

KORPUS LINGVISTIKASINING AHAMIYATI

Abdurahmanova Muqaddas Tursunaliyevna,

Mparfi2005@yandex.ru

O'zbekiston Milliy universiteti Jurnalistika fakulteti

O'zbek tilshunosligi kafedrasida dotsenti, f.f.n

Olimova Muxlisa Parmonovna

olimovamuxlisa566@gmail.com

O'zbekiston Milliy universiteti Jurnalistika fakulteti

Filologiya va tillarni o'qitish (o'zbek tili) yo'nalishi 1-kurs talabasi

Annotatsiya: Maqolada korpus lingvistikasining bugungi kundagi ahamiyati tahlilga tortilgan. Tahlil davomida mavjud korpuslarga ham murojaat etilgan.

Annotation: The article analyzes the current importance of corpus linguistics. Existing corpora also addressed during the analysis. The importance of corpus linguistics is explained using examples.

Kalit so'zlar: *korpus lingvistikasi, kompyuter tilshunosligi, baza, teg, statistik tahlil, platforma*

Bugungi hayotimizni axborot texnologiyalar va kompyuterlarsiz tasavvur etish qiyin. Kompyuter kundan kunga jamiyatimizning barcha sohalariga kirib keldi. Natijada ko'plab zamonaviy yo'nalishlar yuzaga kelishiga zamin yaratilyapti. Xususan, tilshunoslikda kompyuter texnologiyalaridan faol foydalanish kompyuter lingvistikasi shakllanishiga turtki bo'ldi. Professor B. Mengliyev ta'kidlashicha, kompyuter texnologiyalari sohasi tillarning taraqqiyoti va yashab qolishi uchun hal qiluvchi omilga aylanib bo'ldi. Tilning internet tiliga aylanishi matematik lingvistika, uning davomi bo'lgan kompyuter lingvistikasining shakllanganligi va rivojlanish darajasiga bog'liq. Kompyuter tilshunosligi tilning kompyuter nuqtayi nazaridagi talqinidir. Korpus lingvistikasi kompyuter tilshunosligi tarkibida rivojlangan va hozirda alohida yo'nalish sifatida o'rganiladigan soha. Korpus elektron holda saqlanadigan ma'lum til birliklari yig'indisi bo'lib, tilshunoslar tomonidan turli vaziyatlarga munosib tarzda

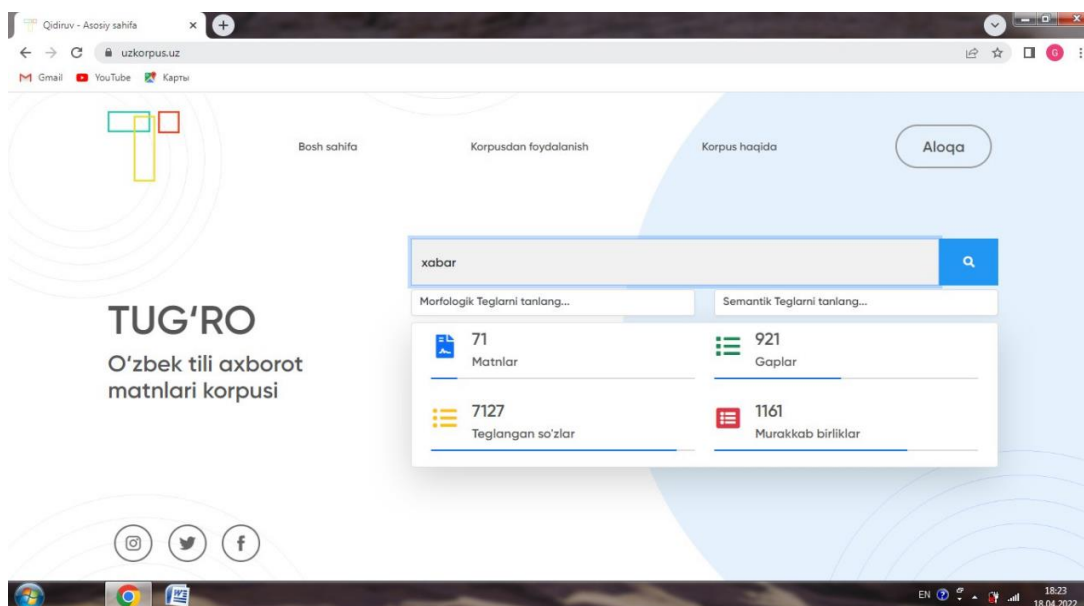
tanlanadi. “Korpus” atamasiga kengroq ta'rif berish uchun izlanishlar davomida har xil izohlarga duch kelish mumkin. Jumladan, E. Finegan darsligida shunday deyilgan: “Korpus—bu odatda mashinada o'qiladigan formatdagi matnlarning reprizentativ to'plami bo'lib, matn yaratilgan vaziyat haqidagi ma'lumotlarni, masalan, ma'ruzachi, muallif haqidagi ma'lumotlarni o'z ichiga oladi”. Vikipediya korpusni statistik tahlil va gipotezalarni tekshirish, til qoidalarini tasdiqlash yoki asoslash uchun foydalaniladigan katta va tuzilgan matnlar to'plami sifatida baholaydi. T. Makenire va E. Uilson: “Korpus til modeli sifatida foydalanish uchun aniq til mezonlariga muvofiq tanlangan til bo'laklari yig'indisidir” deya ta'kidlashadi. J. Sinkler esa korpusni tilning xilma-xilligini ko'rsatish, turli ko'rinishdagi til materiallarini tasniflash maqsadida tanlangan tabiiy holda tuziladigan asl matnlar yig'indisidir deya e'tirof etgan. Korpusga oid ilk odimlar o'tgan asrning 60-yillarida tashlandi. Shundan buyon turli maqsad va vazifalarga xoslangan korpuslar yaratilib kelinmoqda. Misol tariqasida Britaniya milliy korpusi¹, Amerika milliy korpusi², ingliz tili bank sohasi uchun korpus³, rus tili milliy korpusi⁴ kabilarni keltirib o'tamiz, ular nutq turi, matn tili, janri, ixtisoslashuvi, kirish usulidek bir qancha xususiyatlarga ko'ra farqlanadi. Hozirda har bir til o'z korpusini yaratish masalasi ustida izlanishda. Kompyuter sohasidagi so'nggi yillarda erishilayotgan yutuqlar katta hajmli, keng imkoniyatli korpuslarni kashf etilishiga asos bo'lib kelyapti. Shu o'rinda savol tug'ilishi tabiiy: korpus bugungi kunda qanchalik muhim? Korpus lingvistikasiga oid manbalarda ularning ahamiyatligi haqida so'z yuritilar ekan, qo'llanish doirasi kengligiga urg'u berib o'tiladi. Haqiqatan ham, korpusdan o'qituvchi, talaba, tadqiqotchi, adabiyotshunos, tarjimon, til o'rganuvchi, tarixchilar – barchasi birdek foydalana oladi. Xuddi shunday keng auditoriyaga mo'ljallangan korpus sifatida “Tug'ro” platformasini ayta olamiz. Platformada so'z, so'z-teg, teg qidiruvlari mavjud. So'z qidiruvida istalgan birlikni yozishimiz mumkin. Izlanayotgan so'z bazaga kiritilgan bo'lsa, u grammatik ko'rsatkichlari bilan namoyon bo'ladi. Agar foydalanuvchiga so'zning boshqa bo'laklar bilan birikuv holati kerak bo'lsa, KWIC formatiga o'tkazish kifoya. Masalan, **xabar** so'zini qidiramiz

¹www.natcorp.ox.ac.uk.

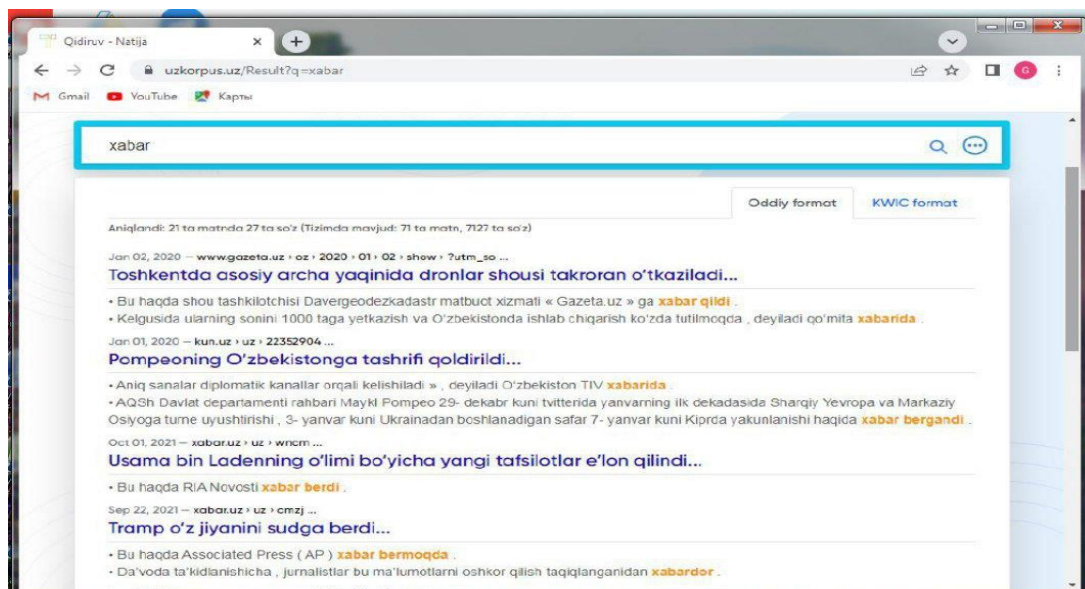
²www.americannationalcorpus.org

³www.collins.co.uk/corpus/corpussearch

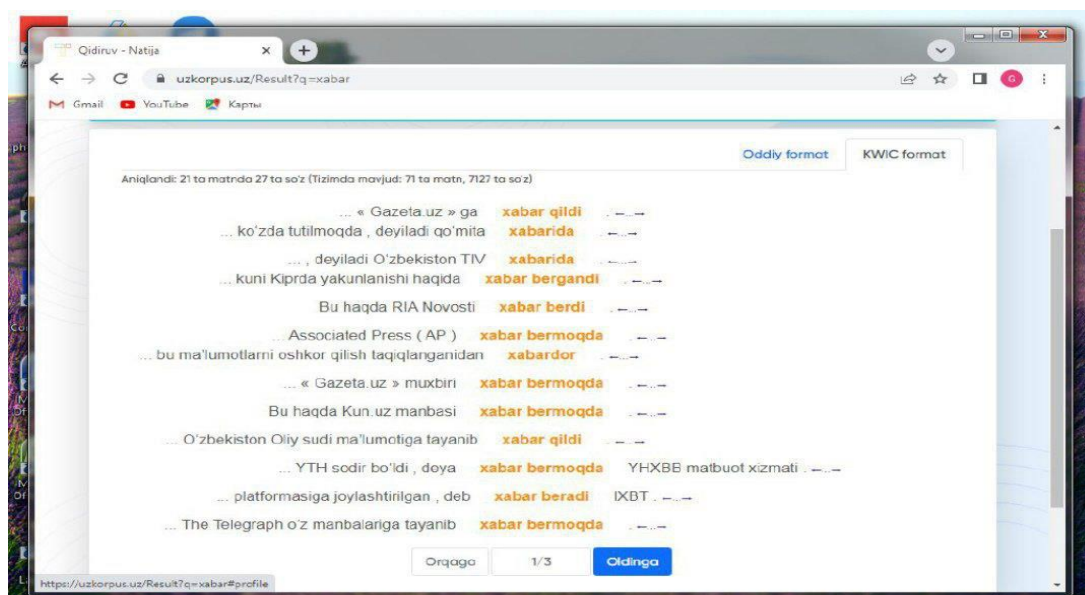
⁴www.ruscorpora.ru



Tizimda saqlangan 71 ta matn, 7127 so'zdan 21 ta matnda 27 ta so'z aniqlanadi. Aniqlangan so'zlarga e'tibor bersak, faqat **xabar** so'zining o'zini emas, balki qo'shma fe'l hosil qilingandagi holati (**xabar bermoq**, **xabar qilmoq**), sifat yasalgandagi shakli (**xabardor**), shakl yasovchi qo'shimchalar bilan kelgandagi ko'rinishlarini uchratamiz:



Xabar so‘zining KWIC formatini ham ko‘rib o‘tamiz:



Ushbu so'z so'z-teg qidiruvi orqali izlansa, morfologik-semantik xususiyatlari oydinlashadi. Tadqiqotchi uchun muayyan so'z turkumidagi qaysidir morfologik yoki semantik ko'rsatkich kerak bo'lganda teg qidiruvidan foydalanadi. Bu qidiruvlar axborot matnlariga asoslangan bo'lib, ulardan ko'zlangan asosiy maqsad tilshunoslikda tadqiqotlar olib borishni osonlashtirish, ularni haqiqiy va asosli misollar bilan ta'minlashdir. “Tug'ro” platformasini bugungi kunda bimalol statistik tahlillar, til



hodisalarini tekshirishda qo'llay olamiz.¹ Umuman olganda, korpuslar ko'pgina yo'nalishlar uchun foydali : uslubiyatni aniqlashda, lug'atshunoslikda, lingvodidaktikada muhim tayanch sifatida, ona tili va chet tillarni o'rganishda, so'zlarning qo'llanish davrini aniqlashda, lug'at tarkibida ro'y berayotgan o'zgarishlarni bilishda, darslik va qo'llanmalar tuzishda. Til o'rganishda o'rganuvchilar ma'nosi bir-biridan farq qilmaydigan, ammo matn tarkibida muayyan jihatiga ko'ra tafovut hosil qiladigan so'zlarni yaxshiroq tushunishi uchun korpusdan foydalansalar o'rinli bo'ladi. O'rganilayotgan tilda chalkashlik tug'dirayotgan so'zlarni, masalan, ingliz tilida **pretty** va **comely** so'zlarining semantikasini aniqroq anglash uchun korpusda qidiruvga beriladi. Biz bu birliklarni til o'rganish uchun maxsus yaratilgan “**Learning corpus of English**”dan ham izlay olamiz. Qidiruv natijalariga ko'ra mazkur so'zlar ishtirokida bir nechta gaplar beriladi va taqqoslab o'rganish imkoni tug'iladi. Bu esa o'rganuvchining ko'nikmasini oshiradi, mustaqil o'rganish salohiyatini o'stiradi. Ma'lumki, til ta'limiga oid darsliklarga beriladigan misollar faqat badiiy asarlar emas, huquqshunoslik, siyosat, iqtisodiyot, kundalik turmushda talaba nutqida qo'llaniladigan, jonli muloqot namunalariga oid birliklardan ham iborat bo'lishi lozim. Chunki shundagina mavzuni keng doirada mushohada etib o'rganish mumkin bo'ladi. Buning uchun baza ko'lami keng, mavzu doirasi rang-barang, janri turlicha bo'lgan korpuslarga murojaat etiladi.

¹ Tug'ro nomli korpus platformasi “O'zbek tilining internet axborot matnlari korpusini shakllantirishning nazariy va amaliy asoslari ” mavzusidagi PhD doirasida yaratildi.

O'zbekistonda bir qancha universitetlardagi mutaxassisliklarda hozirgi o'zbek tili o'qitiladi. Yo'nalishlarda o'tiladigan mavzular bir xil bo'lishi mumkin, ammo har bir mutaxassislik uchun alohida, ixtisoslashgan matnlar asosida darslar beriladi. Bir-birlarinikiga o'xshamaydigan, takrorlanmaydigan manbalarni tez va oson topishda korpusdek samaralisi yo'q. Xulosa qilib aytishimiz mumkinki, yuqorida keltirilgan korpusning bir nechtagina muhim tomonlari orqali ham uning vazifasi katta va inkor etib bo'lmas ekanligi namoyon bo'ldi. Zero, tilimizni rivojlangan xalqaro tillari qatoriga



ko‘rishni istar ekanmiz, uning korpus bazasini kengaytirishimiz, ko'paytirishimiz lozim. Chunki, dunyo kompyuter lingvistikasida milliy til korpusining mavjudligiga tilning yashovchanlik va kompyuter tiliga aylanish mezoni sifatida qaralmoqda.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. Kompyuter lingvistikasi. Toshkent: Nodirabegim, 2021
2. Богданова С. Ю, Захаров В. П., Корпусная лингвистика. Учебное пособие – Санкт-Петербург, 2020
3. Hamroyeva SH. M. Korpus lingvistikasi atamalarining qisqacha izohli lug‘ati. – Buxoro – 2018
4. Xolmanova Z. Kompyuter lingvistikasi. – Toshkent – 2019
5. <https://uz.m.wikipedia.org>
6. <https://uzkorpus.uz/>