

## KORPUS LINGVISTIKASI VA MILLIY KORPUS HAQIDA

Abduvohidov Behruz

Termiz davlat universiteti

O'zbek filologiyasi fakulteti 3-bosqich 421-guruh talabasi.

Abdualimova G.H.

Ilmiy rahbar.

<https://doi.org/10.5281/zenodo.10407029>

**Annotatsiya.** Mazkur maqolada korpus lingvistikasining shakllanishi va taraqqiyoti, asosiy tushuncha va masalalar, korpuslarni tuzish, korpuslarning lingvistik bazasini shakllantirish talablari, o'zbek tilining ta'limiy korpusi, korpus lingvistikasini rivojlanish bosqichlari, olimlarimizni korpus lingvistikasi uchun qo'shgan hissalarini, milliy korpusning xususiyatlari haqida ma'lumot berilgan.

**Kalit so'zlar:** Korpus, korpus lingvistikasi, milliy korpus, poetik matnlar korpusi, ta'limiy korpus, mualliflik korpusi.

### ON CORPUS LINGUISTICS AND THE NATIONAL CORPUS

**Abstract.** In this article, the formation and development of corpus linguistics, the main concepts and issues, the creation of corpora, the requirements for the formation of the linguistic base of corpora, the educational corpus of the Uzbek language, the stages of development of corpus linguistics, the contributions of our scientists to corpus linguistics, the characteristics of the national corpus reported.

**Key words:** corpus, corpus linguistics, national corpus, corpus of poetic texts, educational corpus, corpus of authorship.

### О КОРПУСНОЙ ЛИНГВИСТИКЕ И НАЦИОНАЛЬНОМ КОРПУСЕ

**Аннотация.** В данной статье рассматриваются становление и развитие корпусной лингвистики, основные понятия и вопросы, создание корпусов, требования к формированию языковой базы корпусов, учебного корпуса узбекского языка, этапы развития корпусная лингвистика, представлен вклад наших ученых в корпусную лингвистику, дана характеристика национальной корпусной информации.

**Ключевые слова:** корпус, корпус лингвистики, национальный корпус, корпус поэтических текстов, учебный корпус, корпус авторства.

#### Kirish.

Bugungi kunda yurtimizda o'zbek tiliga qaratilayotgan e'tibor nihoyatda katta. Bu esa tilshunoslarimiz, shuning bilan birga informatiklarimizning oldiga yana bir muhim vazifani qo'yadi-ki, u ham bo'lsa o'zbek tilining korpus lingvistikasini yaratishdir. Bugungi kunda dunyo tillari bo'lmish rus, ingliz, ispan, arab tili kabi barcha tillarda mavjud bo'lgan ushbu baza bizning tilshunosligimizda ham bor bo'lishini kutib turgan asosiy masaladir.

Korpus lingvistikasi tilshunoslik kompyuter texnologiyalaridan foydalangan holda lingvistik ma'lumotlar korpusidan foydalanish hamda ularni yaratishning umumiy prinsiplarini ishlab chiqarish bilan shug'ullanadi. Leksikaning korpus tavsifi leksikografiyada eng ko'p qo'llaniladi. Deyarli barcha zamonaviy ingliz lug'atlar korpuslar asosida yaratilgan. Korpus tilda

soʻzlar toʻplamini ishonchli tarzda aniqlashga yordam beradi, yangi soʻz shakllarining paydo boʻlishini koʻrsatadi, bir soʻzning turli maʼnolari va uning chastotasini aniqlash uchun ishlatiladi.

Korpus haqidab soʻz borar ekan, uning rivojlanishi zaminida tilshunos olimlarga berayotgan nafi juda kata. Jumladan tillarni saqlab qolish, uni keying avlodga toʻla toʻkis yetkazish, tabiiy tilning morfologik xususiyatlarini saqlab qolish va hokozolarni korpusni vazifalari qilib koʻsatsak boʻladi. Jahonni yetakchi mamlakatlarida korpus lingvistikasi ustida qilinayotgan ularning Milliy korpusini boyitmoqda.

Korpus termini qoʻllanilganda doim Milliy korpus birikmasi ishlatiladi. Milliy korpus – maʼlum bir tilning muayyan bir bosqichini ifodalaydigan barcha janrlar, uslublar, ijtimoiy hamda hududiy lahjalar va boshqalarda aks ettiruvchi elektron shakldagi matnlar toʻplami. Milliy korpus boʻlishi uchun imkon qadar tilda mavjud yozma va ogʻzaki matnlarning barcha turlarini oʻz ichiga olishi, korpusga kiritilib maxsus izohiga ega boʻlishi kerak. Albatta bunda korpus hajmi boshqa korpus turlariga nisbatan oshib ketadi. Masalan, Ingliz tilining milliy korpusi 2.4 milliarddan oshiq, rus tilining milliy korpusi 200 millionga yaqin, turk tilining milliy korpusi 50 milliondan oshiq soʻz va soʻz shakllariga ega. Bugungi kunda dunyo tillarining juda koʻpchiligi murakkab qidiruv tizimiga ega boʻlgan, ilmiy va amaliy lingvistik tadqiqotlar uchun bir –biridan farqli tahlil funksiyalariga ega boʻlgan milliy korpusiga ega.

#### **Oʻzbek tilshunosligida korpus.**

Oʻzbek tili uchun ham milliy korpus yaratish tadqiqotlari tilshunoslar tomonidan boshlab yuborilgan. Bu borada oʻzbek olimlaridan A.Poʻlatov, B.Mengliyev, N.Abdurahmonova, Sh.Hamroyeva, A.Eshmoʻminovlarning kompyuter lingvistikasi va korpus lingvistikasi sohalarida olib borayotgan ilmiy izlanishlarini eʼtirof etish mumkin. Ammo koʻpchilik ishlar natijasi nazariyalarni oʻrganish va tahlil qilish bilan cheklanib qolmoqda. Bir til doirasida milliy korpus yaratish murakkab va uzoq vaqtni oladigan jarayon ekanligi hech bir tilshunosga sir emas. Ammo kichik tadqiqotlar doirasida muayyan bir uslublar yoki janrlar boʻyicha kichik korpus namunalarini yaratish kerak. Lekin shu oʻrinda aytishkerakki, foydalanuvchi soʻzning kontekstlarda toʻgʻri qoʻllanilishini tekshirishi yoki tabiiy til nutq birikmalarini izlashi, tilda tez-tez uchraydigan til qoliplarini aniqlashi va ulardan foydalanishi uchun kattaroq hajmdagi korpusga ehtiyoj sezadi.[1;43-46]

Oʻzbekistonda korpus yaratish maqsadida qilingan ishlardan “”hamda “Taʼlimiy korpus”ni misol qilsak boladi.Ikkala korpus ham oʻz nihoyasiga yetmagan bolsada korpusda alohida alohida mavzuchalar tuzulgan. Ularni toʻldirish biz yoshlarni ishi deb oʻlayman.Shu oʻrinda Nilufar Abdurahmonovanning korpus haqidagi fikrlarini eslab oʻtmoq joiz: ”Biz tuzgan korpusda tilimizni oʻz holida saqlab qolish, uni kelgusi avlodga yetkazish va tilimizni imkon darajasida boyitish,korpusimizni soʻz hajmini kengaytirish,bizning kelajakdagi maqsadimiz.”- deya oʻz nutqini tamomlaydi.

Oʻzbek tillining Milliy korpusi ustida ish olib borayotgan professor B.Mengliyev korpusga quydagicha tarif beradi: "Korpus-til birliklarining xususiyatlarini aniqlash maqsadida qidiruv dasturiga boʻysundirilgan matnlar majmui,tabiiy tildagi elektron shaklda saqlanadigan yozma yoki ogʻzaki, kompyuterlashtirilgan qidiruv tizimiga dasturiy taʼminot asosida joylashtirilgan matnlar jamlammasi.Til korpuslari-til boʻyicha tadqiqot va amaliy topshiriqlar yechimi uchun inkor etib boʻlmas ish quroli" deb tarif beradi. Milliy korpusi nima? degan savol tuguladi.[2;4-5]

Milliy til korpusi — til birliklarining xususiyatlarini aniqlash maqsadida elektron qidiruv imkoniyati mavjud tizim, tabiiy tilning raqamlashgan yozma va og‘zaki matnlar jamlanmasi. Milliy korpus — til birli gining o‘zgarishi, eskirishi, yangilarining paydo bo‘lishi, ma’nosining kengayishi va torayishi, yangi iboralarning paydo bo‘lishini kuzatish, an’anaviy va zamonaviy lug‘atlar tuzishda keng imkoniyatli dasturlashtirilgan tizim. Korpusdagi milliy so‘zi nafaqat tilning, balki korpus tuzilishi va tarkibining ham o‘ziga xosligini anglatadi. Milliy korpusning boshqa tur til korpuslaridan farqi ta’limiy korpus, mualliflik korpusi, poetik matnlar korpusi, ilmiy va rasmiy matnlar korpusi, og‘zaki matnlar, badiiy matnlar va gazetalar hamda dialektlar korpusi kabi ichki korpuslarni o‘z tarkibiga qamrab olganligi bilan belgilanadi.

Milliy korpusning o‘ziga xos ikki muhim xususiyati mavjud. Birinchidan, u me’yorlashtirilgan muayyan tarkibga ega ekanligi bilan xarakterlanadi. Bu korpus ma’lum tilda berilgan (turli badiiy janrlar: publitsistik, o‘quv, ilmiy, ish yuritish, so‘zlashuv, ilmiy, ish yuritish, so‘zlashuv, shevaviy kabi), ularning barchasi imkon darajasida ma’lum doiraga oid ma’lumotlarning proporsional matnlari hisoblanadigan og‘zaki va yozma ko‘rinishlarining barchasini o‘z qamroviga oladi degani.

Ikkinchidan, korpus unga kiritilayotgan matnlarning o‘ziga xos jihatlari to‘g‘risida muhim ko‘shimcha ma’lumotlarni saqlaydi. Kiritma–korpusning asosiy bosh ma’lumotnomasi bo‘lib, u korpusni zamonaviy internet tarmog‘ida mavjud bo‘lgan oddiy matnlar to‘plamidan ajratib turadi.

#### **Xulosa**

Xulosa o‘rnida shuni aytmoqlik kerakki, mamlakatimizda kompyuter savodxonligini oshirish maqsadida qilinayotgan ishlar juda ko‘p. Ulardan birinchi prezidentimiz Islom Karimovning 2002-yil 30-maydagi

-Kompyuterlashtirishni yanada rivojlantirish va axborot-kommunikatsiya texnologiyalarni joriy etish to‘g‘risidagi farmoni;

O‘zbekiston Respublikasining 2003-yil 11-dekabirdagi

-Axborotlashtirish to‘g‘risidagi qonuni.

keying yillarda chiqarilayotgan qonun va farmoyishlarni bunga yorqin misol qilib ko‘rsatsak bo‘ladi.

Gap o‘rnida shuni takidlash kerakki korpus lingvistikasini o‘zbek tilshunosligiga kirib kelishi juda kata yangilik bo‘ladi deb hisoblayman.

#### **REFERENCES**

1. Abdurahmonova N. Mashina tarjimasining lingvistik asoslari (o‘quv qo‘llanma). Toshkent: Akademnashr, 2012.- B 43-46
2. Abduraxmonova N. "O‘zbek tili elektron korpusining elektron modellari" maqolasi.-B 4-5.
3. Raximov A. Kompyuter lingvistikasi asoslari,-Toshkent "Akademik"-2011-B 89-90.