



## KORPUS HAQIDA UMUMIY TUSHUNCHA

Ma'ripov Jalolkhan Kamoliddin ugli

Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti Jizzax filiali  
Psixologiya fakulteti Xorijiy tillar kafedrasasi assistenti.  
<https://doi.org/10.5281/zenodo.7936666>

## ARTICLE INFO

Qabul qilindi: 10-May 2023 yil  
Ma'qullandi: 13-May 2023 yil  
Nashr qilindi: 15-May 2023 yil

## KEY WORDS

*korpus, matn, janr, maxsus, monolingual, bilingual, multilingual.*

## ABSTRACT

*Korpus lingvistikasi ko'p o'lchovli sohadir. Bu tilning o'zaro ta'siri, aloqa va tushunishning barcha sohalarida tildan foydalanishning barcha xilma-xilligini qamrab oladigan keng spektrga ega sohadir. Tilni o'rganish va qo'llashda korpusning kiritilishi tilshunoslikka yangi yo'nalishni kiritdi.*

“Korpus” atamasi Lotincha “corpus” so'zidan olingan bo'lib, “tana, vujud” degan ma'noni anglatadi. Hozirgi kunda bu tushuncha lingvistik tahlil uchun olingan ma'lum bir til, dialekt yoki tilning boshqa kichik matnlar to'plamlarini ifodalaydi. Agar bu atamani yanada kengroq tushunishga harakat qilsak, korpus:

1. Har qanday matnni o'zida ifoda etishi mumkin;
2. Kompyuter matnining asosi bo'lishi mumkin;
3. Til yoki xilma-xillikni ifodalash uchun namuna sifatida olingan kompyuter matnining cheklangan to'plami bo'lishi mumkin (McEnery va Vilson 1996:218)

Korpus o'zaro lingvistik munosabatga ega bo'lgan turli xil tillarni qamrab oluvchi matnlarning katta namunalarini o'z ichiga qamrab olishi mumkin. Nazariy jihatdan qaraydigan bo'lsak “corpus” atamasi quyidagicha ta'riflanishi mumkin: C (capable) O (of) R (representing) P (potentially) U (unlimited) S (selections) of texts, ya'ni, “matnning cheksiz tanlovini ifodalashga qodir potensial kuch”. Korpus kompyuterda foydalanish uchun, tadqiqotlarda va amaliyotda qo'llash uchun, manba sifatida keltirilayotgan til vakili sifatida, inson tomonidan qayta ishlashga va cheksiz ma'lumotlarni shakllantirish va taqdim etish uchun qulay hisoblanadi (Dash 2005:35).

**Korpusning asosiy xususiyatlari:**

- 1. Miqdor va hajm:** Korpus og'zaki va yozma ma'lumotlarni o'z ichiga olgan katta hajmli bo'lishi kerak. Uning hajmi tarkibiy qismlari, ya'ni, komponentlarininig yig'indisidir.
- 2. Sifat:** Barcha matnlar yozma va og'zaki nutqning haqiqiy namunalaridan olinishi kerak. Bu jarayonda tilshunos muhim vazifani bajaradi. U keltirilgan ma'lumotlar oddiy muloqotdan to'planganligini tekshirishi kerak.
- 3. Til vakili sifatida:** Korpus turli xil matnlardan ko'p miqdordagi namunalarini o'zida jamlagan bo'lishi kerak. Tildagi xilma-xillikni ifodalash uchun tilning barcha sohalarida teng miqdorda bo'lishi kerak.
- 4. Oddiylik.** U asosan soddagina ko'rinishdagi oddiy matnlarni o'z ichiga olishi lozim. Bu shuni

anglatadiki, foydalanuvchi hech qanday lingvistik tushuntirishlarsiz izchil va uzluksiz ma'lumot olishini xohlaydi va korpus aynan mana shu ehtiyojni qondirishi zarur.

**5. Tenglik.** Korpusda ishlatiladigan matnlar hajm va miqdor jihatdan teng bo'lishi lozim. Biroq, bu masala barcha tilshunoslar tomonidan ham bir xil qabul qilinmaydi va birdek amal qilinmaydi.

**6. Qayta ishlash imkoniyati.** Zarurat tug'ilgan holda foydalanuvchilar sanalar, ma'lumotlar, namunalar va havolalarni hech qanday qiyinchiliklarsiz yuklab olishiga imkoniyat yaratilishi kerak. Zamonaviy texnologiyalar bizga kompyuterda korpus yaratish va uni saqlash imkonini beradi, shunda biz istalgan vaqtda kerakli ma'lumotni osongina olishimiz mumkin.

**7. Ishonchlilik.** Yaratilgan korpus har qanday tekshiruvlardan o'ta olishi va talabga javob bera olishi kerak. Biz korpusdan hech qanday ikkilanishlarsiz uning ishonchli ekanligini tekshirish uchun foydalanishimiz mumkin.

**8. Muntazam rivojlanish.** Mavjud korpus doimiy ravishda yangilanishi va to'ldirib borilishi lozim. Bu vaqt o'tishi bilan tilda sodir bo'ladigan lingvistik o'zgarishlardan va yangiliklardan boxabar bo'lish imkonini beradi.

**9. Hujjatlashtirish.** Korpusni tashkil qilgan tarkibiy qismlar haqidagi tegishli hujjatlar matndan alohida saqlanishi lozim. Bunday yondashuv bizga korpusdan foydalanish jarayonida yengillik beradi.

#### **Matn janriga ko'ra korpus shakllari.**

Nilardi Sekhar Dashning "Korpus lingvistikasi" maqolasida keltirilishicha korpus matnining quyidagicha janrlari mavjud:

**Yozma korpus:** Yozma korpus (masalan, TDIL korpusi) o'z janriga ko'ra faqat turli yozma, bosma, nashr etilgan va electron manbalardan to'plangan til ma'lumotlarini o'z ichiga oladi;

**Nutq korpusi:** Nutq korpusi (masalan, Yangi Zelandiyada ingliz tilidagi Vellington korpusi) barcha rasmiy va norasmiy munozaralar, tahlillar, oddiy suhbatlar, dialoglar, monologlar, turli xil suhbatlar, onlayn diktantlarni va ommaga qilingan murojaatlarni o'z ichiga oladi. Bunday matnlarda ommaviy axborot vositalarining aralashuvi kuzatilmaydi;

**Og'zaki korpus:** Og'zaki korpus (masalan, London-Lund ingliz og'zaki korpusi) nutq korpusining texnik kengayishi, og'zaki nutq matnlarini o'z ichiga oladi. Bunday korpusda nutq transkripsiyadan tashqari hech qanday o'zgarishsiz yozma shaklda ifodalanadi. U fonetik transkripsiya shakli yordamida ifodalanadi.

#### **Ma'lumotlar sohasiga ko'ra korpusning turlari.**

**Umumiy korpus:** Umumiy korpus (masalan, Britaniya Milliy Korpusi) turli fanlar, janrlar va mavzu sohaslariga tegishli umumiy matnlarni o'z ichiga oladi. Shakl va foydalilik jihatidan olib qaralsa, u matnlar to'plamining soni bo'yicha cheklangan. Bu shuni anglatadiki, matn turlari soni va undagi so'z va jumlar soni cheklangan. Biroq vaqt o'tishi bilan u o'sib borishi va yangi matnlar mavjudligi bilan yangi ma'lumotlarni qo'shish imkoniyatiga ega. U hajmi jihatdan juda katta, xilma-xillikka boy va foydalanish doirasi keng.

**Maxsus korpus:** Maxsus korpus (masalan, CHILDES ma'lumotlar bazasi) o'rganilayotgan mavzuning ma'lum xususiyatlariga e'tibor qaratib, til, dialekt va mavzuning muayyan xilma-xilligi uchun umumiy korpusda namuna olingan matnlardan tuzilgan korpusdir. Maxsus korpus qo'llanilish maqsadiga ko'ra hajm va tarkib jihatdan farqlanishi mumkin. Bola tili korpusi, ona tili bo'lmaganlar korpusi, turli dialektda suhbatlashuvchilar

korpusi va maxsus aloqa (suhbat) sohalari (masalan, auksion, tibbiy suhbatlar, qimor o'yinlari, sud muhokamasi va boshqalar) maxsus korpusga misol bo'la oladi. Maxsus korpusni o'zi tegishli bo'lgan sohadan tashqari qo'llash deyarli imkonsiz ya'ni, keng ko'lamda foydalanib bo'lmaydi. Maxsus korpusning asosiy afzalligi shundaki, matnlar shunday tanlanadiki, siz izlayotgan ma'lumot umumiy korpusga qaraganda tez-tez uchraydi. Aynan shu xislati bilan bir xil turdagi ma'lumotlarni ta'minlovchi maxsus korpus umumiy korpusdan kichikroqdir (Sinclair, 1996b).

**Tilosti korpusi:** Bu korpus ma'lum bir tilning faqat bitta matn turidan iborat bo'ladi. Uning tuzilishi va lug'at boyligining bir xilligi odatda ma'lumotlar miqdori kichikroq bo'lishiga imkon beradi.

**Namuna korpusi:** Namuna korpusi (masalan, Ingliz gazetalarining Syurix korpusi) maxsus korpus toifalaridan biri bo'lib, u juda ehtiyotkorlik bilan tanlangan va batafsil o'rganilgan cheklangan matnlar to'plamini o'z ichiga olgan namunalardan iborat bo'ladi. Namuna korpusi ishlab chiqilgandan so'ng, unga hech qanday o'zgartirish kiritilmaydi (Sinclair 1991:24), chunki har qanday kichik o'zgarish ham namunalarga bo'lgan tadqiqot talablarini buzadi. Namunalar matnlarga nisbatan oz sonli va doimiy hajmga egaligi uchun ular matn sifatida tasniflanmaydi.

**Adabiy korpus:** Namuna korpusining alohida toifasi adabiy korpus bo'lib, uning bir qancha turlari mavjud. Bunday korpusni yaratish uchun tasniflash mezonlari orasida muallif janr (masalan, she'rlar, qissalar, fantastika va boshqalar), davr (nasalan, 15-asr, 18-asr va boshqalar), guruh (masalan, romantik shoirlar, Viktoriya romanchilari va boshqalar), mavzu (masalan, inqilobiy yozuvlar, oilaviy hikoyalar, sanoatlashtirish va boshqalar) va shu kabi boshqa mezonlarga asoslanadi.

**Monitor korpusi:** Monitor korpusi (masalan, Ingliz Bank Korpusi) bu tildagi o'zgarishlarni aks ettiruvchi ma'lumotlarni doimiy ravishda ko'paytirish imkoniyatiga ega bo'lgan o'sib borayotgan, cheksiz bo'lmagan matnlar to'plamidir. Tilning o'zgarishi korpusning doimiy yangilanib borishida o'z aksini topadi. Monitor korpusining asosi faqatgina bir yilda aytilgan yoki yozilgan matnlarga asoslanadi (Sinclair 1991:21). Monitor korpusi orqali biz yangi so'zlarni topishimiz, qo'llanishdagi o'zgarishlarni ko'rishimiz, ma'no o'zgarishini kuzatishimiz va keng doiradagi leksik ma'lumotlarni olishimiz mumkin. Yangi ma'lumotlar paydo bo'lishi bilan kam miqdorda berilgan ma'lumotlar ko'payishi mumkin va bu monitor korpusining tarkibiy qismlari balansi o'zgarishiga sabab bo'lishi mumkin.

#### **Matn turiga ko'ra korpus shakllari:**

**Monolingual (bir tilli) korpus:** Monolingual korpus (masalan, ISI Bengal Korpusi) ma'lum bir tilning bir davrda yoki bir necha davrlarda ifodalangan matnlarini o'z ichiga oladi. Bu korpus yozma va og'zaki matn namunalardan iborat bo'ladi va ular bir-biriga mutanosib tarzda rivojlanadi.

**Bilingual (ikki tilli) korpus:** Ikki tilli korpus (masalan, TDIL Bengal-Oriya Korpusi) ikkita o'zaro bog'liq yoki bog'liq bo'lmagan tillar bir qolipga solinganda paydo bo'ladi. Agar bu tillar genetic va tipologik jihatdan bog'liq bo'lsa, ular parallel korpusga aylanadi va oldindan belgilangan ayrim parametrlar bo'yicha tartiblanadi. Hajmi, mazmuni va qamrab olgan maydoni jihatdan bir korpus ikkinchisidan farq qilishi mumkin, lekin bunga parallel korpusda ruxsat berilmaydi.

**Multilingual (ko'p tilli) korpus:** Ko'p tilli korpus (masalan, Krater Korpusi) ikkitadan

ortiq tillardagi matnlar to'plamini o'z ichiga oladi. Odatda, bu korpus turida ham, ikki tilli korpusda ham matnlar turli tillarga tegishli bo'lsa-da, o'xshash matn toifalari va bir xil namuna olish tartib-qoidalariga amal qilinadi.

### **Qurilish uslubi (dizayni) ga ko'ra korpus turlari.**

**Izohsiz (Un-annotated) korpus:** Izohsiz korpus (masalan, TDIL Korpusi) bu – qo'shimcha lingvistik va non-lingvistik ma'lumotlarsiz oddiy matnlarning oddiy holatini ifodalaydi. Bu turdagi korpus til o'rganishda keng qo'llaniladi, biroq korpusning foydaliligi uning izohga ega ekanligi bilan yanda oshadi.

**Izohli (Annotated) korpus:** Izohli korpus (masalan, Britaniya Milliy Korpusi) matnlarga qo'shimcha ma'lumotlarni (analitik belgilar, nutq qismlari belgilari, grammatik kategoriya ma'lumotlari va boshqalar) yozib olish uchun dizaynerlar tomonidan tashqaridan kiritilgan teglar va kodlarni o'z ichiga oladi. Izohsiz korpusdan farqli o'laroq, izohli korpus til texnologiyasining turli vazifalarida foydali bo'lgan tegishli ma'lumotlarni taqdim etish uchun ko'proq mos keladi, shu jumladan morfologik ishlov berish, jumalarni tahlil qilish, ma'lumotlarni qidirish, so'z ma'nosini aniqlash, mashina tarjimasiga imkon yaratadi.

### **Foydalanilgan adabiyotlar ro'yxati:**

1. Tolibovna A. K. et al. Functions of Allusion and Allusion as a Marker of Intertextuality and Precedence //European Multidisciplinary Journal of Modern Science. – 2022. – T. 6. – C. 485-487.
2. Ma'ripov J. K. A BRIEF INFORMATION ABOUT TENSES //O'ZBEKISTON RESPUBLIKASI OLIY VA O'RTA. – C. 464.
3. Tolibovna A. K. et al. Features Of Anthropocentric Study Of Sacred Texts //JournalNX. – T. 8. – №. 1. – C. 5-10.
4. Ma'ripov J. ANTROPOSENTRIZM–TILSHUNOSLIKNING ZAMONAVIY YONALISHI SIFATIDA //Zamonaviy dunyoda innovatsion tadqiqotlar: Nazariya va amaliyot. – 2022. – T. 1. – №. 28. – C. 62-68.
5. Nasiba, P. (2022). THE TERM "STREAM OF CONSCIOUSNESS" IN ENGLISH LITERATURE. PEDAGOGICAL SCIENCES AND TEACHING METHODS, 13(2), 221-223.
6. Fayzullaeva, U., & Parmonova, N. (2021). THE EVOLUTION OF FEMALE WRITERS IN THE 20TH CENTURY: FROM THE LATE 19TH CENTURY TOWARDS 21ST CENTURY. Mental Enlightenment Scientific-Methodological Journal, 2021(02), 75-83.
7. Parmonova Nasiba Amriddin qizi. (2022). HOZIRGI ZAMON TALABALARINI RAQAMLI TEXNOLOGIYALAR YORDAMIDA O'QITISH SAMARADORLIGI. International Journal of Contemporary Scientific and Technical Research, 1(2), 664–667. Retrieved from <http://journal.jbnuu.uz/index.php/ijcstr/article/view/302>
8. Abduraxmanova, Z., & Mamurova, M. (2021). THEORETICAL APPROACH TO SPEECH DISFLUENCIES IN SIMULTANEOUS INTERPRETATION. In МОЛОДОЙ ИССЛЕДОВАТЕЛЬ: ВЫЗОВЫ И ПЕРСПЕКТИВЫ (pp. 43-45).
9. Toshtemirovna, B. X. (2022). CHARACTERISTICS OF THE USE OF WORDS RELATED TO SOME ADJECTIVES IN THE WORK " KUTADGU BILIG" BY YUSUF KHOS HAJIB. Web of Scientist: International Scientific Research Journal, 3(11), 976-981.