

KORPUS TURLARI HAMDA UNING TIPOLOGIYASI

Unarova Dilafruz Abdimajit qizi

SamDCHTI akademik litseyi

ona tili va adabiyot fani o'qituvchisi

dilafruzunarova1@gmail.com

Annotatsiya. Ushbu maqolada korpus turlari va uning tipologiyasi haqida ma'lumotlar berilgan. Lingvistik korpus va elektron kutubxona uning amaliy ahamiyati, korpus menedjeri va konkordanslar haqida ma'lumotlar berilgan. Korpus annotatsiyasi, korpus hajmi va klassifikatsiyasi. Korpusning nutqqa ko'ra turi: og'zaki, yozma, aralash korpus va matn tili, parallelligi kabi masalalar talqin etilgan. javob bera oladigan, til hodisasi bilan shug'ullanadigan, soha oldiga avval qo'yilmagan dolzarb muammolarni qo'ya oladigan tizim ekanligi maqolada batafsil yoritib berilgan.

Kalit so'zlar: on-line va off-line korpus, og'zaki, yozma, aralash korpus, matn tili, parallelligi, matnning ixtisoslashuvi va janriga ko'ra korpus turlari, kirish usuli, maqsadi, izoh xususiyati, o'zgaruvchan va turg'unligi, strukturasi, matn hajmi, xronologik jihatiga ko'ra turlari tavsifi.

Korpusning tuzilishi, maqsadi, turg'un, o'zgaruvchanligi va boshqa jihatlariga ko'ra bir qancha turlarga ega bo'lib, bu mutaxassislar tomonidan turlicha tasniflangan. Olimlar maqsadi hamda oldiga qo'yilgan vazifasiga qarab korpus turlarining boshqacha tasnifini ham keltiradi: saqlash shakliga ko'ra (ovozli, yozma, aralash) (a); matn tiliga ko'ra (bir va ko'p tilli) (b); janriy mansubligiga ko'ra (adabiy, dialektal, og'zaki, publitsistik, aralash) (v); korpusga kirish imkoniyatiga ko'ra (erkin, tijorat korpuslari, yopiq) (g); maqsadiga ko'ra (tadqiqiy, illyustrativ) (d); dinamikligiga ko'ra; qo'shimcha axborotga egaligiga ko'ra annotatsiyalangan (teglangan) va teglanmagan. V.V.Rikov korpus tasnifiga boshqacha yondashadi.

III.Ko'I. Tashkil qilish darajasi va qurilishiga ko'ra:

1. Elektron arxiv – elektron tashuvchi qurilmadagi matnlar jamlanmasi; faqat shakli standartlanmagan, bir xir shaklga keltirilmagan.
2. Elektron kutubxona – bir turdagi matnlar keltirilgan; standart shaklga keltirilgan.
3. Matnlar korpusi – standartlangan, bir xir shaklga keltirilgan; matnlar lingvistik voqelikning ma'lum qismini aks ettirishga mo'ljallangan.
4. Subkorpus – ma'lum bir korpusning avtonom qismi.

II. Xronologik belgisiga ko'ra: sinxron, monitor (tilning hozirgi holatini kuzatuvchi), diaxron.

rsatkichning mavjudligiga ko'ra: oddiy, annotatsiyalangan.

IV. Tiliga ko'ra: bir, ikki, ko'p tilli.

V. Qo'llash, foydalanish usuliga ko'ra: tadqiqot korpuslari, illyustrativ, parallel.

VI. Mavjudlik usuliga ko'ra: dinamik, turg'un. Barcha tillar uchun korpusni loyihalash, qo'llashning bir xil metodikasi mavjud emas, chunki til, an'ana, texnologik jarayonlar bir-biridan farq qiladi. Shunday bo'lsa-da, bu borada V.V. Rikov quyidagi asosiy talablarni sanab o'tadi:

1. Korpus foydalanuvchisi (yakka shaxs, guruh, lingvistik jamiyat).

2. Korpus ko'zlagan mantiqiy maqsad.

3. Korpus tuzishda ishlanadigan ma'lumotlar bazasi hajmi hamda buning real, zarurlik darajasi.

4. Mantdan foydalanish usuli (parcha, to'liq ko'rinish yoki har ikkisi).

Korpusga matn tanlash amaliyoti har xil maqsadlarda turlicha bo'ladi:

1. Til materialini tadqiq etish.

2. Matnni skanerlash.

3. To'liq shakllantirish, korpusni tuzish. Yuqoridagi jadvalda ko'rinish turganidek, eng avvalo, yozma, og'zaki manbaga asoslanishiga ko'ra korpus 2 turga bo'linadi. Korpusning dastlabki avlodi to'liq yozma korpus sanalardi. Chunki yozma matnlarni yig'ish ancha qulayroq bo'ladi. Yozma matnlarni kompyuterga kiritishning 3 usuli mavjud:

1. Matnlarni qayta kiritish.

2. Elektron shaklda mavjud matnlardan foydalanish.

3. Bosma matnni skanerlash (lekin bunda ko'plab xatoni tuzatish kerak bo'ladi).

Tajriba shuni ko'rsatadiki, korpus lingvistikasi kamida 2 tipdagi obyekt (korpus) ustida amaliyot bajaradi. Korpuslarning birinchi tipi universal, ular o'zida nutqiy faoliyatning barcha xilma-xilliklarini aks ettiradi. Ikkinchi turdagi korpus jamiyatda nutqiy amaliyotning ma'lum bir lingvistik yoki madaniy fenomenning odatga aylanishi (yoki aylanganligi)ni aks ettirib, ular ma'lum maqsadni ko'zlab tuziladi. Bunday korpus sifatida maqollar korpusi, gazeta tilida siyosiy metaforalar korpusini sanab o'tishimiz mumkin. Korpus: yozma va og'zaki, bir tilli va ko'p tilli. Mutaxassislar nutq turiga ko'ra korpusning yozma, og'zaki, aralash shaklini ajratishadi. Yozma korpusda og'zaki nutq aks ettirilmagan bo'lsa (Braun korpusi, LOB), og'zaki korpus faqat og'zaki matndan iborat bo'ladi; aralash korpus, odatda, til me'yorlari aksi bo'lgan milliy korpuslardir.

Publitsistik korpus sifatida XX asr oxiri rus tili gazeta matnlari korpusi, siyosiy metaforalar korpuslarini sanash mumkin. Foydalanuvchi uchun eng asosiy mezon korpusdan foydalanish imkonining erkin yoki chegaralanganligi sanaladi. Shunga ko'ra erkin, tijorat (pullik) korpusi hamdayopiq korpuslar mavjud. Erkin foydalanish

mumkin bo'lgan korpus on-line rejimda xohlagan paytda istalgan matnning to'liq shakli bilan ishlash imkonini beradi. Ba'zan erkin korpus ham korpus ma'lumotlarining ma'lum qismidagina foydalanish bilan chegaralangan bo'ladi. Tijorat korpusi (pullik korpus)dan foydalanish on-line rejimda ishlash ruxsatnomasini talab qiladi yoki foydalanuvchi korpusning disk variantini sotib olishi ham mumkin. Korpusni olishdan oldin korpus annotatsiyasi bilan tanishish, hatto sinov variantda korpusning ma'lum qismi bilan ishlab ko'rish mumkin.

Korpusning bunday tasniflanishi yana boshqa atamalar (masalan, indekslangan, indekslanmagan, annotatsiyalangan, annotatsiyalanmagan, rus tilida таггированные и нетаггированные kabilar) bilan ham nomlanadi. Teglangan korpusda teg xarakteridan kelib chiqqan holda so'z hamda gapga morfologik, sintaktik, semantik, prosodik va h. izoh (korpus tili bilan teg) qo'shilgan bo'ladi. Dinamik korpusning maqsadi doimiy ravishda kompyuter xotirasini matn bilan to'ldirib, korpusni boyitib borish, asosiy belgisi esa yangilanish, to'ldirilib borilishidir. Bu korpus leksikografiya rivoji uchun xizmat qiladi, chunki til doimiy harakatda, so'zlar yangi-yangi ma'nolar kasb etishi, neologizm, arxaizm, istorizmlarning paydo bo'lishini kuzatib borish imkonini beradi.

Ko'pincha milliy korpuslar shunday korpus sirasiga kiradi. Matn hajmiga ko'ra to'liq matnli yoki lavhali (fragment matnli) korpuslar bor. Braun, Lankaster-Oslo-Bergen korpusiga qo'yilgan dastlabki talab qat'iy ravishda har biri 2000 so'zdan (so'z qo'llash holati) iborat matnlardan tuzilishi lozim edi. Tabiiyki, bu mezonga qat'iy amal qiluvchi korpus yo'q. Shunday ekan, bu korpuslar fragment, ya'ni lavha matnli hisoblanadi.

Foydalanilgan adabiyotlar ro'yxati:

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2013. – 161 с.
2. Грудева В.А. Корпусная лингвистика. Учебное пособие. 2-е изд., стереотип. – М.: Флинта, 2012. – 165 с.
3. Недошивина Э.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – СПб., – 2006. – 26 с.
4. Хамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли лугати. – Тошкент: Камалак, 2018. – 96 б.
5. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
6. <http://corpus.leeds.ac.uk/>
7. <http://corpus.leeds.ac.uk/ruscorpora.html>