

MATNDAN KORPUSGA: O'ZBEK SHE'RIYATI MUALLIFLIK KORPUSINI YARATISHDAGI METODOLOGIK MUAMMOLAR

Maxammadiyeva Diyora Raxmatovna,
Tayanch doktorant
diyoramaxammadiyeva0@gmail.com
ToshDO'TAU

Annotatsiya. Mazkur maqolada o'zbek she'riyati mualliflik korpusini yaratish jarayonida yuzaga keladigan metodologik muammolar tizimli ravishda tahlil qilingan. Poetik matnlar oddiy matnlar to'plami sifatida emas, balki murakkab, ko'p qatlamli til material sifatida talqin qilinadi. Tadqiqotda poetik korpus yaratishning asosiy muammolari – matn chegarasini aniqlash, janr aralashuvi, mualliflik masalasi, xronologik noaniqlik hamda ortiqcha elementlar muammosi batafsil yoritiladi. Shuningdek, ushbu muammolarni hal etish uchun filtrlash, normalizatsiya, annotatsiya va muvozanatlashga asoslangan konseptual model taklif etiladi. Tadqiqot natijalari poetik korpusni shakllantirish jarayonini standartlashtirish, uning reprezentativligini oshirish va keyingi lingvostatistik hamda stilistik tahlillar uchun ishonchli empirik asos yaratishga xizmat qiladi.

Kalit so'zlar: *Poetik korpus, korpus lingvistikasi, mualliflik korpusi, matn tanlash, reprezentativlik, annotatsiya, konkordans, korpus dizayni.*

Abstract. This article provides a systematic analysis of the methodological challenges involved in constructing an authorial corpus of Uzbek poetry. Poetic texts are treated not as a simple collection of texts, but as complex and multi-layered linguistic material. The study examines key issues in poetic corpus construction, including the definition of text boundaries, genre hybridity, authorship attribution, chronological uncertainty, and the problem of extraneous elements (“noise”). In addition, a conceptual model based on filtering, normalization, annotation, and balancing is proposed to address these challenges. The findings contribute to the

standardization of poetic corpus design, enhance its representativeness, and provide a reliable empirical foundation for subsequent linguostatistical and stylistic analyses.

Keywords: *Poetic corpus, corpus linguistics, authorial corpus, text selection, representativeness, annotation, concordance, corpus design.*

Zamonaviy tilshunoslikda empirik yondashuvlar ustuvor metodologik yo'nalish sifatida shakllanib, til birliklarini real foydalanish asosida o'rganish imkonini kengaytirdi. Ayniqsa, korpus lingvistikasi tilni kuzatish, tavsiflash va modellashtirishda katta hajmdagi matnlar asosida statistik va kontekstual tahlilni amalga oshirish imkonini beruvchi muhim ilmiy paradigma sifatida rivojlanmoqda. Ushbu yondashuv til birliklarining nafaqat chastotaviy xususiyatlarini, balki ularning real diskursdagi funksional va kombinator xususiyatlarini aniqlashga xizmat qiladi. Shuningdek, matnlarni katta hajmda yig'ish, saqlash va avtomatik tahlil qilish imkoniyatlarini yanada kengaytiradi. Bu yo'nalish badiiy matnlarni, xususan poetik diskursni yangi metodlar asosida o'rganishga yo'l ochdi. Poetik matnlarni ilmiy o'rganish an'anaviy ravishda metrika, ritmika, qofiya va stilistik tahlil doirasida olib borilgan bo'lib, bu yo'nalishda rus she'riyatshunosligida muhim nazariy asoslar yaratilgan. Xususan, M.Gasparov [4] tomonidan ishlab chiqilgan metrika va ritmika nazariyalari she'riy tizimni formal-struktur jihatdan tahlil qilishda fundamental ahamiyat kasb etadi. Uning ishlari she'riy nutqning ritmik tomonini tarixiy va tipologik nuqtai nazardan izohlashga xizmat qilib, keyingi tadqiqotlar uchun metodologik asos bo'lib xizmat qilgan. Shuningdek, A.Andreyeva [2], K. Vlasov [3] kabi tadqiqotchilar tomonidan Iosif Brodskiy she'riyati misolida tonik she'r, metrika hamda ritm-sintaksis o'zaro munosabatlari chuqur tahlil qilingan. Zamonaviy bosqichda korpus lingvistikasi poetik tahlilga empirik va kvantitativ yondashuvni olib kirdi. Xususan, E.Grishina, K.Korchagin, V.Plungyan va D.Sichinava [5] tomonidan yaratilgan poetik korpus modeli annotatsiyalangan matnlar asosida stilistik naqshlarni aniqlash imkonini berdi. Eng mashhur poetik korpuslardan biri

Russian National Corpus tarkibida esa poetik subcorpus tuzilgan bo'lib, korpus semantik va morfologik annotatsiyani taqdim etadi [9]. Turkiy tillardan qozoq tilshunoslari tomonidan Abay Qunanboyev she'riyati asosida yaratilgan poetik korpusda chastota tahlili, fonetik-ritmik annotatsiya hamda NLP asosidagi semantik modellash usullari qo'llanilgan [1:405-410]. O'zbek tilshunosligida ham korpus lingvistikasi so'nggi yillarda faol rivojlanayotgan yo'nalishlardan biri bo'lib, bu borada mualliflik va milliy korpuslarni yaratish masalalari alohida ilmiy ahamiyat kasb etmoqda. Xususan, Sh.Hamroyeva tadqiqotida [8] o'zbek tili mualliflik korpusini tuzishning lingvistik asoslari ishlab chiqilib, korpusni shakllantirishda matn tanlash, annotatsiya va struktur tashkil etish masalalari tizimli yoritilgan. Ushbu ish mualliflik korpusining nazariy modeli va uning lingvistik parametrlarini belgilashda muhim metodologik asos bo'lib xizmat qiladi. G.Toirova [7] esa o'zbek tili milliy korpusini yaratishning nazariy va amaliy jihatlarini tahlil qilib, korpusning reprezentativligi, balanslanganligi va funksional imkoniyatlarini ta'minlash muammolariga e'tibor qaratadi. Tadqiqotda korpus dizayni, matnlarni tanlash me'zonlari hamda ularni tizimlashtirish prinsiplari keng yoritilgan. Shu bilan birga, mavjud tadqiqotlarda poetik matnlar asosida mualliflik korpusini yaratishning o'ziga xos metodologik muammolari – she'riyatning yagona tip emasligi, ya'ni unda lirika, epik she'r, dramatik she'r, doston, muallif izohlari kabi turli stilistik-struktur xususiyatga ega shakllarning mavjudligi va matn chegaralarining noaniqligi masalalari yetarlicha chuqur tahlil qilinmagan. Mazkur tadqiqot aynan shu bo'shliqni to'ldirishga qaratilgan.

Mavjud tadqiqotlarda poetik korpuslar yaratish masalasi ko'pincha texnik bosqichlar, asosan, raqamlashtirish, tozalash va annotatsiya doirasida ko'rib chiqilgan bo'lib, matn tanlash, ya'ni korpusga qaysi she'rlar, qaysi janrlar va qaysi matn komponentlarini kiritish masalasi yetarli darajada nazariy asoslanmagan. Masalan, poetik matn chegarasi qanday aniqlanadi, muallif izohlari va aralash janr

elementlari korpusga kiritiladimi degan savollar ochiq qolmoqda. Mazkur holat poetik korpus yaratishda matn tanlash jarayonini metodologik jihatdan qayta ko'rib chiqish zaruratini yuzaga keltiradi.

Poetik matnlarni korpus materiali sifatida saralash – bu oddiy “matn yig‘ish” jararayoni emas. Sababi she'riy matnlarning semantik ko'pqatlamliligi, yuqori darajadagi metaforikligi va kontekstga kuchli bog'liqligi ularni standart korpus dizayn tamoyillari asosida modellashni qiyinlashtiradi. Shu bois poetik korpus yaratishda boshqa janrlarga qaraganda ko'proq noaniqlik va metodologik muammolarni keltirib chiqaradi. Avvalo, poetik matnlarning chegarasini aniqlash masalasi muhim metodologik muammolardan biridir. Tadqiqot doirasida faqat asosiy she'riy matnlarni korpusga kiritish va paratekstual elementlarni (izohlar, bag'ishlovlar, nasriy fragmentlar, tarjimalar) chiqarib tashlash stilistik tahlilning tozaligini ta'minlashga xizmat qiladi. Bu yondashuv konkordans natijalarida leksik birliklarning haqiqiy poetik funksiyasini aniqlash imkonini beradi. Poetik matnlarning semantik zichligi, struktur erkinligi va janriy o'zgaruvchanligi korpus dizaynida an'anaviy yondashuvlarni qayta ko'rib chiqishni talab qiladi. Quyida poetik korpus yaratishda yuzaga keladigan asosiy metodologik muammolar va ularning yechimlari tizimli ravishda bayon etiladi.

1. Matn chegarasi muammosi.

Poetik matnlar bilan ishlashda eng birlamchi muammolardan biri sarlavha, epigraf, muallif izohlari va bag'ishlovlar kabi elementlar matn tarkibiga aralashib ketishidir. Mualliflik poetik korpusini tuzishda, bizningcha, asosiy she'riy matn **asosiy matn qatlami** sifatida ajratilishi kerak. Sarlavha **metadata elementi sifatida**, epigraf va izohlar **esa paratext** sifatida alohida qatlamga chiqariladi. Konkordans va chastota tahlili faqat asosiy matn qatlami asosida amalga oshiriladi. Bu yondashuv konkordans namunalarini turli “shovqin”dan ajratishga imkon beradi.

2. Janr aralashuvi.

Poetik korpuslarda ko‘pincha sof lirika bilan bir qatorda dramatik yoki epik elementlar ham uchraydi (masalan, A.Oripov “Sohibqiron” she‘riy drama yoki M.Yusuf “Qora quyosh” nasriy aralashgan she‘riy doston). Bunday matnlar stilistik jihatdan notekis korpus hosil qiladi. Shuning uchun korpus **janr bo‘yicha filtrlanadi**. Aralash janr matnlari alohida subkorpus sifatida belgilanadi Bu qiyosiy tahlillarning metodologik tozaligini ta‘minlaydi.

3. Mualliflik masalalari.

Poetik matnlarda mualliflik har doim ham aniq emas. Xususan, tarjima she‘rlar, qayta ishlangan matnlar, bir nechta variantlar. Bu esa mualliflik idiositili tahlilida noto‘g‘ri xulosa chiqarishga olib keladi. Bizningcha, tarjimalar asosiy korpusdan chiqarilib, subkorpus sifatida kiritilishi to‘g‘ri qaror bo‘ladi. Variantlar ichidan nashrlar yillariga qarab yakuniy versiya tanlanadi. Har bir matn uchun **mualliflik turi: asl/ tarjima** belgisi kiritiladi. Bu muallif idiositili haqida to‘g‘ri xulosa chiqarishga imkon beradi.

4. Xronologik noaniqlik.

Ko‘plab she‘rlarning yozilgan aniq sanasi mavjud emas yoki mualliflarning ayrim she‘rlarida yil belgilangan, ayrim she‘rlarida esa yillar ko‘rsatilmagan. Bunday hollarda shoir ijodini vaqt bo‘yicha taqsimlash subyektiv yoki noaniq bo‘lib qoladi. She‘riy korpusga metama‘lumot tayyorlashda yozilish yili noaniq bo‘lgan she‘rlarga nisbatan she‘r nashr qilingan to‘plamning yilini belgilashni qo‘shimcha parametr sifatida qo‘llash mumkin.

Yuqoridagi muammolar shuni ko‘rsatadiki, poetik korpus yaratish oddiy matn yig‘ish jarayoni emas, balki murakkab metodologik tanlovlar tizimidir. Har bir qaror lingvostatistik, stilistik tahlil natijalariga bevosita ta‘sir qiladi. Shu sababli poetik korpus dizayni nazariy asoslangan va qat‘iy nazorat qilingan bo‘lishi lozim. Muammo va taklif etilgan yechimlar 1-jadvalda ifodalangan.

1-jadval. Poetik korpus yaratishdagi metodologik muammolar va ularning yechimlari.

№	Muammo	Tavsif	Taklif etilgan yechim
1.	Matn chegarasi muammosi	She'r matni va paratekst (sarlavha, epigraf, izoh) o'zaro aralashib ketadi	Asosiy matn va metadata qatlamlarga ajratish
2.	Janr aralashuvi	Lirik, dramatik va epik elementlar bir korpusda birlashadi	Janr bo'yicha filtr qo'llash yoki subkorpuslarga ajratish
3.	Mualliflik noaniqligi	Tarjima, variant va qayta ishlangan matnlar mavjud	Mualliflik turini annotatsiya qilish (original / tarjima)
4.	Xronologik noaniqlik	She'rlarning aniq yozilgan yili noma'lum	Taxminiy vaqt metama'lumotlarini kiritish (davr / nashr yili)

Poetik korpusni shakllantirishning konseptual modeli. Poetik korpusni shakllantirish jarayoni izchil, bosqichma-bosqich amalga oshiriladi. Jarayon xom matnlar to'plamidan boshlanadi. Bu bosqichda she'rlar bilan bir qatorda izohlar, aralash janrlar (she'r-nasr, she'riy drama va boshqalar), nashr parametrlari, muharrir izohlari hamda turli metama'lumotlar birgalikda mavjud bo'ladi. Shuning uchun keyingi filtrlash bosqichi muhim ahamiyat kasb etib, unda janr seleksiyasi amalga oshiriladi hamda mualliflik masalalari aniqlashtiriladi (asl matn, tarjima). Keyingi bosqich – matn chegarasini aniqlash. Poetik korpus yaratishdagi eng nozik nuqtalardan biri bo'lib, unda asosiy matn (she'rning o'zi) paratekst elementlaridan (sarlavha, izoh, epigraf va boshqalar) ajratiladi. Shundan so'ng oldindan ishlov berish bosqichi amalga oshirilib, matn texnik jihatdan tozalanadi va yagona formatga keltiriladi. Annotatsiya bosqichi korpusning ilmiy qiymatini belgilovchi

asosiy qatlam hisoblanadi. Bu yerda har bir matnga muallif, janr, davr, matn turi kabi metama'lumotlar biriktiriladi. Natijada muvozanatlangan tanlanma shakllanib, korpus tarkibi nazorat ostida, reprezentativ va tahlil uchun yaroqli holatga keltiriladi. Yakuniy bosqichda tozalangan va strukturaviy jihatdan tashkil etilgan poetik korpus hosil bo'ladi. Ushbu korpus keyingi tahlil bosqichi uchun asos bo'lib xizmat qiladi, bunda chastota tahlili, konkordans, kollokatsion naqshlar va stilistik xususiyatlar empirik asosda aniqlanadi.

Konseptual model. Xom matnlar to'plami (she'rlar + izohlar + aralash janrlar + metadata) → Filtrlash bosqichi (janr tanlash, mualliflik nazorati) → Matn chegarasini aniqlash (asosiy matn va paratekstni ajratish) → Oldindan ishlov bosqichi (tozalash, normalizatsiya qilish) → Annotatsiya (metadata: muallif, janr, davr, matn turi) → Muvozanatlangan tanlanma matnlar (nazorat qilingan korpus tarkibi) → Yakuniy poetik korpus (tozalangan, tuzilgan, tahlilga tayyor) → Tahlil bosqichi (chastota → konkordans → kollokatsiya → stilistika).

Tadqiqot natijalari shuni ko'rsatadiki, poetik korpus yaratish jarayonida janrni tasniflash, mualliflikni belgilash hamda vaqtga oid metama'lumotlarni aniqlashtirish kabi muammolar mavjud. Shu bois korpusni shakllantirishda qat'iy filtrlash mezonlari, aniq annotatsiya modeli va muvozanatlangan tanlanma strategiyasidan foydalanish zarur. Taklif etilgan konseptual model poetik korpusni tizimli ravishda qurish imkonini berib, uning reprezentativligi va ilmiy ishonchliligini ta'minlaydi. Natijada yaratilgan korpus nafaqat lingvistik, balki adabiyotshunoslik tadqiqotlari uchun ham samarali empirik baza vazifasini bajaradi. Ushbu yondashuv o'zbek tilida korpus lingvistikasi va raqamli adabiyotshunoslik yo'nalishlarining rivojiga xizmat qilib, kelgusida boshqa mualliflik korpuslarini yaratish uchun metodologik asos bo'lib xizmat qilishi mumkin.

Foydalanilgan adabiyotlar ro'yxati

1. Pirmanova K., Madiyeva G., Serikbayeva A. et al., 2025. Creation of a Small Poetic Corpus Based on Abai's Poetry. Forum for Linguistic Studies. 7(9): 402–420. DOI: <https://doi.org/10.30564/fls.v7i9.9547>
2. Андреева А. Эволюция тонического стиха в поэзии Иосифа Бродского. Дисс... канд. филолог. наук. М., 2003.
3. Власов К. Система стиха И.А.Бродского (Проблемы метра и ритма). Дисс... канд. филол. наук. Орел, 2005.
4. Гаспаров М. Очерк истории русского стиха. Метрика. Ритмика. Рифма. Строфика. Изд. 2 (доп.). М.: Фортуна Лимитед, 2002.
5. Гришина Е., Корчагин К., Плунгян В., Сичинава Д. Поэтический корпус в рамках Национального корпуса русского языка: общая структура и перспективы использования.//Нацио-нальный корпус русского языка: 2006–2008. Новые результаты и перспективы / Отв. ред. В.А. Плунгян. СПб.: Нестор-История, 2009. С. 71–114.
6. Орлицкий Ю. Стих и проза в русской литературе. М.: РГГУ, 2002.
7. Тоирова Г. Ўзбек тили миллий корпусини яратишнинг назарий ва амалий масалалари. Филология фанлари доктори (DSc) диссертацияси автореферати. – Бухоро, 2021.
8. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Фил.фан.бўйича фалсафа доктори (PhD)...дис. – Бухоро, 2018. – 250 б.;
9. <https://ruscorpora.ru/en/corpus/poetic>