

## “BOBURNOMA” MATNI PARALLEL KORPUSINI YARATISH MASALALARI

**Muhammadiyeva Dilafuz Axtamovna,**

*Alisher Navoiy nomidagi O‘zbek tili va adabiyoti universiteti dotsenti,  
filologiya fanlari doktori (DSc)*

*e-pochta: [d.muhammadiyeva@mail.ru](mailto:d.muhammadiyeva@mail.ru)*

*ORCID ID: <https://orcid.org/0000-0001-6761-462X>*

**Annotatsiya.** Mazkur maqolada “Boburnoma” asarini zamonaviy raqamli va lingvistik texnologiyalar asosida o‘rganish masalalari tahlil etiladi. Xususan, raqamlashtirish, korpus lingvistikasi, parallel matnlar yaratish, sun‘iy intellekt va mashinaviy o‘rganish usullarining qo‘llanilishi ilmiy jihatdan asoslab beriladi. Asarning lingvistik, tarixiy va madaniy qatlamlarini kompleks tadqiq etishda elektron korpuslarning ahamiyati ochib beriladi. Shuningdek, parallel korpuslar yordamida tarjimashunoslik, qiyosiy va chog‘ishtirma tilshunoslik sohalarida yangi ilmiy natijalarga erishish imkoniyatlari ko‘rsatib o‘tiladi. Maqolada “Boburnoma” asosida yaratiladigan korpusning nazariy va amaliy ahamiyati, uni shakllantirish bosqichlari hamda metodologik tamoyillari yoritilgan.

**Kalit so‘zlar:** Boburnoma, korpus lingvistikasi, parallel korpus, raqamlashtirish, sun‘iy intellekt, tarjimashunoslik, matnshunoslik, lingvistik tahlil.

**Abstract.** This article explores the issues of studying the “Baburnama” using modern digital and linguistic technologies. It provides a scientific analysis of digitization processes, corpus linguistics, the creation of parallel corpora, and the application of artificial intelligence and machine learning methods. The study highlights the importance of electronic corpora in the comprehensive investigation of linguistic, historical, and cultural layers of the text. Furthermore, it demonstrates how parallel corpora contribute to advancements in translation studies, contrastive and comparative linguistics. The paper also discusses the theoretical and practical significance of constructing a corpus based on the “Baburnama,” including its development stages and methodological principles.

**Keywords:** Baburnama, corpus linguistics, parallel corpus, digitization, artificial intelligence, translation studies, text analysis, linguistic research.

**Аннотация.** В данной статье рассматриваются вопросы изучения произведения «Бабурнама» с использованием современных цифровых и лингвистических технологий. Научно обоснованы процессы оцифровки, корпусной лингвистики, создания параллельных корпусов, а также применения методов искусственного интеллекта и машинного обучения. Раскрывается значение электронных корпусов в комплексном исследовании лингвистических, исторических и культурных слоёв текста. Кроме того, показаны возможности использования параллельных корпусов в переводоведении, сопоставительной и сравнительной лингвистике. В статье освещаются теоретическая и практическая значимость создания корпуса на основе «Бабурнама», а также этапы его формирования и методологические принципы.

**Ключевые слова:** Бабурнама, корпусная лингвистика, параллельный корпус, оцифровка, искусственный интеллект, переводоведение, текстология, лингвистический анализ.

**Kirish.** “Boburnoma” yaratilgan davrdan boshlab turli davrlarda yashab, turfa tillarda so‘zlashgan sharqshunoslar diqqatini o‘ziga tortib kelgan va bir necha asrlar davomida kitobxonlarning sevimli kitobiga aylanib ulgurgan. Asar dunyo ilm ahli tomonidan tilshunoslik, adabiyotshunoslik yo‘nalishlarida, geografik va tarixiy aspektlarda o‘rganilgan.

“Boburnoma” tarixiy xotira, badiiy tafakkur va til boyligining noyob uyg‘unligini aks ettiradi. Shu bois uni faqat an‘anaviy filologik usullar orqali emas, balki zamonaviy raqamli va analitik texnologiyalar yordamida tadqiq etish zarurati tobora ortib bormoqda.

Bunday yondashuv asarning mazmun-mohiyatini yanada chuqurroq ochish, undagi tarixiy, ijtimoiy va madaniy qatlamlarni kompleks ravishda tahlil qilish imkonini beradi.

Keyingi yillarda o'zbek tilshunosligida korpus lingvistikasi borasidagi tadqiqotlar miqdori ortdi. Korpus lingvistikasining nazariy asoslari B. Zaxarov, K. Boyarskiy, o'zbek tilshunoslaridan A. Po'latov, S. Muhamedova, B. Mengliyev, Z. Xolmanova, D. O'rinboyeva ishlarida tavsiflangan. Sh. Hamrayeva, N. Abdurahmonova, A. Eshmo'minov, D. Axmedova, R. Karimov, N. Atabayev, A. Begmatova, O. Xidirov, G. Begmatova, O. Xoliyorov, Sh. Gulyamova izlanishlarida o'zbek tili mualliflik korpusi, o'zbek-ingliz tili mashina tarjimasini, til korpusi lingvistik bazasini tuzish tamoyillari, o'zbek tili atov birliklarini semantik teglashning 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 etilgan. Korpus lingvistikasi, parallel korpus yaratish bo'yicha amalga oshirilgan izlanishlarda lingvistik baza yaratish tamoyillari tavsiflangan.

Tarixiy-badiiy matn tarjimasini uchun mo'ljallangan parallel korpusning lingvistik va tarjima asoslarini yaratish masalasi monografik tadqiqot mavzusi sifatida o'rganilmagan.

**Tadqiqotning metodologiyasi.** Tadqiqotda qiyosiy-tarixiy, statistik metod, modellashtirish, komponent, konseptual tahlil metodlaridan foydalanilgan.

**Qiyosiy-tarixiy metod** parallel korpusga kiritilgan matnlarning turli davrlarda yaratilgan nusxalari hamda tarjimalarini o'zaro taqqoslashda;

**Statistik metod** korpusdagi til birliklarining qo'llanish chastotasini aniqlash, eng faol leksik birliklar, terminlar va grammatik konstruksiyalarni hisoblashda;

**Modellashtirish metodi** parallel korpusning tuzilishini ishlab chiqish, matnlarni segmentlarga ajratish, ularning o'zaro bog'lanish sxemasini yaratish hamda korpusning elektron modelini loyihalash jarayonida;

**Komponent tahlil metodi** "Boburnoma" matnidagi so'z va iboralarning ma'no tarkibini tahlil qilishda, ayniqsa tarixiy-madaniy tushunchalarning tarjimadagi semantik mosligini aniqlashda;

**Konseptual tahlil metodi** asarda aks etgan milliy-madaniy konseptlarni (davlat, hukmdor, vatan, tabiat, urf-odat va boshqalar) aniqlash hamda ularning turli tillardagi ifodalanish xususiyatlarini o'rganishda qo'llanildi.

**Tahlil va natijalar.** Zamonaviy texnologiyalarning eng muhim yo'nalishlaridan biri bu raqamlashtirish jarayonidir. Qo'lyozma va eski nashrlarning yuqori aniqlikdagi elektron nusxalarini yaratish orqali "Boburnoma"ning turli variantlarini bir platformada jamlash imkoni paydo bo'ladi. Bu esa matnshunoslik nuqtai nazaridan juda katta ahamiyatga ega bo'lib, turli nusxalar o'rtasidagi farqlarni aniqlash, matnning asl ko'rinishini tiklash va tahrir jarayonlarini ilmiy asosda olib borishga xizmat qiladi. Raqamli bazalar orqali tadqiqotchilar dunyoning istalgan nuqtasidan turib ushbu manbaga

murojaat qila oladi, bu esa ilmiy hamkorlikni kengaytiradi va tadqiqotlar samaradorligini oshiradi.

Shu bilan birga, korpus lingvistikasi asosida “Boburnoma” matnini o‘rganish tilshunoslikda yangi imkoniyatlar yaratadi. Maxsus elektron korpuslar yordamida matndagi so‘zlarning qo‘llanish chastotasi, grammatik shakllar, sintaktik konstruktsiyalar va stilistik vositalar aniq statistik ko‘rsatkichlar asosida tahlil qilinadi. Bu yondashuv orqali Bobur tilining o‘ziga xos xususiyatlari, uning davriga xos leksik qatlamlar hamda turkiy til taraqqiyotidagi o‘rni ilmiy jihatdan asoslab beriladi. Natijada, “Boburnoma” nafaqat tarixiy manba, balki til taraqqiyotini o‘rganishda muhim lingvistik korpus sifatida ham xizmat qiladi.

Sun‘iy intellekt va mashinaviy o‘rganish texnologiyalarining joriy etilishi esa asarni o‘rganishda sifat jihatidan yangi bosqichni boshlab berdi. Semantik tahlil, avtomatik tarjima, matnni mavzular bo‘yicha guruhlash va nomlangan obyektlarni aniqlash kabi usullar yordamida asardagi murakkab ma‘lumotlar tizimlashtiriladi. Masalan, asarda tilga olingan shaxslar, joy nomlari va tarixiy voqealar avtomatik ravishda ajratilib, ular o‘rtasidagi bog‘liqliklar aniqlanadi. Bu esa tadqiqotchilarga katta hajmdagi ma‘lumotni qisqa vaqt ichida tahlil qilish imkonini beradi hamda inson omili tufayli yuzaga kelishi mumkin bo‘lgan xatoliklarni kamaytiradi.

Bundan tashqari, geografik axborot tizimlari yordamida “Boburnoma”da tasvirlangan hududlar va voqealar interaktiv xaritalar ko‘rinishida tiklanadi. Boburning safarlari, janglari va siyosiy faoliyati aniq geografik koordinatalar asosida vizualizatsiya qilinib, tarixiy jarayonlarning makon bilan bog‘liqligi yanada ravshan ko‘rinish oladi. Bu yondashuv nafaqat ilmiy tadqiqotlar uchun, balki ta‘lim jarayonida ham katta ahamiyatga ega bo‘lib, talabalar va keng jamoatchilik uchun asarni yanada tushunarli va qiziqarli qiladi. Shuningdek, zamonaviy texnologiyalar “Boburnoma”ni keng auditoriyaga yetkazishda ham muhim rol o‘ynaydi. Elektron kutubxonalar, mobil ilovalar va onlayn platformalar orqali asarning turli tarjimalari va sharhlari ommaga taqdim etiladi. Bu esa milliy merosni global miqyosda targ‘ib qilish, uni turli madaniyat vakillariga tanishtirish va ilmiy muomalaga keng jalb etish imkonini beradi. Ayniqsa, multimediya vositalari – audio, video va interaktiv materiallar yordamida asarning mazmuni yanada boyitilib, uning qabul qilinish darajasi oshirilmoqda.

Umuman olganda, zamonaviy texnologiyalar asosida “Boburnoma”ni o‘rganish an’anaviy ilmiy yondashuvlarni inkor etmaydi, balki ularni to‘ldiradi va yangi ilmiy ufqlarni ochadi. Bu jarayon natijasida asarning matni, mazmuni va tarixiy ahamiyati yanada chuqurroq anglanadi, ilmiy xulosalar esa yanada aniq va asosli bo‘ladi. Shu tariqa, “Boburnoma”ni o‘rganish nafaqat o‘tmishni anglash, balki zamonaviy ilm-fan yutuqlarini amaliyotga tatbiq etishning yorqin namunasi sifatida namoyon bo‘lmoqda.

Tilshunoslikda korpusga asoslangan yondashuv zamonaviy lingvistik tahlilning ajralmas qismi hisoblanadi. Biroq bu metodologiya o‘zbek tili, adabiyoti va madaniyati

uchun mutlaqo yangi hodisa emas. Tarixiy manbalardan ko‘rinadiki, o‘zbek tilining shakllanishi va rivojlanish jarayonida korpus xususiyatiga ega bo‘lgan namunalar allaqachon yuzaga kelgan va amaliyotda qo‘llanilgan. Garchi bu korpuslar bugungi kunda tushunilayotganidek elektron shaklda bo‘lmasa-da, mazmun va maqsad jihatidan zamonaviy til korpuslariga yaqinlashadi. Z.Xolmanova qarashlari [Xolmanova, 2021,56] asosida shuni aytishimiz mumkinki, Mahmud Koshg‘ariy tomonidan XI asrda yaratilgan “Devonu lug‘otit turk” asari qadimgi turkiy til leksikasi va matnshunosligining muhim manbai bo‘lib, unda turkiy so‘zlar bilan bir qatorda ularning kontekstdagi qo‘llanilishi, matn parchalari, hikmatli so‘zlar va xalq og‘zaki ijodi namunalari jamlangan. Ushbu asar mazmunan va funksional jihatdan korpus tilshunosligining dastlabki namunalaridan biri sifatida qaralishi mumkin.

Matn asosidagi parallel korpuslar tarjimashunoslik, chog‘ishtirma hamda qiyosiy tilshunoslik uchun beqiyos ilmiy imkoniyatlar yaratadi. Ular lingvistik birliklarning turli tillarda ishlatilishini real kontekstda kuzatish, til strukturasi va semantikasining tarjimada qanday o‘zgarishini aniqlash, shuningdek, tarjima strategiyalarining qanday amalga oshirilishini sistematik tahlil qilishga xizmat qiladi.

Tarjimashunoslikda parallel korpuslardan foydalanish ikki va undan ortiq tillarning kontrastiv tahlilini samarali amalga oshirish imkonini beradi. Ular tarjimonlarga manba va maqsadli til o‘rtasidagi tarjima ekvivalentlarini aniqlashda yordam beradi, xususan, to‘g‘ridan to‘g‘ri muqobili mavjud bo‘lmagan birliklar uchun mos kontekstual yechimlar topishga xizmat qiladi. Bundan tashqari, parallel korpuslar so‘zlar, birikmalar, frazeologik birliklar, metaforik ifodalar va stilistik ko‘chimlarning har ikki til sistemasi doirasidagi ishlatilish chastotasi haqida aniqlovchi ma‘lumotlar beradi. Ushbu empirik asos tarjimonlarga kontekstga mos tarjima strategiyalarini ishlab chiqishda real vosita bo‘lib xizmat qiladi.

So‘nggi yillarda parallel korpuslardan foydalanish doirasi yanada kengayib, avtomatik tarjima tizimlarini ishlab chiqish, neyron tarmoqlar asosida mashina tarjimasini o‘rgatish hamda sun‘iy intellekt ilovalari uchun o‘quv bazasini shakllantirishda asosiy manbalardan biriga aylangan. Bu esa parallel korpuslarni nafaqat an‘anaviy lingvistika va tarjimashunoslik, balki kompyuter lingvistikasi, tabiiy tilni qayta ishlash va til texnologiyalari kabi zamonaviy yo‘nalishlar uchun ham strategik ahamiyatga ega resursga aylantirmoqda [Musulmankulova, 2024,35].

Parallel korpuslar qiyosiy tilshunoslik uchun grammatik va sintaktik strukturalarning o‘zaro mos kelish yoki farqlanish xususiyatlarini ochib beradi. Bu orqali ikki qardosh til orasidagi grammatik parallellik, strukturaviy o‘zgarishlar, leksik variantlar, stilistik o‘zgarishlar va funksional tenglik kabi hodisalarni aniqlash imkoniyati tug‘iladi. Korpuslar yordamida aniqlangan bu tafovutlar nazariy til modellari va tarjima tiplari bo‘yicha aniq, empirik dalillarga asoslangan umumlashmalar chiqarish imkonini beradi.

Chog‘ishtirma tilshunoslikda esa parallel korpuslar ikki yoki undan ortiq tillar o‘rtasida yuzaga keladigan semantik, sintaktik va stilistik tafovutlarni kontekstual jihatdan aniqlashga imkon beradi. Real matnlarda til birliklarining qanday ishlatilgani, ularning tarjimada qanday transformatsiyaga uchrashi va bu o‘zgarishlarning sabablarini tahlil qilish chog‘ishtirma metodologiyaning asosiy yondashuvlarini boyitadi. Ayniqsa, madaniy tafovutlar, pragmatik belgilar va stilistik moslashuvlar nuqtayi nazaridan olib borilgan tadqiqotlar uchun parallel korpuslar muhim empirik maydon yaratadi.

Tarjimashunoslikda parallel korpuslar tarjima nazariyasini amaliy misollar bilan to‘ldirish, tarjima strategiyalarining toifalarini empirik tarzda tasdiqlash, tarjima sifati, moslik, stilistik ekvivalentlik, kompensatsiya hodisalarini tadqiq qilishda ishonchli poydevor bo‘lib xizmat qiladi. Ayniqsa, tarjimaning stilistik va emotsional tafovutlarini tahlil qilishda real kontekst muhim bo‘lib, bu jarayonda parallel korpuslarning hissasi alohida e‘tiborga loyiq.

Bundan tashqari, sun‘iy intellekt asosidagi avtomatik tarjima tizimlarining rivojlanishi bilan bir qatorda, parallel korpuslar mashina tarjimasi modellarini o‘rgatish, test qilish va tahlil qilishda ham asosiy lingvistik resurs sifatida foydalanilmoqda. Bu esa korpus lingvistikasini nafaqat an‘anaviy ilmiy tadqiqotlar bilan bog‘labgina qolmay, balki uni texnologik innovatsiyalar, tabiiy tilni qayta ishlash va sun‘iy aql sohalari bilan integratsiyalash imkonini beradi.

Parallel hizalanish morfologik darajada ham tekshiriladigan natijalar beradi. Eski o‘zbek tilidagi qo‘shimchalar tizimi zamonaviy shakllar bilan qiyoslanganda, qaysi affikslar funksional yukini saqlab qolgan, qaysilari esa boshqa vositalar bilan almashtirilganini aniq ko‘rish mumkin bo‘ladi. Bu kuzatuvlar taxminiy xulosalarga emas, balki ko‘p martalik takrorlanuvchi mosliklarga tayanadi. Natijada diaxron o‘zgarishlar “tendensiya” sifatida emas, balki statistik jihatdan tasdiqlangan hodisa sifatida tavsiflanadi.

Tarjima matnlari bilan hizalangan segmentlar tarjima strategiyalarini ham bevosita matn darajasida ochib beradi. Qaysi o‘rinlarda tarjimon asl struktura va leksikani maksimal darajada saqlagan, qaysi joylarda esa izohlovchi yoki adaptiv yondashuvni tanlaganini aniqlash mumkin bo‘ladi. Ayniqsa, tarixiy va madaniy realiyalar tarjimasida ko‘p variantlilik kuzatiladi: ayrim hollarda transliteratsiya, boshqa hollarda tavsifiy tarjima qo‘llanadi. Parallel korpus bunday farqlarni umumiy mulohaza sifatida emas, balki aniq segmentlar asosida ko‘rsatadi va shu orqali tarjima nazariyasidagi “ekvivalentlik” tushunchasini empirik material bilan boyitadi.

Bunday korpus kompyuter lingvistikasi uchun ham tayyor o‘quv materiali vazifasini bajaradi. Mashina tarjimasi modellarini o‘qitishda tarixiy matnlar asosidagi parallel ma’lumotlar ayniqsa qimmatli, chunki ular zamonaviy til modellari uchun odatda mavjud bo‘lmagan grammatik va leksik shakllarni o‘z ichiga oladi. Model shu orqali nafaqat zamonaviy tilni, balki uning tarixiy qatlamlarini ham qisman o‘zlashtiradi. Bu esa past

resursli tillar uchun sifatli tarjima tizimlarini yaratishda sezilarli natija beradi. Shu bilan birga, morfologik tahlil yoki sintaktik parsning kabi vazifalarda ham bunday korpusdan foydalangan holda yanada aniqroq algoritmlar ishlab chiqish mumkin.

Matnning stilistik qatlamini tahlil qilishda ham parallel korpus muhim rol o'ynaydi. Asl matndagi tasvir vositalari, metaforalar yoki sintaktik qurilmalar tarjimada qanday aks etgani segmentlar kesimida solishtiriladi. Bu jarayonda muallif uslubining qaysi elementlari saqlanib qolganini, qaysilari esa yo'qolgan yoki transformatsiyaga uchraganini aniqlash mumkin bo'ladi. Shunday qilib, adabiy tarjimada "uslubiy adekvatlik" tushunchasi konkret dalillar bilan izohlanadi.

"Boburnoma"ning parallel korpusi boburshunos olimlar uchun matnni talqin qilishda subyektivlik darajasini kamaytiradigan aniq dalillar maydonini yaratadi. Asarning turli tarjimalari va nashr variantlari bir tizimda jamlangani sababli, tadqiqotchi muayyan parchani yakka holatda emas, balki boshqa talqinlar bilan yonma-yon ko'rish imkoniga ega bo'ladi. Bu esa, ayniqsa, bahsli yoki ko'p ma'noli o'rinlarda muallif niyatini aniqlashda muhim metodik ustunlik beradi. Natijada sharh yozish jarayoni individual taxminga emas, balki qiyosiy kuzatuvga tayangan holda amalga oshiriladi.

Korpus boburshunoslikda matn variantlarini tekshirish va tanqidiy nashr tayyorlash jarayonini ham sezilarli darajada yengillashtiradi. Turli qo'lyozmalar yoki nashrlar o'rtasidagi tafovutlar avtomatik ravishda aniqlanib, qaysi variantning ishonchliroq ekani haqida asosli xulosa chiqarish imkoniyati paydo bo'ladi. Shu yo'l bilan matnshunos olimlar uchun qo'lda solishtirishga ketadigan vaqt qisqaradi va e'tibor mazmuniy tahlilga ko'proq qaratiladi.

Shuningdek, bunday korpus tarixiy shaxslar, joy nomlari va madaniy tushunchalarning izchil identifikatsiyasini ta'minlaydi. "Boburnoma"da uchraydigan ko'plab toponimlar va antroponimlar turli tarjimalarda turlicha shaklda berilgan bo'lishi mumkin; parallel tuzilma ularni yagona tizimda ko'rib chiqish va standartlashtirish imkonini beradi. Bu esa tarixiy-geografik tadqiqotlar olib borayotgan boburshunoslar uchun muhim ahamiyat kasb etadi, chunki ma'lumotlarning tarqoqligi kamayadi va izchillik ta'minlanadi.

Parallel korpus boburshunoslikdagi intertekstual tadqiqotlarni ham yangi bosqichga olib chiqadi. Olimlar "Boburnoma"dagi ayrim motivlar, iboralar yoki tarixiy tasvirlarni boshqa manbalar bilan tez va aniq qiyoslay oladi. Shu orqali asarning boshqa tarixiy yoki adabiy manbalar bilan aloqalari chuqurroq o'rganiladi, Boburning dunyoqarashi va manba bazasi haqida yanada asosli xulosalar chiqariladi.

Bundan tashqari, korpus ilmiy hamkorlikni kuchaytiradigan infratuzilma vazifasini bajaradi. Turli mamlakatlarda faoliyat yuritayotgan boburshunoslar yagona raqamli platforma orqali bir xil ma'lumotlar bilan ishlash imkoniga ega bo'ladi. Bu esa tadqiqot natijalarining taqqoslanishini osonlashtiradi, bahs-munozaralarni aniq dalillar asosida olib borishga yordam beradi va sohada umumiy metodologik standartlarning shakllanishiga

turtki beradi. Buning natijasida “Boburnoma” parallel korpusi boburshunos olimlar uchun oddiy yordamchi vosita emas, balki tadqiqot metodologiyasini o‘zgartiradigan, talqin aniqligini oshiradigan va ilmiy muloqotni yangi darajaga olib chiqadigan asosiy ilmiy manbaga aylanadi.

“Boburnoma”ning parallel matnini yaratish oddiy matnlarni yonma-yon qo‘yish emas, balki qat’iy ilmiy tamoyillarga asoslangan murakkab jarayon hisoblanadi. Eng avvalo, asos qilib olinadigan matnning ishonchliligi hal qiluvchi ahamiyatga ega bo‘ladi. Asarning qaysi nashri yoki qo‘lyozmasi bazaviy matn sifatida tanlanayotgani aniq belgilanishi kerak, chunki turli nashrlar o‘rtasidagi kichik farqlar keyingi hizalanish jarayonida katta tafovutlarga olib kelishi mumkin. Shu sababli tanqidiy nashrga tayangan holda ishlash yoki kamida bir nechta variantni solishtirib, asosiy matnni aniqlab olish zarur.

Keyingi muhim bosqich – matnni bo‘laklarga ajratish. Bu oddiy ko‘rinadi, lekin amalda ancha murakkab. Parallel korpus yaratishda matnni gap, jumla yoki mazmuniy bo‘laklarga qanday ajratish masalasi bevosita natijaning sifatiga ta’sir qiladi. Agar segmentlar juda yirik bo‘lsa, aniq mosliklarni yo‘qotish xavfi paydo bo‘ladi; haddan tashqari mayda bo‘lsa, mazmuniy yaxlitlik buziladi. “Boburnoma” kabi tarixiy asarda gap tuzilishi zamonaviy til me’yorlaridan farq qilgani uchun segmentatsiya mexanik emas, balki filologik tahlil asosida amalga oshirilishi kerak.

Moslashtirish jarayoni (ya’ni asl matn bilan tarjimani yonma-yon qo‘yish) eng mas’uliyatli qism hisoblanadi. Hizalanish (alignment) jarayonida semantik moslikni saqlash ustuvor mezon bo‘lishi lozim. Tarjimalar ko‘pincha so‘zma-so‘z emas, balki mazmuniy jihatdan erkin bo‘lganligi sababli, har bir segmentni faqat tashqi shakliga qarab emas, balki ma’nosiga ko‘ra moslashtirish talab etiladi. Bu yerda yarim avtomatik usullar eng samarali hisoblanadi: dastlab algoritmlar yordamida taxminiy mosliklar aniqlanadi, so‘ngra mutaxassis tomonidan qo‘lda tekshiriladi va tuzatiladi. To‘liq avtomatik hizalanish tarixiy matnlar uchun ko‘pincha xatolik darajasini oshiradi.

Imlo va transkripsiya masalalari ham alohida e’tibor talab qiladi. Eski o‘zbek tilidagi matnlar arab yozuvida bo‘lgani sababli ularni lotin yoki kirill yozuviga o‘tkazishda yagona standart tanlanishi kerak. Aks holda, bir xil so‘zning turli yozuv variantlari korpus ichida chalkashlik keltirib chiqaradi va qidiruv hamda tahlil jarayonini murakkablashtiradi. Shu bilan birga, asl yozuv shaklini ham saqlab qolish maqsadga muvofiq, chunki u filologik tadqiqotlar uchun muhim manba hisoblanadi.

Leksik va grammatik belgilash (annotatsiya) bosqichi korpusning ilmiy qiymatini sezilarli darajada oshiradi. Har bir segment yoki so‘z uchun morfologik tahlil, so‘z turkumi, asosiy shakl (lemma) kabi ma’lumotlar kiritilsa, korpus oddiy matn to‘plamidan ilmiy tadqiqot vositasiga aylanadi. Biroq bu jarayonda tarixiy til xususiyatlarini hisobga oladigan maxsus belgilash tizimini ishlab chiqish zarur, chunki zamonaviy o‘zbek tiliga mo‘ljallangan annotatsiya standartlari to‘liq mos kelmasligi mumkin.

Texnik jihatdan korpusni standart formatlarda (masalan, XML yoki TEI) saqlash zarur. Bu nafaqat ma'lumotlarni uzoq muddat saqlash, balki boshqa tadqiqotchilar bilan almashish va turli dasturiy vositalar orqali qayta ishlash imkonini ham beradi. Metadata – ya'ni matn manbasi, tarjimon, nashr yili, hizalanish darajasi kabi qo'shimcha ma'lumotlar ham to'liq kiritilishi kerak, chunki ular keyingi ilmiy tahlil uchun kontekst yaratadi.

Sifat nazorati jarayoni doimiy bo'lishi kerak. Korpus bir marta tuzilib, o'zgarmas holatda qoladigan mahsulot emas; u tekshirib boriladigan, xatolari tuzatiladigan va boyitiladigan dinamik tizim hisoblanadi. Shu sababli ekspertlar tomonidan muntazam ravishda audit o'tkazish, foydalanuvchilardan kelgan taklif va tuzatishlarni inobatga olish muhim.

**Xulosa.** “Boburnoma” asosida parallel korpus yaratishning ilmiy ahamiyati, eng avvalo aniq, tekshiriladigan til birliklari o'rtasidagi mosliklarni sistematik ravishda qayd etish imkoniyati bilan belgilanadi. Asarning asl matni segmentlarga ajratilib, ularning zamonaviy o'zbek va boshqa tillardagi tarjimalari bilan qat'iy hizalanishi natijasida “birlik–ekvivalent” munosabati sun'iy emas, balki real matn kontekstida kuzatiladi. Shu asosda tarixiy leksik birliklarning semantik diapazoni kengayganmi yoki torayganmi, ma'no ko'chishlari qaysi kontekstlarda yuz bergani aniq misollar orqali ko'rsatiladi. Masalan, bir so'zning turli joylarda turlicha tarjima qilinishi uning polisemiyasini yoki kontekstga bog'liqligini dalillaydi; bu esa lug'at tuzishda yoki tarixiy semantika rekonstruksiyasida muhim dalil bo'lib xizmat qiladi. “Boburnoma” parallel matnini yaratish filologik aniqlik, texnik standartlash va metodologik izchillik uyg'unligini talab qiladigan jarayon bo'lib, har bir bosqichdagi e'tiborsizlik yakuniy korpusning ilmiy qiymatini pasaytirishi mumkin.

#### Foydalanilgan adabiyotlar ro'yxati:

1. Abdullayeva O. O'zbek tilining internet axborot matnlari korpusini shakllantirishning nazariy va amaliy asoslari. Filol. fan.b. fals. dok.(PhD) ...diss. – Toshkent, 2022. – 142 b.
2. Каримов Р. Ўзбек-инглиз параллел корпусини тузишнинг лингвистик ва дастурий масалалари. Филол. фан. б. фалс. док. (PhD) ...дисс. автореф. – Бухоро, 2022. – 63 б.
3. Rahimov A. Kompyuter lingistikasi asoslari. –Toshkent: Akademnashr, 2011. – 155 б.
4. Рахманова А. Ўзбек тили миллий корпусини яратишда компьютер усуллари. Филол. фан. б. фалс. док. (PhD) ... дисс. – Тошкент, 2021. – 62 б.
5. Жўраева Н., Шаропова С. Параллел корпус тузиш муаммолари // O'zbek tilining milliy korpusi: muammo va vazifalar. – Toshkent, 2022.
6. – B. 293–298.
7. Musulmankulova Sh. O'zbekcha va turkcha frazemalar korpusining lingvistik ta'minoti. Filol. fan. b. fals. dok. (PhD) ... diss. – Toshkent, 2024. – B. 35.
8. Xolmanova Z. O'zbek milliy va ta'limiy korpuslarini yaratishning nazariy hamda amaliy masalalari // Xalqaro ilmiy-amaliy konferensiya materiallari. – Toshkent, 2021. – B. 56.