ZAMONAVIY TILSHUNOSLIKDA TIL KORPUSLARINING O‘RNI VA NAZARIY-AMALIY AHAMIYATI

Munisa Xudoyberganova
TDTU assistent o‘qituvchisi
munisaxudayberganova13@gmail.com

Annotatsiya: *Ushbu maqolada zamonaviy tilshunoslik va axborot texnologiyalarining tutashgan nuqtasi bo‘lgan korpus lingvistikasining amaliy hamda nazariy ahamiyati atroflicha yoritilgan. Tadqiqotda zamonaviy leksikografiyada lug‘atlar yaratish, so‘zlarning real chastotasi va kontekstdagi semantik dinamikasini aniqlashda korpuslarning hal qiluvchi o‘rni tahlil qilinadi. Shuningdek, korpus texnologiyalari orqali tilning grammatik qoidalaridan chetga chiqish holatlari, nutqiy variantdorlik hamda yangi sintaktik konstruksiyalarning paydo bo‘lish tendensiyalari ko‘rsatib berilgan. Maqolada til korpuslarining tabiiy tilni qayta ishlash (NLP), katta til modellari (LLM), neyron mashina tarjimasi hamda lingvodidaktika kabi amaliy sohalardagi roli tizimlashtirilgan holda dalillangan.*

Kalit so‘zlar: *korpus lingvistikasi, leksikografiya, chastotali lug‘at, kollokatsiya, nutqiy variantdorlik, tabiiy tilni qayta ishlash (NLP), katta til modellari (LLM), lingvodidaktika.*

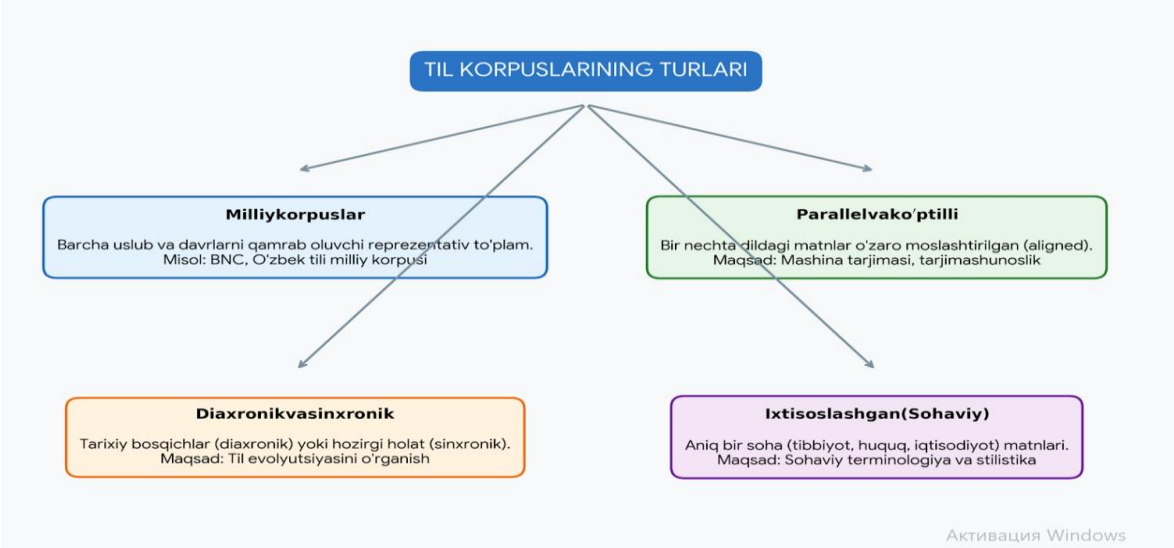
Axborot texnologiyalarining jadal rivojlanishi va raqamli transformatsiyalar davrida tilshunoslik fani ham sifat jihatidan yangi bosqichga ko‘tarildi. Xususan, Amaliy va kompyuter lingvistikasining eng jadal rivojlanayotgan sohalaridan biri bo‘lgan Korpus lingvistikasi zamonaviy ilmiy tadqiqotlarning metodologik asosiga aylanib bormoqda.

Til korpusi — bu ma'lum bir tilga oid matnlarning muayyan tizim va qoidalar asosida jamlangan, mashina o‘qiy oladigan (machine-readable),

strukturalangan va annotatsiyalangan raqamli to‘plamidir. An'anaviy tilshunoslikda til birliklari ko‘pincha intuitiv va kichik hajmdagi empirik materiallar asosida tahlil qilingan bo‘lsa, korpus lingvistikasi ulkan hajmdagi (milionlab va hatto milliardlab so‘z qo‘llanishlaridan iborat) real matn massivlari ustida aniq statistik hamda sifat tahlillarini o‘tkazish imkonini beradi. Ushbu tezisda til korpuslarining zamonaviy tilshunoslikdagi o‘rni, ularning strukturaviy xususiyatlari, lisoniy tadqiqotlar va amaliy sohalardagi ahamiyati atroflicha yoritiladi.

1. ***Til korpusining turlari va tuzilish tamoyillari.*** Til korpuslari o‘zining maqsadi, matnlarning xarakteri va qamrov doirasiga ko‘ra bir necha asosiy turlarga bo‘linadi:

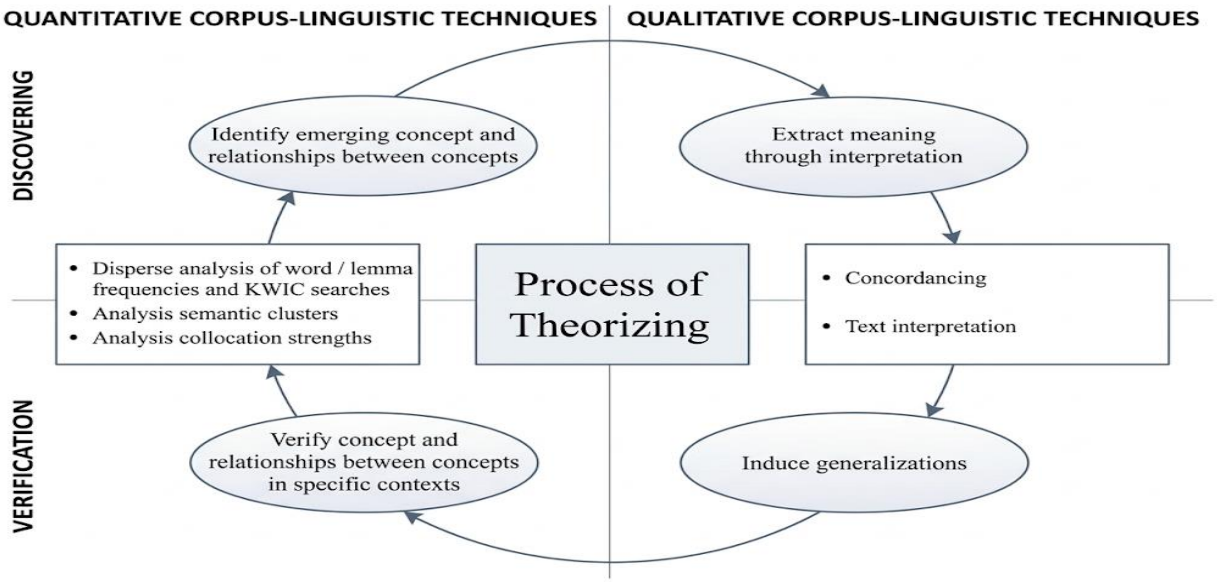
a. *Milliy korpuslar:* Muayyan bir tilning barcha uslublari (badiiy, ilmiy, rasmiy, publisistik, so‘zlashuv) va davrlarini qamrab oluvchi reprezentativ to‘plam (masalan, *British National Corpus*, *Rassiya Milliy korpusi*, *O‘zbek tili milliy korpusi*).
b. *Parallel va ko‘p tilli korpuslar:* Bir tildagi matnlar va ularning boshqa tillardagi tarjimalari o‘zaro moslashtirilgan (aligned) holda saqlanadigan korpuslar. Bu turdagi korpuslar mashina tarjimasi va tarjimashunoslik uchun beqiyos manba hisoblanadi.



c. *Diaxronik (tarixiy) va sinxronik korpuslar:* Tilning ma'lum bir tarixiy bosqichini yoki hozirgi holatini aks ettiruvchi matnlar to‘plami.

d. *Ixtisoslashgan (sohaviy) korpuslar:* Aniq bir soha (tibbiyot, huquq, aviatsiya, iqtisodiyot) yoki muayyan uslubga oid matnlardan tashkil topgan korpuslar.

Korpusning eng muhim xususiyatlaridan biri uning annotatsiyalanganligi (belgilanganligi)dir. Annotatsiya matndagi har bir so‘z va birikmaga morphologik (POS-tagging), sintaktik, semantik va pragmatik ma'lumotlarni biriktirish jarayonidir.

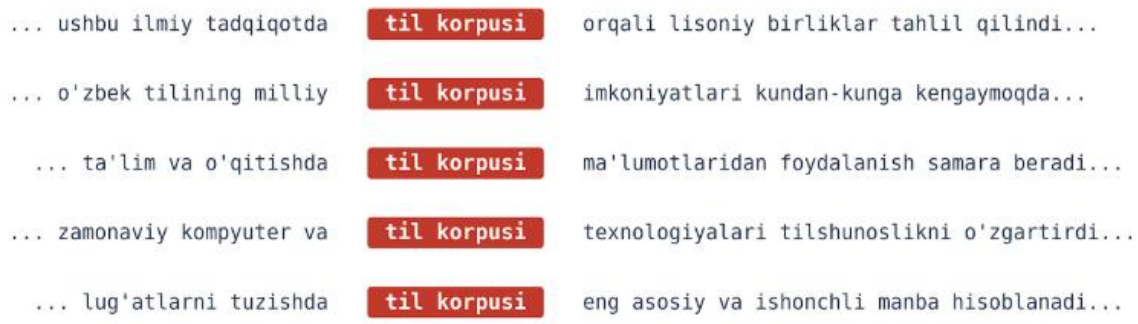


1-***rasm. Korpus lingvistikasida sifat va miqdoriy tahlil doirasi.***

2. ***Til korpuslarining ilmiy va nazariy ahamiyati.*** Til korpuslari tilshunoslikning turli sohalarida quyidagi nazariy va metodologik muammolarni hal etishda asosiy vosita vazifasini o‘taydi:

- *Lisoniy dalillarning obyektivligi va empirik asos.* An'anaviy tilshunoslikdagi "subyektiv intuitsiya" o‘rnini korpuslarda empirik va miqdoriy (konkordans, chastota, kollokatsiya) ko‘rsatkichlar egallaydi. Tadqiqotchi u yoki bu lisoniy hodisaning tilda qanchalik faol qo‘llanishini

statistik ma'lumotlar orqali aniq isbotlab beradi. Masalan: an'anaviy tilshunoslikda *"Chiroyli"* va *"Go'zal"* синонимларining qo‘llanilish farqini aniqlash tadqiqotchining shaxsiy sezgisi (intuitsiyasi)ga tayanib tushuntirilgan bo‘lsa, Korpus lingvistikasida bu holat empirik va statistik usulda isbotlanadi: ***Chastota (Frequency):*** Millionlab so‘zli matnlar bazasida *"chiroyli"* so‘zi qaysi uslubda nechta marta (masalan, 12,450 marta), *"go'zal"* so‘zi esa qaysi uslubda (masalan, 5,320 marta) qo‘llangani aniq raqamlarda ko‘rsatiladi. ***Kollokatsiya (Birikuvchanlik):*** Korpus birikmalar tahlili orqali *"go'zal"* so‘zi ko‘proq vaqt, qiz, kelajak (*go'zal kelajak, go'zal qiz*) so‘zlari bilan, *"chiroyli"* so‘zi esa buyum va kiyim-kechaklar (*chiroyli kiyim, chiroyli uy*) bilan ko‘proq birikib kelishini miqdoriy va foizli ko‘rsatkichlar orqali isbotlab beradi.



2-rasm. Konkordans (KWIC - Key Word in Context) amaliy misoli.

- ***Leksikografiya va lug‘atchilik.*** Zamonaviy lug‘atlarni yaratish jarayonini korpussiz tasavvur etib bo‘lmaydi. Korpus yordamida: So‘zlarning haqiqiy chastotasi (tezauruslar uchun) aniqlanadi. So‘z ma'nolarining kontekstdagi haqiqiy dinamikasi va semantik o‘zgarishlari kuzatiladi. So‘z birikuvchanligi (kollokatsiyalar va frazeologizmlar) tabiiy matnlar misolida shakllantiriladi.
- ***Grammatika va sintaksis tahlili.*** Korpuslar tilning grammatik qoidalaridan chetga chiqish holatlarini, nutqiy variantdorlikni va yangi grammatik

konstruksiyalarning paydo bo‘lish tendensiyalarini aniqlashga yordam beradi.

- ***Amaliy sohalar va raqamli texnologiyalardagi o‘rni.*** Til korpuslarining amaliy ahamiyati faqat tilshunoslik bilan cheklanib qolmay, axborot texnologiyalari va ta'lim sohalariga ham keng kirib borgan hamda bu quyidagi jadvalda ifodaladik:

Soha / Yo'nalish	**Texnologiya va yondashuv**	**Korpusning hal qiluvchi o'rni va natijasi**
Sun'iy intellekt va NLP	**Katta til modellari (LLM)** *(ChatGPT, Gemini va b.)*	Gigant ko'lamli korpuslar orqali modellarga tilning grammatikasi, konteksti va semantikasi o'rgatiladi.
Mashina tarjimasi	**Neyron tarjima tizimlari (NMT)** *(Google Translate, Yandex va b.)*	Sifatli parallel korpuslar (ikki/ko'p tilli) bazasida aniq va tabiatga yaqin tarjima modellari shakllantiriladi.
Nutq texnologiyalari	**AIT va sintez** *(Ovozli yordamchilar, Transkripsiya)*	Akustik va matnli korpuslar yordamida nutqni matnga o'girish hamda matndan tabiiy ovoz yaratish amalga oshiriladi.

Lingvodidaktika	**Korpusga asoslangan ta'lim (DDL)**	Real kontekstli misollar, chastotali so'zliklar va o'quv lug'atlari orqali tilni jonli va samarali o'rgatish imkonini beradi.
Mualliflik va ekspertiza	**Glottometriya va stilometriya**	Matnlar uslubini statistik tahlil qilish orqali mualliflikni aniqlash va sud-lingvistik ekspertizasida yuqori aniqlik beradi.

Xulosa qilib aytganda, til korpuslari zamonaviy tilshunoslikning intellektual va texnologik infratuzilmasidir. Ular tilni faqat nazariy qoidalar to‘plami sifatida emas, balki jonli, dinamik va rivojlanuvchi tizim sifatida o‘rganishga keng imkoniyat yaratadi. Ayniqsa, davlat tilining raqamli makondagi mavqeini mustahkamlash, milliy korpuslarni rivojlantirish va ularni avtomatik qayta ishlash darajasiga olib chiqish bugungi kunning eng dolzarb vazifalaridan biridir. Sifatli annotatsiyalangan til korpuslarining mavjudligi milliy tilning axborot jamiyati va sun'iy intellekt asrida to‘liq va erkin yashashining bosh omilidir.

FOYDALANILGAN ADABIYOTLAR

1. Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

2. McEnery, T., & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.

3. Abdurohmanov, E., & Abdujalilova, SH. (2021). *Kompyuter lingvistikasi va milliy korpus masalalari*. Toshkent.

4. Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.