



KORPUS LINGVISTIKASINING ZAMONAVIY YO‘NALISHLARI VA TILSHUNOSLIKDAGI AMALIY AHAMIYATI

Ro‘zimurodova Gulasal Rustam qizi

*Buxoro davlat pedagogika instituti, o‘zbek tili va adabiyoti yo‘nalishi,
1-bosqich magistranti*

Annotatsiya: Ushbu maqolada korpus lingvistikasi fanining shakllanishi, asosiy tushunchalari, yo‘nalishlari va ularning o‘zbek tilshunosligidagi tatbiqi o‘rganiladi. Korpus yaratish, teglash, menejment tizimlari, korpusdagi reprezentativlik kabi nazariy va amaliy masalalar. Korpus tilshunosligining dolzarb yo‘nalishlari, o‘zbek tilining ta’limiy va mualliflik korpuslari misolida bayon qilinadi.

Kalit so‘zlar: til korpuslari, CLAWS, TreeTagger, UDPipe, interfeys

Абстрактный: В статье рассматривается становление дисциплины корпусной лингвистики, ее основные понятия, направления и их применение в узбекском языкознании. Теоретические и практические вопросы, такие как создание корпуса, тегирование, системы управления и репрезентативность в корпусе. На примере учебных и авторских корпусов узбекского языка описываются современные тенденции в корпусной лингвистике.

Ключевые слова: языковые корпуса, CLAWS, TreeTagger, UDPipe, интерфейс

Annotation: This article studies the formation of the discipline of corpus linguistics, its main concepts, directions and their application in Uzbek linguistics. Theoretical and practical issues such as corpus creation, tagging, management systems, representativeness in the corpus. Current trends in corpus linguistics are described using the example of educational and authorial corpora of the Uzbek language.

Keywords: language corpora, CLAWS, TreeTagger, UDPipe, interface

Kirish

Zamonaviy lingvistikada yangi texnologiyalarga asoslangan tahlil metodlari keng qo‘llanilmoqda. Ana shunday yondashuvlardan biri korpus lingvistikasi bo‘lib, u lingvistik tadqiqotlarda katta hajmdagi raqamlashtirilgan matnlar korpusidan foydalanishga asoslanadi. Bu yo‘nalish matnlarni avtomatik tarzda analiz qilish, qoidalarni aniqlash, lingvistik xulosalar chiqarish, nutq birligini statistik o‘rganish imkonini beradi. Abjalovanning fikricha, korpus lingvistika – bu “kompyuter texnologiyalaridan foydalangan holda lingvistik korpuslarni yaratish va ulardan foydalanishning umumiy tamoyillarini ishlab chiqish bilan shug‘ullanuvchi soha”dir¹.

E.Finegan darsligida korpus – matn yaratilgan vaziyat va unga tegishli ma’lumotlar, jumladan, so‘zlovchi, muallif, adresat yoki auditoriya haqidagi ma’lumotlarni o‘z ichiga olgan, mashina o‘qiydigan formatga ega reprezentativlangan matnlar majmuasi deya ta’rif beriladi. Vikipediya lingvistik, til yoki matnlar korpusi – annotatsiyalangan va



strukturlangan, mashina o'qiydigan formatda taqdim etilgan, muayyan lingvistik masalalarni hal etishga mo'ljallangan til ma'lumotlarining filologik kompetentli massivi, deb ta'riflanadi. [2]. T.Mak-Eneri va E.Vilson shunday ta'rif beradi: korpus – til modeli sifatida foydalanish uchun aniq til mezonlari bo'yicha tanlangan til fragment (qism / parcha)lari to'plami. J.Sinkler – korpus lingvistikasining asoschilaridan biri, korpusni “tilni yoki tilning o'zgarishini to'liq aks ettirish uchun tashqi mezonlarga muvofiq tanlangan elektron shakldagi matnlarning parchalari to'plami” deb ta'riflaydi. Rus tili milliy korpusida korpus “ma'lum bir tilda elektron shaklda matnlar to'plamiga asoslangan axborot-ma'lumot tizimi” deya ta'riflanadi. [5] XXI asrga kelib dunyo tilshunosligida, korpus lingvistikasini ilmiyamaliy jihatdan o'rganish va amaliyotga tatbiq etish jarayonlari kegaydi. Shunga muvofiq ravishda kompyuter lingvistikasida tilga oid so'zlarni lemmalash va teglash nazariyasi, korpus bazasini yaratish, matnlarni tahlil qilish va tabiiy tilni qayta ishlash jahon tilshunosligida dolzarb masalaga aylandi. Rossiyada KorpLing rivojlanishi 2000-2004-yilda yaratilib, 2004-yilda foydalanuvchilarga taqdim etilgan rus tilining milliy korpusi (Национального корпуса русского языка / НКРЯ)[6] bilan bog'liq.

O'zbekistonda KorpLing shakllanishi 2018-yillarda nazariy tadqiqotlar bilan boshlanib, 2021-yilda O'zbek tilining ta'limiy korpusi [7] yaratilishi bilan rivojlandi.

Rus korpusshunosligida V.P.Zaxarov, A.K.Kutuzov, E.V.Nedoshivina, V.Rikov, V.Plungyan, M.Kopotev korpus, uning turlari, xususiyatlari, korpusning ijtimoiy ahamiyati va korpusni yaratish tamoyillarini tadqiq etgan. Mualliflik korpuslari bo'yicha maxsus Tadqiqotlar O.Kukushkina, A.Polikarpov va E.Surovseva asarlarida mavjud. [7]

O'zbek tilshunosligida korpus lingvistikasi nazariy tadqiqotlari jadallashgan, amaliy ishlanmalari foydalanuvchilar e'tiboriga taqdim etilayotgan, o'zbek kompyuter lingvistikasi zamirida shakllangan navqiron soha hisoblanadi. O'zbek korpusshunosligida yirik matnlar ustida ishlash, maxsus yaratilgan elektron matnlar majmuining ko'plab tadqiqotlar va ta'limiy jarayonda asosiy manba bo'lishi borasidagi fikrlar dastlab A.Po'latov [8], A.Rahimov [9]lar tomonidan ilgari surilgan bo'lsa, keyinchalik B.Mengliyev, Sh.Xamroyeva, M.Abjalova, A.Eshmo'minov, D.Axmedova, O'.Xoliyorov, N.Abdurahmonova, G.Toirova, D.O'rinboyeva, B.Elov, A.Raxmanova tomonidan til korpuslari hamda uning turlari tadqiq qilingan.

1. Korpus lingvistikasining shakllanishi

Korpus lingvistikasi XX asrning 60-yillaridan boshlab ilmiy soha sifatida shakllandi. Avvaliga bu kompyuter lingvistikasi tarkibida ko'rilgan bo'lsa-da, 1980–1990 yillarda mustaqil yo'nalishga aylandi. Braun korpusi (1 million so'zshakldan iborat), LOB korpusi, keyinchalik Britaniya milliy korpusi (100 million so'z) bu yo'nalishning asosiy yutuqlari bo'ldi. O'zbek tilshunosligida bu yo'nalish 2018-yildan boshlab jadal rivojlana boshladi. Til korpuslari bir nechta turlarga bo'linadi: milliy korpuslar, parallel korpuslar, ta'limiy korpuslar, mualliflik korpuslari va boshqa xususiy korpuslar. Ularning har biri o'z funksiyasi, maqsadi, qamrovi va foydalanuvchi auditoriyasiga ega. Milliy korpuslar biror tilning zamonaviy yoki tarixiy shaklini keng ko'lamda o'rganishga xizmat qiladi. Parallel



korpuslar tarjimaviy tahlil, kontrastiv lingvistika uchun ishlatiladi. Mualliflik korpuslari muayyan yozuvchi yoki shoirning lingvopoetikasini o'rganish imkonini beradi.

2. Korpus turlari va ularni yaratish prinsiplari

Teglash va annotatsiyalash korpus yaratishning asosiy bosqichlaridan biri hisoblanadi. Bu jarayonda har bir so'z yoki so'z shakliga grammatik, sintaktik, semantik ma'lumotlar biriktiriladi. Masalan, kelmoqda – fe'l turkumiga oidligini bildiradi. Bunday tizimlashtirish so'zlar orasidagi bog'lanishni, ularning kontekstdagi vazifasini o'rganishda muhim ahamiyatga ega. Korpus tilshunosligida CLAWS, TreeTagger, UDPipe kabi dasturlar keng qo'llaniladi. O'zbek tilida bu jarayon hali bosqichma-bosqich shakllanmoqda, lekin ta'limiy korpuslarda bu ishlar boshlangan.

3. Korpuslarni teglash va izohlash

Korpus lingvistikasining nazariy va amaliy yo'nalishlari mavjud. Nazariy yo'nalishlar korpus tushunchasining metodologik asoslarini, korpus turlarining ta'rifi va tasnifini o'z ichiga oladi. Amaliy yo'nalishlar esa korpus yaratish texnologiyalari, menejer interfeyslari, avtomatik teglash va izoh tizimlarini ishlab chiqishga qaratilgan. Lingvodidaktik korpuslar ta'lim tizimida til o'rgatishda, sun'iy intellektda mashina tarjimasini, ovoqli tanish tizimlarida qo'llaniladi. Shu bilan birga, tilni o'rgatishda grammatik qoidalarni korpus asosida tushuntirish, real kontekstda foydalanish imkonini beradi.

4. Korpus lingvistikasining asosiy yo'nalishlari

O'zbek tilshunosligida korpus lingvistikasi sohasida bir qancha muhim tadqiqotlar amalga oshirilgan. Jumladan, Sh.Xamroyeva mualliflik korpuslari tuzish asoslarini ishlab chiqqan, M.Abjalova o'zbek tilining semantik teglash tizimlarini ishlab chiqqan, A.Eshmo'minov sinonim birliklarni korpusda teglash bo'yicha tadqiqotlar olib borgan. Bundan tashqari, Alisher Navoiy korpusi, ta'limiy korpus, til ontologiyasi asosida yaratilgan korpuslar tilshunoslik, tarjimashunoslik va lingvodidaktik tadqiqotlar uchun boy manba bo'lib xizmat qilmoqda. Ular orqali o'zbek tilining lug'aviy boyligi, grammatik vositalari, uslubiy imkoniyatlari statistik tahlil qilinmoqda.

Foydalanilgan adabiyotlar

1. Abjalova, M. Korpus lingvistikasi (Uslubiy qo'llanma). – Toshkent: Bookmany Print, 2022. – 103 b.
2. https://ru.wikipedia.org/wiki/Корпусная_лингвистика (2022).
3. Xamroyeva, Sh. O'zbek tili mualliflik korpusini tuzishning lingvistik asoslari. – Qarshi, 2018.
4. Abjalova, M. Tahrir va tahlil dasturlarining lingvistik modullari. – Toshkent, 2020.
5. Eshmo'minov, A. O'zbek tili milliy korpusining sinonim so'zlar bazasi. – Qarshi, 2019.
6. www.arxiv.uz
7. <http://ruscorpora.ru>

8. <http://uzschoolcorpara.uz/>

9. Rus tili milliy korpusi [Elektron resurs].

10. Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011. – 520 б.

11. Rahimov A. Kompyuter lingvistikasi asoslari. – Toshkent: Akademnashr. 2011. – 160 b

