

KORPUS TURLARI VA TIPOLOGIYASI

*Unarova Dilafruz Abdimaqit qizi**SamDCHTI akademik litseyi ona tili va adabiyot fani o'qituvchisi**dilafruzunarova1@g.mail.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.

Abstract. This article provides information about corpus types and their typology. Linguistic corpus and electronic library, its practical importance, corpus manager and concordances are given information. Corpus annotation, corpus size and classification. Issues such as the type of corpus according to speech: oral, written, mixed corpus and text language, parallelism are interpreted. able to answer, dealing with the phenomenon of language, before the field

The article states that it is a system that can solve unsolved problems explained in detail.

Абстрактный. В этой статье представлена информация о типах корпусов и их типологии. Приводятся сведения о лингвистическом корпусе и электронной библиотеке, его практическом значении, менеджере корпуса и согласованиях. Аннотация корпуса, размер и классификация корпуса. Раскрываются такие вопросы, как тип корпуса по речи: устный, письменный, смешанный корпус и язык текста, параллелизм. способен ответить, имея дело с явлением языка, перед полем

В статье говорится, что это система, способная решать нерешенные проблемы.

объяснил подробно.

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.

Ключевые слова: онлайнный и офлайнный корпус, устный, письменный, смешанный корпус, язык текста, параллелизм, виды корпуса по специализации и жанру текста, способ доступа, назначение, особенность

аннотации, изменение и устойчивость, структура, размер текста, описание типов по хронологическому аспекту.

Keywords: on-line and off-line corpus, oral, written, mixed corpus, text language, parallelism, types of corpus according to specialization and genre of the text, access method, purpose, annotation feature, change and stability, structure, text size, description of types according to the chronological aspect.

Korpusning tuzilishi, maqsadi, turg'un, o'zgaruvchanligi va boshqa jihatlariga ko'ra bir qancha turlarga ega bo'lib, bu mutaxassislar tomonidan turlicha tasniflangan (Qarang: 1.1-jadval). 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 (dinamik (monitor), turg'un) (d); qo'shimcha axborotga egaligiga ko'ra annotatsiyalangan (teglangan) va teglanmagan (e). V.V.Rikov korpus tasnifiga boshqacha yondashadi.

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.

III. Ko'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‘rinib 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. Zamonaviy katta hajmli korpuslar, odatda, aralash: yozma matn bilan birga og‘zaki matnni ham qamrab oladi. Hatto Britaniya milliy korpusi (BNC) matnlarining 10 foizi og‘zaki matnlar yozuvidan iborat. Til, asosan, og‘zaki shaklda yashaydi, uning yozma shakli ikkilamchi. Shuning uchun og‘zaki korpus juda muhim, hech bo‘lmaganda, korpusning aralash turi til tadqiqi uchun maqsadga muvofiq hisoblanadi. Adabiylik mezoniga ko‘ra korpus adabiy, dialektal, og‘zaki, terminologik hamda aralash kabi turlarga ajratiladi. Og‘zaki korpusga Sankt-Peterburgda tuzilgan “Одинъ речевоу день” korpusini misol keltirish mumkin. Terminologik korpus – bevosita jonli til asosida terminlar lug‘ati yaratishni maqsad qilgan korpus. Janriga ko‘ra adabiy, folklor, dramaturgiyaga xos hamda publitsistik korpus kabi turlarga ajratiladi.

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.

Yopiq korpuslar o'ziga xos tor (xususiy) maqsadda tuziladi va ular ommaviy foydalanishga mo'ljallanmaydi. Maqsadiga ko'ra korpusning tadqiqot, tavsifiy yoki illyustrativ kabi ko'rinishlari mavjud. Tadqiqiy korpus til harakati (mavjudligi)ning turli aspektini o'rganish maqsadida tuziladi. Korpusning bu tipi keng miqyosdagi lingvistik topshiriqlarni bajarishga mo'ljallangan. Illyustrativ korpus tadqiqot natijasi sifatida paydo bo'ladi, uning vazifasi faqat yangi faktlarni topish emas, balki u yoki bu tilga oid faktni avvalroq boshqa usulda topilgan misollarni dalillashning lingvistik manbayi sifatida xizmat qiladi. Illyustrativ korpusning tipik namunasi ma'lum bir so'zning aniq matndagi ma'nosini izohlashga mo'ljallangan "Путеводитель по дискурсивным словам русского языка" korpusidir. Korpusga qo'shimcha izoh mavjudligiga ko'ra qo'shimcha izohli (razmetkalangan), izohsiz (razmetkalanmagan) turlari farqlanadi.

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. Korpus tarkibining to'ldirib borilishiga ko'ra turg'un va dinamik (o'zgaruvchan) korpus turi mavjud. Dastlabki korpuslar tilning turg'un holati aksi, tilning ma'lum bir davrini qamrab olgan. Nisbatan ahamiyatli zamonaviy korpus sanalgan Britaniya milliy korpusi (BNC) ham turg'un korpusdir. Bunday korpusning tipik namunasi avtorlik korpusi, ya'ni bir yozuvchi asarlari yig'indisi sanaladi. Ba'zi lingvistik amallarni bajarish uchun (masalan, so'z chastotasini aniqlash) turg'un korpus talabga javob berolmay qoladi. Chunki faqat chegaralangan davr qamrab olingan bo'ladi. Shu ehtiyojdan kelib chiqqan holda keyinchalik to'ldirib boriladigan dinamik (monitoring) korpus ham paydo bo'la boshladi. 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. Dinamik korpus katta davrning yozma manbalarini aks ettirib, ular turli diaxronik tadqiqotlarni bajarishga mo'ljallangan. Bunga Cobuild Project korpusi misol bo'la oladi. Bunday korpuslar "ochiq korpus" deb ataladi va ularning kamchiligi shundaki, har doim ham tilni bir xilda (adekvat holatda) aks ettirmaydi, matn tanlashning aniq bir mezoniga amal qilmaydi. Yaratilish maqsadiga ko'ra universal, maxsus korpus ko'rinishi ham mavjud.

Universal korpus, odatda, turli janrdagi matnlarni qamrab oladi.

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. To'liq matnli korpusga ma'lum muallif asarlari korpusi, qisqa matnlar korpusini misol keltirish mumkin. Gazeta sarlavhalari korpusi ham shunga misol bo'ladi. Korpusga qo'shimcha izoh mavjudligiga ko'ra qo'shimcha izohli (razmetkalan), izohsiz (razmetkalanmagan) turlari farqlanadi. Korpusning bunday tasniflanishi yana boshqa atamalar (masalan, indekslangan, indekslanmagan, annotatsiyalangan, annotatsiyalanmagan, rus tilida таггированный и нетаггированный 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.

Korpus tarkibining to'ldirib borilishiga ko'ra turg'un va dinamik (o'zgaruvchan) korpus turi mavjud. Dastlabki korpuslar tilning turg'un holati aksi, tilning ma'lum bir davrini qamrab olgan. Nisbatan ahamiyatli zamonaviy korpus sanalgan Britaniya milliy korpusi (BNC) ham turg'un korpusdir. Bunday korpusning tipik namunasi avtorlik korpusi, ya'ni bir yozuvchi asarlari yig'indisi sanaladi. Ba'zi lingvistik amallarni bajarish uchun (masalan, so'z chastotasini aniqlash) turg'un korpus talabga javob berolmay qoladi. Chunki faqat chegaralangan davr qamrab olingan bo'ladi³⁶. Shu ehtiyojdan kelib chiqqan holda keyinchalik to'ldirib boriladigan dinamik (monitoring) korpus ham paydo bo'la boshladi. 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. Dinamik korpus katta davrning yozma manbalarini aks ettirib, ular turli diaxronik tadqiqotlarni bajarishga mo'ljallangan. Bunga Cobuild Project korpusi misol bo'la oladi. Bunday korpuslar "ochiq korpus" deb ataladi va ularning kamchiligi shundaki, har doim ham tilni bir xilda (adekvat holatda) aks ettirmaydi, matn tanlashning aniq bir mezoniga amal qilmaydi. Yaratilish maqsadiga ko'ra universal, maxsus korpus ko'rinishi ham mavjud. Universal korpus, odatda, turli janrdagi matnlarni qamrab oladi. Ko'pincha milliy korpuslar shunday korpus sirasiga kirad

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. Ҳамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли лугати. – Тошкент: Камалак, 2018. – 96 б.
5. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
6. [http:// corpus.leeds.ac.uk/](http://corpus.leeds.ac.uk/)
7. <http://corpus.leeds.ac.uk/ruscorpora.html>