texts in different languages, helping to understand intercultural features, was used to write scripts for the programs “Zamondosh” and “Milliy ta’lim” on the channel “Madaniyat va ma’rifat” (reference № 02-02/125 dated June 1, 2023). As a result, the scientific significance and popularity of these TV programs increased.

Structure and volume of the dissertation. The dissertation comprises an introduction, 3 chapters, a conclusion, list of referenced literature and an application, and its total volume is 162 pages..

E'LON QILINGAN ISHLAR RO'XATI
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

1. Avezov S. S. A corpus-based approach to phraseology research // Scientific reports of Bukhara State University. Bukhara 2023. -№ 1 (95) – С. 189-193. [10.00.00. №.1].

2. Avezov S. S. Corpus linguistics: historical and linguodidactic aspects // Scientific reports of Bukhara State University. Bukhara 2022. -№ 3 (91) – С. 86-94. [10.00.00. №.1].

3. Аvezов С.С. Особенности и различия параллельного корпуса русского и узбекского языков // Преподавание языка и литературы. – Ташкент, 2023. - №3. – С.66-69. [10.00.00. №.9].

4. Avezov S. S. Utilizing parallel corpora for electronic dictionary development: an examination of the uzbek-russian languages // Innovative Technologica: Methodical Research Journal (IT). – 2023. – Т. 2. – №. 07. – С. 43-48. SJIF = 7.247 (2023) [№23].

5. Аvezов С., Юсупова А. Процесс обработки узбекского параллельного корпуса в условиях недостаточности данных //Eurasian Journal of Social Sciences, Philosophy and Culture. – 2023. – Т. 3. – №. 3. – С. 49-58. SJIF = 7.164 (2023) [№23].

6. Аvezов С. Корпусная лингвистика: новые подходы к анализу языка и их приложения в обучении иностранным языкам //International Bulletin of Applied Science and Technology. – 2023. – Т. 3. – №. 7. – С. 177-181. SJIF = 6.904 (2023) [№23].

7. Аvezов С.С., Маринина Ю.А. Электронные Корпусы: Инновационный Подход К Обучению Переводу //Periodica Journal of Modern Philosophy, Social Sciences and Humanities. – 2023. – Т. 16. – С. 7-13. SJIF = 6.527 (2023) [№23].

8. Avezov S. S. Machine translation to align parallel texts //International Scientific and Current Research Conferences. – USA-2022. – С. 64-66.

9. Аvezов С.С. Применение параллельных корпусов в создании многоязычных переводных словарей: проблемы и перспективы // “O‘zbek tilining milliy korpusi: muammolar va vazifalar” mavzusidagi xalqaro ilmiy-amaliy anjumani, 25-marta. – Самарканд, 2023. – С.65-68.

10. Аvezов С. Применение методов NLP в корпусных исследованиях: особенности и ограничения //«Узбекские национальные образовательные здания теоретическое и практическое создание вопросы» Республиканская научно-практическая конференция. – 2023. – Т. 2. – №. 2. – С.216-220.

11. Аvezов С. Параллельный корпус узбекского и русского языков: применение ИИ и NLP в образовании //«Ta'limda raqamli texnologiyalarni tadbiq etishning zamonaviy tendensiyalari va rivojlanish omillari» mavzusidagi respublika ilmiy-amaliy anjumani. – 2023. – Т. 2. – №. 2. – С.216-220.

II bo‘lim (II часть, Part II)

12. Avezov S. S. Development of a Parallel Corpus of the Uzbek and Russian Languages // Vital Annex: International Journal of Novel Research in Advanced Sciences. – 2022. – Т. 1. – №. 5. – С. 152-155.

13. Аvezов С.С. О корпусной лингвистике, трудностях перевода и принципах организации параллельных корпусов текстов // «O‘zbek tilining milliy korpusi: muammo va vazifalar» mavzusidagi xalqaro ilmiy-amaliy konferensiya, 31-март. – Ташкент, 2022. – С.274-279

14. Аvezов С.С. On the use of corporals in linguistic research // VII международный научно-методический онлайн-семинар «Совершенствование методики обучения языкам площадка обмена прогрессивной практикой», – Казань, 2023. – С.2-4.

15. Аvezов С.С. Параллельный корпус как инструмент в работе переводчика // Международной научно-практической конференции «Актуальные проблемы сопоставительного языкознания, сравнительного литературоведения и переводоведения», 22-декабря. – Навои, 2022. – С.252-256.

Avtoreferat Farg‘ona davlat universiteti
“Lingvistik tahrir va tarjimashunoslik” markazida tahrirdan o‘tkazildi

Bosishga ruxsat etildi: 2024 y. Nashriyot bosma tabog‘i – 3,375.
Shartli bosma tabog‘i – 1,68. Bichimi 84x108 1/16. Adadi 100.
«Poligraf Super Servis» MCHJ
150114, Farg‘ona viloyati, Farg‘ona shahar, Aviasozlar ko‘chasi 2-uy.