

**O‘ZBEKISTON DAVLAT JAHON TILLARI UNIVERSITETI
HUZURIDAGI ILMY DARAJALAR BERUVCHI
DSc.03/30.12.2019.FIL.27.01 RAQAMLI ILMY KENGASH
ASOSIDAGI BIR MARTALIK ILMY KENGASH**

O‘ZBEKISTON DAVLAT JAHON TILLARI UNIVERSITETI

XAITQULOV ZAHRIDDIN XAZRATQUL O‘G‘LI

**TARJIMANI AVTOMATLASHTIRISHDA PARALLEL KORPUS
KONTEKSTUAL QIDIRUV TIZIMINING FUNKSIONAL MOHIYATI**

10.00.11 – Til nazariyasi. Amaliy va kompyuter lingvistikasi

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

Toshkent – 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) in
philological sciences**

Xaitqulov Zahriddin Hazratqul o‘g‘li

Tarjimoni avtomatlashtirishda parallel korpus kontekstual qidiruv tizimining funksional mohiyati3

Хаиткулов Захриддин Хазраткул угли

Функциональная особенность параллельной корпусной контекстуальной поисковой системы в автоматизации перевода29

Khaitkulov Zahriddin Khazratkul ugli

The Functional Features of an Automated Translation of Parallel Corpus Contextual Search57

E‘lon qilingan ishlar ro‘uxati

Список опубликованных работ
The list of publications.....62

**O‘ZBEKISTON DAVLAT JAHON TILLARI UNIVERSITETI
HUZURIDAGI ILMY DARAJALAR BERUVCHI
DSc.03/30.12.2019.FIL.27.01 RAQAMLI ILMY KENGASH
ASOSIDAGI BIR MARTALIK ILMY KENGASH**

O‘ZBEKISTON DAVLAT JAHON TILLARI UNIVERSITETI

XAITQULOV ZAHRIDDIN XAZRATQUL O‘G‘LI

**TARJIMANI AVTOMATLASHTIRISHDA PARALLEL KORPUS
KONTEKSTUAL QIDIRUV TIZIMINING FUNKSIONAL MOHIYATI**

10.00.11 – Til nazariyasi. Amaliy va kompyuter lingvistikasi

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

Toshkent – 2024

Falsafa doktori (PhD) dissertatsiyasi mavzusi O'zbekiston Respublikasi Oliy attestatsiya komissiyasida B2023.1.PhD/Fil3237 bilan ro'yxatga olingan.

Dissertatsiya O'zbekiston davlat jahon tillari universitetida bajarilgan.

Dissertatsiya avtoreferati uch tilda (o'zbek, rus, va ingliz (rezyume)) Ilmiy kengashining veb-sahifasida (www.uzswlu.uz) va "ZiyoNet" Axborot ta'lim portalida (www.ziynet.uz) joylashtirilgan.

Ilmiy rahbar:

Ergasheva Guli Ismoil qizi
filologiya fanlari doktori, dotsent

Rasmiy opponentlar:

Xayrullayev Xurshid Zayniyevich
filologiya fanlari doktori, professor

Ataboyev Nozimjon Bobojon o'g'li
filologiya fanlari doktori, (PhD) dotsent

Yetakchi tashkilot:

Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti

Dissertatsiya himoyasi O'zbekiston davlat jahon tillari universiteti huzuridagi ilmiy darajalar beruvchi DSc.03/30 12.2019. Fil.27.01 raqamli ilmiy kengashining "26" aprel 2024-yil soat 10:00 dagi majlisida bo'lib o'tadi (Manzil Toshkent shahri, Uchtepa tumani, Kichik halqa yo'li ko'chasi, G-9A mavzesi, 21-uy. Tel.: (71) 230-12-91; faks: (99871) 230-12-92; e-mail: uzswlu_info@mail.ru).

Dissertatsiya bilan O'zbekiston davlat jahon tillari universiteti Axborot-resurs markazida tanishish mumkin (1 raqami bilan ro'yxatga olingan). Manzil: 100138, Toshkent sh., Kichik halqa yo'li ko'chasi, G-9A mavzesi, 21-uy. Tel.: (71) 230-12-91; faks: (99871) 230-12-92.

Dissertatsiya avtoreferati 2024-yil "13" aprel kuni tarqatildi.
(2024-yil "13" aprel dagi 1 raqamli reyestr bayonnomasi).



G.H. Bakiyeva

Ilmiy darajalar beruvchi Bir martalik
ilmiy kengash raisi, filologiya fanlari
doktori, professor

X.B. Samigova

Ilmiy darajalar beruvchi Bir martalik
ilmiy kengash ilmiy kotibi, filologiya
fanlari doktori professor

J.A. Yakubov

Ilmiy darajalar beruvchi Bir martalik
ilmiy kengash qoshidagi Bir martalik
ilmiy seminar raisi, filologiya fanlari
doktori, professor

KIRISH (falsafa doktori (PhD) dissertatsiyasi annotatsiyasi)

Tadqiqot mavzusining dolzarbligi va zarurati. Jahon tilshunosligida tabiiy tilga ishlov berish, asliyat tilni kompyuter yordamida tarjima qilish matematik lingvistika va mashina tarjimasi amaliyotida keng qo'llana boshlandi. Axborot texnologiyalarini tilshunoslikka tatbiq etish bilan bog'liq istiqbolli ilmiy yo'nalishlarda axborot-qidiruv tizimi, korpus lingvistiksi va mashina tarjimasida bir qancha ikki hamda ko'p tilli elektron tarjimon dasturlarini yaratish, korpus turlari orqali tarjima dasturlarining mexanizmlarini kengaytirish uni amliyotga tatbiq qilish zaruriyatini yuzaga keltirdi. Zero, sun'iy intellektning parallel korpus orqali turli xil ko'zlangan tarjimon darsturlarini yaratishga, avtomatlashtirilgan lug'atlar, parallel matnlarni ikki va ko'p tillarda tahlil hamda tahrir qilish borasidagi imkoniyatlarining amaliy ahamiyatini belgilaydi.

Dunyo tilshunosligida tarjimani avtomatlashtirish, ikki va ko'p tilli elektron tarjimon dasturlarini korpus turlari yordamida yaratish, tarjima sifatini yaxshilash, ikki tilli parallel matnlar banki mexanizmini yaratish kabi masalalarni ilmiy asoslash zarurati kun tartibiga qo'yilgan. Axborot-qidiruv tizimida ikki va ko'p tilli elektron lug'atlar, so'zma so'z tarjima qiluvchi, qoidaga asoslangan tarjimon dasturi, statistik ma'lumotlarga tayanuvchi tarjimon dasturi, korpus turlari orqali matnlarni tahlil qiluvchi dasturlar, shuningdek, tarjima matnlarni lingvostatistik o'rganish dasturlari va ilovalarini ishlab chiqish zarurati mavjud. Elektron tarjimon dasturlarini parallel va boshqa korpus turlaridan tashkil topgan matnlarni tahlil qilish ilovalari, kontekstual tarjimon dasturlari, katta va kichik hajmdagi ikki tilli ma'lumotlar bazasi, parallel korpusda avtomatik tarjima platformalari va ko'p tilli korpuslarni o'zida aks etgan yirik platformalarning yaratilishi hamda ularning jadal rivojlanishi kompyuter lingvistikasining hamda korpusga asoslangan tarjimashunoslikning ilmiy-nazariy muammolarini tadqiq etish muhim ahamiyatga ega.

Respublikamizda barcha sohalarda bo'lgani kabi tarjimashunoslikka korpus tahlillari, korpus turlari, avtomatik tarjima, statistik mashina tarjimasi, gibridd mashina tarjimasi, qoidaga asoslangan mashina tarjimasi, kompyuter lingvistikasi, ikki tilli va ko'p tilli korpuslar hamda tarjimon dasturlar ishlab chiqishda ingliz tilini o'zbek tilidagi tarjimalarini yanada chuqurroq tushunish va qayta ishlash yuzasidan ilmiy izlanishlar olib borish kabi masalalar kun tartibiga qo'yildi. "Davlat tilining sofligini saqlash, uni boyitib borish va aholining nutq madaniyatini oshirish; davlat tilining zamonaviy axborot texnologiyalari va kommunikatsiyalariga faol integratsiyalashuvini ta'minlash"¹ bugungi kunda tilshunoslik va tarjimashunoslik oldida turgan asosiy vazifalardan biridir. Shunga ko'ra, ingliz va o'zbek tillarini kompyuter tiliga aylantirish, kontekstual tarjimon dasturlarini parallel matnlar orqali korpusga solish, matnlarning leksik birlikari, semantik xususiyatlarini tarjimon dasturlaridan foydalanuvchilar uchun maxsus korpus yaratish singari masalalar muhim ahamiyat kasb etadi.

¹ 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>

O‘zbekiston Respublikasi Prezidentining 2017 yil 7 fevraldagi PQ-4947-son “O‘zbekiston Respublikasini yanada rivojlantirish bo‘yicha Harakatlar strategiyasi to‘g‘risida”gi qarori, 2017 yil 11avgustdagi 610-son “Ta’lim muassasalarida chet tillarini o‘qitishning sifatini yanada takomillashtirish chora-tadbirlari to‘g‘risida”gi, 2019 yil 8 oktyabrdagi PF-5847-son “O‘zbekiston Respublikasi oliy ta’lim tizimini 2030 yilgacha rivojlantirish konsepsiyasini tasdiqlash to‘g‘risida”gi Farmoni 2019 yil 21 oktyabrdagi PF-5850-son “O‘zbek tilining davlat tili sifatidagi nufuzi va mavqeini tubdan oshirish choratadbirlari to‘g‘risida”, 2020 yil 20 oktyabrdagi PF-6084-son “Mamlakatimizda o‘zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to‘g‘risida”gi farmonlari hamda O‘zbekiston Respublikasi Prezidentining Oliy Majlis palatalariga 2020-yilning 24-yanvardagi Murojaatnomasi hamda mavzuga oid boshqa me‘yoriy-huquqiy hujjatlarda belgilangan vazifalarni amalga oshirishda ushbu dissertatsiya tadqiqoti muayyan darajada xizmat qiladi.

Tadqiqotning respublika fan va texnologiyalari rivojlanishining ustuvor yo‘nalishlariga mosligi. Tadqiqot respublika fan va texnologiyalari rivojlanishining I. “Axborotlashgan jamiyat va demokratik davlatni ijtimoiy, huquqiy, iqtisodiy, madaniy, ma’naviy-ma’rifiy rivojlantirish, innovatsion iqtisodiyotni rivojlantirish” ustuvor yo‘nalishi doirasida bajarilgan.

Muammoning o‘rganilganlik darajasi. Tarjimani avtomatlashtirishda parallel korpus kontekstual qidiruv tizimi ko‘plab tilshunoslarning tadqiqot obyektiga aylanganligi bejiz emas. Buni so‘nggi yillarda tarjima sohasiga bo‘lgan katta e‘tibor va qiziqish, avtomatlashtirilgan tarjima dasturlarining rivojlanishi bilan tushuntirish mumkin.

Jahon tilshunosligida tarjimaning avtomatlashtirilgan xususiyatlari, mashina tarjimasining yaratilishi, uni yuritish tizimlari, dasturlari, mashina va inson tarjimasi, o‘ziga xos metodlari hamda METEO, ALPS, ATLAS, SMARTCAT, SYSTERON, LINGUEE, NTRON dasturlarining yaratilishi va ularning o‘ziga xos funksional imkoniyatlari bir qancha ingliz hamda rus tilshunos olimlar ishlarida o‘z ifodasini topgan. Dastlab G.V.Leybnits va R.Deskart² tomonidan jumla hamda so‘zlarning o‘zaro bog‘liqligini kodlash masalasi ilgari surilgan. Shuningdek, mashina tarjimasining kelib chiqishi, ishlash tizimi, tarjimaning asosiy xususiyatlari, mashina va inson tarjimasining shakllanish yo‘li hamda tarjima dasturlarining yaratilish tadqiqotlari J.S.Zager, J.A.Alonso, P.Benett, R.T.Bell, R.Jakobson, J.B.Larj, V.Lavson kabi ingliz olimlari tomonidan amalga oshirilgan³. Tabiiy tilga ishlov berish muammolari A.Keyler, K.V.Linden, N.Vard⁴,

² Deskart R. History of machine translation. – Paris: Pergamon Press, 1983 // Article for translators: History of machine translation (translationdirectory.com).

³ Sager J.C. Language engineering and translation consequences of automation. – Philadelphia: John Benjamin’s publishing company, 1994. – P. 243-264; Alonso J.A. “Machine translation technology: on the way of market introduction” // Siemens review, 1987. – P. 54-60; Bennett P. The translation unit in humans and machines. – Manchester: CCL/UMIST, 1993. – P. 41-67; Bell R.T. Translation and translating, theory and practice. – London: Longman press, 1991. – P. 32-54. Jacobson R. “On linguistic aspects of translation”. R.A.Brower (Ed). – London, 1966. – P. 62-68; Large J.B. The foreign language barrier. – London, 1983. – P. 102; Lawson V. Practical experience of machine translation. – Amsterdam: NorthHolland press, 1982. – P. 105.

⁴ Jurafsky D.S. and. Martin J.H. Speech and language processing. Contributing writers: Prentice Hall, Englewood cliffs. – New Jersey, 2000. – P. 950.

M.A.Mohri⁵ tomonidan keng yoritilgan. Rus olimlaridan K.K.Boyarskiy kompyuter lingvistikasi haqida umumiy tushuncha, M.V.Xoxlova mashina tarjimasining asosiy tizimlari va ushbu dasturlarning kamchiliklari, I.G.Ovchinnikova kompyuter tarjimasini usullaridan samarali foydalanish va tarjimani avtomatlashtirishning yangi imkoniyatlari, P.V.Xonkina mashina tarjimasida idioma tarjimasining qo'llanilishi, L.I.Chovichning ingliz va rus tillarida matn tarjimasining xususiyatlarini tadqiqot ishlarida o'z aksini topgan⁶.

O'zbekistonda ham tarjimaning avtomatlashtirilgan nazariyasi va mashinaviy tarjima dasturlarining ishlash tizimlari bir qancha olimlar tomonidan o'rganilgan. Jumladan, B.Mengliyev hamda D.Ahmedovalar⁷ korpus tilshunosligidagi semantik teg toifalari hamda onlayn korpuslarni leksik-grammatik, leksik-mavzuviy va uslubiy baholash xarakteristikasini ochib berishgan. G.Sobirova BNC korpusini tillarni o'qitishda qo'llashga e'tibor qaratgan. J.Jumabayeva tildagi graduonimiyalarni ingliz tili korpusi yordamida o'rganib, semantik xususiyatlarini aniqlagan. Mualliflik korpusini yaratishning lingvistik asoslari Sh.Xamroyeva⁸ tomonidan tadqiq etilgan. N.Atoboyev korpuslar tasnifini maxsus prinsiplarga asoslangan holda keltirish mumkinligi va ularni ma'lum ma'noda qo'llash sohalariiga ko'ra turkumlash shartli ravishda yakuniy shaklga keltirilganligini o'z ilmiy tadqiqotida keltirib o'tgan⁹. N.Abdurahmonova¹⁰ tadqiqotida esa mashina tarjimasining lingvistik taminot yaratish masalalari yoritilgan. R.Karimov¹¹ o'zbek-ingliz parallel korpusini tuzishning lingvistik va dasturiy masalalarini, parallel korpuslar funksionalining obyektivligi va adekvat mohiyatini ta'minlashda aslyatdan tarjima tiliga bevosita o'girilgan matnlarni janrlarga muvofiq tanlash strategiyasi va matn elementlarini parallel teglashda tarjima transformatsiyalari chegaralarini belgilash ustuvorlik qilishini ilmiy dalillagan.

Tadqiqotning dissertatsiya bajarilgan oliy ta'lim muassasasining ilmiy-tadqiqot ishlari rejalari bilan bog'liqligi. Dissertatsiya tadqiqoti O'zbekiston davlat jahon tillari universitetining ilmiy-tadqiqot ishlari rejasiga muvofiq № I-204-4-5 «Ingliz tili mutaxassislik predmetlaridan axborot-kommunikatsion texnologiyalari asosida virtual resurslar yaratish va o'quv jarayoniga joriy etish» mavzusidagi innovatsion loyiha doirasida bajarilgan.

⁵ Mohri M.A. Finite-state transducers in language and speech processing. Computational linguistics. – New Mexico, 1994. – P. 269-312.

⁶ Боярский К.К. Введение в компьютерную лингвистику. – Санкт-Петербург, 2013. – С. 4-5; Холхова М.В. Анализ особенности машинного перевода. – Санкт-Петербург, 2016. – С. 7-12; Овчинникова И.Г. Использование компьютерных переводческих инструментов: новые возможности новые ошибки. Russian journal of linguistics; Хонькина П.В. Перевод идиоматических выражений в системах машинного перевода // Вестник московского университета. – Москва, 2010. – С. 68-71; Чович Л.И. Перевод как интерпретация и импровизация на материале сербских переводов мемуаров // Вестник московского университета. – Москва, 2012. – С. 68-71.

⁷ Mengliyev B., Ahmedova D. Semantic Tag Categories in Corpus Linguistics: Experience and Examination // International Journal of recent technology and engineering (IJRTE) ISSN: 2277-3878, 2019.

⁸ Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Филол. фан. бўйича фалсафа докт. (PhD) ... дисс. автореф. – Қарши, 2018. – 52 б.

⁹ Атабоев Н.Б. Инглиз тили корпусининг функционал хусусиятлари (СОСА мисолида). Филол. фан. докт. ... дисс. – Тошкент, 2020. – 14 б.

¹⁰ Абдурахмонова Н. Машина таржимасининг лингвистик асослари. – Т.: «Akademnashr», 2012. – 25 б.

¹¹ Каримов Р.А. Ўзбек-инглиз параллел корпусини тузишнинг лингвистик ва дастурий масалалари Филол. фан. докт. ... дисс. – Тошкент, 2022. – 7 б.

Tadqiqotning maqsadi tarjimani avtomatlashtirishda parallel korpus kontekstual qidiruv tizimining funksional imkoniyatlarini ingliz va o‘zbek tillari misolida (Sketch engine platformasi orqali) ochib berishdan iborat.

Tadqiqotning vazifalari:

tarjimani avtomatlashtirishda til nazariyasi, amaliy va kompyuter lingvistikasi sohasi nazariy asoslarini tadqiq etish;

ingliz va o‘zbek tillaridagi ilmiy hamda publitsistik matnlarning leksik xususiyatlarini izohlash va tarjimada kontekstual tahlil qilish;

parallel matnlar korpusi asosida ingliz va o‘zbek tillaridagi ilmiy hamda publitsistik matnlarning tarjima jarayonida qo‘llaniladigan usullarini aniqlash;

ilmiy hamda publitsistik uslubga xos mantlarda qo‘llanilgan desemantlashgan so‘zlarning leksik ma’nosi va kontekstual tarjimasini tahlilga tortish;

kontekstual tarjimon dasturlarini yaratishda korpus turlarining mohiyatini o‘rganish;

tarjima amaliyotida parallel korpusdan samarali foydalanish usullarini aniqlash va Sketch engine platformasining til korpuslarini yaratishdagi ahamiyatini ochib berish;

Linguee va Reverso kontekstual tarjimon dasturlarining mushtarak hamda farqli jihatlarini ochib berish asosida tarjimani avtomatlashtirishda parallel matnlar korpusining funksional vazifalarini aniqlash;

Sketch engine platformasida til korpusi va tarjimon dasturi yaratish bosqishlarini o‘rganish hamda ushbu platformada ingliz hamda o‘zbek tillari matnlarini parallel korpusga joylashtirish orqali kontekstual tarjimon dasturini yaratish.

Tadqiqotning obyekti sifatida ingliz va o‘zbek tillaridagi parallel korpus kontekstual tizimi olingan.

Tadqiqotning predmetini ingliz va o‘zbek tillari parallel korpusi kontekstual tizimining tarjima xususiyatlari va funksional aspekti tashkil qiladi.

Tadqiqotning usullari. Tadqiqotda empirik tahlil, kuzatuv tahlili, korpus tahlili, tavsiflash, tasniflash, kontekstual tahlil va komponent tahlil usullaridan foydalanilgan.

Tadqiqotning ilmiy yangiligi quyidagilardan iborat:

ingliz va o‘zbek tillarida tarjimani avtomatlashtirishga oid kontekstual dasturni yaratishda mashina tarjimasining Statistik (Statistic machine translation) va Neyron (Neyron machine translation) tizimlarining tabiiy tilni tushunish, uni avtokorrektor bilan tekshirish hamda yirik ma’lumotlar bazasidan keng foydalanish imkoniyatlari HMT (Hybrid machine translation), RBMT (Rule based machine translation) mashina tarjima tizimlaridan funkcionallik jihatidan samaradorligi asoslangan;

ilmiy va publitsistik matnlarda tez-tez uchraydigan desemantlashgan so‘zlarning kontekstual ma’nosining tarjimani avtomatlashtirishda mashina tarjimasining Rule based tizimi hamda dasturiy algoritmlarni qo‘llanilishini asoslash orqali o‘zbek tili parallel korpusini Sketch engine platformasiga kiritish mexanizmi CATning Trados tarjima xotirasi funksiyalarini joriy qilish orqali ishlab chiqilgan;

Sketch engine platformasida ingliz va o'zbek tillarida ilmiy va publitsistik matnlardagi kollokatsiyalar kontekstual tarjimon dasturini parallel korpusda yaratish bosqichlari *Adhog, Diy, comparable, monolingval, diaxronik, sinxron va parallel* korpusda so'z, ibora, kollokatsiya, idiomatik birlik, jumla va matn tanlash, uni elektron tizimga aylantirish, kontekstual qidiruv tizimini yaratish hamda dasturga yuklash bosqichlarini amalga oshirish orqali takomillashtirilgan;

avtomatlashtirilgan tarjimon dasturlariga oid umumiy talablarni amalda qo'llashning tizimli usullari parallel korpusda yaratishni takomillashtirish orqali Sketch engine platformasida ingliz va o'zbek tillarida kontekstual tarjimon dasturiga oid ilmiy va publitsistik matnlardan tashkil topgan elektron namunaviy baza yaratish bo'yicha eksperimental amaliy tavsiyalar ishlab chiqilgan.

Tadqiqotning amaliy natijalari quyidagilardan iborat:

tarjimani avtomatlashtirishda korpus turlarining metodlari, vazifalari va resurslari takomillashtirilgan;

parallel korpus asosida ilmiy va publitsistik matnlardan iborat bo'lgan kontekstual tarjima dasturining ixchamlashtirilgan bazasi shakllantirilgan;

Sketch engine platformasidan foydalangan holda ilmiy va publitsistik matnlarning ikki tilli (inglizcha-o'zbekcha) parallel korpusda kontekstual tarjimon dasturi mexanizmi yaratilgan;

ingliz va o'zbek tillaridagi ilmiy hamda publitsistik matnlarga oid kollokatsiyalarning ma'lumotlar bazasi ishlab chiqilgan.

Tadqiqot natijalarining ishonchliligi o'rganilgan manba materiallarning tamoyillari va undagi qo'llanilgan usullar, nazariy ma'lumotlarning rasmiy manbalardan olinganligi ingliz, rus va o'zbek tillari tabiatidan kelib chiqqan holda xulosalar qilishga yordam bergani, ularning asosli ekanligi, xorijiy olimlarning nazariy qarashlarining tadqiqot ishida qamrab olingani, keltirilgan tahlillar, korpus turlari, tahlili va metodlari, tarjimani avtomatlashtirishda qo'llaniladigan bir qancha usullarning ishonchliligi, xulosa taklif va tavsiyalarining amaliyotda joriy qilingani, olingan natijalarning amalda isbotlangan manbalarga tayangani bilan izohlanadi.

Tadqiqot natijalarining ilmiy va amaliy ahamiyati. Tadqiqot natijalarining ilmiy ahamiyati parallel ilmiy matnlar korpusi kontekstual tarjimon dasturini yaratilishi va uni mukammallashtirishning nazariy asoslarini ishlab chiqilishi tarjimashunoslik, korpus lingvistikasi hamda mashina tarjmasi yo'nalishida tadqiqotlar olib borishda hamda kontekstual tarjimon dasturlari uchun ingliz tilidan o'zbek tiliga, tarjima qilish metodlarini tadbiiq qilinishi tarjimashunoslik sohasiga doir metodologik nazariyalarni tahlil qilish bilan belgilanadi.

Tadqiqot natijalarining amaliy ahamiyati til nazariyasi, amaliy tilshunoslik, kompyuter lingvistikasi, korpus lingvistikasi, mashina tarjimasini fanlari o'qitilish jarayonida dastur va reja tuzish, mavzularni bayon etishda manba vazifasini o'tash, tarjimaning avtomatlashtirilgan dasturlaridan keng foydalanish, shuningdek, tarjima amaliyotida kontekstual ikki tilda (ingliz-o'zbek tillarida) elektron lug'at tuzish va yaratish bilan izohlanadi.

Tadqiqot natijalarining joriy qilinishi. Tarjimani avtomatlashtirishda parallel korpus kontekstual qidiruv tizimining funksional mohiyatini ochib berish bo'yicha olingan ilmiy natijalar va amaliy takliflar asosida:

ingliz va o'zbek tillarida tarjimani avtomatlashtirishga oid kontekstual dasturni yaratishda mashina tarjimasining Statistik (Statistic machine translation) va Neyron (Neyron machine translation) tizimlarining tabiiy tilni tushunish, uni avtokorrektor bilan tekshirish hamda yirik ma'lumotlar bazasidan keng foydalanish imkoniyatlari HMT (Hybrid machine translation), RBMT (Rule based machine translation) mashina tarjima tizimlaridan funkcionallik jihatidan samaradorligiga oid xulosalardan universitetda bajarilgan Erasmus+dasturining ICM, KA-107 – International Credit Mobility between European Union and Uzbekistan HEIs, through staff and student mobility exchange projects between Uzbekistan State University of World Languages and Ruhr University Bochum (Germany) – 2020-1-DE01-KA107-005433 (Yevropa Ittifoqi hamda O'zbekiston Oliy Ta'lim muassasalari o'rtasida xalqaro tajriba almashinuvi doirasida O'zbekiston davlat jahon tillari universiteti hamda Boxumdagi Rur universiteti (Germaniya) o'rtasidagi kredit safarbarligi) loyihasida foydalanilgan (O'zbekiston davlat jahon tillari universitetining 2023-yil 17-iyul № IRD/23-156-sonli ma'lumotnomasi). Natijada, ushbu dissertatsion ishda tadqiq etilgan kontekstual tarjimon dasturining kichik bazasidan loyiha doirasida qo'llanilgan;

ilmiy va publitsistik matnlarda tez-tez uchraydigan desemantlashgan so'zlarning kontekstual ma'nosining tarjimani avtomatlashtirishda mashina tarjimasining Rule based tizimi hamda dasturiy algoritmlarni qo'llanilishini asoslash orqali o'zbek tili parallel korpusini Sketch engine platformasiga kiritish mexanizmi CATning Trados tarjima xotirasi funksiyalarini joriy qilishga oid tavsiyalardan Erasmus+dasturining ICM, KA-107 – International Credit Mobility between European Union and Uzbekistan HEIs, through staff and student mobility exchange projects between Uzbekistan State University of World Languages and Ruhr University Bochum (Germany) – 2020-1-DE01-KA107-005433 (Yevropa Ittifoqi hamda O'zbekiston Oliy Ta'lim muassasalari o'rtasida xalqaro tajriba almashinuvi doirasida O'zbekiston davlat jahon tillari universiteti hamda Boxumdagi Rur universiteti (Germaniya) o'rtasidagi kredit safarbarligi) loyihasida amaliy foydalanilgan (O'zbekiston davlat jahon tillari universitetining 2023-yil 3-oktabr № 06/556-sonli ma'lumotnomasi). Natijada, yuqorida qayd etilgan korpus turlaridan dastur doirasida tajriba sifatida foydalanilgan;

avtomatlashtirilgan tarjimon dasturlariga oid umumiy talablarni amalda qo'llashning tizimli usullari parallel korpusda yaratishni takomillashtirish orqali Sketch engine platformasida ingliz va o'zbek tillarida kontekstual tarjimon dasturiga oid ilmiy va publitsistik matnlardan tashkil topgan elektron namunaviy baza yaratish bo'yicha eksperimental amaliy tavsiyalardan, shuningdek parallel korpusda Sketch engine platformasida ingliz va o'zbek tillarida kontekstual tarjimon dasturiga oid ilmiy va publitsistik matnlardan tashkil topgan elektron namunaviy bazasidan yoshlarni tarjimaga bo'lgan qiziqishlaridan kelib chiqqan holda ilmiy hamda publitsistik matnlarni parallel korpusda tarjima qilishga oid tanlovlarini o'tkazish chora-tadbirlarini amalga oshirishda O'zbekiston

Yozuvchilar uyushmasining xalqimizning qadimiy tarixi va boy madaniyatini, el-yurtimizning fidokorona mehnati bilan erishilayotgan tub o'zgarishlarni teran aks ettirishga foydalanilgan (O'zbekiston Yozuvchilar uyushmasining 2023-yil 28-noyabrdagi 01-03/3150-ma'lumotnomasi). Natijada, Sketch engine platformasida ilk marotaba ingliz va o'zbek tillari parallel korpusi kontekstual tarjimon dasturi mexanizmini shakllantirish bo'yicha taklif va tavsiyalar ishlab chiqilganligi borasidagi tadqiqotchining umumlashma xulosalari yosh yozuvchilar uchun ilmiy hamda publitsistik matnlarni tarjima qilish uchun amaliy-nazariy manba sifatida xizmat qilgan.

Tadqiqot natijalarining aprobatsiyasi. Tadqiqot natijalari bo'yicha 10 ta ilmiy maqola chop etilgan bo'lib, shundan 4 ta xalqaro va 6 ta respublika ilmiy-amaliy anjumanlarda qilingan ma'ruzalarda jamoatchilik muhokamasidan o'tkazilgan.

Tadqiqot natijalarining e'lon qilinganligi. Dissertatsiya mavzusi bo'yicha jami 10 ta ilmiy ish nashr ettirilgan, jumladan, O'zbekiston Respublikasi Oliy attestatsiya komissiyasining doktorlik (PhD) dissertatsiyalari asosiy ilmiy natijalari chop etish tavsiya qilingan ilmiy nashrlarda 5 ta maqola ulardan 4 tasi respublika va 1 tasi xorijiy jurnalda e'lon qilingan.

Dissertatsiya hajmi va tuzilishi. Dissertatsiya kirish, 3 bob, xulosa, foydalanilgan adabiyotlar ro'yxatidan iborat. Tadqiqot hajmi 178 sahifani tashkil etadi.

DISSERTATSIYANING ASOSIY MAZMUNI

Kirish qismida ilmiy tadqiqotning mavzusi va uning dolzarbligi asoslangan bo‘lib, tadqiqot ishining maqsadi va vazifalari, obyekti va predmeti aniqlangan, ilmiy ishning fan va texnologiyalarni rivojlantirishning muhim yo‘nalishlarga mosligi ko‘rsatib o‘tilgan. Shu bilan birga, dissertatsiyaning ilmiy yangiligi, amaliy ahamiyati, erishilgan natijalarining amaliyotga joriy etilishi, ilmiy nashrlarda e‘lon qilinganligi, ishning tuzilishi borasida ma‘lumotlar keltirilgan.

Dissertatsiyaning “**Amaliy tilshinoslikda tarjimani avtomatlashtirish konsepsiyasi**” deb nomlangan birinchi bobida mashina tarjimasining tarixi, etimologiyasi, ixtiro qilinish bosqichlari, tarjimani avtomatlashtirishning tasniflash muammosiga oid nazariy tahlillar hamda mashina tarjimasini mexanizmlarining funksional bosqichlari tahlilga tortilgan.

Etimologik jihatdan “mashina tarjimasini” atamasi dastlab faqatgina inson aralashuvisiz ishlaydigan avtomatik tizimlarni belgilash uchun qo‘llanilgan. Yevropa mashina tarjimasini uyushmasi (The European Association for Machine Translation EAMT)¹² mashina tarjimasini “matnni asl tildan maqsadli tilga o‘tkazish uchun kompyuterdan foydalanish” sifatida izohlaydi. O‘z navbatida, xalqaro mashina tarjimasini assotsiatsiyasi (International Association for Machine Translation IAMT)¹³ mashina tarjimasini “bir vaqtning o‘zida to‘liq jumlaning kiritish orqali tarjima variantini hosil qilish” deb hisoblagan¹⁴.

A.V.Semenovga ko‘ra, “mashinali tarjima tizimlarining asosiy maqsadi manba tilini maqsadli tilning ekvivalent matniga yuqori sifatli tarjima qilishdir. Biroq avtomatlashtirilgan tarjima tizimi tomonidan amalga oshirilgan tarjima mukammal emas, shuningdek, tarjima qilingan matn keyingi tahrirni talab qiladi. Ba‘zan tarjima qilingan matnlarga tarjimaning aniqligi, taqdimotning o‘ziga xos uslubi uchun yuqori talablar qo‘yilmaydi, asosan, qilingan tarjima umumiy “*mazmun*”ni o‘quvchiga “*hisobot*” tariqasida taqdim qilish bilan kifoyalanadi. Natijada, tarjima qilingan “*xom*” matn qabul qilinadi”¹⁵.

Ch.Bebbij, asosan, XIX asrning o‘rtalarida raqamli analitik mashina loyihasi ustida ishlagan va ushbu raqamli analitik mashinalar loyihasini 1855-yilda Parijdagi ko‘rgazmada omma oldiga taqdim qilgan. Ch.Bebbij bunday mashinadan lug‘atlarni saqlash uchun ham foydalanish mumkin degan xulosaga kelgan¹⁶.

1933-yilda lingvistik arifmometr muallifi P.Smirnov-Troyanskiy Sobiq Rossiya Fanlar akademiyasiga o‘zi ishlab chiqqan “bir tildan ikkinchi tilga

¹² EAMT – The European Association for Machine Translation // <https://eamt.org/>

¹³ IAMT – International Association for Machine Translation.

¹⁴ Семенова А.В., Курейчик В.М. Обзор и анализ состояния проблемы обработки текстовой информации в системах машинного перевода / Информатика, вычислительная техника и инженерное образование. – М., 2014. – С. 2.

¹⁵ Семенова А.В., Курейчик В.М. Обзор и анализ состояния проблемы обработки текстовой информации в системах машинного перевода / Информатика, вычислительная техника и инженерное образование. – М., 2014. – С. 2-3.

¹⁶ Bromley A.G. “Charles Babbage’s Analytical Engine, 1838”. Annals of the history of computing. – Cambridge-Massachusetts: MIT Press, 1990. – P. 196-217.

tarjima qilishda soʻzlarni tanlash va chop etish mashinasi” loyihasi bilan murojaat qilgan¹⁷. Shuningdek, P.P.Smirnov-Troyanskiy tomonidan *Esperanto* tili asosida tillar oʻrtasidagi grammatik boshqaruv usuli bilan taqsimlash metodi va bilingval lugʻatni oʻz ichiga oluvchi tizim 1933-yilning soʻngi choragida yaratildi. Ushbu tizim uch bosqichni oʻz ichiga olgan: *birinchi bosqichda* muharrir asosiy tildagi soʻzlarni tartibga solib, soʻz birliklarning mantiqiy shakllarni hamda sintaktik funksiyalarini belgilaydi; *ikkinchi bosqichda* mashina leksik shakllarni tarjima qilinayotgan tilga oʻgiradi; *uchinchi bosqichda* muharrir natijani meʼyorlashtiradi¹⁸.

Mashina tarjimasi 1946-yilga kelib ilmiy fan sifatida shakllandi. Bunga Rokfeller fondining Rokfeller ilmiy departamenti direktori U.Uiver, E.But va N.Viner bilan yozishmalarda birinchi marta mashina tarjimasi tushunchasini asos qilib koʻrsatilgan¹⁹. Shu bilan birga, ushbu tadqiqotlar mashina tarjimasiga boʻlgan yangi bosqichni boshlab berdi.

I.A.Melchuk va Yu.Apresyanlarning metatil bu tilga oid tushunchalarning modeli, inson psixikasida tushunchalarni yozib oluvchi va saqlovchi gipotetik til ekanligi oʻz qarashlarida ifodasini topgan²⁰. Aynan shu yillar mobaynida matematik hamda mexaniklardan tarkib topgan mashina tarjimasi uchun eksperimental laboratoriya Rossiyaning Leningraddagi ilmiy tadqiqot institutida tashkil etildi.

Dunyoda mashina tarjimasi tizimlarini rivojlanishiga ikkita asosiy omil mavjud: 1) sof ilmiy; bu kompyuter tarjimasini modellashtirishning murakkabligi va koʻp funktsionalligi bilan belgilanadi; 2) mashina tarjimasining ijtimoiy maqsadda tillararo kommunikatsiyani taʼminlashga boʻlgan imkoniyatning kengligi hisoblanib, uning hajmi har yili ortib borayotgan tillararo muloqotning zaruriy sharti sifatida zamonaviy dunyoda mashina tarjimasi amaliyotining oʻsib borayotgan roli bilan bogʻliq.

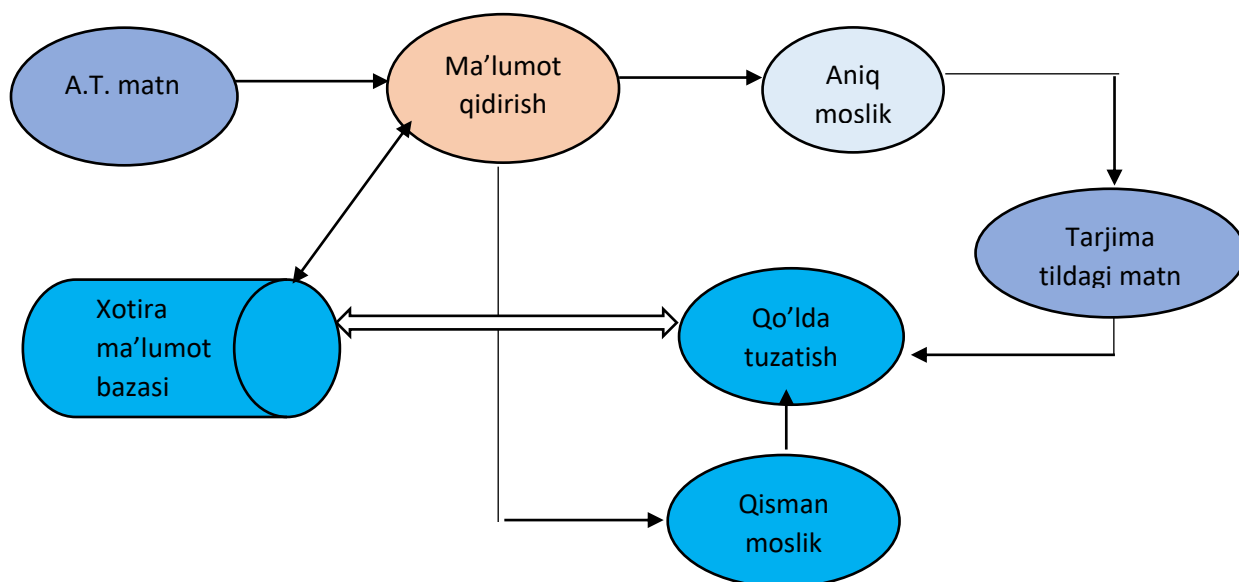
Olimlarning MT tizimlari ustidan tinmay tadqiqot olib borishlari natijasida mashina tarjimasini sunʼiy intellekt sinfiga kirishiga muvaffaq boʻlishdi. Zotan, mashina tarjimasi birikmasidagi mashina soʻzi ***sunʼiy intellekt*** maʼnosida insonga xos xususiyatlar: oʻqish, tekshirish, tarjima va tahrir qilish jarayonlarining bajaruvchisi sifatida qaraladi. Kompyuter tarjimasining asosiy qismlaridan biri bu xotira maʼlumotlar bazasining mavjudligidir. Shuningdek, CAT, kompyuter yordamida tarjima qilish, mashina yordamida tarjima qilish (MAT machine-assisted translation) yoki inson tomonidan avtomatlashtirilgan tarjima MAHT (machine-aided human translation) deb ham nomlanadi. Quyidagi koʻrsatilgan rasmda CATning ishlash mexanizmi koʻrsatilgan:

¹⁷ Колосёнок П. Системы машинного перевода. // Системы машинного перевода — Lingva (msu.ru) – P. 2019.

¹⁸ Troyanskii P.P. A forgotten pioneer of mechanical translation. Machine translation. – Moscow, 2000. – P. 187-221.

¹⁹ Norbert W.C. Control and communication in the animal and the machine. – Cambridge, 1961. – P. 94-95 // https://ru.Cybernetics:_Or_Control_and_Communication_in_the_Animal_and_the_Machine.

²⁰ Апресян Ю.Д. О языке для описания значений слов. – Москва: Известия АН СССР. Серия языка и литературы, 1966. – С. 415-429.



1-rasm. Kompyuter yordamida tarjimaning ishlash mexanizmi

Neyron mashina tarjimasi NMT (Neural Machine Translation) – bu mashina tarjimasining eng yangi bosqichi sun’iy intellektni rivojlantirishdan iborat. Uning nazariyasi tabiiy tilni tushunish (NLU – natural language understanding), tabiiy tilni qayta ishlash (NLP – natural language processing), mashina tarjimasi MT tarjima xotirasi va statistikaga asoslangan mashina tarjimasi (SMT – statistics-based machine translation) hamda chuqur o’rganish nazariyasi va texnikasiga asoslanadi²¹.

Hozirda MTning, asosan, to’rt turi mavjud bo’lib, ular:

1) elektron lug’atlar yordamida tarjima (machine-assisted human translation – MAHT) – inson tomonidan matnni tarjima qilishda kompyuter elektron lug’atlar va unga kiritilgan ko’rsatmalaridan foydalanish;

2) kompyuter va dasturiy ilovalar yordamidagi tarjima (computer assisted-translation – CAT) – matnni kompyuter vositasida tarjima qilish va unda anglashilgan asosiy tushunchalarni ajratishga xizmat qiladi;

3) muharrir yordamidagi mashina tarjimasi (human-assisted machine translation – HAMT) kompyuter dasturiy ta’minoti orqali amalga oshiriladigan tarjimaga inson resursi tomonidan tahrir qilish vazifasini yuklash;

4) to’liq avtomatik mashina tarjimasi (fully automatic machine translation – FAMT) – tarjimon dasturi orqali matnni tarjima hamda tahrir qilish vazifalarini amalga oshirish.

Mashina tarjima tizimlari muayyan jihatlarga ko’ra xarakterlanadi:

1) *Tillarning ishtiroki: a) ikki tilli; b) ko’p tilli; 2) tanlangan mavzuning miqdori: a) muayyan uslub va sohaga aloqador; b) bir uslub hamda ko’p sohalarga tegishli; 3) turli uslub va bir necha sohalar bilan bog’liq*²².

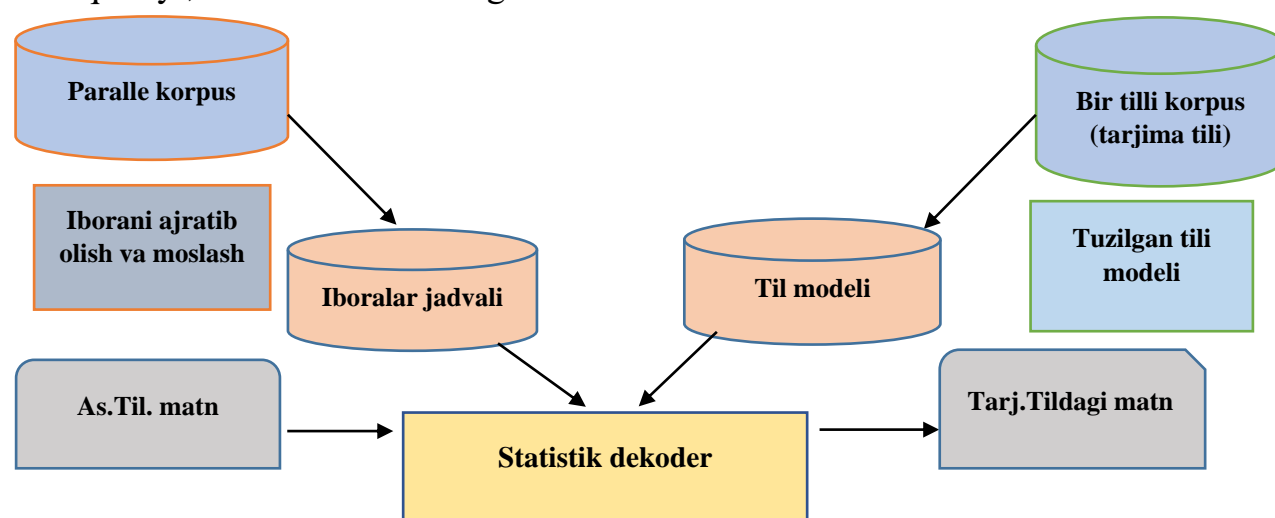
²¹ Halliday M. Linguistics and machine translation. In J.Webster computational and quantitative studies. Beijing: Peking University Press. – China, 2007. – P. 166.

²² Галактионов В.А., Мусатов А.М. Система машинного перевода “Кросслятор 2.0” и анализ ее функциональности для задачи трансляции знаний. – Москва, 2007. – С. 13.

Hind olimi R.Sankaravelayutan va J.Xatchins mashina tarjimasini quyidagicha ajratadi²³:

1. Qoidaga asoslangan MT – (Rule-Based MT): 1) transfer tizim (transfer-based machine translation); 2) interlingval tizim (interlingual machine translation); 3) lug‘at asosidagi tizim (dictionary-based machine translation); 2. Korpus asosidagi mashina tarjimasini – (Corpus based MT: 1) statistikaga asoslangan MT (statistics-based); 2) namuna asosidagi MT (example-based); 3. Gibrid mashina tarjimasini (Hybrid machine translation HMT). Gibrid mashina tarjimasini qoida hamda statistikaga asoslangan tarjima texnologiyasidan foydalanuvchi tizimdir. Hozirda bir nechta mashina tarjimasini kompaniyalari (Asia Online, LinguaSys, Systran, PangeaMT, UPV, Logos) mana shu tizimdan foydalanadi²⁴.

Statistik mashina tarjimasining g‘oyasi, asosan, axborot nazariyasidan kelib chiqadi. Statistik tarjima tizimlari bir vaqtning o‘zida ikkita til versiyasi (parallel korpus) tomonidan taqdim etilgan matnlar massivlaridan foydalanish bilan tavsiflanadi. A.B.Smenovaga²⁵ ko‘ra, ushbu model kelajakda hujjatni tarjima qilishda tegishli chastotalarni hisobga olgan holda turli tillarda eng ko‘p ishlatiladigan so‘zlardan foydalanish orqali tarjimini yaxshilash imkoniyatlarini beradi. Statistik mashina tarjimasini texnologiyasiga asoslangan mashinali tarjima tizimlarini ishlab chiquvchilari *Google, Microsoft, Language Weaver, Yandex* va boshqa sayt, dasturlarni o‘z ichiga oladi.



2-rasm. Statistik mashina tarjima tizimlarining tipik arxitekturasini PROMT kompaniyasi

Tadqiqot jarayonida MT tizimlarining bosqichma-bosqich rivojlanish strategiyalarini ilmiy nazariy jihatdan tahlil qilindi. Shunday ekan, yillar davomida MT vazifasiga bir nechta yondashuvlar va tizimlar rivojlanishini kuzatish mumkin.

²³ Rajendran S. Prof. English to Tamil Machine Translation System Using Parallel Corpus. Amrita Universiy – Coimbatore, 2019. – P. 45.

²⁴ Hutchins J. Machine translation: past, present, future. – London, 1986. – P. 118.

²⁵ Семенова А.В., Курейчик В.М. Обзор и анализ состояния проблемы обработки текстовой информации в системах машинного перевода / Информатика, вычислительная техника и инженерное образование. – М., 2014. – С. 9.

Bugungi kunda eng rivojlangan tizimlar qoidaga asoslangan MT, (RBMT), neyron MT tizimlari va statistik MTlar hisoblanadi. Garchi ular qisman bir-biriga mos keladigan tizim sifatida tushunilsada, umumiy nuqtai nazarga ko'ra, ham statistik, ham lingvistik tizimli yondashuvlar birlashishi oqibatida MT yanada rivojlanadi. Natijada esa dasturlash va kompyuter texnologiyalarining yangi imkoniyatlari ham mashina tarjimai nazariyasi hamda amaliyotini takomillashtirish imkoniyatiga ega bo'ladi.

Tadqiqotning **“Tarjimani avtomatlashtirishda korpus lingvistikasining o‘rni”** deb nomlangan ikkinchi bobida kontekstual tarjimon dasturlarida korpus turlarining o‘rni va ularning qo‘llanilish konsepsiyasi, KL metodologiyasi, korpus turlarining tahlil metodlari va tarjima amaliyotida parallel korpus resurslaridan foydalanish imkoniyatlari, shuningdek, Linguee va Reverso kontekst tarjimon dasturlarining funksional xususiyatlari tadqiq qilingan.

Hozirda korpusning bir qancha turlari yaratilgan bo‘lib ular tarjima sohasida turli xil funksiyalar orqali qo‘llaniladi. Masalan *parallel korpus*, *monolingual (bir tilli)*, *bilingual (ikki tilli)*, *general (umumiy)*, *specialized (maxsus, ixrisoslashtirilgan)*, *Ad hog*, *DIY korpuslari* kabi korpuslar funksional jihatdan boshqa turdagi korpuslardan o‘zaro farq qiladi.

M.Beyker²⁶, G.Turi, M.Shlesinger, K.Malmkjaer²⁷ va boshqa olimlar tarjimaning imkoniyatlarini o‘rganish uchun parallel korpusdan foydalanish lozimligi haqidagi g‘oyani ilgari surgan va tarjimaning universalligi haqidagi quyidagi gipotezani belgilagan: a) *parallel korpusni soddalashtirish*, bunda tarjimonlar til va ma’lumotni ongsiz ravishda soddalashtiradi; b) *parallel korpusni oson tushuntirish*, bunda tarjimonlar tarjimada mazmunni yanada aniqroq qilishga intiladi; v) *parallel korpusni me‘yorlashtirish*, bunda tarjima periferiyaga emas, balki oraliq bo‘lishga intiladi.

Parallel korpusni yaratadigan tizimlar qisman kerakli misollarni emas, balki parallel jumalarni hattoki butun paragraflarni topa olish imkoniyatiga ega. Misol qilib, *“ParaConc”* parallel korpus qidiruv tizimini tavsiflash mumkin, ushbu parallel korpus so‘zlarning chastotasi va ularning birikmalari haqidagi ma’lumotlarni alohida ko‘rsatadi. Kontekstlarni tartiblovchi tizimlar tarjima holatlari statistikasidan foydalanadi. Bunday tizimlar odatda leksikograflar va professional tarjimonlar uchun mo‘ljallangan. Bunga misol qilib parallel *ingliz-xitoy* korporatsiyasini qidiradigan ikkita vebga asoslangan *DOMCAT* va *TotalRecall* tizimi yaratilgan. *DOMCAT* tizimida *Dyes* statistik testi yordamida eng mos kontekstlarni topadi (*Sorensen-Dyes koeffitsiyenti*).

$$QS = \frac{2 | X \cap Y |}{|X| + |Y|},$$

²⁶ Baker M. Corpora in translation studies: An Overview and Some Suggestions for Future Research. Target. – Amsterdam, 1995. – P. 223-243.

²⁷ Malmkjaer K. Linguistics and the language of translation. – USA, 2005.

$|X|$ va $|Y|$ mos ravishda soʻzning chastotalari, tarjima ekvivalenti hamda $|X \cap Y|$ – ushbu soʻzlarning qoʻshma paydo boʻlish chastotasini belgilaydi.

Biroq veb-interfeys bazasi reyting va tizim qanday ishlashi haqida hech qanday maʼlumotlarni oʻzida jamlamaydi.

Ikkinchi tizim *TotalRecall*-ushbu tizim ikkita imkoniyatni taqdim etadi – chastota boʻyicha va *log-ehtimollik* nisbati (log-likelihood ratio, LLR) boʻyicha tartiblash. Ushbu ikki hodisa, masalan, soʻzning paydo boʻlishi va uning tarjimasi oʻrtasidagi munosabatlarga nisbatan oxirgi oʻlchov T.Dunning²⁸ nazariyasi sifatida hisoblanadi.

$$LL(a, b) = \sum_{i \in \{a, \neg a\}, j \in \{b, \neg b\}} 2C(i, j) \log \frac{C(i, j)N}{C(i)C(j)}$$

Bu yerda $C(i, j)$ soʻz va tarjimaning qoʻshma chastotasi, $C(i)$, $C(j)$ esa mos ravishda i va j chastotalari hisoblanadi.

Hozirgi vaqtda *DOMCAT* va *TotalRecall* parallel korpuslari tarjima uchun aniq boʻlgan kontekstdagi tarjima variantlarini istemolchiga sifatli yetqazib berish maqsadida yaratilgan eng zamonaviy dizaynga ega tarjima korpuslaridan biri hisoblanadi.

Linguee-bu ikki tilli parallel korpus orqali yaratilgan kontekstual tarjimon dastur sanaladi. Lingueeda lugʻatni avtomatik yaratish quyidagi bosqichlarni oʻz ichiga oladi: 1) parallel korpusni oldindan qayta ishlash (iboralar jadvalini chiqarish, morfosintaktik tahlil, lemmatizatsiya, tarjima ekvivalentlarini filtrlash); 2) har bir soʻz uchun tarjima ekvivalentlarini sinonimlar qatoriga birlashtirish; 3) grammatik belgilar qoʻshish; 4) teskari tarjimalarni, yaʼni kirish soʻzning sinonimlarini hosil qilish. Bunday holda, ushbu bosqich toʻgʻridan-toʻgʻri oʻtkazmalar asosida amalga oshiriladi; 5) parallel korpusdan foydalanish misollari bilan tarjimalarning har bir guruhining illyustratsiyasi.

Linguee tarjimonining Reverso kontekstdan afzalliklaridan biri bu foydalanuvchi tomonidan murakkab leksik soʻzlar, kollokatsiyalar yoki iboralarni qidirganida misol tariqasidagi jummalarning muayyan koʻzga tashlanib turishi hisoblanadi va ushbu kollokatsiya hamda iboralarni ajratib olish, oʻqish va undan foydalanish istemolchi uchun oson kechadi. Shuningdek, mazkur lugʻat koʻrsatilgan misol tariqasidagi jummalari ustidagi boʻsh joy minimal darajaga qisqartiriladi

Olim A.Segunning fikriga koʻra, “Reverso kontekst bu til vositalarining portali, jumladan, *frantsuz, ispan, nemis, rus, portugal, xitoy, arab yapon* va boshqa tillar kabi turli til birikmalariga tarjima qilish vazifasini bajaruvchi tarjimon dasturi”²⁹ deb hisoblaydi. Reverso kontekst bilan Lingueening oʻrtasidagi funksional farqlar shundaki, Reverso avtomatlashtirilgan tarjima va

²⁸ Dunning, T. Accurate methods for the statistics of surprise and coincidence. Computational linguistics. MIT Press. – Massachusets 1999. – P. 61–74

²⁹ Segun.A... La traduction automatique par opposition à la traduction humaine: analysed’un corpus de productions réelles. International journal of humanities and social science (IJHSS). – Nigeria, 2014. – P.213-222.

til o'rganish bo'yicha ixtisoslashgan tarjimon veb-sayti hisoblanadi. Linguee esa, asosan, *yuridik, diplomatik, rasmiy matnlar bazasidan tshkil topgan kontekstual tarjimaga ixtisoslashtirilgan.*

Uchinchi bob "**Parallel korpus kontekstual qidiruv tizimining funksional imkoniyatlari**" deb nomlanadi. Unda parallel korpusda kontekstual dastur yaratish tamoyillari, parallel korpusda kollokatsiyalarni joylashtirish, qidirish usullari, undan foydalanish yo'riqnomalari hamda ilmiy va publitsistik matnlarning leksik xususiyatlari hamda kontekstual tarjima tahlillari, ingliz va o'zbek tillarida kontekstual tarjimon dasturini yaratish bosqichlari, Sketch engine platforfamasida parallel korpus yaratish strategiyalari tahlil etilgan. Tadqiqot obyekti hamda predmeti hisoblangan parallel korpus kontekstual qidiruv tizimlari, ingliz va o'zbek tillari ilmiy hamda publitsistik parallel matnlar korpusi kontekstual tarjima xususiyatlariga e'tibor qaratildi. Shunga ko'ra, parallel korpusda avtomatlashtirilgan kontekstual tarjimon dasturi mexanizmlarining kichik bazasini shakllantirish uchun Sketch engine platformasi tanlandi.

Kollokatsiya ikki va undan ortiq so'z bilan birikishi natijasida hosil bo'ladigan hamda alohida komponentlarni qayta ko'rib chiqish natijasida so'zlarning ma'nolarini o'zgartirish jarayonida hosil bo'ladi. O'.Dell va M.Kartilarning fikriga ko'ra, kollokatsiya so'zlarning tabiiy birikmasi sifatida anaiqlanadi. Bu ingliz tilidadgi so'zlarning bir-biri bilan chambarchas bog'liqligini anglatadi. Masalan, *go hand in hand, as do commit and crime, blond goes with hair, and rain is heavy.* O'.Dell va M.Kartil kollokatsiyani uchta toifaga ajratishgan. Ular *kuchli birikmalar (strong collocations), qat'iy belgilangan birikmalar (fixed collocations) va zaif birikmalar (weak collocations).*

P.Neymark kollokatsiyani "*matn va leksikaning nervlari*" deb hisoblaydi"³⁰. Kollokatsiyalar *leksik va grammatik* kategoriyalarga bo'linadi. Leksik va grammatik kollokatsiyalarni quyidagi usullar orqali bir-birdan ajratish mumkin. Leksik kollokatsiyalar *faqat fe'llar sifatlar, otlar va qo'shimchalardan (verbs, adjectives, nouns, and adverbs)* va turli xil birikmalar asosida tuziladi. Grammatik kollokatsiyalarning tuzilishi *fe'l, sifat yoki ot (verbs, adjectives, or nouns combined with a preposition or a grammatical structure)* kabi so'zlarni bosh gap yoki grammatik tuzilishi bilan birlashtiradi.

Shunday ekan, tadqiqoqtimizda ilmiy va publitsistik matnlardan tashkil topgan matnlarning ingliz tilidan o'zbek tiliga tarjima qilingan maqolaning annotatsiya bo'limidagi "*deal with*" kollokatsiyaning matndagi ma'no mushtarakligi tahlil qilindi. Bunda tarjima jarayonida mazkur kollokatsiya bir necha ma'noda kontekstual tarjima qilinganligini kuzatishimiz mumkin. "***Deal with***" kollokatsiyasining tarjima variantlarini quyidagi 1-jadvalda ko'rish mumkin:

³⁰ Newmark P.A. Textbook of Translation. – New York: Prentice Hall International press, 1988. – P. 186.

“deal with” kollokatsiyasining kontekstual tarjimasini

<i>The article deals with</i> the aesthetic views of English writer Oscar Wilde	<i>Maqolada</i> ingliz yozuvchisi Oskar Vayldning estetik qarashlari <i>o‘rganilgan.</i>
<i>This article deals with</i> the issues of fiction texts, their specific features and multidimensional structure	<i>Ushbu maqola</i> badiiy matnlar va ularning ko‘p qirrali tarkibi tahlili masalalariga <i>bag‘ishlangan.</i>
<i>This study deals with</i> the determination of pedagogical methods, development of the technology of formation of aesthetic qualities of a person with the help of a foreign language during the educational process and outside the educational process; moreover, the criteria for the formation of civic qualities of students were determined.	<i>Mazkur maqolada</i> aynan texnika oliy o‘quv yurtlari talabalarida yuksak insoniy fazilatlarni rivojlantirish tamoyillari va estetik tarbiyani shakllantirishda yuzaga keladigan muammolarni bartaraf etish yo‘llari hamda ishlab chiqilgan texnologiyalar, mezonlar xususida batafsil <i>so‘z yuritiladi.</i>
<i>The article deals with</i> the information on religiously marked allusions, their symbolic aspects in English literary texts and analysis from the perspectives of cognitive linguistics	<i>Ushbu ishda</i> diniy markerlangan allyuziyalar, ularning ingliz badiiy matnlaridagi ramziy ko‘rinishlari haqida ma’lumot berilgan hamda ushbu til birliklari kognitiv tilshunoslik nuqtayi nazaridan <i>tahlil qilingan.</i>
<i>The study deals with</i> the meanings and features of the terms “dialogue” and “dialogic speech”.	<i>Maqola</i> “muloqot” va “dialogik nutq” atamalarining mazmuni hamda xususiyatlarini <i>ochib bergan.</i>
<i>This article deals with</i> the conditions for the emergence of psycholinguistics, that is, the direction of lingu-psychology, its uniqueness and its necessity today.	<i>Ushbu maqolada</i> psixolingvistik, ya’ni lingvopsixologiya yo‘nalishining yuzaga kelish shartlari, o‘ziga xosligi va bugungi kundagi zarurati xususida <i>so‘z boradi.</i>

Yuqorida keltirilgan **“deal with”** kollokatsiyasi bilan keltirilgan tahlilda ilmiy maqolada tarjimaning bir nechta tarjima ma’nolarini ko‘rish mumkin. Ya’ni **“deal with”** kollokatsiyasi ilmiy maqolada **“so‘z yuritmoq, o‘rganmoq, bag‘ishlamoq, tahlil qilmoq va ochib bermoq”** kabi kontekstual tarjima usulidan foydalangan hamda ushbu kollokatsiyaning bir necha ma’nolarini berish orqali tarjima variantlarini muallif o‘z maqolasida aks ettirgan. Yuqoridagi tahlildan kelib chiqib, **“deal with”** kollokatsiyasini toifaga ajratadigan bo‘lsak, **“deal with”** grammatik *cheklangan* kollokatsiya toifasiga kiradi. Cheklangan kollokatsiyalar faqat bir tarkibiy qismdan iborat bo‘lib bir fe’lning o‘zi bir necha ot yoki predloglar bilan birlashadi. Shuningdek, kollokatsiyani tarjima amaliyotida bir necha muhim xususiyatlarini ko‘rishimiz mumkin.

Mazkur tadqiqot ishimizning asosiy vazifalaridan biri hisoblangan desemantlashgan **“point”, “matter”, “thing”, “discuss” “analyse”, “examine” “deal”** ot va fe’l so‘z turkumlarining leksik ma’nosi va kontekstual tahlillarini parallel korpusda ochib berishdan iborat. Desemantizatsiya bu – **“lotincha de-prefiksidan, ajratish, bekor qilish-t-yunoncha semantikos-bildiruvchi”** degan ma’noni bildirib, asosan, tarjima jarayonida leksik-semantik ma’noni yo‘qotish, o‘zgarib borish kabi ma’noni anglatadi.

Ilmiy uslubga mansub maqolalarning annotatsiya bo‘limi hamda publitsistik uslubda qo‘llaniladigan so‘zlarning leksik ma’nosi va kontekstual tarjimasini

tahlilga tortish nazarda tutilgan. Shu nuqtayi nazardan mazkur soʻzlarni tarjima qilishda kontekstning oʻrni juda muhim hisoblanadi. Buni turli **“point”, “matter”, “thing”, “discuss” “analyse”, “examine” “deal”** ot va feʼl soʻz turkumlarining kontekstual tarjima jarayonida kuzatish mumkin. Ayrim hollarda mazkur soʻzlar faqat grammatik vazifani bajaradi va bular tayanch soʻzlar deb ataladi. Masalan, **“poor (little) thing”** birikmasida, **“thing”** oti faqat asosiy soʻz boʻlib, tayanch soʻz sifatida ishlatiladi. **“Poor”** kombinatsiyaning oʻzi oʻzbek tiliga **“bechora”** deb bir soʻz bilan tarjima qilinadi. Xuddi shu gap ingliz tilida **“just the thing”** degan maʼnoni anglatuvchi oʻzbek tilidagi tarjimasi **“shunday narsa, kerak boʻlgan narsa”** kombinatsiyasiga nisbatan ham toʻgʻri keladi. Bu turdagi maʼnolarning noaniqligi sababli mazkur soʻzlar juda keng moslikka ega boʻladi, bu esa, oʻz navbatida, tarjima jarayonida yanada keng imkoniyat yaratadi. Shuning uchun ularning tarjimasi, qoida tariqasida, ular ishlatilgan kontekstga bogʻliq. Shu nuqtayi nazardan ham kontekstual tarjima ushbu jarayondagi duch keladigan muammolarni yechishga yordam beradi. Shuningdek, tarjimaning aniq variantlarini foydalanuvchiga taqdim qiladi. Bunga yuqoridagi soʻzlarning kontekstual tarjimalarini tahlil qilish orqali koʻrish mumkin. **“Point”** ot soʻz turkumining kontekstual tarjimasi parallel matnlarda quyidagi 2-jadvalda tarjima turlari tahlil qilindi.

2-jadval

“Point” ot soʻz turkumining kontekstual tarjimasi

<i>The lawyers may be right in questioning the legality of some minor points.</i>	<i>Yuristlar baʼzi kichik qonun hujjatlarining qonuniyligini haqli ravishda shubha ostiga qoʻyishadi.</i>
<i>This point became particularly apparent in analyses of the Chicago school data.</i>	<i>Bu savol Chikago maktabidagi maʼlumotlarini tahlil qilishda yaqqol namoyon boʻldi.</i>
<i>From their point of view, the main function of the law is to limit the power of the authorities and the courts to inflict evils upon the individual citizen and to restrict his freedom.</i>	<i>Ularga koʻra, qonunning asosiy vazifasi hokimiyat va sudlarning har bir fuqaroga yomonlik qilish hamda uning erkinligini cheklash vakolatlarini chegaralashdan iborat.</i>
<i>Race is, of course, a sensitive point of reference in American society, and a reference point that can become cute for minority youth during early to middle adolescence.</i>	<i>Poyga, shubhasiz, Amerika jamiyatida nozik nuqta hisoblanib, u erta yoshdan oʻrta oʻsmirlik davridagi ozchilik yoshlar orasida keskin kuzatilishi mumkin boʻlgan nozik jihatlardan biridir.</i>
<i>Patterson’s commitment to the ideal of integration implies his belief that such a tipping or turning point exists.</i>	<i>Pattersonning integratsiya idealiga sodiqligi uning bunday burilish nuqtasi mavjudligiga ishonchi komil ekanligini bildiradi.</i>

Ushbu **“matter”** leksik birlikning ham bir qancha maʼnolarni anglatuvchi kontekstual tarjima misollarini keltirish mumkin. **“Matter”** leksik birlikning oʻzbek tilidagi tarjimasi **“vaziyat”, “holat”, “mazmun”, “ish” “materiya”, “modda”** va **“savol”** kabi maʼnolarga ega (*wisdom dictionary*). Biroq quyida keltirilgan misollardan koʻrinadiki, koʻpincha maʼlum bir desemantlashgan soʻzning kontekstual maʼnosi yuqoridagi maʼnolarning birortasiga mos kelmaydi. Koʻpincha adekvat tarjimaga erishish uchun ushbu leksik birlik oʻzbek tiliga toʻgʻridan toʻgʻri tarjima

qilinmaydi, ular kontekstual tarjima qilinadi. Mazkur **matter** leksik birlikning kontekstual tarjima misollari parallel matnlarda 3-jadvalda quyidagi keltririlgan:

3-jadval

“Matter” leksik birlikning kontekstual tarjimasi

<i>It would be a delicate matter to express an opinion about the creativity of younger men.</i>	<i>Yosh xodimlarning ijodi haqida fikr bildirish juda nozik masala.</i>
<i>It has much less experience in this matter than the classical countries of bureaucracy, France, Germany, Austria, and Russia, acquired.</i>	<i>Bu borada Fransiya, Germaniya, Avstriya va Rossiya kabi klassik byurokratiya davlatlariga qaraganda bu davlatning tajribasi ancha kam.</i>
<i>As head of a bureau, he may have the power to alter some minor rules and some matters of internal procedure</i>	<i>U Agentlik rahbari sifatida muayyan qoidalarga va ichki tartib-qoidalarga taalluqli bo‘lgan moddalarga kichik o‘zgartirishlar kiritish huquqiga ega bo‘lishi mumkin.</i>
<i>In violating the law you are perhaps worse than a good many of the racketeers, no matter how good your intentions may be.</i>	<i>Qanchalik yaxshi niyatda bo‘lishingizdan qat’i nazar agar qonunni buzsangiz o‘sha reketlardek bo‘lgan bo‘lasiz.</i>
<i>But it is another matter with the local chief of a government agency.</i>	<i>Mahalliy davlat organi rahbari bilan vaziyat butunlay boshqacha.</i>

“Thing” leksik birligining kontekstual tarjima tahliliga keladigan bo‘lsak, yuqorida ko‘rib chiqilgan “matter” leksik birlikga qaraganda o‘zining kontekstual ma’nosi matnlarda yanada aniqroq bo‘ladi. Lug‘atda ushbu “thing” so‘zi asosiy “narsa”, “mavzu”, “ish” (*wisdom dictionary*) kabi tarjimalarni beradi. Biroq kontekstga asoslanib, aksariyat hollarda berilgan so‘z odatda mustaqil birlik sifatida tarjima qilinmaydi. Buni bir nechta misollar orqali 4-jadvalda quyidagicha tahlil qilish mumkin:

4-jadval

“Thing” leksik birlikning kontekstual tarjimasi

<i>Within a business concern, such things can be left without hesitation to the discretion of the responsible local manager</i>	<i>Biznes doirasida, mahalliy hokimiyatlar bu kabi savollarni ikkilanmasdan ko‘rib chiqishi mumkin.</i>
<i>The same thing is essentially valid for a democratic government.</i>	<i>Demokratik hukumat uchun ham xuddi shunday bo‘ladi.</i>
<i>How can these two things, one good, the other bad, be linked together?</i>	<i>Biri yaxshi va boshqasi yomon kabi bu ikki qarama qarshilikni qanday bir-biriga bog‘lash mumkin?</i>
<i>Government efficiency and industrial efficiency are entirely different things</i>	<i>Hukumat faoliyati va sanoat ko‘rsatkichi bir-biridan butunlay farqlidir.</i>
<i>It may well be that the greatest thing in Atlantis is its better system of government.</i>	<i>Atlantisning eng katta yutug‘i uning yaxshi ishlaydigan boshqaruv tizimi bo‘lishi mumkin.</i>
<i>There are many things about government administration that need to be reformed</i>	<i>Hukumat boshqaruvida isloh qilinishi kerak bo‘lgan juda ko‘p tizimlar mavjud.</i>
<i>However, it is quite a different thing if an officeholder evades the duly promulgated laws of the nation for the benefit of the “State”.</i>	<i>Agar mansabdor shaxs “davlat” manfaati uchun tegishli tarzda e‘lon qilingan qonunlardan qochsa, unda bu butunlay boshqa vaziyat.</i>

Tadqiqotning ikkinchi tahlilidan iborat bo‘lgan ilmiy maqolalardan tarkibidan tashkil topgan annotatsiyalarning ingliz va o‘zbek tillaridagi kontekstual tarjimalari tahlil qilindi. Bunga ilgari zikr qilingan *“discuss”* *“analyse”*, *“deal”* kabi so‘zlarning kontekstual tarjima variantlari tahlilga tortildi. *“Discuss”* fe‘lining lug‘atda *“muhokama qilmoq”*, *“tahlil qilmoq”*, *“ko‘rib chiqmoq”* kabi tarjima variantlari mavjud (wisdom dictionary). Ko‘pincha ilmiy matnlarning tarjima variantlarida *“discuss”* so‘zi yuqoridagi lug‘atda keltirilgan ma‘nolarda tarjima qilinadi, biroq ba‘zan mazkur so‘z ilmiy matnlarda boshqa ma‘nolarda kelishi kuzatildi. Ushbu so‘zning kontekstual varianni 5-jadvalda quyidagi ma‘nolarda tarjima qilingan.

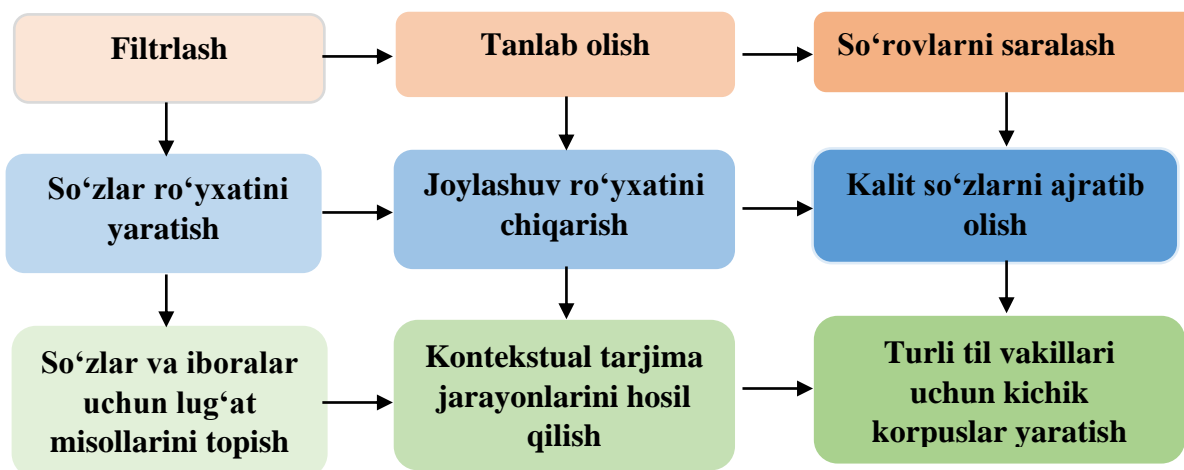
5-jadval

“Discuss” fe‘lining kontekstual tarjimasi

<i>This article discusses the economic efficiency of the world’s most popular interactive TV projects.</i>	<i>Mazkur maqolada jahonda keng ommalashgan interaktiv teleloyihalarning samaradorlik masalalari ko‘rib chiqildi.</i>
<i>The work discusses the views on cognitive linguistics, its subject, and tasks, the essence of the concept of “cognitivism”, the problem of “concept” and “cognition”, and the features of cognitive linguistics in contrast to traditional linguistics.</i>	<i>Mazkur maqolada so‘nggi yillardagi tub burilish bilan bog‘liq holda yuzaga kelgan kognitiv tilshunoslikka doir fikrlar, uning o‘rganish obyekti va vazifalari, “kognitivizm” tushunchasining mohiyati, unda “tushuncha” va “konsept” muammosi hamda kognitiv tilshunoslikning an‘anaviy tilshunoslikdan farqli belgilari, kognitiv tilshunoslikning yo‘nalishlari xususida so‘z yuritiladi.</i> https://journal.fledu.uz
<i>This paper discusses the work of A. Semyonov was a scientist who made a significant scientific and practical contribution to the development of the science of the Uzbek and Tajik peoples.</i>	<i>Ushbu maqolada o‘zbek va tojik xalqlari ilm-fani rivojiga ilmiy-amaliy jihatdan katta hissa qo‘shgan olim A.Semyonovning Navoiyshunoslik, manbashunoslik faoliyati o‘ganildi.</i> https://journal.fledu.uz

Shunday qilib, ilmiy maqolaning tarjima sifati muloqot maqsadiga erishish darajasiga, asl tildagi matnning maqsadli tildagi tarjimasiga maksimal mazmun-mohiyatini kontekstual tarjima usuli orqali amlaga oshirish tarjima aniqligi va ilmiy hamda publitsistik matnning ma‘no-mazmunining umumiylikiga bog‘liq degan xulosaga keldik. Mazkur tarjima matnlarni parallel korpusda kontekstual tarjimon dasturining kichik bazasini shakllantirish muhim vazifalardan biri hisoblanadi. Mazkur ikki tilli matnlarni parallel korpusini yaratish uchun Sketch Engine platformasida joylashtirish orqali samarali natijaga erishish mumkin.

Sketch Engine bu korpus bo‘yicha yetakchi platformalarinig biri hisoblanib, lug‘atshunoslikda keng qo‘llanilib kelmoqda. Sketch engine ikki xil funksiyadan iborat: *dasturiy ta‘minot hamda veb-xizmat*. Mazkur, dasturiy ta‘minot va veb-xizmat ko‘p sonli oldindan yuklangan hamda foydalanishga tayyor korpuslarni va erkin korpus yaratish, o‘rnatish va boshqarish vositalarini hamda vazifalarni o‘z ichiga oladi. Sketch Engine, asosan, quyidagi xizmatlarni taqdim etadi va ular 3-rasmda quyidagicha keltirilgan:



3-rasm. Sketch engine platformasining vazifalari

Shu bilan birga, Sketch Engine ko‘p tillarda teglash vositalarini birlashtirish, terminalogiyani qo‘shish, kollkatsiyalar, idiomalar bilan ishlash, so‘z eskizlari shuning bilan bir qatorda tezaurslarni yaratsih va uni foydalanuvchilar uchun ishlash mexanizmini qulay tarzda ochib berish ustida ko‘plab imkoniyatlari mavjud. Hozirda yuqorida keltirib o‘tilgan imkoniyatlar o‘z natijasini bermoqda, bunga hozirda juda ko‘plab foydalanuvchilar tomonidan qo‘llanilayotgan *Brown* korpusi va *enTenTen* korpusini misol qilishimiz mumkin.

Hozirda Sketch Engine platformasida joylashtirilgan bracha ma’lumotlar quyidagi raqamalarda keltirilib o‘tilgan. Ulardan:

Tokens – 17,329,301
So‘zlar – 15, 513,307
Jumlalar – 615, 181
Paragraflar – 223,394
Hujjatlar – 8,016

Tadqiqotda tarjimani avtomatlashtirishda parallel korpusda kontekstual tarjima dasturini Sketch Engine platformasida yaratish quyidagi bosqichlarda amalga oshirildi:

1) ilmiy va publitsistik matnlar to‘plangan ikki tilli ingliz va o‘zbek tillaridan tashkil topgan ma’lumotlar bazasi *Microsoft Exel* jadval ko‘rinishida shakllantirildi, mazkur matnlarni shakllantirishdan maqsad *Microsoft Exelda*, asosan, ikki tilli matnlardan yig‘ilgan xotira bazasi yaratish ko‘zda tutildi;

2) yig‘ilgan ma’lumotlarni tarjima muvofiqligi qayta tekshirib chiqildi;

3) tekshirilgan tarjimalar muvofiqligi bir biriga mos kelgach kompyuterda ma’lumotlar bazasi yaratildi.

Ma’lumotlar formati parallel korpus uchun *data* formatida tasvirlangan formatda bo‘lishi kerak. Shunday ekan, biz ma’lumotlat formatini *data* ko‘rinishida keltirib olamiz. Ushbu vazifalar bajarilgandan so‘ng ikki tilli parallel korpusni yaratish uchun bizga keyingi bosqichlarni amalga oshirishiz muhim hisoblanadi. Ular quyidagilar: 1) korpus asboblari panelida “*new corpus-yangi*

korpus” tugmasini belgilash orqali; 2) multilingual tugmasini belgilash orqali; 3) korpus nomini kiritish va faylni tanlash orqali; 4) keyingi ekranda tillar to‘g‘riligini tekshirish orqali; 5) “*next*” buyrug‘ini belgilash orqali; 6) yaratilgan ikki tilli parallel korpusni *My corpus* bo‘limidan aniqlash orqali yaratilgan korpusini tekshirish orqali jarayonni tugallanadi.

Bundan tashqari Sketch engine platformsi boshqa turdagi formatlar bilan ham amalga oshirish imkoniyati mavjud. Masalan. *XLIFF* (v. 2.0 va undan yuqori) va *CSV*, *TSV*, *XLSX* bular jadval ma’lumotlarini saqlash uchun ishlatiladigan oddiy matnli hujjat turi hisoblanadi.

Tadqiqotimiz natijasida yuqoridagi ko‘rsatmalardan foydalanib ingliz va o‘zbek tillarida ilmiy hamda publitsistik matnlarning kontekstual tarjima parallel korpusi interfeysi yaratildi. Ushbu dasturning interfeysi va undan foydalanisusullari quyidagi 4-5 ramlarda ko‘rsatib o‘tilgan.



4-rasm. Parallel korpus interfeysi



5-rasm. Ingliz va o‘zbek tillaridagi ilmiy matnlardan tashkil topgan parallel korpus interfeysi

Ingliz va o‘zbek tillarida ilmiy hamda publitsistik matnlardan iborat bo‘lgan kontekstual tarjima dasturi parallel korpus orqali Sketch Engine platformasining mexanizmi shakllantirildi. Ushbu dasturdan foydalanish yo‘riqnomasi quyidagicha belgilandi: 1) *Sketch Engine platformasi saytiga kirish*; 2) *saytga elektron manzil orqali a‘zo bolish*; 3) *platformadan “parallel corpora” bo‘limiga o‘tish*; 4) *“parallel corpora bo‘limidan “English Uzbek contextual translation” bo‘limiga o‘tish*; 5) *“English Uzbek contextual translation” interfeysidan so‘z qidirish*.

Yuqoridagi yaratilgan parallel korpusda kontekstual tarjimon dasturining yaratilishi negizida mashina tarjimasining statistik usuli muhim o‘rin tutadi.

Kontekstual tarjimon dasturlarini yaratish va undan samarali foydalanish uchun mashina tarjimasining statistik usulidan foydalanishdan sabab Sketch Engine platformasida yaratilgan parallel korpus ma‘lumotlari ingliz hamda o‘zbek tillarida internetda kontenti va *“e-library”* mavjud emas. Shu sabab ham barcha ma‘lumotlarni *Microsoft Excel*da kichik baza xotirasini yaratish maqsadida platformaga joylashtirildi.

Mazkur parallel korpusda umumiy qiymati 1658 so‘zdan iborat ingliz va o‘zbek tillarida ilmiy hamda publitsistik matnlardan tashkil topgan matn hamda uning tarjimalari dasturda joylandi. Shundan, 742 so‘z ingliz tilidan tashkil topgan qolgan 916 ta so‘z esa o‘zbek tilidagi matnlarda tashkil topgan. Ingliz tilidagi so‘zlar sonining o‘zbek tilidagiga qaraganda kamligi shundan iboratki, o‘zbek tilidagi leksik so‘zlarning tarjima jarayonida bir necha ma‘noni hosil qilishiga bog‘liq bo‘ladi.

Tarjimoni o‘rgatishda korpusga asoslangan yondashuv so‘nggi o‘n yil ichida jadal rivojlandi, hozirda parallel korpus, ko‘p tilli korpus (multilingual corpus) va taqqoslanadigan korpuslardan (comparable corpus) foydalangan holda tadqiqotlar natijasida ko‘plab ma‘lumotlar paydo bo‘ldi. Ayniqsa, parallel korpus, tarjima tadqiqotlari va ta‘lim sohasida qimmatli ma‘lumot manbalari ekanligini isbotlaydi.

XULOSA

1. Mashina tarjimasi, tarixi va yaratilishi bosqichlari ilmiy nazariy jihatdan keng miqyosda o‘rganilishi dolzarblilik kasb etadi. Bunda mashina tarjima tizimlari muayyan jihatlarga ko‘ra xarakterlanadi. Ular: 1) tillarning ishtiroki: a) ikki tilli; b) ko‘p tilli; 2) tanlangan mavzuning miqdori: a) muayyan uslub va sohaga aloqador; b) bir uslub hamda ko‘p sohalarga tegishli; 3) turli uslub va bir necha sohalar bilan bog‘liqlik jihatlari bilan belgilanadi. Shunga ko‘ra, mashina tarjimasi quyidagi tizimlarga ajratilgan hamda har bir tizim o‘zining kichik va tarjima jarayonida o‘ta muhim vazifani bajaradigan turlari mavjud. Ular, 1. Qoidaga asoslangan MT – (Rule-Based MT); 2. Korpus asosidagi mashina tarjimasi – (Corpus based MT); 3. Gibrud mashina tarjimasi – (Hybrid MT). Bugungi kunda statistik MT tizimidan samarali foydalanilmoqda, lekin kelajakda istiqboli sun‘iy intellekt hamda neyro MT tizimlaridan foydalanish ehtimoli mavjud.

2. MT, asosan, uch bosqichni, dastlab lug‘atga mos keladigan mashina tarjimasi, korpus statistikasiga asoslangan kompyuter tarjimasi, so‘ngra so‘ngi

yillarda asosiy texnologiya sifatida sun'iy intellekt bilan neyro-mashina tarjimasigacha bo'lgan bosqichlaridan tashkil topadi. Hozirda lug'at asosida tarjima qilinadigan mashina tarjimasidan foydalanish darajasi ancha pasaygan, uning o'rniga korpus statistikasiga asoslangan kompyuter tarjimasidan foydalanish jadal rivojlanmoqda. Kelajakda asosiy texnologiya sifatida sun'iy intellekt hamda neyro-mashina tarjimasidan foydalanish insoniyat uchun tildagi va tarjimadagi muammolarni yechimi sifatida xizmat qiladi.

3. MT tizimlarining jadal rivojlanish mobaynida korpusning paydo bo'lishi va tarjima bozorining tarjima samaradorligiga bo'lgan talabining ortishi bilan CAT tarjima xotiralari hamda terminologiya ma'lumotlar bazalarini boshqarishni yaxshilash uchun mashinali tarjima dasturlarini korpus platformalari orqali parallel tarzda qo'llanilishi natijasida KL hamda MT ning rivoji uchun burilish nuqtasi bo'lganligini alohida inobatga olish zarur.

4. Tadqiqotda korpus turlarining bir qancha turlari, xususan, DIY korpus, comparable korpus, stable korpus, ad hoc korpus va boshqa korpus turlarining funksional jihatlari ilmiy nazariy jihatdan keng yoritilib, mazkur korpuslarning bir-biriga o'xshash va farqli jihatlari keng yoritildi. Xususan, DIY korpusi internet hujjatlari to'plami, aniqrog'i *HTML* tilidagi veb-sahifalardan tuziladi. DIY korpusi ochiq korpus hisoblanib undan farqli o'laroq Ad hoc korpusi esa tarjimon tomonidan ma'lum bir tarjima paytida duch kelgan muayyan muammoni tekshirish uchun qisqa vaqtda tuziladi. Ad hoc korpusi odatda ko'p matnlarni o'z ichiga olmaydi. DIY va Ad hoc korpuslarning o'xshash jihatlari bu ikki korpus tarjimini amalga oshirish uchun tuziladigan virtual korpusdir. Comparable corpus ikki yoki undan ortiq bir tilli tarjimalar to'plamidagi korpus bo'lib, odatda, har tabiiy til boshqa tilda bir xil printsiplarga asoslanib muvofiq quriladi. Ushbu korpusning ahamiyatli jihatlardan biri matnlardagi ma'lumotlar bir xil bo'lishi mumkin, lekin ko'zlangan tarjimalar natijasi boshqa tillar bilan solishtirganda har xil hosil bo'ladi. Stable korpus barqaror korpuslar turlariga mansub, unda ko'pincha faqat o'z interfeyslariga kiritilgandan so'ng taqdim etiladi. Stable korpora DIY, AD hoc, Comparable va boshqa turdagi korpuslardan farqli jihati shundaki, stable korpus tarjimonlarning ehtiyojlariga moslashtirilmaydi hamda tarjimadan foydalanuvchilar uchun maxsus qurilmaydi va tarjima istemolchilariga matnlarni tarjima qilishda ushbu stable korpusdan foydalanish tavsiya etilmaydi.

5. Korpus lingvistikaning eng asosiy vazifalaridan biri bu konkordansda kollakatsiyalarni (*collocations*) qidirish hisoblanadi. Kollakatsiyani korpusdagi muhim o'rin tutishi uch turga strong (kuchli), fixed (qat'iy belgilangan) va weak (kuchsiz) kollokatsiya turlari orqali aniqlanib, uning asosiy vazifalaridan biri bu tarjimadagi tarkibiy qismlaridan birining lug'aviy ma'nosini saqlab qolishdir. Tadqiqotda ingliz tilidagi kollokatsiyalarni o'zbek tiliga o'girishda ilmiy va publitsistik matnlardagi materilallardan foydalangan holda tahlil qilinganligi natijasida ingliz hamda o'zbek tillarida kontekstual tarjimon dasturini parallel korpusda kollokatsiya ikki tildagi jummalarni aniq tarjima ma'nosini topishda muhim vazifani amalga oshiradi.

6. Ilmiy va publitsistik uslubda yozilgan matnlar tarjimada e'tibordan chetda qoldirilmaydigan bir qator xususiyatlari mavjud. Shunga qarab, ilmiy va publitsistik matnlardagi asliyat tildagi matnning ko'zlangan tildagi tarjimasi muqobiligi aniqlanadi. Ilmiy va publitsistik matnlarni tarjima qilishdagi asosiy muammolardan biri bu terminalogiyadir. Tarjimon uchun bitta so'zdan iborat bo'lgan termini tarjima qilish ancha mashaqqatli bo'lishi mumkin. Ushbu muammoning yechimi quyidagi tematik lug'atlar, xalqaro belgilar, va umume'tirof etilgan atamalar bilan birlashtirib tekshirishi kerak. Ilmiy va publitsistik matnlarning atamalardan tashqari tarjima amaliyotida quyidagi vazifalari mavjud. 1) Bir ma'nolilik – ilmiy va publitsistik matnlarda so'z hamda atamalarni ikki marta talqin qilishga yo'l qo'ymaslik; 2) turli olmoshlardan, foydalansih – ingliz tilidagi maqolalarda ko'pincha birinchi shax birlikdagi olmoshlarni toppish mumkin. O'zbek tilida esa “biz” olmoshini qo'llash mumkin; 3) aniqlilik – ilmiy va publitsistik matnlar iloji boricha aniqroq, manbaga yaqinroq tarjima qilinishi maqsadga muvofiq bo'ladi. Ushbu vazifalar ilmiy va publitsistik matnlarni tarjima jarayonida duch kelgan muammolarga yechim sifatida xizmat qiladi.

7. Ilmiy va publitsistik matnlarda tez-tez uchraydigan desemantlashtirilgan so'zlarning leksik ma'nosi hamda kontekstual tarjimasi tarjimani avtomatlashtirishda faol qo'llanilib, asosan, leksik-semantikaga oid ot va fe'l so'z turkumlarida ko'proq uchraydi. Desemantizatsiya fe'l va ot so'z turkumlaridagi leksik ma'noning asta-sekin yo'qolishi natijasida so'zning leksik ma'nosini grammatizatsiyaga olib keladi. Ayrim hollarda ot va fe'l so'z turkumlarini kontekstual tarjima qilganda faqat grammatik vazifani amalga oshirib tayanch so'zlar sifatida qaraladi. Desemantlashgan fe'l va ot so'z turkumlari kontekstual tarjimani amalga oshirishda muhim ahamiyat kasb etadi.

8. Parallel korpusning afzalligi shundaki, ular standartlashtirilgan ikki tilli terminologiyaga to'la ixtisoslashgan parallel hujjatlardir. Korpusdan olingan qo'shma so'zlar ko'p jihatdan terminologik elementlardir. Shuning uchun tarjima birliklari va tarjima ekvivalentlari o'rtasida juda izchil munosabat mavjud. Bu korpusni tahlil qilishni osonlashtiradi. Parallel korpus yaratishning eng oddiy usuli bu dastlab ma'lumotlarni *Excel elektron* jadvalda joylashtirish orqali tarjima xotirasini yaratish bilan boshlanadi. Bunga sabab ingliz va o'zbek tillarida ilmiy hamda publitsistik matnlarning tarjima variantlari internet libraryda mavjud emasligidadir. Trajima xotirasini yaratishning ikkinchi metodi bu tarjima xotirasi formati hisoblanib, ular *TMX* yoki *XLIFF* yordamida tarjima xotirasi yartiladi. *XLIFF ning TMX, CSV, TSV va XLSX lardan farqli jihati XLIFF* maxsus dasrturlarda ochiladi va v. 2.0 va undan yuqori formatdagi ma'lumotlarni saqlash uchun foydalaniladi. *TMX bilan CSV* esa to'g'ridan-to'g'ri *Microsoft Excel, Open Office, Google Documents* va boshqa veb platformalarda ochiladi. *XLIFF* tilni avtomatlashtirishda ikki tildagi matnlarni kodlash bilan belgilanadi.

9. Tadqiqotda parallel korpus funksiyasini ilmiy nazariy o'rgangan holda Sketchengine platformasida tarjimon dastur yaratishning mezonlari ilmiy jihatdan o'rganilib, mazkur platformada korpus va parallel korpus yaratish xususiyatlari aniqlanib, parallel korpusda aynan kontekstual tarjimon dasturlarini yaratish

bosqichlari ilmiy asoslandi. Sketchengine platformasi kelajak korpuslar dunyosi talqinida ko‘rib kelinayotganligi til korpuslari hamda tarjima dasturlarini o‘zida jamlaydigan yagona yirik platformaga ega bo‘lishiga umid berayotganiga alohida e‘tibor qaratish zarur.

10. Tadqiqotimiz natijasida <https://auth.sketchengine.eu> platformasida parallel korpusda kontekstual tarjimon dasturining ilk mexanizmi yaratildi. Mazkur dasturning kontenti ilmiy va publitsistik uslubga doir matnlarning kontekstual ingliz hamda o‘zbek tillariga tarjima qilingan parallel bazasi shakllantirildi. Ingliz va o‘zbek tillarida ilmiy hamda publitsistik matnlardan tashkil topgan parallel korpusda kontekstual tarjimon dasturi Sketch Engine platformasiga joylashtirildi. Shuningdek, mazkur dasturdan foydalanish yo‘riqnomasi tuzildi.

11. Sketch Engine platformasidagi 3-ta yo‘riqnomasidan farqli ravishda quyidagi tadqiqotda Sketch Engineni parallel korpusda kontekstual tarjimon dasturining 5 ta: Sketch Engine platformasiga a‘zo bo‘lish, dastur interfeysidan turli korpus va tarjimon dasturlaridan parallel korpus bo‘limini tanlash, parallel korpus bo‘limidan, *english-uzbek contextual translation* dasturini izlash, mazkur tarjimon dastur platformaning chap tomondagi burchak qismidan ajratgan holda platforma interfeys markaziga olib o‘tish va dasturdan foydalanish bosqichlari ishlab chiqildi. Sketch Engine platformasida yaratilgan parallel korpusda kontekstual tarjimon dasturi kompyuter lingvistikasi sohasida tarjimadan samarali foydalanish, insoniyat uchun hamda ba’zi soha vakillari uchun manba va asosiy tarjimon dastur sifatida samarali xizmat qiladi.

**РАЗОВЫЙ НАУЧНЫЙ СОВЕТ ПРИ НАУЧНОМ СОВЕТЕ
DSc.03/30.12.2019.Fil.27.01 ПО ПРИСУЖДЕНИЮ УЧЕНЫХ
СТЕПЕНЕЙ ПРИ УЗБЕКСКОМ ГОСУДАРСТВЕННОМ
УНИВЕРСИТЕТЕ МИРОВЫХ ЯЗЫКОВ**

**УЗБЕКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ МИРОВЫХ
ЯЗЫКОВ**

ХАЙТКУЛОВ ЗАХРИДДИН ХАЗРАТКУЛ УГЛИ

**ФУНКЦИОНАЛЬНАЯ ОСОБЕННОСТЬ ПАРАЛЛЕЛЬНОЙ
КОРПУСНОЙ КОНТЕКСТУАЛЬНОЙ ПОИСКОВОЙ СИСТЕМЫ В
АВТОМАТИЗАЦИИ ПЕРЕВОДА**

10.00.11 – Теория языка. Прикладная и компьютерная лингвистика

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

Ташкент – 2024

Тема диссертации доктора философии (PhD) зарегистрирована в Высшей аттестационной комиссии Республики Узбекистан за № В2023.1.PhD/Fil3237.

Диссертация выполнена в Узбекском государственном университете мировых языков.

Автореферат диссертации на трёх языках (узбекском, русском и английском (резюме)) размещен на веб-странице Научного совета (www.uzswlu.uz) и Информационно-образовательном портале «Ziyounet» (www.ziyounet.uz).

Научный руководитель: Эргашева Гули Исмоил кизи
доктор филологических наук, доцент

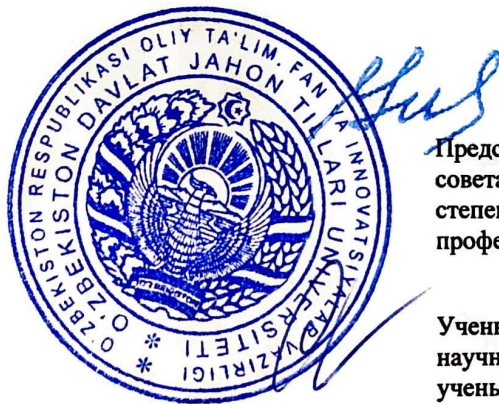
Официальные оппоненты: Хайруллаев Хуршид Зайниевич
доктор филологических наук, профессор
Атабоев Нозимжон Бобожон угли
доктор философии по филологическим наукам (PhD), доцент

Ведущая организация: Национальный университет Узбекистана
имени Мирзо Улугбека

Защита диссертации состоится "26" апреля 2024 года в 10:00 часов на заседании Научного совета DSc.03/30.12.2019.Fil/Ped.27.01 по присуждению ученых степеней при Узбекском государственном университете мировых языков (Адрес: 100138, город Ташкент, Учтепинский район, улица Кичик халка йули, дом 21А. Тел.: (+99871) 230-12-91; факс: (99871) 230-12-92; e-mail: uzswlu_info@mail.ru).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Узбекского государственного университета мировых языков (зарегистрирована за № 1). (Адрес: 100138, город Ташкент, Учтепинский район, улица Кичик халка йули, дом 21А. Тел.: (+99871) 230-12-91; факс: (+99871) 230-12-92, e-mail: www.uzswlu_info@mail.ru).

Автореферат диссертации разослан "13" апреля 2024 года.
(Реестр протокола рассылки № 1 от "13" апреля 2024 года).



Г.Х. Бакнева
Председатель Разового научного совета по присуждению ученых степеней, доктор филол. наук, профессор

Х.Б. Самигова
Ученый секретарь Разового научного совета по присуждению ученых степеней, доктор филол. наук, профессор

Ж.А. Якубов
Председатель Разового научного семинара при Разовом научном совете по присуждению ученых степеней, доктор филол. наук, профессор

ВВЕДЕНИЕ (аннотация диссертации доктора философии (PhD))

Актуальность и востребованность темы диссертации. В мировом языкознании обработка естественного языка, компьютерный перевод языка оригинала стали широко использоваться в практике математической лингвистики и машинного перевода. Создание ряда двуязычных и многоязычных электронных программ-переводчиков в информационно-поисковой системе, корпусной лингвистике и машинном переводе по перспективным научным направлениям, связанным с применением информационных технологий в лингвистике, расширение механизмов перевода программ за счет корпусных типов, обусловили необходимость его внедрения в практику. Ведь это определяет практическую значимость возможностей искусственного интеллекта по созданию различных нацеленных программ перевода посредством параллельных корпусов, автоматизированных словарей, анализа и редактирования параллельных текстов на двух и более языках.

В мировом языкознании на повестку дня поставлена необходимость научного обоснования таких вопросов, как автоматизация перевода, создание двуязычных и многоязычных электронных программ-переводчиков с использованием корпусных типов, повышение качества перевода, создание механизма двуязычного параллельного банка текстов. Информационно-поисковая система имеет необходимость разработки электронных двуязычных и многоязычных словарей, дословного переводчика, программы переводчика на основе правил, программы переводчика на основе статистических данных, программы анализа текстов по типам корпусов, а также программы и приложения лингвостатистического изучения переводных текстов. Создание приложений для анализа текста, состоящих из параллельных и других типов корпусов электронных переводчиков, приложений контекстных переводчиков, больших и малых двуязычных баз данных, платформ автоматического перевода в параллельном корпусе и больших платформ, содержащих многоязычные корпуса, а также их быстрое развитие имеет важное значение для исследования научно-теоретических проблем компьютерной лингвистики и корпусного переводоведения.

В нашей республике на повестку дня в переводоведении, как и в других сферах, ставятся такие вопросы, как анализ корпуса, типы корпуса, автоматический перевод, статистический машинный перевод, гибридный машинный перевод, машинный перевод на основе правил, компьютерная лингвистика, двуязычные и многоязычные корпуса и проведение научных исследований по более глубокому пониманию и обработке переводов английского языка на узбекский язык при разработке переводческих программ поставлено. Ведь «...сохранение чистоты государственного языка, обогащение его и повышение речевой культуры населения; обеспечение активной интеграции государственного языка в современные

информационные технологии и коммуникации»¹ является одной из главных задач, стоящих сегодня перед лингвистикой и переводоведением. Соответственно большое значение приобретают такие вопросы, как превращение английского и узбекского языков в компьютерные языки, помещение программ контекстуального перевода в корпус через параллельные тексты, создание специального корпуса для пользователей программ перевода с лексическими единицами и семантическими особенностями текстов.

Наряду с этим данное диссертационное исследование в определенной степени служит для реализации задач, поставленных в Указах Президента Республики Узбекистан от 7 февраля 2017 № УП-4947 «О стратегии действий по дальнейшему развитию Республики Узбекистан», от 8 октября 2019 № УП-5847 «Об утверждении концепции развития системы высшего образования Республики Узбекистан до 2030 года», от 21 октября 2019 года № УП-5850 «О мерах по кардинальному повышению роли и авторитета узбекского языка в качестве государственного языка», от 20 октября 2020 года «О мерах по дальнейшему развитию узбекского языка и совершенствованию языковой политики в стране», Постановлении Кабинета Министров Республики Узбекистан от 11 августа 2017 года № 610 «О мерах по дальнейшему совершенствованию качества преподавания иностранных языков в образовательных учреждениях», а также в Послании Президента Республики Узбекистан палатам Олий Мажлиса от 24 января 2020 года и в других нормативно-правовых документах, относящихся к данной деятельности.

Соответствие исследования приоритетным направлениям развития науки и технологий республики. Исследование выполнено в соответствии с приоритетным направлением развития науки и технологий республики I. «Пути формирования и реализации системы инновационных идей в социально-правовом, экономическом, культурном, духовно-просветительском развитии информированного общества и демократического государства».

Степень изученности проблемы. Неслучайно система контекстного поиска в параллельном корпусе в области автоматизации перевода стала объектом исследований многих лингвистов. Это можно объяснить большим вниманием и интересом к сфере перевода и развитию программ автоматического перевода в последние годы.

Особенности автоматизации перевода в мировой лингвистике, создание машинного перевода, систем управления им, программ, машинного и человеческого перевода, конкретных методов, а также создание программ METEO, ALPS, ATLAS, SMARTCAT, SYSTERON, LINGUEE, NTRON и их уникальные функциональные возможности выражены в работах ряда

¹ 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>

английских и российских лингвистов. Первоначально Г.В. Лейбниц и Р.Декарт² выдвинули вопрос о кодировании взаимообусловленности предложений и слов. Также исследования о происхождении машинного перевода, рабочей системе, основных особенностях перевода, способах формирования машинного и человеческого перевода, создании программ перевода, были осуществлены такими английскими учеными, как Дж.С. Загер, Дж.А. Алонсо, П. Беннетт, Р.Т. Белл, Р. Джейкобсон, Дж.Б. Лардж, В. Лоусон³. Проблемы обработки естественного языка широко освещены А. Кейлером, К.В. Линденом, Н. Уордом⁴, М.А. Мори⁵. Из русских ученых в исследовательских раюотах К.К. Боярского нашло свое отражение общее понимание компьютерной лингвистики, М.В. Холховой – основные системы машинного перевода и недостатки этих программ, И.Г. Овчинниковой – новые возможности эффективного использования методов компьютерного перевода и автоматизации перевода, П.В. Хонькиной – использование идиоматического перевода в машинном переводе, Л.И. Човича – особенности перевода текста в английском и русском языках⁶.

В Узбекистане автоматизированную теорию перевода и рабочие системы программ машинного перевода изучалась рядом ученых. Например, Б.Менглиев и Д.Ахмедова⁷ выявили категории семантических тегов в корпусной лингвистике и особенности лексико-грамматической, лексико-тематической и методологической оценки онлайн-корпусов. Г. Собирова остановилась на использовании корпуса ВНС в обучении языку. Дж. Джумабаева изучила градуонимы в языке с использованием корпуса английского языка и определила их семантические свойства. Лингвистические основы создания авторского корпуса исследовала Ш.Хамроева⁸. Н. Атабоев в своих научных исследованиях отмечал, что

² Deskart R. History of machine translation. – Paris: «Pergamon Press» 1983. Article for translators: History of machine translation (translationdirectory.com).

³ Sager J. C. Language engineering and translation consequences of automation. – Philadelphia: «John Benjamin's publishing company» 1994. – P. 243-264. J.A. Alonso “Machine translation technology: on the way of market introduction”. Siemens review. 1987. – P. 54-60. P. Bennett, P. The translation unit in humans and machines. – Manchester: «CCL/UMIST».1993. – P.41-67. R. T. Bell. Translation and translating, theory and practice. – London, «Longman press». 1991. – P. 32-54. R. Jacobson. “On linguistic aspects of translation”. R. A. Brower (Ed). – London, 1966. – P. 62-68. J. B. Large. The foreign language barrier. – London, 1983. – P.102. V. Lawson. Practical experience of machine translation. – Amsterdam: «NorthHolland press». 1982. – P.105.

⁴ Jurafsky D. S. and. Martin J. H. Speech and language processing. Contributing writers: Prentice Hall, Englewood cliffs. – New Jersey, 2000. – P. 950.

⁵ Mohri M.A. Finite-state transducers in language and speech processing. Computational linguistics. – New Mexico. 1994. – С.269-312.

⁶ Боярский К.К. Введение в компьютерную лингвистику. – Санкт-Петербург, 2013. – С. 4-5. Холхова М.В. Анализ особенностей машинного перевода. – Санкт-Петербург. 2016. – С. 7-12. Овчинникова И.Г. Использование компьютерных переводческих инструментов: новые возможности новые ошибки. «Russian journal of linguistics». Хонькина П.В. Перевод идиоматических выражений в системах машинного перевода. Вестник московского университета. – Москва, 2010. – С. 68-71. Л.И.Чович. Перевод как интерпретация и импровизация на материале сербских переводов мемуаров. - вестник московского университета. – Москва, 2012. – С. 68-71.

⁷ Mengliyev B., Ahmedova D. Semantic Tag Categories in Corpus Linguistics: Experience and Examination. International Journal of recent technology and engineering (IJRTE) ISSN: 2277-3878, 2019.

⁸ Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа докт. (PhD) ... дисс. автореф. – Қарши, 2018. – 52 б.

классификация корпусов может основываться на особых принципах и что их категоризация по областям применения в определенном смысле условно доведена до окончательного вида⁹. В исследовании Н. Абдурахмоновой¹⁰ освещены вопросы создания лингвистического обеспечения машинного перевода. Р.Каримов¹¹ научно исследует лингвистические и программные вопросы создания узбекско-английского параллельного корпуса, стратегию отбора текстов, непосредственно переведенных с оригинала на язык перевода, в соответствии с жанрами.

Связь диссертационного исследования с научно-исследовательскими работами высшего образовательного учреждения, где выполнена диссертация. Диссертационная работа выполнена в рамках темы «Теория языка. Направления прикладной и компьютерной лингвистики» научного проекта «Научно-исследовательские, научно-творческие работы, проводимые среди докторантов, самостоятельных соискателей и студентов» плана научно-исследовательских работ Узбекского государственного университета мировых языков.

Цель исследования – выявить функциональные возможности параллельной корпусной системы контекстуального поиска при автоматизации перевода на примере английского и узбекского языков (через платформу Sketch engine).

Задачи исследования:

исследовать теоретические основы области теории языка, прикладной и компьютерной лингвистики в сфере автоматизации перевода;

интерпретировать лексические особенности научных и публицистических текстов в английском и узбекском языках и контекстуально проанализировать в переводе;

на основе корпуса параллельных текстов определить методы, используемые в процессе перевода научных и публицистических текстов в английском и узбекском языках;

проанализировать лексическое значения и контекстуальный перевод десемантизированных слов, употребляемых в научном и публицистическом стиле;

изучить сущность корпусных типов при создании программ-контекстуальных переводчиков;

выявить методы эффективного использования параллельных корпусов в переводческой практике и раскрыть значение платформы Sketch engine при создании языковых корпусов;

⁹ Атабоев Н.Б. Инглиз тили корпусининг функционал хусусиятлари (СОСА мисолида): Филол. фан. докт. ... дисс. – Тошкент, 2020. – 14 б.

¹⁰ Абдурахмонова Н. Машина таржимасининг лингвистик асослари. – Т.: «Akademnashr», 2012. – 25 б.

¹¹ Каримов Р.А. Ўзбек-инглиз параллел корпусини тузишнинг лингвистик ва дастурий масалалари Филол. фан. докт. ... дисс. – Тошкент, 2022. – 7 б.

определить функциональные задачи корпуса параллельных текстов в автоматизации перевода на основе выявления общего и отличий программ контекстуального переводчика Linguee и Reverso;

изучить этапы создания языкового корпуса и программы-переводчика на платформе Sketch engine и создать программы контекстуального переводчика путем размещения английских и узбекских текстов в параллельном корпусе на этой платформе.

Объектом исследования является контекстуальная система параллельного корпуса в английском и узбекском языках.

Предметом исследования являются переводческие особенности и функциональный аспект контекстуальной системы параллельного корпуса английского и узбекского языков.

Методы исследования. При освещении темы исследования были использованы методы эмпирического, наблюдательного, корпусного анализов, методы описания, классификации, контекстуального анализа и компонентного анализа.

Научная новизна исследования заключается в следующем:

обоснована эффективность с точки зрения функциональности способностей Статистической (Statistic machine translation) и Нейронной (Neuron machine translation) систем машинного перевода понимать естественный язык, проверять его автокорректором и использовать большую базу данных, чем системы НМТ (Hybrid machine translation), RBMT (Rule based machine translation) машинного перевода;

разработан механизм включения параллельного корпуса узбекского языка в платформу Sketch engine путем обоснования применения системы машинного перевода Rule based и программных алгоритмов контекстного значения десемантизированных слов, часто встречающихся в научных и публицистических текстах, путем внедрения функций памяти перевода Cat Trados;

совершенствованы этапы создания контекстной программы-переводчика коллокаций в научных и публицистических текстах в английском и узбекском языках на платформе Sketch engine в параллельном корпусе, на основе создания контекстуальной поисковой системы, превращения в электронную систему в таких корпусах, как *Adhog*, *Diy*, *Comparable*, *монолингвальный*, *диахронический*, *синхронный*, *параллельный*, слов, выражений, коллокаций, идиоматических единиц, фраз и выборки текста, а также реализации этапов загрузки в программу;

разработаны экспериментальные практические рекомендации по созданию электронной образцовой базы данных на платформе Sketch engine, состоящей из научных и публицистических текстов по программе контекстного переводчика на английском и узбекском языках путем совершенствования создания методов системного использования на практике общих требований к программам автоматизированного перевода.

Практические результаты исследования заключаются в следующем: усовершенствованы методы, функции и ресурсы типов корпуса в автоматизации перевода;

на основе параллельного корпуса сформирована сжатая база программы контекстного перевода, состоящая из научных и публицистических текстов;

с использованием платформы Sketch engine создан программный механизм контекстного переводчика в двуязычном (англо-узбекском) параллельном корпусе научных и публицистических текстов.;

разработана база коллокаций, относящихся к научным и публицистическим текстам в английском и узбекском языках..

Достоверность результатов исследования определяется принципами изученных исходных материалов и использованными в них методами, тем, что теоретическая информация, полученная из официальных источников, помогла сделать выводы, исходя из характера английского, русского и узбекского языков, охватом теоретических взглядов зарубежных ученых в освещении исследовательских работ, представленными анализами, надежностью ряда методов, используемых при автоматизации перевода, тем, что выводы и рекомендации реализованы на практике, а полученные результаты основаны на практически проверенных источниках.

Научная и практическая значимость результатов исследования. Научная значимость результатов исследования заключается в том, что создание программы-контекстного переводчика корпуса параллельных научных текстов и разработка теоретических основ ее совершенствования служат научно-теоретическим источником для исследований в области переводоведения, корпусной лингвистики и машинного перевода. Также применение методов перевода с английского на узбекский язык для программ контекстного перевода определяется анализом методологических теорий в области переводоведения.

Практическая значимость результатов исследования заключается в том, что они могут послужить источником при создании программ и планов в процессе преподавания теории языка, прикладной лингвистики, компьютерной лингвистики, корпусной лингвистики, машинного перевода, а также при описании тем, широком использовании программ автоматического перевода, а также практика контекстного перевода на двух языках (в английском-узбекском языках) объясняется составлением и созданием электронного словаря.

Внедрение результатов исследования. На основе научных результатов и практических предложений по раскрытию функциональной сущности системы контекстного поиска параллельного корпуса в автоматизации перевода:

выводы об эффективности с точки зрения функциональности способностей Статистической (Statistic machine translation) и Нейронной (Neuron machine translation) систем машинного перевода понимать естественный язык, проверять его автокорректором и использовать

большую базу данных, чем системы НМТ (Hybrid machine translation), RBMT (Rule based machine translation) машинного перевода были использованы в проекте кредитной мобилизации между Узбекским государственным университетом мировых языков и Рурским университетом в Бохуме (Германия) в рамках международного обмена опытом между Европейским Союзом и высшими учебными заведениями Узбекистана (ICM, КА-107 – International Credit Mobility between European Union and Uzbekistan HEIs, through staff and student mobility exchange projects between Uzbekistan State University of World Languages and Ruhr University Bochum (Germany) - 2020-1-DE01-КА 107-005433 программы Erasmus+) (Справка № IRD/23-156 от 17 июля 2023 года). В результате малая база программы контекстного переводчика, исследованная в данной диссертации, используется в рамках проекта в практическом аспекте;

рекомендации о механизме включения параллельного корпуса узбекского языка в платформу Sketch engine путем обоснования применения системы машинного перевода Rule based и программных алгоритмов контекстного значения десемантизированных слов, часто встречающихся в научных и публицистических текстах, путем внедрения функций памяти перевода Cat Trados были применены в проекте кредитной мобилизации между Узбекским государственным университетом мировых языков и Рурским университетом в Бохуме (Германия) в рамках международного обмена опытом между Европейским Союзом и высшими учебными заведениями Узбекистана (ICM, КА-107 - International Credit Mobility between European Union and Uzbekistan HEIs, through staff and student mobility exchange projects between Uzbekistan State University of World Languages and Ruhr University Bochum (Germany) - 2020-1-DE01-КА107-005433 программы Erasmus+) (Справка № 06/556 от 3 октября 2023 года). В результате вышеупомянутые типы корпусов были использованы в качестве эксперимента в рамках программы;

экспериментальные практические рекомендации по созданию электронной образцовой базы данных на платформе Sketch engine, состоящей из научных и публицистических текстов по программе контекстного переводчика на английском и узбекском языках путем совершенствования создания методов системного использования на практике общих требований к программам автоматизированного перевода, были использованы при реализации мероприятий по конкурсам переводов научных и публицистических текстов в параллельном корпусе, направленных Союзом писателей Узбекистана стремиться глубоко отразить древнюю историю и богатую культуру нашего народа, коренные изменения, достигнутые самоотверженным трудом нашей страны, учитывая интересы молодежи к переводу (Справка Союза писателей Узбекистана № 01-03/3150 от 28 ноября 2023 года). В результате общие выводы исследования по разработке предложений и рекомендаций по созданию программного механизма контекстного переводчика параллельного корпуса английского и узбекского

языков впервые на платформе Sketch engine служат практическим примером, теоретическим ресурсом по переводу научных и публицистических текстов для молодых писателей.

Апробация результатов исследования. По результатам исследования опубликованы 10 научных статей, из них 4 обсуждались на международных и 6 республиканских научно-практических конференциях.

Опубликованность результатов исследования. По теме диссертации опубликовано 10 научных работ, в том числе 5 статей в изданиях, рекомендованных Высшей аттестационной комиссией Республики Узбекистан для публикации основных научных результатов докторских диссертаций, из них 4 – в республиканских и 1 – в зарубежном журнале.

Структура и объем диссертации. Диссертация состоит из введения, трех глав, заключения и списка использованной литературы. Общий объем исследования составляет 178 страницы.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обосновываются актуальность и востребованность темы исследования, определяются цель и задачи, объект и предмет, методы исследования, а также его научная новизна, практическая значимость, характеризуются степень изученности проблемы, достоверность и обоснованность полученных результатов, приводятся сведения о внедрении результатов исследования в практику, об апробации и опубликованный результатов исследования, структуре и объеме диссертации.

В первой главе диссертации под названием **«Концепция автоматизации перевода в прикладной лингвистике»** анализируются история машинного перевода, этимология, этапы его изобретения, теоретический анализ проблемы классификации автоматизации перевода, функциональные этапы механизмов машинного перевода.

Этимологически термин «машинный перевод» изначально использовался для обозначения только автоматических систем, работающих без вмешательства человека. Европейская ассоциация машинного перевода (ЕАМТ) определяет машинный перевод как «использование компьютера для перевода текста с исходного языка на целевой язык»¹². В свою очередь, Международная ассоциация машинного перевода (International Association for Machine Translation IAMT)¹³ рассматривала машинный перевод как «создание варианта перевода путем одновременного ввода полного предложения»¹⁴.

По А.В. Семенову, «основная цель систем машинного перевода – качественно перевести исходный язык в эквивалентный текст целевого языка. Однако перевод, выполненный системой автоматизированного перевода, не идеален и переведенный текст требует дальнейшего редактирования. Иногда к переведенным текстам не предъявляются высокие требования к точности перевода и уникальному стилю изложения; в основном этого перевода достаточно, чтобы представить читателю общее «содержание» в виде «отчета». В результате переведенный «сырой» текст принимается»¹⁵.

Ч. Бэббидж в основном работал над проектом цифровой аналитической машины в середине XIX века и представил публике этот проект цифровой аналитической машины в 1855 году на Парижской

¹² ЕАМТ – The European Association for Machine Translation - <https://eamt.org/>

¹³ IAMT– International Association for Machine Translation

¹⁴ Семенова А.В. Курейчик В.М. Обзор и анализ состояния проблемы обработки текстовой информации в системах машинного перевода. Информатика, вычислительная техника и инженерное образование. – М. 2014. – С. 2.

¹⁵ Семенова А.В. Курейчик В.М. Обзор и анализ состояния проблемы обработки текстовой информации в системах машинного перевода. Информатика, вычислительная техника и инженерное образование. – М. 2014. – С. 2-3.

выставке. Ч. Бэббидж пришел к выводу, что такую машину можно также использовать для хранения словарей¹⁶.

В 1933 году автор лингвистического арифмометра П. Смирнов-Троянский обратился в Бывшую Российскую Академию наук с разработанным им проектом «машины для подбора и печатания слов при переводе с одного языка на другой»¹⁷. Также П.П. В последнем квартале 1933 года Смирнов-Троянский создал систему на основе языка эсперанто, включающую метод распределения между языками с методом грамматического управления и двуязычный словарь. Эта система включает в себя три этапа: *на первом этапе* редактор систематизирует слова основного языка и определяет логические формы и синтаксические функции словесных единиц; *на втором этапе* машина преобразует лексические формы в переводимый язык; *на третьем этапе* редактор нормализует результат¹⁸.

Машинный перевод как научная дисциплина возник в 1946 году. Директор Рокфеллеровского научного департамента фонда Рокфеллера У.Уивер в переписках с Э.Бутом и Н.Винером впервые упоминается понятие машинного перевода¹⁹. В то же время эти исследования открыли новый этап в машинном переводе.

Во взглядах И.А. Мельчука и Ю. Апресяна нашло своё отражение, что метаязык – это модель языковых понятий, гипотетический язык, записывающий и хранящий понятия в психике человека²⁰. Именно в эти годы в Российском научно-исследовательском институте в Ленинграде была создана экспериментальная лаборатория машинного перевода, состоящая из математиков и механиков.

Существует два основных фактора, лежащих в основе развития систем машинного перевода в мире: 1) чисто научный; это определяется сложностью и многофункциональностью моделирования компьютерного перевода; 2) учитывая широту возможностей машинного перевода для обеспечения межъязыкового общения в социальных целях, его размеры связаны с растущей ролью практики машинного перевода в современном мире как необходимого условия межъязыкового общения, которая возрастает с каждым годом.

В результате непрерывных исследований учёных систем МП им удалось ввести машинный перевод в класс искусственного интеллекта. Уже сейчас слово машина в сочетании «машинный перевод»

¹⁶ Bromley A. G. “Charles Babbage’s Analytical Engine, 1838”. Annals of the history of computing. Cambridge. – Massachusetts: «MIT Press», 1990. – P.196-217.

¹⁷ Колосёнок П. Системы машинного перевода. – Р.: 2019. Системы машинного перевода — Lingva (msu.ru).

¹⁸ Troyanskii P. P. A forgotten pioneer of mechanical translation. Machine translation. – Moscow, 2000. – P.187-221.

¹⁹ Norbert W. C. Control and communication in the animal and the machine. – Cambridge, 1961. – P. 94-95. https://ru.Cybernetics:_Or_Control_and_Communication_in_the_Animal_and_the_Machine.

²⁰ Апресян Ю.Д. О языке для описания значений слов. – Москва.: «Известия АН СССР». Серия языка и литературы, 1966. – С. 415-429.

рассматривается как исполнитель процессов чтения, проверки, перевода и редактирования в смысле *искусственного интеллекта*. Одним из ключевых компонентов компьютерного перевода является наличие базы данных в памяти. Также, называется CAT – компьютерный перевод, машинный перевод (MAT machine-assisted translation) или автоматизированный перевод человека МАНТ (machine-aided human translation). В аналитической схеме ниже показан механизм работы CAT:

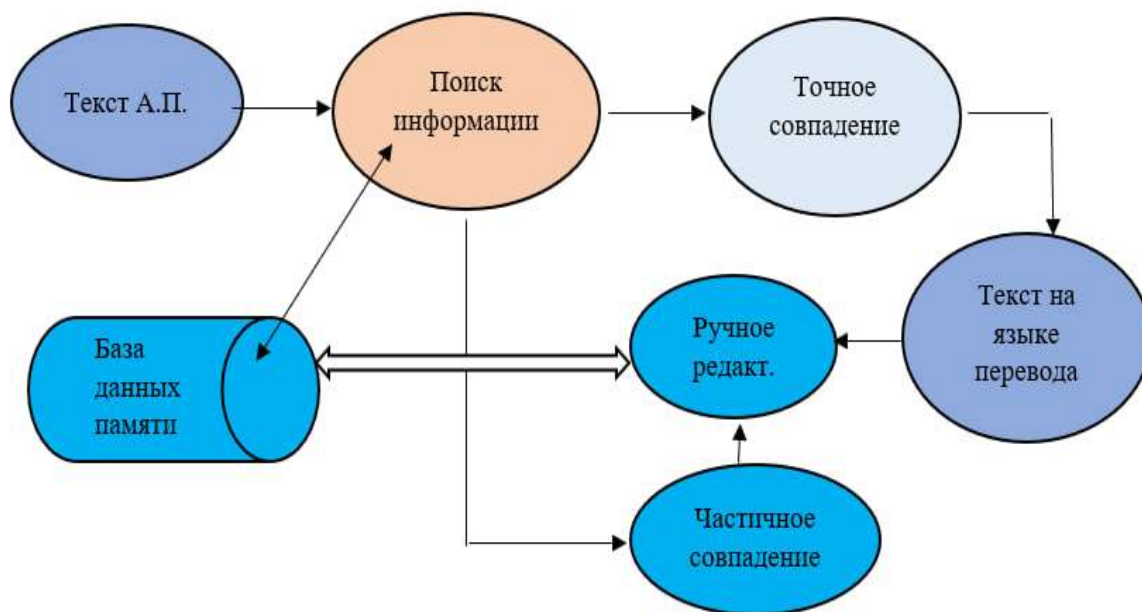


Рисунок 1. Механизм работы компьютерного перевода.

Нейронный машинный перевод NMT (Neural Machine Translation) — новейший этап машинного перевода, представляющий собой развитие искусственного интеллекта. Его теория основывается на понимании естественного языка (NLU – natural language understanding), обработке естественного языка (NLP – natural language processing), памяти перевода машинного перевода МП и машинном переводе на основе статистики (SMT – statistics-based machine translation) и теории и технике глубокого обучения²¹.

В настоящее время существует четыре основных типа МП:

- 1) перевод с использованием электронных словарей (machine-assisted human translation – МАНТ) – использование компьютерных электронных словарей и входящих в них инструкций при переводе текста человеком;
- 2) перевод с использованием компьютерных и программных приложений (computer assisted-translation – CAT) – служит для перевода текста с помощью компьютера и выделения основных понятий в нем понятий;

²¹ Halliday M. Linguistics and machine translation . In J. Webster computational and quantitative studies. Beijing: «Peking University Press». – China. 2007. – P.166.

3) машинный перевод с редактором (human-assisted machine translation – НАМТ) назначение задачи редактирования человеческим ресурсом переводу, выполняемому компьютерным программным обеспечением;

4) полностью автоматический машинный перевод (fully automatic machine translation – FAMT) – выполнение задач по переводу и редактированию текста через программу-переводчик.

Системы машинного перевода характеризуются определенными аспектами:

1) участие языков: а) двуязычный; б) многоязычный; 2) объем выбранной темы: а) относящийся к определенному стилю и области; б) относящийся к одному стилю и множеству областей; 3) относящийся к разным стилям и областям²².

Индийские учёные Р. Шанкаравелаутан и Дж. Хатчинс различают системы машинного перевода следующим образом²³:

1. МП, основанный на правилах (Rule-Based MT): 1) трансферная система (transfer-based machine translation); 2) интерлингвальная система (interlingual machine translation); 3) система, основанная на словаре (dictionary-based machine translation).

2. Корпусный машинный перевод (Corpus based MT): 1) МП, основанный на статистике (statistics-based); 2) образцовый МП (example-based);

3. Гибридный машинный перевод (Hybrid machine translation НМТ). Гибридный машинный перевод – это система, которая использует технологию перевода, основанную на правилах и статистике. Эту систему в настоящее время используют несколько компаний машинного перевода (Asia Online, LinguaSys, Systran, PangeaMT, UPV, Logos)²⁴.

Идея статистического машинного перевода в основном исходит из теории информации. Системы статистического перевода характеризуются одновременным использованием массивов текстов, представленных двумя языковыми версиями (параллельный корпус). По мнению А.В. Семеновой²⁵, данная модель дает возможности улучшить перевод за счет использования наиболее часто употребляемых слов на разных языках с учетом соответствующих частот при переводе документа. К разработчикам систем машинного перевода на основе технологии статистического машинного перевода относятся *Google, Microsoft, Language Weaver, Яндекс* и другие сайты и программы.

²² Галактионов В.А., Мусатов А.М. Система машинного перевода “Кросслятор 2.0” и анализ ее функциональности для задачи трансляции знаний. – Москва, 2007. – С. 13.

²³ Prof. Rajendran S. English to tamil machine translation system using parallel corpus. Amrita University. – Coimbatore, 2019. – P. 45.

²⁴ Hutchins J. Machine translation: past, present, future. – London, 1986. – P.118.

²⁵ Семенова А.В. Курейчик В.М..Обзор и анализ состояния проблемы обработки текстовой информации в системах машинного перевода. Информатика, вычислительная техника и инженерное образование. – М. 2014. – С. 9.

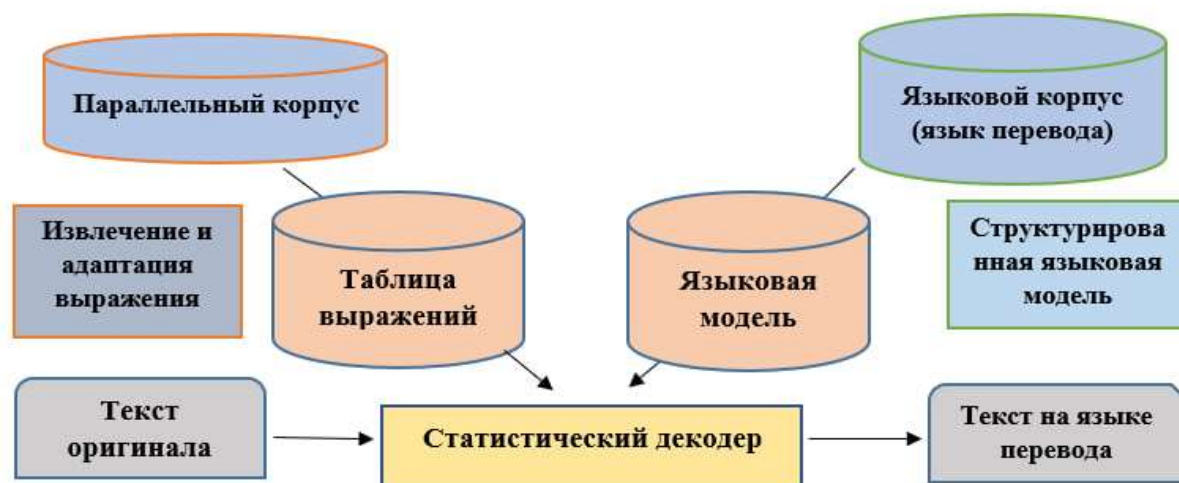


Рисунок 2. Типичная архитектура систем статистического машинного перевода, Компания PROMT

В ходе исследования были проанализированы стратегии поэтапного развития систем МП с научно-теоретической точки зрения, поэтому можно наблюдать развитие нескольких подходов и систем к задаче МП на протяжении многих лет.

Сегодня наиболее развитыми системами являются МП на основе правил (RBMТ), системы нейронного МП и статистические МП. Хотя они понимаются как частично совпадающие системы, общая точка зрения гласит, что МП получает дальнейшее развитие в результате объединения как статистического, так и лингвистического систематического подхода. В результате новые возможности программирования и компьютерных технологий дадут возможность усовершенствовать теорию и практику машинного перевода.

Во второй главе под названием «**Место корпусной лингвистики в автоматизации перевода**» рассмотрены роль корпусных типов в программах-контекстных переводчиках и концепция их использования, методология КЛ, методы анализа корпусных типов и возможности использования параллельных корпусных ресурсов в переводческой практике, а также изучаются функциональные характеристики программ-контекстных переводчиков *Linguee* и *Reverso*.

В настоящее время создано несколько типов корпусов, которые используются в сфере перевода с разными функциями. Например, *параллельный корпус*, *одноязычный*, *двуязычный*, *общий*, *специализированный*, *корпус Ad Hoc*, *корпус DIY* функционально отличаются от корпусов других типов.

М. Бейкер²⁶, Г. Тури, М. Шлезингер, К. Мальмкьер²⁷ и другие ученые выдвинули идею использования параллельного корпуса для изучения возможностей перевода и определили следующую гипотезу об универсальности

²⁶ Baker, M. Corpora in Translation studies: An Overview and Some Suggestions for Future Research. Target. – Amsterdam, 1995. – P. 223-243.

²⁷ Malmkjaer K. Linguistics and the language of translation. 2005.

перевода: а) *упрощение параллельного корпуса*, при котором переводчики неосознанно упрощают язык и информацию; б) *простое объяснение параллельного корпуса*, при котором переводчики стараются сделать содержание более понятным в переводе; в) *нормализация параллельного корпуса*, в котором перевод имеет тенденцию быть промежуточным, а не периферийным.

Системы, создающие параллельные корпуса, способны находить параллельные предложения или даже целые абзацы, а не только частично необходимые примеры. Например, можно описать параллельную корпусную поисковую систему «*ParaConc*», которая отдельно отображает информацию о частотности слов и их сочетаний. Системы ранжирования контекстов используют статистику состояний перевода. Такие системы обычно предназначены для лексикографов и профессиональных переводчиков. Например, для поиска параллельных англо-китайских корпусов были созданы две веб-системы – *DOMCAT* и *TotalRecall*. *DOMCAT* находит наиболее релевантные контексты с помощью статистического теста *Dyes* (коэффициент *Sorensen-Dyes*).

$$QS = \frac{2 |X \cap Y|}{|X| + |Y|},$$

$|X|$ и $|Y|$ определяют частоты слов, эквиваленты перевода соответственно, а также $|X \cap Y|$ – определяет частоту совместного появления этих слов.

Однако база данных веб-интерфейса не содержит никакой информации о рейтинге и о том, как работает система.

Вторая система – *TotalRecall* – эта система предлагает две опции: сортировку по частоте и по *логарифмическому* коэффициенту правдоподобия (*log-likelihood ratio*, LLR). Эти два явления, по теории Т. Данна²⁸, являются окончательным измерением отношения между появлением слова и его переводом.

$$LL(a, b) = \sum_{i \in \{a, -a\}, j \in \{b, -b\}} 2C(i, j) \log \frac{C(i, j)N}{C(i)C(j)},$$

Здесь $C(i, j)$ – совокупная частота слова и перевода, а $C(i)$, $C(j)$ – частоты i и j соответственно.

В настоящее время параллельные корпуса *DOMCAT* и *TotalRecall* являются одними из наиболее современных корпусов переводов, созданных для того, чтобы предоставить потребителю высококачественные варианты перевода в конкретном контексте перевода.

Linguee – это программа контекстного перевода, построенная на двуязычном параллельном корпусе. Автоматическое создание словаря в *Linguee* включает следующие шаги: 1) параллельная предварительная доработка корпуса (извлечение таблицы фраз, морфосинтаксический анализ,

²⁸ Dunning, T. Accurate methods for the statistics of surprise and coincidence. Computational linguistics. MIT Press. – Massachusetts 1999. – P. 61–74

лемматизация, фильтрация эквивалентов перевода); 2) объединение переводных эквивалентов каждого слова в ряд синонимов; 3) добавление грамматических знаков; 4) создание обратных переводов, то есть синонимов вступительного слова. В этом случае данный этап осуществляется на основе прямых переводов; 5) иллюстрация каждой группы переводов с примерами использования параллельного корпуса.

Одним из преимуществ переводчика Linguee над Reverso Context является то, что когда пользователь ищет сложные лексические слова, коллокации или фразы, примеры предложений становятся более заметными, и потребителю легче извлекать, читать и использовать эти коллокации и фразы. Кроме того, пустое пространство над примерами предложений, показанными в этом словаре, сокращено до минимума.

По словам ученого А. Сегуна, «Reverso Context – это портал языковых инструментов, включающий программу-переводчик, выполняющую задачу перевода на различные языковые комбинации, такие как французский, испанский, немецкий, русский, португальский, китайский, арабский, японский и другие языки»²⁹. Функциональные различия между Reverso Context и Linguee заключаются в том, что Reverso – это вебсайт-переводик, специализирующийся на автоматическом переводе и изучении языков. А Linguee в основном специализируется на контекстуальном переводе, формируемом на основе *юридических, дипломатических и официальных текстов*.

Третья глава называется **«Функциональные возможности системы контекстуального поиска параллельного корпуса»**. В ней проанализированы принципы создания контекстуальной программы в параллельном корпусе, размещение коллокаций в параллельном корпусе, методы поиска, инструкции по ее использованию, а также лексические особенности и анализы контекстуального перевода научных и публицистических текстов, этапы создания программы-контекстного переводчика в английском и узбекском языках, стратегии создания параллельного корпуса на платформе Sketch engine. Обращено внимание на особенности параллельных корпусных систем контекстуального поиска, контекстуального перевода корпусов научно-публицистических параллельных текстов английского и узбекского языков, которые являются объектом и предметом исследования. Согласно этому, для формирования механизмов и небольшой базы программы автоматизированного контекстного переводчика в параллельном корпусе была выбрана платформа Sketch engine.

Коллокация образуется в результате соединения двух и более слов и в процессе изменения значений слов в результате пересмотра отдельных компонентов. По мнению О. Делла и М. Карти, коллокация определяется как естественное сочетание слов. Это означает, что слова в английском языке тесно связаны друг с другом. Например, *go hand in hand, as do commit and crime, blond goes with hair, and rain is heavy*. О. Делл и М. Карти разделили коллокацию на

²⁹ Segun.A... La traduction automatique par opposition à la traduction humaine : analysed'un corpus de productions réelles. International journal of humanities and social science (IJHSS). – Nigeria, 2014. – P.213-222.

три категории. Это *сильные сочетания (strong collocations)*, *фиксированные сочетания (fixed collocations)* и *слабые сочетания (weak collocations)*.

П. Ньюмарк считает словосочетание «нервами текста и лексики»³⁰. Коллокации делятся на лексические и грамматические категории. Лексические и грамматические коллокации можно отличить друг от друга по следующим признакам. Лексические коллокации создаются только на основе *глаголов, прилагательных, существительных и наречий (verbs, adjectives, nouns, and adverbs)*. Структура грамматических коллокаций объединяет такие части речи, как *глаголы, прилагательные или существительные (verbs, adjectives, or nouns combined with a preposition or a grammatical structure)*, с главным предложением или грамматической структурой.

Поэтому в нашем исследовании была проанализирована коллокация «*deal with*» в аннотационной части статьи, переведенной с английского на узбекский язык, состоящей из научных и публицистических текстов. В этом случае мы можем наблюдать, что эта коллокация контекстуально переводится в нескольких значениях в процессе перевода. Варианты перевода коллокации «*deal with*» можно увидеть в 1- таблице ниже:

Таблица 1

Контекстуальный перевод коллокации “*deal with*”

<i>The article deals with</i> the aesthetic views of English writer Oscar Wilde	<i>Maqolada</i> ingliz yozuvchisi Oskar Vayldning estetik qarashlari <i>o‘rganilgan</i> .
<i>This article deals with</i> the issues of fiction texts, their specific features and multidimensional structure	<i>Ushbu maqola</i> badiiy matnlar va ularning ko‘p qirrali tarkibi tahlili masalalariga <i>bag‘ishlangan</i> .
<i>This study deals with</i> the determination of pedagogical methods, development of the technology of formation of aesthetic qualities of a person with the help of a foreign language during the educational process and outside the educational process; moreover, the criteria for the formation of civic qualities of students were determined.	<i>Mazkur maqolada</i> aynan texnika oliy o‘quv yurtlari talabalarida yuksak insoniy fazilatlarni rivojlantirish tamoyillari va estetik tarbiyani shakllantirishda yuzaga keladigan muammolarni bartaraf etish yo‘llari va ishlab chiqilgan texnologiyalar, mezonlar xususida batafsil <i>so‘z yuritiladi</i> .
<i>The article deals with</i> the information on religiously marked allusions, their symbolic aspects in English literary texts and analysis from the perspectives of cognitive linguistics	<i>Ushbu ishda</i> diniy markerlangan allyuziyalar, ularning ingliz badiiy matnlaridagi ramziy ko‘rinishlari haqida ma’lumot berilgan hamda ushbu til birliklari kognitiv tilshunoslik nuqtai nazaridan <i>tahlil qilingan</i>
<i>The study deals with</i> the meanings and features of the terms “dialogue” and “dialogic speech”.	<i>Maqola</i> «muloqot» va «dialogik nutq» atamalarining mazmuni va xususiyatlarini <i>ochib bergan</i> .
<i>This article deals with</i> the conditions for the emergence of psycholinguistics, that is, the direction of lingu-psychology, its uniqueness and its necessity today.	<i>Ushbu maqolada</i> psixolingvistika, ya’ni lingvopsixologiya yo‘nalishining yuzaga kelish shartlari, o‘ziga xosligi va bugungi kundagi zarurati xususida <i>so‘z boradi</i> .

³⁰ Newmark P.A. Textbook of translation. – New York.: «Prentice Hall International press». 1988. – P. 186.

В приведенном выше анализе коллокации «*deal with*» можно увидеть несколько значений перевода в научной статье. То есть в коллокации использовано в научной статье контекстуальный прием перевода типа «*so‘z yuritmoq, o‘rganmoq, bag‘ishlamoq, tahlil qilmoq, ochib bermoq*», и за счет придания нескольких значений этой коллокации автором были созданы варианты перевода в статье. На основании вышеизложенного анализа, если классифицировать коллокацию «*deal with*», то «*deal with*» относится к категории грамматически ограниченных коллокаций. Ограниченные коллокации состоят только из одного компонента, один глагол сочетается с несколькими существительными или предлогами. Также мы можем увидеть несколько важных особенностей коллокации в переводческой практике.

Раскрытие лексического значения и контекстуального анализа в параллельном корпусе десемантизированных групп существительных и глаголов “*point*”, “*matter*”, “*thing*”, “*discuss*” “*analyse*”, “*examine*” “*deal*”, является одним из основных задач данной исследовательской работы. Десемантизация – «от лат. приставки *de-* - отделять, отменять, от греч. *semantikos* – означающий» и в основном означает утрату, изменение лексико-семантического значения в процессе перевода.

Предназначен анализ лексического значения и контекстуального перевода слов аннотационного раздела научных статей, а также используемых в публицистическом стиле. С этой точки зрения роль контекста при переводе этих слов очень важна. Это можно наблюдать в процессе контекстуального перевода таких существительных и глаголов, как “*point*”, “*matter*”, “*thing*”, “*discuss*” “*analyse*”, “*examine*” “*deal*”. В некоторых случаях эти слова выполняют только грамматическую функцию и называются вспомогательными, базовыми словами. Например, в сочетании “*poor (little) thing*”, существительное “*thing*” являясь главным словом, используется как вспомогательное слово. Комбинация “*poor (little) thing*” переводится на узбекский язык как “*bechora*» одним словом. То же самое можно сказать и о узбекском переводе комбинации «*just the thing*», которая на английском языке означает «*такая вещь; то, что нужно*». Благодаря неоднозначности этого типа значений, эти слова имеют широкий диапазон сочетаемости, что, в свою очередь, создает более широкие возможности в процессе перевода. Поэтому их перевод, как правило, зависит от контекста, в котором они употребляются. С этой точки зрения контекстуальный перевод помогает решить проблемы, возникающие в этом процессе. Он также предоставляет пользователю точные варианты перевода. В этом можно убедиться, проанализировав контекстуальные переводы приведенных выше слов. Проанализированы следующие виды контекстуального перевода имени существительного “*point*” в параллельных текстах.

Таблица 2

Контекстуальный перевод слова “Point”

<i>The lawyers may be right in questioning the legality of some minor points.</i>	<i>Yuristlar ba'zi kichik qonun hujjatlarining qonuniyligini haqli ravishda shubha ostiga qo'yishadi.</i>
<i>This point became particularly apparent in analyses of the Chicago school data.</i>	<i>Bu savol Chikago maktabidagi ma'lumotlarini tahlil qilishda yaqqol namoyon bo'ldi</i>
<i>From their point of view, the main function of the law is to limit the power of the authorities and the courts to inflict evils upon the individual citizen and to restrict his freedom.</i>	<i>Ularga ko'ra qonunning asosiy vazifasi hokimiyat va sudlarning har bir fuqaroga yomonlik qilish va uning erkinligini cheklash vakolatlarini chegaralashdan iborat.</i>
<i>Race is, of course, a sensitive point of reference in American society, and a reference point that can become cute for minority youth during early to middle adolescence.</i>	<i>Poyga, shubhasiz, Amerika jamiyatida nozik nuqta hisoblanib, u erta yoshdan o'rta o'smirlik davridagi ozchilik yoshlar orasida keskin kuzatilishi mumkin bo'lgan nozik jihatlardan biridir.</i>
<i>Patterson's commitment to the ideal of integration implies his belief that such a tipping or turning point exists.</i>	<i>Pattersonning integratsiya idealiga sodiqligi uning bunday burilish nuqtasi mavjudligiga ishonchi komil ekanligini bildiradi.</i>

Можно привести примеры контекстуального перевода лексической единицы «*matter*», обозначающей несколько значений. В переводе на узбекский язык лексическая единица «*matter*» обозначает “*vaziyat*”, “*holat*”, “*mazmun*”, “*ish*” “*materiya*”, “*modda*” и “*savol*” («*ситуация*», «*состояние*», «*содержание*», «*работа*», «*материя*», «*статья*» и «*вопрос*») (*wisdom dictionary*). Однако следующие примеры показывают, что зачастую контекстуальное значение данного десемантизированного слова не соответствует ни одному из приведенных выше значений. Зачастую – чтобы добиться адекватного перевода, эта лексическая единица не переводится на узбекский язык напрямую, а переводится контекстуально. Примеры контекстуального перевода лексической единицы «*matter*» в параллельных текстах даны следующим образом:

Таблица 3

Контекстуальный перевод лексической единицы “Matter”

<i>It would be a delicate matter to express an opinion about the creativity of younger men.</i>	<i>Yosh xodimlarning ijodi haqida fikr bildirish juda nozik masala.</i>
<i>It has much less experience in this matter than the classical countries of bureaucracy, France, Germany, Austria, and Russia, acquired.</i>	<i>Bu borada Fransiya, Germaniya, Avstriya va Rossiya kabi klassik byurokratiya davlatlariga qaraganda bu davlatning tajribasi ancha kam.</i>
<i>As head of a bureau, he may have the power to alter some minor rules and some matters of internal procedure</i>	<i>U Agentlik rahbari sifatida muayyan qoidalarga va ichki tartib-qoidalarga taalluqli bo'lgan moddalarga kichik o'zgartirishlar kiritish huquqiga ega bo'lishi mumkin.</i>
<i>In violating the law you are perhaps worse than a good many of the racketeers, no matter how good your intentions may be.</i>	<i>Qanchalik yaxshi niyatda bo'lishingizdan qat'iy nazar, agar qonunni buzsangiz o'sha reketlardek bo'lgan bo'lasiz.</i>
<i>But it is another matter with the local chief of a government agency.</i>	<i>Mahalliy davlat organi rahbari bilan vaziyat butunlay boshqacha.</i>

Что касается контекстуально-переводческого анализа лексической единицы «*thing*», то ее контекстуальное значение в текстах более четкое, чем рассмотренной выше лексической единицы «*matter*». В словаре слово «*thing*» переводится в основном как “*narsa*”, “*mavzu*”, “*ish*” («*вещь*», «*предмет*», «*работа*») (*wisdom dictionary*). Однако, исходя из контекста, в большинстве случаев данное слово обычно не переводится как самостоятельная единица. Это можно проанализировать следующим образом на нескольких примерах:

Таблица 4

Контекстуальный перевод лексической единицы “*Thing*”

<i>Within a business concern, such things can be left without hesitation to the discretion of the responsible local manager</i>	<i>Biznes doirasida, mahalliy hokimiyatlar bu kabi savollarni ikkilanmasdan ko‘rib chiqishi mumkin.</i>
<i>The same thing is essentially valid for a democratic government.</i>	<i>Demokratik hukumat uchun ham xuddi shunday bo‘ladi.</i>
<i>How can these two things, one good, the other bad, be linked together?</i>	<i>Biri yaxshi va boshqasi yomon kabi bu ikki qarama qarshilikni qanday bir biriga bog‘lash mumkin?</i>
<i>Government efficiency and industrial efficiency are entirely different things</i>	<i>Hukumat faoliyati va sanoat ko‘rsatkichi bir biridan butunlay farqlidir.</i>
<i>It may well be that the greatest thing in Atlantis is its better system of government.</i>	<i>Atlantisning eng katta yutug‘i uning yaxshi ishlaydigan boshqaruv tizimi bo‘lishi mumkin.</i>
<i>There are many things about government administration that need to be reformed</i>	<i>Hukumat boshqaruvida isloh qilinishi kerak bo‘lgan juda ko‘p tizimlar mavjud.</i>
<i>However, it is quite a different thing if an officeholder evades the duly promulgated laws of the nation for the benefit of the “State”.</i>	<i>Agar mansabdor shaxs “davlat” manfaati uchun tegishli tarzda e‘lon qilingan qonunlardan qochsa, unda bu butunlay boshqa vaziyat.</i>

Проанализированы контекстуальные переводы аннотаций на английский и узбекский языки, являющиеся частью второго анализа исследования. Проанализированы контекстуальные варианты перевода ранее упомянутых слов, таких как “*discuss*” “*analyse*”, “*deal*”. Глагол “*discuss*” имеет в словаре (*wisdom dictionary*) такие варианты перевода, как “*muhokama qilmoq*”, “*tahlil qilmoq*”, “*ko‘rib chiqmoq*” («*обсуждать*», «*анализировать*», «*рассматривать*»). Чаще всего в переводных вариантах научных текстов слово «*discuss*» переводится в значениях, приведенных выше, но иногда это слово встречается в научных текстах и в других значениях. Контекстуальный вариант этого слова переводится в следующих значениях:

Контекстуальный перевод глагола “Discuss”

<i>This article discusses the economic efficiency of the world’s most popular interactive TV projects.</i>	<i>Mazkur maqolada jahonda keng ommalashgan interaktiv teleloyihalarning samaradorlik masalalari ko‘rib chiqildi</i>
<i>The work discusses the views on cognitive linguistics, its subject, and tasks, the essence of the concept of “cognitivism”, the problem of “concept” and “cognition”, and the features of cognitive linguistics in contrast to traditional linguistics.</i>	<i>Mazkur maqolada so‘nggi yillardagi tub burilish bilan bog‘liq holda yuzaga kelgan kognitiv tilshunoslikka doir fikrlar, uning o‘rganish obyekti va vazifalari, “kognitivizm” tushunchasining mohiyati, unda “tushuncha” va “konsept” muammosi hamda kognitiv tilshunoslikning an‘anaviy tilshunoslikdan farqli belgilari, kognitiv tilshunoslikning yo‘nalishlari xususida so‘z yuritiladi. https://journal.fledu.uz</i>
<i>This paper discusses the work of A. Semyonov was a scientist who made a significant scientific and practical contribution to the development of the science of the Uzbek and Tajik peoples.</i>	<i>Ushbu maqolada o‘zbek va tojik xalqlari ilm-fani rivojiga ilmiy-amaliy jihatdan katta hissa qo‘shgan olim A. Semyonovning Navoiyshunoslik, manbashunoslik faoliyati o‘ganildi. https://journal.fledu.uz</i>

Таким образом, мы пришли к выводу, что качество перевода научной статьи зависит от степени достижения коммуникативной цели, максимальной содержательной сущности перевода текста с языка оригинала на целевой язык посредством метода контекстуального перевода, точности перевода и общности смыслового содержания научно-публицистического текста. Одной из важных задач является формирование небольшой базы программы контекстного переводчика в параллельном корпусе этих переведенных текстов. Эффективного результата можно добиться, разместив эти двуязычные тексты на платформе Sketch Engine для создания параллельного корпуса.

Sketch Engine считается одной из ведущих платформ для этого корпуса и широко используется в лексикографии. Sketch Engine состоит из двух функций: *программного обеспечения и веб-сервиса*. Это программное обеспечение и веб-сервис включают в себя большое количество предварительно загруженных и готовых к использованию корпусов, а также бесплатные инструменты и задачи для создания, установки и управления корпусами. Sketch Engine в основном предоставляет следующие услуги:

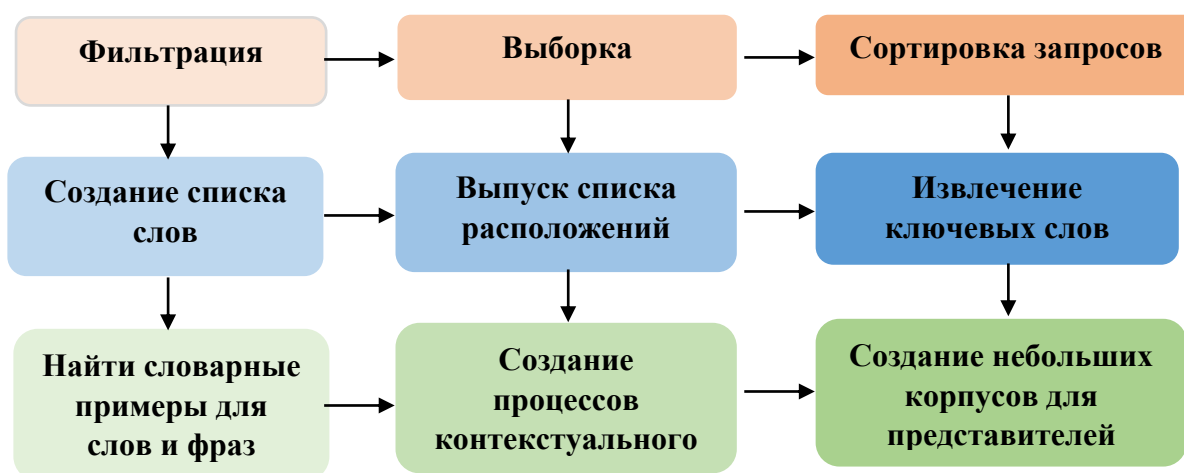


Рисунок 3. Задачи платформы Sketch engine

В то же время Sketch Engine имеет множество возможностей для объединения инструментов разметки на нескольких языках, добавления терминологии, работы с словосочетаниями, идиомами, создания эскизов слов и тезаурусов, а также удобного предоставления пользователям своего рабочего механизма. Все они сейчас благополучно работают, например, корпус *Brown* и корпус *enTenTen*, которые в настоящее время используются многими пользователями.

Вся информация, размещенная в Sketch Engine, представлена в цифрах

Токены - 17,329,301
Слова - 15, 513,307
Фразы - 615, 181
Параграфы - 223,394
Документы - 8,016

В нашем исследовании создание программы контекстного перевода на платформе Sketch Engine в параллельном корпусе для автоматизации перевода осуществлялось в следующие этапы:

1) создана база данных, состоящая из двуязычного английского и узбекского языков с научными и публицистическими текстами в виде таблицы *Microsoft Excel*, целью создания этих текстов было создание базы данных памяти, в основном собранной из двуязычных текстов, в *Microsoft Excel*;

2) была перепроверена совместимость перевода собранных данных;

3) после того, как проверенные переводы совпадали друг с другом, была создана база данных на компьютере.

Формат данных должен быть в формате *data*, описанном в разделе формат данных для параллельного корпуса. Поэтому представим формат информации в виде *data*. После выполнения этих задач важно предпринять следующие шаги по созданию параллельного двуязычного корпуса: 1) нажав кнопку «*new corpus-новый корпус*» на панели инструментов корпуса;

2) нажав мультиязычную кнопку; 3) введя название корпуса и выбрав файл; 4) проверив правильность языков на следующем экране; 5) указав команду «next»; 6) завершается процесс проверкой созданного корпуса путем идентификации созданного двуязычного параллельного корпуса из раздела *My corpus*.

Кроме того, платформа Sketch engine может быть реализована с другими типами форматов. Например, *XLIFF* (версия 2.0 и выше), а также *CSV*, *TSV*, *XLSX* – это типы документов в виде обычного текста, используемые для хранения табличных данных.

В результате нашего исследования с использованием вышеизложенных рекомендаций был создан интерфейс параллельного корпуса контекстуального перевода научных и публицистических текстов на английский и узбекский языки. Интерфейс этой программы и способы ее использования показаны на следующих картинках.



Рисунок 4. Интерфейс параллельного корпуса



Рисунок 5. Интерфейс параллельного корпуса, состоящего из научных текстов на английском и узбекском языках

Механизм платформы Sketch Engine сформирован за счет параллельного корпуса программы контекстного перевода, состоящего из научных и публицистических текстов в английском и узбекском языках. Инструкции по использованию этой программы следующие: 1) *войти на сайт платформы Sketch Engine.*; 2) *зарегистрироваться на сайте через электронный адрес*; 3) *перейти с платформы в раздел «parallel corpora»*; 4) *перейти с раздела «parallel corpora» на раздел «English Uzbek contextual translation»*; 5) *поиск слов в интерфейсе «English Uzbek contextual translation».*

Статистический метод машинного перевода играет важную роль при создании программы контекстного переводчика в созданном выше параллельном корпусе.

В связи с использованием статистического метода машинного перевода для создания программ-контекстных переводчиков и их эффективного использования, на платформе Sketch Engine в английском и узбекском языках отсутствует онлайн-контент и «*e-library*» параллельных корпусных данных, созданных на платформе Sketch Engine. По этой причине все данные были размещены на платформе для создания небольшой базы данных в *Microsoft Excel*.

В этом параллельном корпусе в программе размещен текст, состоящий из научных и публицистических текстов на английском и узбекском языках общим объёмом 1658 слов и их переводы. Из них 742 слова составлены из английского текста, а остальные 916 слов составлены из узбекских текстов. Тот факт, что количество слов в английском языке меньше, чем в узбекском, объясняется тем, что лексические слова в узбекском языке в процессе перевода имеют несколько значений.

Корпусный подход к обучению переводу быстро развивался за последнее десятилетие, и в результате исследований с использованием параллельных корпусов, многоязычных корпусов (multilingual corpus) и сопоставимых корпусов (comparable corpus) появилось множество данных. В частности, параллельные корпуса оказываются ценными источниками информации в области переводоведения и образования.

ЗАКЛЮЧЕНИЕ

1. Широко актуально научно-теоретическое изучение машинного перевода, истории и этапов создания. При этом системы машинного перевода характеризуются определенными аспектами: 1) участие языков: а) двуязычный; б) многоязычный; 2) объем выбранной темы: а) *относящийся к определенному стилю и области*; б) *относящийся к одному стилю и множеству областей*; 3) *относящийся к разным стилям и областям*. Соответственно, машинный перевод делится на следующие системы, и каждая система имеет свои малые и очень важные функции в процессе перевода: 1. МП, основанный на правилах (Rule-Based MT), 2. МП, основанный на корпусе (Corpus based MT), 3. Гибридный МП (Hybrid MT).

Сегодня эффективно используется статистическая система МП, но в будущем есть возможность использования систем искусственного интеллекта и нейро-МП.

2. МП в основном состоит из трех этапов: сначала машинный перевод со словарем, компьютерного перевод на основе корпусной статистики, а затем до нейромашинного перевода с искусственным интеллектом в качестве основной технологии в последние годы. В настоящее время использование машинного перевода на основе словарей значительно сократилось, вместо этого быстро развивается использование компьютерного перевода на основе корпусной статистики. В будущем использование искусственного интеллекта и нейромашинного перевода в качестве основных технологий послужит решению проблем языка и перевода для человечества.

3. В ходе быстрого развития систем машинного перевода, с появлением корпусов и растущим спросом рынка переводов на эффективность перевода, параллельное применение программного обеспечения машинного перевода через корпусные платформы для улучшения управления памятью переводов САТ и терминологическими базами данных привело к развитию КЛ и МП.

4. В исследовании были научно и теоретически освещены функциональные аспекты нескольких типов корпусов, в частности, корпус *DIY*, *comparable corpus*, *stable corpus*, *ad hoc corpus* и других типов корпусов, а также были широко освещены сходства и различия между этими типами корпусов. В частности, корпус *DIY* – это совокупность интернет-документов, точнее, веб-страниц на языке HTML. В отличие от корпуса *DIY*, который считается открытым корпусом, корпус *Ad hoc* создается переводчиком за короткое время для исследования конкретной проблемы, возникшей во время определенного перевода. Корпус *Ad hoc* обычно не содержит большого количества текста. Сходство между корпусами *DIY* и *Ad hoc* заключается в том, что оба корпуса представляют собой виртуальные корпуса, созданные для выполнения перевода. *Comparable corpus* – это корпус двух или более одноязычных переводов, обычно каждый естественный язык построен по тем же принципам, что и другой язык. Одним из важных аспектов этого корпуса является то, что информация в текстах может быть одинаковой, но результаты предполагаемых переводов будут отличаться от других языков. Корпус *Stable* – это тип стабильного корпуса, который часто предоставляется только после включения в его интерфейсы. Корпус *Stable* отличается от корпусов *DIY*, *AD hoc*, *Comparable* и других типов тем, что корпус *stable* не адаптирован к потребностям переводчиков и не создан специально для пользователей перевода, а потребителям перевода не рекомендуется использовать этот корпус *stable* для перевода текстов.

5. Одной из основных задач корпусной лингвистики является поиск коллокаций (*collocations*) в корпусах. Значимость коллокации в корпусе определяется тремя типами коллокаций: *strong* (сильным), *fixed* (фиксированным) и *weak* (слабым), и одной из его основных задач является сохранение лексического значения одного из его компонентов при переводе. В результате анализа перевода коллокаций английского языка на узбекский язык с

использованием в исследовании материалов научных и публицистических текстов, коллокация в параллельном корпусе программы контекстного переводчика на английский и узбекский языки выполняет важную задачу по поиску точного перевода значения предложений на двух языках.

6. Тексты, написанные в научно-публицистическом стиле, обладают рядом особенностей, которые нельзя игнорировать при переводе. На основании этого определяется альтернатива перевода текста на языке оригинала в научных и публицистических текстах на целевой язык. Одной из основных проблем при переводе научных и публицистических текстов является терминология. Переводчику может быть очень сложно перевести термин, состоящий из одного слова. Решение этой задачи должно сочетаться со следующими тематическими словарями, международной символикой и общепризнанными терминами. Помимо терминов, перед практикой перевода научных и публицистических текстов стоят следующие задачи: 1) однозначность – не допускать двойного толкования слов и терминов в научных и публицистических текстах, 2) использование разных местоимения – в английских статьях часто можно встретить форму первого лица множественного числа. В узбекском языке можно использовать местоимение «*biz*», 3) точность – научные и публицистические тексты должны переводиться максимально точно, ближе к первоисточнику. Эти задачи служат решением проблем, возникающих в процессе перевода научных и публицистических текстов.

7. Лексическое значение и контекстный перевод десемантизированных слов, которые часто встречаются в научных и публицистических текстах, активно используются при автоматизации перевода и чаще всего встречаются в форме частей речи существительного и глагола, относящихся к лексико-семантике. Десемантизация приводит к грамматизации лексического значения слова в результате постепенной утраты лексического значения глаголов и существительных. В ряде случаев при контекстуальном переводе существительных и глаголов, они рассматриваются как вспомогательные слова, выполняющие только грамматические задачи. Десемантизированные глаголы и существительные важны при осуществлении контекстуального перевода.

8. Преимущество параллельных корпусов состоит в том, что они представляют собой специализированные параллельные документы, наполненные стандартизированной двуязычной терминологией. Сложные слова корпуса во многом являются терминологическими элементами. Таким образом, существует очень устойчивая связь между единицами перевода и переводными эквивалентами. Это облегчает анализ корпуса. Самый простой способ создать параллельный корпус – сначала создать память переводов, поместив данные в электронную таблицу *Excel*. Причина этого в том, что в интернет-библиотеке отсутствуют переводные версии научных и публицистических текстов на английском и узбекском языках. Второй метод создания памяти переводов – создание памяти переводов, который считается форматом памяти переводов, с использованием *TMX* или *XLIFF*. Разница между *XLIFF* и *TMX*, *CSV*, *TSV* и *XLSX* заключается в том, что *XLIFF* можно открывать в специальных программах и используется он для хранения данных в форматах 2.0 и выше. А

CSV с *TMX* открываются непосредственно в *Microsoft Excel*, *Open Office*, *Google Documents* и других веб-платформах. *XLIFF* определяется как кодирование двуязычных текстов при автоматизации языка.

9. В ходе исследования были научно изучены критерии создания программы перевода на платформе Sketchengine, определены особенности создания корпуса и параллельного корпуса на данной платформе, а также научно обоснованы этапы создания программ контекстного перевода на параллельном корпусе. Стоит отметить, что платформа Sketchengine задумана как интерпретация будущего мира корпусов, в надежде иметь единую крупную платформу, которая интегрирует языковые корпуса и программное обеспечение для перевода.

10. В результате наших исследований был создан первый механизм программы-контекстного переводчика в параллельном корпусе на платформе <https://auth.sketchengine.eu>. По содержанию этой программы была создана параллельная база данных текстов, относящихся к научному и публицистическому стилю, переведенных на контекстуальные английский и узбекский языки. В параллельном корпусе, состоящем из научных и публицистических текстов на английском и узбекском языках, программа контекстного переводчика была размещена на платформе Sketch Engine. Также созданы инструкции по использованию этой программы.

11. В отличие от 3 инструкций на платформе Sketch Engine, в следующем исследовании используются 5 программ контекстного переводчика Sketch Engine в параллельном корпусе: регистрация на платформе Sketch Engine, выбор раздела параллельного корпуса из разных корпусов и программ-переводчиков из интерфейса программы, в разделе параллельного корпуса поиск программы *uzbek contextual translation*, перемещение этой программы-переводчика из левого угла платформы в центр интерфейса платформы и использование программы. Программа-контекстный переводчик в параллельном корпусе, созданная на платформе Sketch Engine, эффективно служит ресурсом и базовой программой-переводчиком для эффективного использования перевода в области компьютерной лингвистики, для всего человечества и для некоторых представителей этой области.

**ONE-TIME SCIENTIFIC COUNCIL UNDER THE
SCIENTIFIC COUNCIL DSc.03/30.12.2019.Fil.27.01 ON AWARDING
SCIENTIFIC DEGREES AT UZBEKISTAN STATE UNIVERSITY OF
WORLD LANGUAGES**

UZBEKISTAN STATE UNIVERSITY OF WORLD LANGUAGES

KHAIKULOV ZAHRIIDIN KHAZRATKUL UGLI

**THE FUNCTIONAL FEATURES OF AN AUTOMATED TRANSLATION
OF PARALLEL CORPUS CONTEXTUAL SEARCH**

10.00.11 – Theory of language. Applied and computational linguistics

**DISSERTATION ABSTRACT
for the doctor of philosophy degree (PhD) in philological sciences**

Tashkent - 2024

The theme of the dissertation for doctor of philosophy (PhD) in philological sciences was registered at the Supreme Attestation Commission of the Republic of Uzbekistan under the number DSc.03/30.12.2019.Fil.27.01.

The doctoral dissertation was carried out at the Uzbekistan State University of World Languages. The abstract of the dissertation is posted in three languages (Uzbek, Russian, English (abstract)) on the Scientific Council website www.uzswlu.uz and on the website Ziyonet information and educational portal www.ziyonet.uz.

Scientific supervisor: Ergasheva Guli Ismoil kizi
Doctor of Philological Sciences, Associate Professor

Official opponents: Khayrullaev Khurshid Zaynievich
Doctor of Philological Sciences, Professor

Ataboev Nozimjon Bobojon ugli
Doctor of Philological Sciences, Associate Professor (PhD)

Leading organization: National University of Uzbekistan named after Mirzo Ulugbek

The defense of the dissertation will take place on «26» april 2024 at 10⁰⁰ at the meeting of the Scientific Council DSc.03/30.12.2019.Fil.27.01 at Uzbekistan State University of World Languages (Address: 100138, 21A, Kichik halqa yoli street, Uchtepa district, Tashkent city, Tel:(99871) 230-12-91, fax: (99871) 230-12-92, e-mail: uzswlu_info@mail.ru).

The doctoral dissertation can be reviewed at the Information Recourse Center at Uzbekistan State University of World Languages (registered under the number 1). The address: 100138, 21A Kichik halqa yoli street, Uchtepa district, Tashkent, Uzbekistan. Tel:(99871) 230-12-91, fax: (99871) 230-12-92.

The abstract of the dissertation sent out on «13» april 2024
Mailing report № 1 on _____ of «13» april 2024



G.H. Bakieva
Chairperson of the One-time
Scientific Council on awarding
Scientific degrees Doctor of
Philological Sciences, Professor

Kh.B. Samigova
Scientific Secretary of the One-time
Scientific Council, Doctor of
Philological Sciences, Professor

J.A. Yakubov
Chairman of the One-time Scientific
Seminar under the One-time Scientific
Council, Doctor of Philological
Sciences, Professor

INTRODUCTION (annotation of PhD thesis)

The purpose of the dissertation is to reveal the functional possibilities of the parallel corpus contextual search system in the automated translation using the examples of English and Uzbek languages (through the Sketch engine platform).

The object of the research is the contextual system of parallel corpus in English and Uzbek languages.

The subject of the research is the translation features and functional aspect of the contextual system of the parallel corpus of English and Uzbek languages.

Tasks of the research:

to research the theoretical foundations of the field of language theory, applied and computational linguistics in automated translation;

to interpret lexical features of scientific and journalistic texts in English and Uzbek and contextual analysis in translation;

to determine the methods used in the translation process of scientific and journalistic texts in English and Uzbek based on the corpus of parallel texts;

to analyze the lexical meaning and contextual translation of desemantized words used in scientific and journalistic style;

to study the nature of corpus types in creating contextual translator programs;

to identify methods of effective use of parallel corpora in translation practice and reveal the importance of the Sketch Engine platform in creating language corpora;

to identify the functional tasks of the corpus of parallel texts in the automation of translation based on revealing the common and different aspects of Linguee and Reverso contextual translator programs;

to study the steps of creating a language corpus and translator program on the Sketch Engine platform and creating a contextual translator program by placing English and Uzbek language texts into a parallel corpus on this platform.

Scientific novelty of the research consists of the following:

the effectiveness from the point of view of the functionality of the abilities of Statistic machine translation and Neuron machine translation systems to understand natural language, verifying with an auto-corrector, and utilize a larger database than HMT (Hybrid machine translation), RBMT (Rule-based machine) systems is substantiated translation) machine translation;

a mechanism has been developed for including a parallel corpus of the Uzbek language in the Sketch engine platform by justifying the use of the Rule-based machine translation system and software algorithms for the contextual meaning of desemantized words, every so often found in scientific and journalistic texts, by introducing Cat Trados translation memory functions;

an automated system in corpora such as *Adhog*, *Diy*, *Comparable*, *monolingual*, *diachronic*, *synchronic*, *parallel*, *words*, *expressions*, *collocations*, *idiomatic units*, *phrases*, and *text selection*, as well as the implementation of loading stages into the program, are the steps of improving a contextual program-

translator of collocations in scientific and journalistic texts in English and Uzbek languages on the Sketch engine platform in a parallel corpus;

experimental practical recommendations have been developed for the creation of an electronic sample database on the Sketch engine platform, consisting of scientific and journalistic texts using the contextual translator program in English and Uzbek languages, by improving the creation of methods for systematically using in practice the general requirements for automated translation programs.

Scientific and practical significance of the research results. The scientific significance of the research results lies in the fact that the creation of a contextual translator program for a corpus of parallel scientific texts and the development of theoretical foundations for its improvement serve as a scientific and theoretical source for research in the field of translation studies, corpus linguistics and machine translation. Also, the application of translation methods from English into Uzbek for contextual translation programs is determined by the analysis of methodological theories in the field of translation studies.

The practical significance of the research results lies in the fact that they can serve as a source when creating programs and plans in the process of teaching language theory, applied linguistics, computational linguistics, corpus linguistics, machine translation, as well as when describing topics, the widespread use of automatic translation programs, as well as the practice of contextual translation in two languages (English-Uzbek) is explained by the compilation and creation of an electronic dictionary.

Implementation of research results. Implementation of research results. Based on scientific results and practical proposals for revealing the functional essence of the parallel corpus context search system in automated translation:

conclusions about the effectiveness in terms of functionality of the abilities of Statistic machine translation and Neuron machine translation systems to understand natural language, check it with an auto-corrector and use a larger database than HMT (Hybrid machine translation), RBMT (Rule based) systems machine translation) machine translation were used in the credit mobilization project between the Uzbek State University of World Languages and the Ruhr University in Bochum (Germany) as part of the international exchange of experience between the European Union and higher educational institutions of Uzbekistan (ICM, KA-107 – International Credit Mobility between European Union and Uzbekistan HEIs, through staff and student mobility exchange projects between Uzbekistan State University of World Languages and Ruhr University Bochum (Germany) – 2020-1-DE01-KA 107-005433 Erasmus+ program) (Certificate No. IRD/23-156 dated 17 July 2023). As a result, the small base of the context translator program were studied in this dissertation and used within the project in a practical aspect;

recommendations on the mechanism for including a parallel corpus of the Uzbek language in the Sketch engine platform by justifying the use of the Rule based machine translation system and software algorithms for the contextual meaning of desemantized words, often found in scientific and journalistic texts, by

introducing Cat Trados translation memory functions were applied in the credit mobilization project between Uzbek State University of World Languages and Ruhr University in Bochum (Germany) as part of the international exchange of experience between the European Union and higher educational institutions of Uzbekistan (ICM, KA-107 – International Credit Mobility between European Union and Uzbekistan HEIs, through staff and student mobility exchange projects between Uzbekistan State University of World Languages and Ruhr University Bochum (Germany) – 2020-1-DE01-KA107-005433 Erasmus+ program) (Certificate No. 06/556 dated October 3, 2023). As a result, the abovementioned types of corpus were used as an experiment within the program;

experimental practical recommendations for creating an electronic sample database on the Sketch engine platform, consisting of scientific and journalistic texts using the contextual translator program in English and Uzbek languages by improving the creation of methods for systematic use in practice of general requirements for automated translation programs, were used in the implementation of activities for competitions for translations of scientific and journalistic texts in a parallel corpus, directed by the Writers' Union of Uzbekistan, strive to deeply reflect the ancient history and rich culture of our people, the fundamental changes achieved by the dedicated work of our country, taking into account the interests of young people in translation (Certificate of the Writers' Union of Uzbekistan No. 01-03/3150 dated November 28, 2023). As a result, the general conclusions of the study on the development of proposals and recommendations for creating a software mechanism for a contextual translator of a parallel corpus of English and Uzbek languages for the first time on the Sketch engine platform serve as a practical example, a theoretical resource for the translation of scientific and journalistic texts for young writers.

The structure and scope of the dissertation. The dissertation consists of an introduction, three chapters, conclusion, the list of references and appendix. The volume of the study is 178 pages.

E'LON QILINGAN ISHLAR RO'YXATI
СПИСОК ОПУБЛИКОВАННЫХ НАУЧНЫХ РАБОТ
LIST OF PUBLISHED SCIENTIFIC WORKS

1. Xaitqulov Z.X. An overview of automated translation and its linguistic problems // *Philology Matters*, 2021. – № 1. – P. 139-149 (10.00.0. № 25).
2. Xaitqulov Z.X. Amaliy tilshunoslikning nazariy va metodologik asoslari // *Ilm sarchashmalari*. – Urganch, 2020. – B. 35-40 (10.00.00. № 13).
3. Xaitqulov Z.X. Tarjima qilishga o'rgatishda elektron korpuslardan foydalanish usullari // *Philology Matters*, 2022. – № 2. – P. 102-118 (10.00.00. № 41).
4. Xaitqulov Z.X. Tarjimani avtomatlashtirishning nazariy va amaliy konsepsiyasi // *O'zMU xabarlari*. – T., 2022. – № 1/9/1. – B. 298-301 (10.00.00. № 1).
5. Xaitqulov Z.X. Mashin atarjimasi, tarixi va yaratilish bosqichlari // *O'zMU xabarlari*. – T., 2022. – № 1/5. – B. 269-270 (10.00.00. № 15).
6. Ergasheva G.I & Xaitqulov Z.X. A study of machine translation theories and methods // *The American Journal of Social Science and Education Innovations*, 2020. ISSN 2689-100X. Vol. 2. – P. 665-670 (Impact Factor 5. № 2).
7. Xaitqulov Z.X. Theoretical aspects of language engineering and translation / "Innovatsion 2020: Ma'naviyatini yuksaltirish va ilm-fan sohasidagi yutuqlari" mavzusidagi Respublika ilmiy-onlayn konferensiyasi materiallari. – Toshkent, 2020. – B. 284-287.
8. Xaitqulov Z.X. Applied linguistics to foreign language teaching and learning / *International Scientific Conference, Sheffield United Kingdom. Science and Civilization*. – UK, 2020. – P. 59-62.
9. Xaitqulov Z.X. The difference between Automated and Machine translation / *International scientific and practical conference, Cutting-Edge Science 2020*. – USA, 2020. – P. 68-70.
10. Xaitqulov Z.X. The concepts and functions of Smartcat and Linguee translators / "Filologiya va tillarni o'qitishning dolzarb muammolari 2020" mavzusidagi Respublika ilmiy-amaliy konferensiya materiallari. – Toshkent, 2020. – B. 38-41.

Avtoreferat "O'zMU xabarlari" jurnali tahririyatida tahrirdan o'tkazildi.



Bosishga ruxsat etildi: 09.04.2024
Bichimi: 60x84 1/16. «Times New Roman»
garniturada raqamli bosma usulda bosildi.
Shartli bosma tabog'i 3,3. Adadi 100. Buyurtma: № 23

«DAVR MATBUOT SAVDO» MChJ
bosmaxonasida chop etildi.
100198, Toshkent, Qo'yliq, 4-mavze, 46.