

ERGASH QILICHEV MUALLIFLIK KORPUSI TAHLILI HAQIDA

Taxirova Marjona Olim qizi,
Magistrant

marjonaxontaxirova@gmail.com

Buxoro davlat universiteti

Annotatsiya. Amaliy tilshunoslikda korpus lingvistikasini ilmiy-nazariy tadqiq qilishga bo'lgan ehtiyojning ortib borishi sohada turli yo'nalishdagi til korpuslarining yaratilishiga olib keldi. Mazkur maqolada korpus lingvistikasi doirasida yaratilgan “Ergash Qilichev mualliflik korpusi” tuzilishi, matnlar leksikasi va teglash yondashuvlari aks ettirilgan.

Kalit so'zlar: *Ergash Qilichev mualliflik korpusi, mualliflik korpusi, lingvistik izohlash, lingvistik va ekstralingvistik razmetka, lemmatizatsiya, KWIC.*

Abstract. The growing need for scientific and theoretical research of corpus linguistics in applied linguistics has led to the creation of language corpora of various directions in the field. This article reflects the structure, text vocabulary, and tagging approaches of the “Ergash Kilichev Author's Corpus” created within the framework of corpus linguistics.

Keywords: *Ergash Kilichev's author's corpus, author's corpus, linguistic interpretation, linguistic and extralinguistic markup, lemmatization, KWIC.*

Kirish. Til insoniyat madaniyatining eng murakkab va ko'p qirrali hodisasi hisoblanadi. Uni to'liq tushunish uchun nafaqat nazariy gipotezalar, balki real matnlar asosida empirik tadqiqotlar ham zarur. Mana shu ehtiyojdan korpus lingvistikasi – tabiiy til matnlarining katta yig'malarini kompyuter texnologiyalari yordamida tahlil qiluvchi fan yo'nalishi – vujudga keldi. Bugungi kunga qadar ushbu sohada bir qator til tadqiqi amalga oshirilgan korpuslar yuzaga kelmoqda. Shu jumladan, biz tahlil qilmoqchi bo'lgan “Ergash Qilichev mualliflik korpusi” ham korpus lingvistikasi asosida yaratilgan til tadqiqoti sanaladi.

Adabiyotlar sharhi. Til korpusi (“corpus” – lotincha “tana” degan ma’noni anglatadi) – elektron holda saqlanadigan ma’lum til birliklari majmuasi bo‘lib, tilshunoslar uchun turli muammolarni hal etish manbasi, foydalanuvchilar uchun esa lingvodidaktik va ta’limiy ahamiyatga ega tizim sanaladi. XX asrning 90-yillaridan boshlab korpus lingvistikasiga oid ilmiy tushunchalar mukammalroq shakllandi, turli xil manbalarda bu yangi yo‘nalish haqida bir-biriga yaqin ta’riflar berila boshlandi. Masalan, J.Sinkler korpusni tilning xilma-xilligini ko‘rsatish, turli ko‘rinishdagi til materiallarini tavsiflash maqsadida tanlangan asl matnlar yig‘indisi deb ta’riflaydi[1:171]. Unga ko‘ra 3 ta muhim jihat qamrab olinadi:

- 
1. Elektronik shakl;
 2. Tanlash mezonlari;
 3. Manba.

- elektronik shakl – ma’lumotlar kompyuter tomonidan qayta ishlanadi;
- tanlash mezonlari –korpus maqsadli tarzda shakllantiriladi;
- manba – haqiqiy til qo‘llanishi natijasida yuzaga kelgan matnlar. Bu tamoyillar mualliflik korpuslari uchun ham to‘liq qo‘llaniladi.

Mualliflik korpusi muayyan bir muallif tili, yozish uslubi, asarlaridagi til, matn xususiyatlari haqida ma’lumot beruvchi tizim sanaladi. Jahon tilshunosligida hozirgi kunga qadar mualliflik korpuslarini yaratish borasidagi tadqiqotlar amalga oshirilmoqda. Xususan, 1956-1961-yillarda Pushkin tilining lug‘ati, shuningdek, alohida asarlarning lug‘atlari, masalan, 1978-yilda Lev Tolstoyning “Urush va tinchlik”[2] romanining chastotali lug‘ati nashr etildi. Rus tilining mualliflik korpuslari jahon, shuningdek, rus tilshunosligida shiddat bilan rivojlanib borayotgan zamonaviy soha hisoblanib, uning taraqqiyoti natijasida juda ko‘plab ijodkorlarning elektron lug‘at va korpuslari yaratildi. Xususan, A.P.Chexov, A.S.Pushkin,

F.M.Dostoyevskiylarning korpuslari, M.V.Lomonosov nomidagi Moskva davlat universiteti olimlari tomonidan I.Krilov masallari tili lug'ati, A.Griboyedov komediyalari tili lug'ati, N.Gogolning “Revizor” komediyasi tili lug'ati, D.Fonvizin komediyalari tili lug'atlari yaratilgan bo'lib, yuqoridagi ishlarga “A.S.Pushkin tili lug'atining[3] tajribasi asos bo'lgan.

Tadqiqot metodologiyasi. Ushbu tadqiqot corpus-based (korpus asosidagi) metodologiya doirasida olib borilgan bo'lib, miqdoriy va sifat tahlil usullarini o'zida birlashtiradi. Tadqiqotda quyidagi asosiy metodologik yondashuvlar qo'llanildi:

Yondashuv	Qo'llanilishi
Korpus asosidagi tahlil	Matnni avtomatik va qo'lda teglash
Stilistik tahlil	Muallif idiositilini leksik va grammatik ko'rsatkichlar orqali aniqlash
Diskursiv tahlil	Ilmiy nutq strategiyalari va janr xususiyatlarini baholash
Qiyosiy tahlil	Turli davrlardagi maqolalar leksikasini taqqoslash
KWIC-konkordans	Kalit so'zlarning kontekstual qo'llanilishi va kollokatsion muhit

Tahlil va natijalar. Konkordans va chastota tahlili orqali muallif diskursida markaziy konseptlar aniqlanib, ularning o'zaro semantik bog'lanishi modellashtirildi. “Uslub – ma'no – nutq – ifoda” konseptual zanjiri muallif ilmiy tafakkurining asosiy paradigmasini tashkil etishi kuzatildi. uzbektilshunos.uz platformasida joylashgan barcha matnlar tadqiqot obyekti sanaladi va bular quyidagi toifalarda tashkil etilgan:

Toifa	Miqdor	Umumiy hajm
Ilmiy maqolalar	30 ta	32.818 so'z
Monografiyalar va kitoblar	6 ta	Taxminan 1.200 bet

Xotiralar va shogirdlar maqolalari	5 tadan ortiq.	Qo'shimcha ma'lumot qatlami
Video materiallar	1 ta	45 daqiqa 15 soniya
Bibliografik metadata	Har bir birlik	Muallif, yil, manba, janr

Ilmiy maqolalar yetti tematik yo'nalishda tasniflangan: uslubshunoslik va nutq madaniyati (5 ta), matn lingvistikasi (3 ta), Alisher Navoiy ijodi (4 ta), fonetika va fonologiya (2 ta), Sadriddin Ayniy maqolalari (10 ta), leksikologiya (6 ta). Bu tasniflash metama'lumot teglashning birinchi qatlamini tashkil etadi.

Korpusni tahlil qilishda quyidagi texnologik va lingvistik vositalardan foydalanildi:

1. KWIC-konkordans moduli – platformaning o'z ichki qidiruv tizimi;
2. Chastotaviy tahlili – muallif maqolalaridagi so'z chastotasini hisoblash;
3. POS-teglash – morfologik kategoriyalarni belgilash uchun qo'lda va yarim avtomatik tahlil;
4. Kollokatsion tahlil – so'zlarning bir-biri bilan birikuvchanligini o'rganish;
5. Diaxron tahlil – 1969-2010-yillar orasidagi leksik evrilishni aniqlash.

Korpus oltita mustaqil moduldan tashkil topgan bo'lib, ularning har biri alohida funksional vazifani bajaradi. Tadqiqot nuqtayi nazaridan eng muhimi – qidiruv moduli: u KWIC asosidagi konkordans qidiruv va morfologik filtrlarga tayanuvchi kengaytirilgan qidiruvni o'z ichiga oladi. “Ergash Qilichev mualliflik korpusi”ning interfeysini ushbu me'yor bo'yicha baholasak, platforma foydalanuvchi uchun tushunarli navigatsiyaga ega ekanligi va qidiruv moduli ishlab chiqilganligi – bu g'oyaga intilishni ko'rsatadi.



1-rasm. “Ergash Qilichev mualliflik korpusi” interfeysi

Mualliflik korpusining KWIC qidiruv moduli yordamida “*arxaizm*” so‘zining kontekstual qo‘llanilishi tahlil qilindi. Bu so‘z Ergash Qilichevning 1969-yilgi dissertatsiya mavzusidan boshlab, keyingi asarlarigacha doimiy ravishda qaytib keladi.

Chap kontekst	Kalit so‘z	O‘ng kontekst
Sadriddin Ayniy asarlarida	ARXAIZM	va istorizmlar alohida o‘rinni egallaydi
grammatik	ARXAIZMLAR	tilning eskirgan qatlamini ifodalaydi
leksik	ARXAIZMLAR	badiiy ta’sirchanlikni kuchaytirish vositasi
eskirgan so‘zlarning	ARXAIZM	turlari va ularni tasniflash masalasi
stilistik vazifadagi	ARXAIZMLAR	Navoiy asarlarida keng qollanilgan

KWIC jadvaldan ko‘rinib turibdiki, Ergash Qilichev “*arxaizm*” tushunchasini doimo “leksik”, “grammatik”, “stilistik” sifatlari bilan birga qo‘llaydi. Bu uning ilmiy metodologiyasidagi izchillikni va “*arxaizm*” atrofidagi kontseptual maydonni ko‘rsatadi.

Xulosa. Tahlillar shuni ko'rsatadiki, korpus infratuzilmasi mustahkam asosga ega, biroq lingvistik teglash qatlamlarini boyitish – bu platformani haqiqiy ilmiy-tadqiqot vositasiga aylantirish uchun zarur keyingi qadam. “Ergash Qilichev mualliflik korpusi” jahon mualliflik korpuslarining zamonaviy modellariga mos keluvchi platforma sifatida o'zbek tilshunosligida yangi sahifa ochdi. Modulli arxitektura, KWIC-qidiruv va to'liq matnli manbalar ko'rinishidagi infratuzilmasi bu platformaning mustahkam poydevorini tashkil etadi.

Mualliflik korpuslarini yaratish – bu zamonaviy tilshunoslikning buyuk strategik vazifalaridan biri. Har bir tilning olim va shoirlari, ularning to'la yozma merosi raqamlashtirilib, zamonaviy NLP vositalari bilan boyitilib, kelajak tadqiqotchilariga meros sifatida qoldirilishi lozim. “Ergash Qilichev mualliflik korpusi” bu ulug'vor yo'lda o'zbek tilshunosligi uchun muhim qadam bo'la oladi.

Foydalanilgan adabiyotlar ro'yxati

1. Sinclair J. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991.
2. Национальный корпус русского языка (Rus tilining milliy korpusi). Электрон ресурс: <http://www.ruscorpora.ru/search>
3. “Словарь языка Пушкина” в четырех томах, создан в Институте русского языка АН СССР под ред. акад. В.В.Виноградова.
4. Jurafsky D., Martin J.H. Speech and Language Processing (3rd ed. draft). Stanford University, 2023.
5. Qilichev E. O'zbek tilining praktik stilistikasi. – Toshkent: O'qituvchi, 1985.
6. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Toshkent, 2020.
7. www.uzbektilshunos.uz