
Methodology for creating a parallel corpus of Oybek's novel "Navoiy" and annotating literary devices used to depict human appearance

Maftuna Mirzayeva Bakhtiyor kizi¹

maftuna@urdu.uz

maftunam842@gmail.com

PhD Student,

Urgench State University named after

Abu Rayhan Biruni

-
- Annotation** *The integration of corpus linguistics and translation studies has expanded the possibilities for the empirical investigation of literary texts. This article describes the principles of constructing a parallel corpus based on Oybek's novel Navoi and its English translation, as well as a multi-layer annotation model for literary devices representing human appearance. The corpus was compiled according to the principles of representativeness, completeness, consistency, and reproducibility. Each portrait unit was treated as an independent observation unit and assigned a unique identification code that records its semantic category, type of stylistic device, linguistic realization, translation equivalent, and translation transformation. Statistical analysis was conducted using the token principle, while the source and target texts were aligned through the sentence alignment method based on semantic equivalence. The proposed methodological model provides a systematic framework for investigating the lexical-semantic and translational characteristics of portrait units in a parallel corpus and establishes a methodological basis for corpus-based and translation-oriented studies of Uzbek literary texts.*
- Keywords** *Parallel corpus, annotation, stylistic devices, portrait description, lexical-semantic relations, translation studies*

Oybekning "Navoiy" romani parallel korpusini yaratish va inson qiyofasini ifodalovchi badiiy tasvir vositalarini annotatsiyalash metodologiyasi

Mirzayeva Maftuna Baxtiyor qizi

maftuna@urdu.uz

maftunam842@gmail.com

Tayanch doktorant (PhD),

Abu Rayhon Beruniy nomidagi Urganch

davlat universiteti

-
- Annotatsiya** *Korpus lingvistikasi va tarjimashunoslikning integratsiyalashuvi badiiy matnlarni empirik asosda tadqiq etish imkoniyatlarini kengaytirmoqda. Mazkur maqolada Oybekning «Navoiy» romani va uning inglizcha tarjimasi asosida yaratilgan parallel korpusni shakllantirish tamoyillari hamda inson qiyofasini ifodalovchi badiiy tasvir vositalarini ko'p qatlamli annotatsiyalash modeli yoritiladi. Tadqiqotda parallel korpus reprezentativlik, to'liqlik, izchillik va takror ishlab bo'ladiganlik tamoyillari asosida shakllantirildi. Har bir portret birligi mustaqil kuzatuv obyekti sifatida kodlanib, uning semantik kategoriyasi, badiiy tasvir vositasi turi, lisoniy realizatsiyasi, tarjimadagi ekvivalenti hamda tarjimaviy transformatsiya turi yagona identifikatsiya tizimi asosida annotatsiyalandi. Statistik tahlilda token tamoyili,*

¹ Mirzayeva M.B. — ORCID: 0009-0003-1898-9355

matnlarni o'zaro moslashtirishda esa sentence alignment usuli qo'llanildi. Natijada portret birliklarining leksik-semantik va tarjimaviy xususiyatlarini tizimli tahlil qilish imkonini beruvchi metodologik model ishlab chiqildi. Ushbu model o'zbek badiiy matnlarini parallel korpus asosida o'rganish, korpus annotatsiyasini takomillashtirish hamda korpus lingvistikasi va tarjimashunoslik yo'nalishidagi keyingi empirik tadqiqotlar uchun metodologik asos yaratadi. Taklif etilgan model badiiy tarjimalarni korpus texnologiyalari asosida qiyosiy tahlil qilish, tarjima strategiyalarini aniqlash hamda lingvistik tadqiqotlar natijalarining ilmiy ishonchliligi oshirishga xizmat qiladi.

Kalit so'zlar *Parallel korpus, annotatsiya, badiiy tasvir vositalari, portret tavsifi, leksik-semantik munosabatlar, tarjimashunoslik*

**Методика создания
параллельного корпуса романа
Ойбека «Навои» и
аннотирования
художественных средств,
используемых для описания
внешности человека**

Мирзаева Мафтуна Бахтиёр кизи

maftuna@urdu.uz

maftunam842@gmail.com

Докторант (PhD),

Ургенчский государственный университет
имени Абу Райхана Беруни

Аннотация *Интеграция корпусной лингвистики и переводоведения существенно расширяет возможности эмпирического исследования художественных текстов и открывает новые перспективы для объективного анализа переводческих решений. В статье рассматриваются принципы формирования параллельного корпуса на материале романа Ойбека «Навои» и его английского перевода, а также предлагается многоуровневая модель аннотирования художественных средств, репрезентирующих внешний облик персонажей. Корпус сформирован на основе принципов репрезентативности, полноты, последовательности, воспроизводимости и единообразия обработки данных. Каждая портретная единица рассматривается как самостоятельный объект анализа и кодируется посредством единой системы идентификации, включающей семантическую категорию, тип художественного средства, языковую реализацию, переводческий эквивалент, вид переводческой трансформации и дополнительные параметры аннотации. Статистическая обработка материала выполняется на основе принципа token, а сопоставление оригинального и переводного текстов осуществляется методом sentence alignment, основанным на семантической эквивалентности соответствующих фрагментов. Предлагаемая методологическая модель обеспечивает системное изучение лексико-семантических, стилистических и переводческих особенностей портретных единиц в параллельном корпусе, способствует повышению объективности результатов исследования и может служить универсальной методологической основой для дальнейших*

корпусных, переводоведческих и лингвистических исследований узбекских художественных текстов.

Ключевые слова *Параллельный корпус, аннотирование, художественные средства, портретное описание, лексико-семантические отношения, переводоведение*

Kirish

So'nggi yillarda korpus lingvistikasi tilshunoslikning eng jadal rivojlanayotgan yo'nalishlaridan biriga aylandi (Hasselgård, 2020). Elektron matnlar bazasini tizimli shakllantirish va ularga ko'p qatlamli annotatsiya qo'llash imkoniyati til birliklarini empirik ma'lumotlar asosida o'rganish, ularning qo'llanish chastotasi, funksional xususiyatlari hamda semantik munosabatlarini obyektiv tahlil qilish uchun yangi metodologik imkoniyatlarni yaratdi (O'Brien, 2012). Ayniqsa, parallel korpuslardan foydalanish original va tarjima matnlari o'rtasidagi mosliklarni aniqlash, tarjimaviy transformatsiyalarni kuzatish hamda badiiy uslubning tarjimada qay darajada saqlanishini baholashda muhim vosita sifatida e'tirof etilmoqda (Baker, 1993).

Biroq o'zbek tilshunosligida korpus texnologiyalaridan foydalanishga bag'ishlangan tadqiqotlar soni ortib borayotgan bo'lsa-da, badiiy matnlarda inson qiyofasini ifodalovchi birliklarni parallel korpus asosida kompleks annotatsiyalash metodologiyasi hali yetarli darajada ishlab chiqilmagan (Abdurakhmonova et al., 2021). Amaldagi tadqiqotlarning aksariyatida badiiy tasvir vositalari stilistik yoki lingvopoetik nuqtai nazardan tavsiflanadi, biroq ularni korpus lingvistikasi prinsiplari asosida tizimlashtirish, ko'p qatlamli kodlash hamda tarjima matni bilan bog'liq holda tahlil qilish masalalari cheklangan darajada yoritilgan (Zanettin, 2014). Natijada bunday birliklarning semantik taqsimoti, funksional faolligi va tarjimadagi realizatsiyasini miqdoriy jihatdan baholash imkoniyati yetarlicha shakllanmagan.

Mazkur muammo ayniqsa tarixiy badiiy romanlarda yaqqol namoyon bo'ladi. Tarixiy romanlarda personajlarning tashqi qiyofasi nafaqat vizual tavsif vazifasini bajaradi, balki ularning individual xarakteri, ruhiy holati, ijtimoiy maqomi va muallifning estetik munosabatini ifodalovchi murakkab lingvopoetik tizimning tarkibiy qismi hisoblanadi (Alves, 2003). Shu bois portret tavsiflarini alohida leksik birliklar emas, balki semantik, stilistik va tarjimaviy qatlamlarni o'zida mujassamlashtirgan kompleks birlik sifatida tadqiq etish zarurati yuzaga keladi.

Mazkur maqolada aynan shu ilmiy ehtiyojdan kelib chiqib, Oybekning "Navoiy" romani va uning ingliz tilidagi tarjimasi asosida parallel korpus yaratish hamda inson qiyofasini ifodalovchi badiiy tasvir vositalarini annotatsiyalashning metodologik modeli taklif etiladi. Tadqiqot korpusni shakllantirish tamoyillari, kuzatuv birligini belgilash, annotatsiya mezonlari, *token/type* hisoblash modeli, *sentence alignment* usullari va statistik ishlov berish bosqichlarini yagona tizimda tavsiflashga qaratilgan (Simard and Plamondon, 1998).

Maqolaning ilmiy yangiligi shundaki, unda portret birliklarini faqat stilistik hodisa sifatida emas, balki parallel korpusda mustaqil kuzatuv obyekti sifatida tavsiflash imkonini beruvchi ko'p qatlamli annotatsiya modeli ishlab chiqiladi. Ushbu model semantik klassifikatsiya, badiiy tasvir vositasi, lisoniy realizatsiya, tarjimaviy transformatsiya va stilistik natijalarni yagona kodlash tizimida birlashtiradi. Natijada portret birliklarining original va tarjima matnlaridagi realizatsiyasini tizimli ravishda kuzatish hamda ularni

miqdoriy va sifat jihatdan baholash imkoniyati yaratiladi.

Maqolaning maqsadi Oybekning *“Navoiy”* romani asosida parallel korpus yaratish va inson qiyofasini ifodalovchi badiiy tasvir vositalarini annotatsiyalashning ilmiy-metodologik asoslarini ishlab chiqish hamda ushbu modelning korpus lingvistikasi va tarjimashunoslik tadqiqotlaridagi amaliy imkoniyatlarini asoslashdan iborat.

Asosiy

qism

Tadqiqot materiali

Mazkur tadqiqotning empirik asosini Oybekning *“Navoiy”* romani va uning ingliz tiliga tarjima qilingan matni asosida shakllantirilgan parallel korpus tashkil etadi. Tadqiqot materiali bitta badiiy asar doirasida tanlangan bo'lsa-da, ushbu tanlov tasodifiy emas. Asar tadqiqot maqsadi, obyekti va predmeti bilan uzviy bog'liq ilmiy mezonlar asosida saralandi.

Oybekning *“Navoiy”* romani o'zbek tarixiy romanchiligining yuksak namunalaridan biri bo'lib, unda personajlarning tashqi qiyofasi nafaqat vizual tavsif, balki ularning ruhiy holati, ijtimoiy mavqei, xarakteri va muallifning estetik munosabatini ifodalovchi muhim lingvopoetik vosita sifatida namoyon bo'ladi. Shu sababli mazkur roman inson qiyofasini ifodalovchi badiiy tasvir vositalarini korpus asosida tadqiq etish uchun reprezentativ material sifatida tanlandi.

Tadqiqotda tanlanma usulidan foydalanilmadi. Aksincha, romanning to'liq matni tahlil qilinib, barcha bob va syujet yo'nalishlari tadqiqot doirasiga kiritildi. Bunday yondashuv muallif uslubining yaxlitligini saqlash, portret birliklarining haqiqiy qo'llanish chastotasini aniqlash hamda statistik natijalarning ishonchligini ta'minlash imkonini berdi.

Parallel korpusning ikkinchi qismi romanning ingliz tilidagi tarjimasi asosida shakllantirildi. Original va tarjima matnlarining birgalikda qo'llanilishi nafaqat portret birliklarining semantik xususiyatlarini, balki ularning tarjima jarayonidagi realizatsiyasi va

transformatsiyalarini ham tizimli ravishda kuzatish imkonini yaratdi.

Parallel korpusni shakllantirish metodologiyasi

Korpusni shakllantirishda korpus lingvistikasida qabul qilingan reprezentativlik, to'liqlik, izchillik va takror ishlab bo'ladiganlik tamoyillariga amal qilindi. Dastlab original va tarjima matnlari elektron formatga o'tkazildi hamda grafik, punktuatsion va texnik jihatdan yagona standartga keltirildi. Keyinchalik matnlar gap chegaralari bo'yicha segmentatsiya qilinib, annotatsiya uchun tayyorlandi.

Mazkur tadqiqotda korpus oddiy elektron matnlar yig'indisi emas, balki ilmiy tadqiqot ehtiyojlari uchun maxsus ishlab chiqilgan parallel korpus sifatida qaraldi. Shu bois korpusning analitik qatlamiga faqat inson qiyofasini ifodalovchi birliklar kiritildi. Tabiat manzaralari, predmet tavsiflari yoki portret funksiyasini bajarmaydigan boshqa tasvirlar tadqiqot obyektidan tashqarida qoldirildi. Bunday yondashuv korpusning aynan tadqiqot predmetini maksimal darajada aks ettirishiga xizmat qildi.

Annotatsiya modeli

Mazkur tadqiqotning metodologik yangiliklaridan biri inson qiyofasini ifodalovchi badiiy tasvir vositalarini ko'p qatlamli annotatsiyalash modelining ishlab chiqilganidir.

Annotatsiya jarayonida gap yoki abzats emas, balki har bir mustaqil portret birligi alohida kuzatuv obyekti sifatida qabul qilindi. Chunki bitta gap tarkibida qahramonning ko'zi, qoshi, soqoli, yuz ifodasi yoki gavda tuzilishini ifodalovchi bir nechta mustaqil birliklar uchrashi mumkin. Agar bunday holatda butun gap yagona birlik sifatida qaralsa, portretning alohida semantik komponentlarini mustaqil tahlil qilish imkoniyati cheklanadi.

Har bir portret birligiga maxsus identifikatsiya kodi birlashtirildi. Kodlash tizimi gap raqami va shu gap tarkibidagi portret birligi tartib raqamidan tashkil topdi (masalan, **BVT 024.01**, **BVT 024.02**). Mazkur tizim

original va tarjima matnlari o'rtasidagi moslikni kuzatish, statistik hisob-kitoblarni avtomatlashtirish hamda natijalarning tekshiriluvchanligini ta'minlash imkonini berdi.

Annotatsiya quyidagi qatlamlarni o'z ichiga oldi:

- portretning semantik kategoriyasi;
- badiiy tasvir vositasi turi;
- lisoniy realizatsiyasi;
- original matndagi kontekst;
- ingliz tilidagi realizatsiyasi;
- tarjima transformatsiyasi;
- leksik-semantik natija;
- stilistik natija.

Mazkur ko'p qatlamli model bir vaqtning o'zida lingvistik, stilistik va tarjimashunoslik tahlillarini amalga oshirish imkonini beradi.

Hisoblash birligi va sentence alignment

Korpus lingvistikasida statistik natijalarning ishonchliligi kuzatuv birligini to'g'ri belgilashga bevosita bog'liq. Shu sababli mazkur tadqiqotda statistik hisoblashning asosiy birligi sifatida **BTV tokeni** qabul qilindi.

Tadqiqotda **type** ma'lum bir portret birligining modeli, **token** esa uning matndagi har bir real qo'llanish holati sifatida talqin qilindi. Statistik hisoblash aynan token tamoyiliga asoslandi. Natijada portret birliklarining qo'llanish chastotasi hamda ularning tarjimada uchraydigan transformatsiyalari obyektiv ravishda aniqlandi.

Parallel korpusni yaratishning muhim bosqichlaridan biri original va tarjima matnlarini o'zaro moslashtirish (*sentence alignment*) bo'ldi. Tadqiqotda formal sintaktik moslik emas, balki **mazmuniy ekvivalentlik** asosiy mezon sifatida tanlandi. Chunki badiiy tarjimada gaplarni bo'lish, birlashtirish yoki sintaktik qayta qurish holatlari tez-tez uchraydi.

Shu bois korpusda 1:1, 1:2, 2:1 hamda ayrim hollarda 1:n ko'rinishidagi mosliklar qayd etildi. Har bir original portret birligi uning tarjimadagi realizatsiyasi bilan bog'landi. Agar portret tavsifi tarjimada tushirib qoldirilgan bo'lsa, ushbu holat transformatsiya sifatida alohida qayd etildi. Bu esa tarjima

strategiyalarini keyingi bosqichlarda miqdoriy va sifat jihatdan tahlil qilish imkonini berdi.

Statistik tahlil

Annotatsiya yakunlangach, barcha ma'lumotlar elektron ma'lumotlar bazasiga kiritildi. Har bir yozuv identifikatsiya kodi, semantik kategoriya, badiiy tasvir vositasi, lisoniy realizatsiya, tarjima ekvivalenti va transformatsiya turi bilan tavsiflandi.

Statistik ishlov berishda tavsifiy statistika usullaridan foydalanildi. Har bir kategoriya bo'yicha absolyut chastota (**n**) va nisbiy ulush (%) hisoblandi. Olingan natijalar jadvallar va diagrammalar yordamida umumlashtirildi hamda keyingi semantik va tarjimashunoslik tahlillari uchun empirik asos sifatida xizmat qildi

Natijalar va Muhokama

Parallel korpus asosida amalga oshirilgan annotatsiya inson qiyofasini ifodalovchi badiiy tasvir vositalarining semantik va stilistik jihatdan murakkab tizimni tashkil etishini ko'rsatdi. Tahlil qilingan materiallarda har bir portret tavsifi mustaqil annotatsiya birligi sifatida qayd etildi. Natijada bitta gap tarkibida bir nechta mustaqil portret komponentlari uchrashi kuzatildi. Masalan, 157-gapda qahramonlarning tashqi ko'rinishi, kiyim-kechagi, psixologik holati va baholovchi tavsiflarini ifodalovchi o'ndan ortiq badiiy birlik alohida kodlangan. Bu holat portret tavsiflarining sintaktik emas, balki semantik birlik sifatida tahlil qilinishi metodologik jihatdan to'g'ri tanlov bo'lganini tasdiqlaydi.

Tahlil natijalari shuni ko'rsatdiki, annotatsiya qilingan birliklar orasida **epitet** eng faol qo'llanilgan badiiy tasvir vositasi hisoblanadi. Xususan, *kichik, qiyg'och ko'zlari, mag'rur beklar, xitoyi shoyi to'nli, dag'al mallabo'z to'nli, xayolchan* kabi birliklar personajning tashqi ko'rinishi yoki ichki sifatlarini aniqlashtiruvchi atributiv tavsiflar sifatida namoyon bo'ladi. Bu holat Oybekning individual uslubida epitetning portret yaratishdagi markaziy o'rnini ko'rsatadi.

Semantik klassifikatsiya nuqtai nazaridan fiziognomik belgilar, kiyim-kechak obrazi,

psixologik portret hamda baholovchi-emotsional obraz kategoriyalari bir-biri bilan uzviy bog'langan holda qo'llanilgan. Ayniqsa, kiyim-kechakka oid tavsiflar oddiy nominativ vazifani emas, balki qahramonning ijtimoiy mavqei, iqtisodiy holati va estetik qiyofasini ochib beruvchi lingvopoetik vosita sifatida xizmat qiladi. Masalan, *xitoyi shoyi to'nli* birikmasi nafaqat kiyimni, balki yuqori ijtimoiy maqomni ham anglatadi.

Parallel korpus asosida olib borilgan qiyosiy tahlil tarjimada portret birliklarini uzatishda turli strategiyalar qo'llanganligini ko'rsatadi. Ayrim birliklar to'liq ekvivalent orqali ifodalangan bo'lsa (*xayolchan* → *thoughtful*), ayrim hollarda qisman ekvivalent (*xitoyi shoyi to'nli* → *Chinese silk robes*) yoki leksik almashtirish strategiyasi (*qahrli sipohilar-navkarlar* → *rude officers*) kuzatildi. Shu bilan birga, portret tavsiflarining muhim qismi tarjimada umuman ifodalanmaganligi ham aniqlandi. Xususan, *mag'rur beklar*, *olifta bekozodalar*, *guldor mahsili*, *sokin* kabi birliklar inglizcha matnda tushirib qoldirilgan.

Tarjima transformatsiyalarining tahlili omissiya strategiyasi portret tavsiflarida sezilarli darajada uchrashini ko'rsatadi. Bunday hollarda original matndagi baholovchi yoki ekspressiv komponentlar tarjimada yo'qoladi. Natijada

nafaqat leksik ma'no qisqaradi, balki muallif yaratgan stilistik obrazning emotsional ta'siri ham pasayadi. Aksincha, transformatsiyasiz ekvivalent realizatsiya kuzatilgan hollarda original matndagi semantik va stilistik yuklama deyarli to'liq saqlanadi.

Bundan tashqari, generalizatsiya va modulyatsiya kabi transformatsiyalar ham kuzatildi. Masalan, *xarob kosib-hunarmandlar* birikmasining *craftsmen* shaklida berilishi baholovchi semani neytrallashtiradi va ekspressivlikni kamaytiradi. Xuddi shuningdek, *juldur kiyimli dehqonlar* birikmasining *peasants on donkeys* tarzida tarjima qilinishi kiyim belgisi haqidagi axborotni yo'qotib, tavsifning semantik markazini boshqa komponentga ko'chiradi.

Umuman olganda, tahlil qilingan materiallar Oybekning portret yaratish uslubida epitetlarning ustun mavqega ega ekanligini hamda portret tavsiflarining ko'p qatlamli semantik tuzilishga egaligini ko'rsatadi. Parallel korpus esa ushbu birliklarning tarjimada qanday saqlanishi, qanday o'zgarishi yoki yo'qolishini miqdoriy va sifat jihatdan kuzatish imkonini beradi. Shu jihatdan ishlab chiqilgan annotatsiya modeli nafaqat mazkur roman, balki boshqa o'zbek badiiy matnlarini korpus asosida tadqiq etishda ham qo'llanishi mumkin.

References:

1. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021). Linguistic functionality of Uzbek electron corpus: Uzbekcorpus.uz. *2021 International Conference on Information Science and Communications Technologies (ICISCT)*, 1–4. <https://doi.org/10.1109/ICISCT52966.2021.9670043>
2. Alves, F. (Ed.). (2003). *Triangulating translation: Perspectives in process-oriented research* (Vol. 45). John Benjamins Publishing Company. <https://doi.org/10.1075/btl.45>
3. Baker, M. (1993). Corpus linguistics and translation studies—Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology*. John Benjamins Publishing Company. <https://doi.org/10.1075/z.64.15bak>
4. Hasselgård, H. (2020). Corpus-based contrastive studies: Beginnings, developments and directions. *Languages in Contrast*, 20(2), 184–208. <https://doi.org/10.1075/lic.00015.has>
5. O'Brien, S. (2012). Translation as human–computer interaction. *Translation Spaces*, 1, 101–122. <https://doi.org/10.1075/ts.1.05obr>

6. Simard, M., & Plamondon, P. (1998). Bilingual sentence alignment: Balancing robustness and accuracy. *Machine Translation*, 13(1), 59–80. <https://doi.org/10.1023/A:1008010319408>
7. Zanettin, F. (2014). *Translation-driven corpora: Corpus resources for descriptive and applied translation studies*. Routledge. <https://doi.org/10.4324/9781315759661>