



UDK:811.512

Dilafroz MUHAMMADIYEVA,
Toshkent davlat o'zbek tili va adabiyoti universiteti dotsenti
E-mail:d.muhammadiyeva@mail.ru

O'zMU professori, f.f.d.M.Abdurahmonova taqrizi asosida

“BOBURNOMA” MATNI O‘ZBEKCHA-TURKCHA-QOZOQCHA PARALLEL KORPUSINING TARJIMAGA ASOSLANGAN EKSTRALINGVISTIK VA INTRALINGVISTIK MA‘LUMOTLAR BAZASI

Annotatsiya

“Boburnoma” nafaqat o‘zbek adabiyoti, balki umumturkiy madaniy merosning namunasi sifatida tarixiy va lingvistik jihatdan benazir manba hisoblanadi. Ushbu asar asosida o‘zbekcha-turkcha-qozoqcha parallel korpus tuzish g‘oyasi lingvistik tipologiya, tarjima nazariyasi va kompyuter lingvistikasi yo‘nalishlarida ilmiy izlanishlarni amalga oshirishga imkon yaratadi. Bu kabi korpusning yaratilishi zamonaviy lingvistikada ko‘ptilli muqobil matnlar bilan ishlash, tarjima modellari va leksik-semantik tadqiqotlar uchun zaruriy infratuzilmani shakllantiradi. Ushbu maqolada “Boburnoma” o‘zbekcha-turkcha-qozoqcha parallel korpusining ma‘lumotlar bazasini mavjud parallel korpuslardagi tajribalar asosida shakllantirish usullari tadqiq etiladi.

Kalit so‘zlar: “Boburnoma”, parallel korpus, ma‘lumotlar bazasi, tarjima nazariyasi, kompyuter lingvistikasi, ko‘ptilli muqobil matnlar.

THE EXTRALINGUISTIC AND INTRALINGUISTIC INFORMATION DATABASE OF THE UZBEK-TURKISH-KAZAKH PARALLEL CORPUS OF THE “BABURNAMA” TEXT

Annotation

The Baburnama is regarded as a unique source not only in Uzbek literature but also within the broader Turkic cultural heritage due to its unparalleled historical and linguistic significance. The idea of constructing an Uzbek-Turkish-Kazakh parallel corpus based on this work provides opportunities for conducting research in linguistic typology, translation studies, and computational linguistics. The creation of such a corpus establishes the necessary infrastructure for working with multilingual parallel texts, translation models, and lexical-semantic studies in modern linguistics. This article examines the methods of forming the database of the Uzbek-Turkish-Kazakh parallel corpus of the Baburnama, drawing on the experience of existing parallel corpora.

Key words: “Baburnama”, parallel corpus, database, translation theory, computational linguistics, multilingual parallel texts.

ЭКСТРАЛИНГВИСТИЧЕСКАЯ И ИНТРАЛИНГВИСТИЧЕСКАЯ ИНФОРМАЦИОННАЯ БАЗА ДАННЫХ УЗБЕКСКО-ТУРЕЦКО-КАЗАХСКОГО ПАРАЛЛЕЛЬНОГО КОРПУСА ТЕКСТА «БОБУРНОМА»

Аннотация

«Бабурнома» считается уникальным источником не только для узбекской литературы, но и для общего тюркского культурного наследия благодаря своей исключительной исторической и лингвистической ценности. Идея создания узбекско-турецко-казахского параллельного корпуса на основе данного произведения открывает возможности для проведения исследований в области лингвистической типологии, теории перевода и компьютерной лингвистики. Формирование такого корпуса создает необходимую инфраструктуру для работы с многоязычными параллельными текстами, переводческими моделями и лексико-семантическими исследованиями в современной лингвистике. В статье рассматриваются методы формирования базы данных узбекско-турецко-казахского параллельного корпуса «Бабурномы» на основе опыта существующих параллельных корпусов.

Ключевые слова: «Бабурнома», параллельный корпус, база данных, теория перевода, компьютерная лингвистика, многоязычные параллельные тексты.

Kirish. Parallel korpus yaratish – zamonaviy tarjimasunoslik, lingvistik tahlil va sun‘iy intellekt sohalarida tobora dolzarb ahamiyat kasb etib borayotgan ilmiy-amaliy yo‘nalishlardan biridir. Bunday korpuslar ikki va undan ortiq tilda yaratilgan matnlarning bir-biriga mazmunan, struktur jihatdan va sintaktik nuqtayi nazardan qanday mos tushishini aniqlashga xizmat qiladi. Parallel korpus tuzishning asosiy vazifasi – har ikki til matnlarini muvofiqlashtirish, ya‘ni ularni tahlil va qidiruv uchun yagona struktur birliklar darajasida birlashtirishdan iborat. Parallel korpusning asosiy mohiyati – bu bir tilga oid matnning boshqa tillardagi tarjimalari bilan sintaktik va semantik jihatdan mos bog‘langan holda raqamli tizimda joylashtirilganidir.

Mavzuga oid adabiyotlar tahlili. Keyingi yillarda o‘zbek tilshunosligida korpus lingvistikasi borasidagi

tadqiqotlar miqdori ortdi. So‘nggi izlanishlarda o‘zbek tili mualliflik korpusini, o‘zbek-ingliz tili mashina tarjimasini, til korpusi lingvistik bazasini tuzish tamoyillari, o‘zbek tili atov birliklarini semantik tegrashning lingvistik asoslari, ingliz tili korpuslari tahlili, ingliz-o‘zbek parallel korpusini tuzishning nazariy asoslari, o‘zbek tili korpuslarini tuzish dasturlari, elektron korpuslarning kompyuter modellari, o‘zbek tili korpusining idiomalar bazasi, o‘zbek tili morfologik, semantik analizatorining lingvistik ta‘minoti, o‘zbek tilining milliy va ta‘limiy korpusini yaratish muammolari tadqiq etildi. Korpus lingvistikasi, parallel korpus yaratish bo‘yicha amalga oshirilgan izlanishlarda lingvistik baza yaratish tamoyillari tavsiflangan.

Tadqiqot metodologiyasi. Tadqiqotda ilmiy tavsiflash, tasniflash, qiyosiy-tarixiy, modellashtirish, tizimli tahlil va talqin etish usullaridan foydalanilgan.

Tahlil va natijalar. Tilshunoslikka oid tadqiqotlarda fakt bilan ish ko'rilganda odatda material yig'ilishi, sistemaga solinishi lozim. Bunday katta hajmli ishni bajarishda korpus vaqt va mehnatni tejaydigan ish quroli vazifasini bajaradi. U texnik jarayonni tezlashtiruvchi vosita bo'libgina qolmay, muayyan tilning zamonaviy shakliga xos axborot tizimi ham bo'lib, kutilmagan savolga javob bera oladigan, til hodisasi bilan shug'ullanadigan soha oldiga avval ko'rilgan dolzarb muammolarni qo'ya oladigan tizim [4].

“Boburnoma” o'zbekcha-turkcha-qozoqcha parallel korpusini tuzishda asosiy uch qatlam – asl matn, turkcha va qozoqcha tarjimalar tanlab olinadi. Bu tarjimalar “Boburnoma”ning asl matniga iloji boricha yaqin bo'lgan, lekin zamonaviy o'quvchiga tushunarli bo'lishi uchun til jihatdan soddalashtirilgan variantlar bo'lishi kerak. Tarjimada uslub mosligi, leksik moslik, semantik aniqlik va sintaktik tuzilmaning muvofiqligi asosiy mezon sifatida qabul qilinadi.

Parallel matnlarni to'g'rilash – bu asliyat va tarjima matnidagi bir-biriga mos keluvchi gap hamda abzaslarni identifikatsiya qilish jarayonidir. Matn qismlarini bunday moslashtirish turlicha shakllarda bo'lishi mumkin. Chunki muallifning maqsadiga qarab qismlar bob, bo'lim, abzas yoki gap darajasida ifodalanadi. Foydalanuvchi uchun esa eng qulay variant – asliyat va tarjima matnlarining gaplar bo'yicha muvofiqlashtirilishi hisoblanadi. Eng mukammal avtomatik tarjima dasturi ham gap chegaralarini to'liq aniqlashda aniq yechim bera olmaydi. Chunki asl matndagi bir fikr bir, ikki yoki hatto uchta gap orqali ifodalanishi yoki aksincha, bir nechta gapning tarjimada bitta gap sifatida berilishi mumkin. Shu sababdan asliyat va tarjima matn qismlarini bitta buyruq orqali tenglashtirish mumkin emas[2].

Parallel korpuslarda gaplar bo'yicha moslikka erishish, avvalo, turli lingvistik amallarni bajarishni taqozo etadi. Tarjima jarayonida esa asl matndagi jumla tarjimada ikki jumlagi bo'linishi, bir nechta gap bitta gap sifatida berilishi yoki gaplarning o'rni almashirilishi ehtimoli mavjud.

“Boburnoma” o'zbekcha-turkcha-qozoqcha parallel korpusini yaratish jarayoni quyidagi bosqichlar ketma-ketligi tarzida bajariladi.

1. Matnlarni tanlash. “Boburnoma” matnining original o'zbekcha versiyasi va uning turkcha va qozoqcha tarjimalari tanlanadi. Har bir matnning to'liq va aniq tarjimasini kerak bo'ladi. Bu jarayonda original matnning davriga va tarjimaning aniqligiga e'tibor qaratish lozim.

2. Matnlarni raqamlashtirish (kompyuter shakliga o'tkazish).

Avvallari matnlarni kompyuterga kiritish murakkab va ko'p vaqt talab etuvchi jarayon bo'lgan bo'lsa, hozirgi kunda bu muammo sezilarli darajada soddalashgan. Bunga quyidagilar sabab bo'ldi:

optik belgi tanish texnologiyalari (ya'ni skanerlash va matnni avtomatik tanib olish);
matnli axborotni qayta ishlash sohasida keng kompyuterlashtirish.

Elektron matnlar turli yo'llar bilan olinishi mumkin: qo'lda kiritish, skanerlash, muallif nusxalari, nashriyotlar tomonidan taqdim etilgan original maketlar va boshqalar. “Boburnoma” o'zbekcha-turkcha-qozoqcha parallel korpusi matnlari kitob shaklidan avval skanerlash vositasida raqamli shaklga keltirildi, so'ngra maxsus dasturlar yordamida parallel korpusga joylash uchun kerakli formatga o'tkazildi.

3. Matnlarni dastlabki ishlovdan o'tkazish.

Ushbu bosqichda barcha matnlar filologik jihatdan tekshiriladi va tahrir qilinadi. Matnlar tahlil qilinib, so'z va

iboralarning tarjimasini, sintaktik strukturalar va semantik bog'lanishlar o'rganiladi. Bu jarayonda leksik, sintaktik va semantik analizlar olib boriladi, bunda biror ibora yoki so'zning turli tillardagi ekvivalentlari aniqlanadi. Shuningdek, matnning bibliografik va ekstralingvistik tavsiflari (masalan, muallif haqida ma'lumot, matn sanasi, janr xususiyatlari) tayyorlanadi.

4. Konvertatsiya va grafematik tahlil.

Ba'zi matnlar dastlabki avtomatik ishlov jarayonidan o'tkaziladi. Bu bosqichda quyidagi ishlar amalga oshiriladi:

kodlash tizimlarini o'zgartirish (agar zarur bo'lsa);

matnga aloqador bo'lmagan elementlarni (chizmalar, jadvallar) olib tashlash yoki mos shaklga o'tkazish;

so'z bo'linmalari, qat'iy satr oxirlari kabi belgilarni bartaraf etish;

matndagi belgilarni (masalan, tire) yagona shaklda yozilishini ta'minlash.

Bu amallar, odatda, avtomatik tarzda bajariladi. Aksariyat hollarda, matnni tarkibiy qismlarga – bo'lim, paragraf, gap va h.k. – ajratish ham shu bosqichda amalga oshiriladi.

5. Matnni teglash (belgilash).

Bu bosqichda matn va uning qismlariga qo'shimcha ma'lumotlar – ya'ni metama'lumotlar birlashtiriladi. Korpusning tilshunoslik va tarjima tizimlari uchun mos keladigan formatda bo'lishi ta'minlanadi. Bu yerda tokenizatsiya, lemmatizatsiya va sintaktik tahlil kabi jarayonlar o'rganiladi. Shuningdek, tahlil qilingan ma'lumotlar mashina o'rganishi va tarjima tizimlariga mos kelishi kerak.

Korpusdagi matnlar metama'lumoti quyidagi tarkibiy qismlarni o'z ichiga oladi:

mazmuniy metama'lumotlar (bibliografik ma'lumotlar, janr va uslubiy belgilari, muallif haqidagi ma'lumotlar);

formal metama'lumotlar (fayl nomi, kodlash parametrlari, teglash tili versiyasi, teglashni bajargan shaxslar).

Bu ma'lumotlar, odatda, qo'lda kiritiladi.

Matnning tarkibiy teglanishi (ya'ni paragraf, gap, so'zlarni ajratish) va lingvistik teglanishi (masalan, morfologik yoki sintaktik belgilash) esa odatda avtomatik tarzda amalga oshiriladi. Metadataning to'g'ri kiritilishi ma'lumotlar bazasining to'g'ri ishlashini ta'minlaydi.

6. Avtomatik teglash natijalarini tuzatish.

Bu bosqichda avtomatik teglash natijasida yuzaga kelgan xatoliklar tuzatiladi va noaniqliklar bartaraf etiladi. Bu jarayon to'liq qo'lda yoki yarim avtomatik tarzda bajarilishi mumkin.

7. Maxsus korpus boshqaruv tizimiga o'tkazish.

Teglangan matnlar lingvistik axborot-izlash tizimi (inglizcha: corpus manager) formatiga o'tkaziladi. Bu tizim foydalanuvchilarga tezkor va ko'pyoqlama izlash imkonini yaratadi hamda statistik tahlilni ta'minlaydi.

8. Korpusga kirish imkonini ta'minlash.

Yakuniy bosqichda foydalanuvchilar uchun korpusdan foydalanish imkoniyati yaratiladi. Bunda foydalanuvchilar turli toifadagi huquq va imkoniyatlarga ega bo'lishi mumkin.

Albatta, amaliyotda har bir loyiha o'ziga xos bo'lib, bu yerda keltirilgan bosqichlar soni va tartibi o'zgarishi mumkin. Ba'zi hollarda, real texnologiya bundan murakkabroq bo'ladi.

“Boburnoma” o'zbekcha-turkcha-qozoqcha parallel korpusining tuzilishi va tarkibi quyidagi komponentlardan iborat bo'ladi:

1. Manba va tarjima matnlar

2. Segmentlar: Korpusdagi matnlar odatda segmentlarga, ya'ni gaplar yoki jumlagi bo'linadi. Har bir segmentning tarjimasini alohida ajratib belgilanadi.

3. Metama'lumotlar: Korpusni yaratishda metama'lumotlar (masalan, til juftligi, matnning kelib chiqishi,

tarixiy davr) muhim rol o'ynaydi. Meta-ma'lumotlar parallel korpusning kontekstual va lingvistik tahlilini osonlashtiradi.

4. Struktura va format: Parallel korpuslar odatda XML, TMX (Translation Memory eXchange), CSV, JSON kabi formatlarda saqlanadi. Bu formatlar matnlarni strukturalash, saqlash va tizimli tarzda tahlil qilish imkonini beradi.

Parallel korpusning texnologik tuzilishi quyidagi ko'rinishda bo'ladi:

O'zbekcha matn	Turkcha tarjima	Qozoqcha tarjima
Farg'ona viloyati beshinchi iqlimdir.	Fergana viloyati beshinchi iqlimdir.	Ферғана уёлаятı бесінші ықылым.
Muxtasar viloyattur, oshlig' va mevasi farovon.	Küçük bir vilayettir. Hububat ve meyvası boldur	Шағын уәлаят. Астық пен жемісі мол.
Turkistondin heyli quyiroyq bu daryo tamom qumg'a singlar, hech daryog'a qotilmas.	Türkistan'dan epeyce aşağıda bu nehir büsbütün kumlar arasında kaybolur ve hiç bir nehre karışmaz.	Әрі өзге ешбір өзенге қосылмастан Түркістаннан әлде жырақтау жерде құмға сипіп кетеді.
Andijonning noshpotisidin yaxshiroq noshpoti bo'lmas.	Endican nashpatisinden daha iyi nashpati bulunmaz.	Әнджанның [1] алмұртынан жақсырақ алмұрт болмас.
Elining lafzi qalam bila rosttur.	Halkın konuşma dili yazı diline uygundur.	Елінің сөйлеу тілі мен жазба тілі бірдей.

Parallel korpuslarga avtomatik ishlov berish jarayoni boshqa turdagi korpuslarga nisbatan murakkabroq hisoblanadi. Uning asosiy texnik muammosi korpusning birliklari sifatida original (asliyat) va tarjima matnini gap, so'z birikmasi hamda so'z darajasida aniqlab belgilash (ya'ni razmetkalash) zaruratidir. Amaldagi avtomatik qidiruv modellari, xususan, nomlashgan birliklar (NP – noun phrase) asosidagi tizimlar bunday razmetkani aniq amalga oshira olmaydi. Yevropa tillari, xitoy va yapon kabi yirik lingvistik resurslarga ega tillar uchun maxsus parallel korpuslar ishlab chiqilgan bo'lib, ularda original va tarjima matn birliklarini tenglashtirish uchun samarali usullar qo'llanilgan.

Gap darajasidagi tenglashtirish jarayonida odatda struktur yondashuv (gap uzunligi, tarkibidagi so'zlar soni, tinish belgilarining joylashuvi kabi belgilar asosida) va statistik yondashuv (matnda so'zlar va iboralarning chastotasi, ularning pozitsion joylashuvi va ehtimollik ko'rsatkichlariga asoslangan metodlar) qo'llaniladi. Ushbu usullar katta hajmdagi lug'at resurslarini talab qilmasligi sababli, ular kichik lingvistik resurslarga ega tillar uchun ham samarali hisoblanadi. Ayniqsa statistik modellarga asoslangan usullar – masalan, IBM alignment models, Gale-Church algoritmi, Brown korpusi modellari – parallel gaplarni aniqlashda keng qo'llaniladi. Bu metodlar yordamida har bir gap yoki so'zga uning tarjimadagi eng mos elementini topish imkoniyati paydo bo'ladi. Bunda, ayniqsa, so'z chastotasi muhim o'rin tutadi: bir tildagi so'z qanchalik ko'p uchrasa, uning tarjimadagi ekvivalenti ham shunchalik ko'p uchrashi lozim bo'ladi. Bunday muvofiqlik orqali avtomatik tenglashtirish aniqligi ortadi.

Gap darajasida amalga oshirilgan tenglashtirishdan so'ng, keyingi bosqichda sintaktik tahlil jarayoni boshlanadi. Bu bosqichda har ikkala til matnlari gapning grammatik tuzilmasi asosida tahlil qilinadi va sintaktik tenglik yoki tafovutlar aniqlanadi. So'nggi yillarda razmetkalangan parallel korpuslarda bu jarayonni avtomatlashtirish yo'lida ko'plab izlanishlar olib borilmoqda. Sintaktik tahlil, avvalo, mavjud grammatik nazariyalarni tekshirish, shuningdek, yangi sintaktik qonuniyatlarni aniqlash imkonini beradi. Bu jarayon tarjimashunoslik va kompyuter lingvistikasi uchun nihoyatda muhim bo'lib, ilmiy adabiyotlarda "formal grammatika mashqi" deb yuritiladi. Bunday tahlillar natijasida yaratilgan

Treebank – ya'ni sintaktik tuzilmasi aniqlangan razmetkalangan korpuslar – tilshunoslikdagi korpus lingvistikasi yondashuvining amaliy mahsulidir.

1. segmentatsiya, ya'ni "Boburnoma" manti va tarjimalarining gap shaklidagi segmentlarga ajratilgan ko'rinishi;

2. tekislashtirish (alignment), ya'ni asos matndagi har bir segmentni turk va qozoq tillaridagi mos segmentlar bilan bog'lash;

3. annotatsiya, ya'ni har bir segmentga lingvistik teglar (morfologik, sintaktik, semantik) birlashtirish;

4. indeksatsiya, ya'ni matnlar bazasida tezkor izlash imkonini beruvchi tizim yaratish.

Ingliz, fransuz, nemis kabi roman-german tillari uchun bunday razmetkalash va sintaktik tahlilni amalga oshira oladigan ilg'or parser tizimlari mavjud. Bu parserlar matndagi har bir gapni sintaktik daraxt ko'rinishida strukturaviy tahlil qiladi va turli lingvistik birliklar o'rtasidagi grammatik munosabatlarni aniqlaydi. Afsuski, o'zbek tili, hattoki turkiy tillar uchun bunday avtomatlashtirilgan parserlar hali to'liq ishlab chiqilmagan. Mavjud xorijiy parserlar o'zbek tilining morfologik murakkabligi, agglutinativligi va erkin so'z tartibi kabi o'ziga xosliklarini hisobga olmaydi. Shu sababli o'zbek tili uchun maxsus moslashtirilgan sintaktik modellar va parserlarni ishlab chiqish muhim ilmiy vazifalardan biri bo'lib qolmoqda.

Parallel korpus yaratish jarayoni har bir bosqichda chuqur lingvistik va texnologik bilimlarni talab etadi. Gap darajasidagi tenglashtirish esa bu jarayonning markaziy bo'g'ini hisoblanadi. Ushbu bosqich sifatli bajarilgandagina keyingi tahlil ishlari, jumladan, so'z darajasidagi muvofiqlik, sintaktik va stilistik moslikni aniqlash, shuningdek, mashina tarjimasining sifatini oshirish mumkin bo'ladi.

Xulosa. "Boburnoma"ning o'zbekcha-turkcha-qozoqcha parallel korpusni yaratish – bu lingvistik, madaniy va texnologik jihatdan o'zaro bog'langan yirik ilmiy loyiha hisoblanadi. Bu korpus tilshunoslar, tarjimashunoslar, tarixchilar, leksikograflar hamda sun'iy intellekt tizimlari bilan ishlovchi mutaxassislar uchun noyob manba vazifasini bajaradi. Bunday korpus nafaqat o'tmish adabiy merosini bugungi zamonga yaqinlashtiradi, balki turkiy tillar o'rtasidagi mustahkamlikni chuqurroq anglash, madaniy birlikni mustahkamlashga ham xizmat qiladi. Ayniqsa, gap shaklidagi alignment – tarjima tahlilida aniqlik va strukturaviy tahlilni osonlashtiruvchi muhim vositadir.

"Boburnoma" asosidagi o'zbekcha-turkcha-qozoqcha parallel korpusning ilmiy asosda tuzilishi ko'ptil lingvistik tadqiqotlarda sifatli natijalarni ta'minlovchi asosiy infratuzilma hisoblanadi. Korpus nafaqat tarjima metodlarini chuqur tahlil qilish, balki tilshunoslikda turkiy tillarning o'zaro grammatik va leksik bog'liqligini empirik dalillar asosida ko'rsatishga xizmat qiladi. Bunday ilmiy yondashuv madaniyatlararo muloqot, tarjima etikasi, lingvistik uyg'unlik va tarixiy til o'zgarishlarini o'rganishda katta ahamiyat kasb etadi.

Gap darajasida tuzilgan parallel korpus tarjimonlar uchun o'ziga xos tarjima laboratoriyasi bo'lib xizmat qiladi. Gap darajasidagi moslik erkin tarjima va literal tarjima orasidagi farqlar, kompensatsiya, izohli tarjima, semantik

kengayish yoki torayish kabi strategiyalarni tahlil qilishni osonlashtiradi. Korpusdagi har bir gapdagi leksik birliklar asosida zamonaviy izohli, tematik yoki sohalar bo'yicha tuzilgan tarjima lug'atlarini yaratish mumkin. Bu esa "Boburnoma" leksikasini chuqur o'rganishga zamin yaratadi.

Gap asosida tuzilgan parallel korpus sun'iy intellekt tizimlari uchun o'quv materiali sifatida xizmat qiladi. Ayniqsa, neyron tarmoqlar asosidagi tarjima tizimlari uchun bu turdagi korpuslar juda foydalidir.

ADABIYOTLAR

1. Elova D. O'zbek tili korpusi birliklarining uslubiy teglarini yaratish tamoyillari va lingvistik ta'minoti Filol.fan.b.fals.dok...diss. Toshkent, 2022.
2. Karimov R. O'zbek-ingliz parallel korpusini tuzishning lingvistik va dasturiy masalalari Filol.fan.b.fals.dok...diss. avtoref. Buxoro, 2022.
3. Musurmankulova Sh. G. O'zbekcha va turkcha frazemalar korpusining lingvistik ta'minoti. Filol.fan.b.fals.dok...diss. avtoref. Toshkent, 2024
4. Primova M. Til korpuslarida matnlarni annotatsiyalash: afzallik va kamchiliklari. O'zbekiston: Til va madaniyat. Kompyuter lingvistikasi, (4), 6. –Toshkent, 2023.
5. Gulyamova Sh. O'zbek tili semantik analizatorining lingvistik asoslari. Filol.fan.b.fals.dok...diss. avtoref. Farg'ona, 2022
6. Xidirov O. Milliy korpus uchun parsing dasturi yaratishning lingvistik asoslari. Filol.fan.b.fals.dok...diss. avtoref. Jizzax, 2021.