

TILLARNI O'RGANISHDAGI KORPUS TURLARI VA ULARDAN FOYDALANISH

Ashirbaeva Madina Nuralievna

*Chirchiq davlat pedagogika universiteti, Lingvistika va tillarni o'qitish metodikasi
kafedrası o'qituvchisi va doktranti*

Annotatsiya: korpuslar, ya'ni til yoki ma'lumotlar to'plamlari tilshunoslik, kompyuter lingvistikasi va boshqa ko'plab sohalarda muhim ahamiyatga ega. Korpuslar turli maqsadlar uchun yaratiladi va ularning foydalanish maqsadiga qarab turli klassifikatsiyalari mavjud. Ushbu maqolada korpuslarning asosiy turlari haqida ma'lumotlar berishga harakat qilingan. Ularning turlari har bir sohaning o'ziga xos xususiyatlari va talablariga javob beradi. Bu korpuslar yordamida tilni o'rganish va o'qitishda ham chuqurroq o'rganilishi mumkin.

Kalit so'zlar: umumiy korpus, maxsus korpus, sinxron/diaxron korpuslar, monitor korpusi, bir tilli/ko'p tilli korpus, parallel korpus, reprezentativlik.

Abstract: Corpora, collections of language or data, are important in linguistics, computational linguistics, and many other fields. Corpora are created for various purposes and have different classifications depending on their intended use. This article attempts to provide information about the main types of corpora. Their types correspond to the specific characteristics and requirements of each field. These corpora can also be used to study the language in more depth in learning and teaching.

Keywords: general corpus, specific corpus, synchronic/diachronic corpora, monitor corpus, monolingual/multilingual corpus, parallel corpus, representativeness.

Аннотация: Корпуса, коллекции языка или данных, важны в лингвистике, компьютерной лингвистике и многих других областях. Корпуса создаются для различных целей и имеют различные классификации в зависимости от их предполагаемого использования. В этой статье предпринята попытка предоставить информацию об основных типах корпусов. Их типы соответствуют конкретным характеристикам и требованиям каждой области. Эти корпусы также могут использоваться для более глубокого изучения языка в процессе обучения и преподавания.

Ключевые слова: общий корпус, специфический корпус, синхронные/диахронические корпуса, мониторинг корпус, одноязычный/многоязычный корпус, параллельный корпус, репрезентативность.

Kirish

Har bir korpusning o'ziga xos maqsadi va o'ziga xosligi borligi haqiqat bo'lsa-da, korpuslarning maqsadi va xususiyatlariga qarab turli toifalariga bo'linishi ham haqiqat. Gatto (2013) korpuslarni ulardagi ma'lumotlarga qarab toifalarga ajratish imkonini beruvchi uch xil mezonni taqdim etadi: "umumiy va ixtisoslashtirilgan, sinxron va diaxronik, bir tilli va ko'p tilli" [Gatto, 2013, 15]. Bu turkumlar bir-birini istisno qilishga moyildir, ya'ni bir tilli korpus bir vaqtning o'zida ko'p tilli bo'lishi mumkin emas va aksincha.

Bir tomondan umumiy korpus, uning nomidan ko'rinib turganidek, iloji boricha reprezentativlik uchun turli sohalarda, janrlar va turli registrlardagi matnlarni o'z ichiga olgan korpusdir. Ushbu turdagi korpus "... til o'rganish va tarjima qilish uchun grammatik kitoblar yoki materiallarni ishlab chiqish" uchun foydalaniladi. [Gatto, 2013, 15]. Ular juda aniq ma'lumotlarni o'z ichiga olganligi sababli, ularni lingvistik shveytsariya pichog'i deb hisoblash mumkin. Umumiy korpuslarga misollar sifatida birinchi kompyuterlashtirilgan Broun korpusi va Britaniya Milliy Korpusini (BNC) keltirish mumkin.

Boshqa tomondan ixtisoslashtirilgan yoki maxsus korpus umumiy korpusdan farqli o'laroq, o'zining kengligini yo'qotadi va faqat ma'lum bir soha, janr, vaqt yoki tilning xilma-xilligini ifodalaydi. Ularning o'ziga xosligi tufayli ular umumiy korpuslarga qaraganda qisqaroq, aniqroq va ulardan foydalanish va o'rganish osonroqdir. Bunday korpuslarga misollar etib, Erta

ingliz tibbiy yozuvlar korpusi (Corpus of Early English Medical Writing/CEEM) yoki Havo harakatini boshqarish nutq korpusini (Air Traffic Control Speech Corpus/ ATCOSIM) keltirish mumkin.

Korpuslar, shuningdek, sinxron va diaxronik korpuslar ortasida farqlanadi. Sinxron korpus ma'lum bir vaqt oralig'ida tilning xususiyatlarini tasvirlashga harakat qiladigan ma'lum bir cheklangan vaqt oralig'idagi ma'lumotlarni o'z ichiga oladi. Sinxronik korpuslarga misollar Helsinki ingliz matnlari korpusi (Helsinki Corpus of English Texts) yoki Zamonaviy Amerika ingliz tili korpusi (Corpus of Contemporary American English, COCA). Diaxronik korpus tilning evolyutsiyasini o'rganishga harakat qiladi, shuning uchun ular barcha asr va davrlarga oid matnlarni o'z ichiga oladi. Sinkler (1991) korpusning ushbu tasnifi ichida monitor korpusi deb nomlangan diaxronik korpusning maxsus turini taklif qiladi. Ushbu turdagi korpusning maqsadi "tildagi o'zgarishlarni kuzatishdir, shuning uchun uni yangilash uchun doimiy ravishda yangi matnlar qo'shib turadi, shu sabab u doimiy ravishda o'sib boradi. Monitor korpusiga misol qilib, doimiy ravishda zamonaviy matnlar bilan yangilanib turadigan Bank of English (BoE) hisoblanadi.

Korpusning so'ngi belgilovchi omili uning tarkibiga kiradigan tillar bo'lib, ular bir tili yoki ko'p tili bo'lishi mumkin. Ko'p tili korpuslar turli tillardagi matnlarni o'z ichiga oladi va ko'p tili korpuslarning eng keng tarqalgan turlari parallel korpus va qiyosiy korpuslardir.

Parallel korpus ikki yoki undan ortiq tildagi bir xil matnlarni (asl matn va tarjimani) o'z ichiga oladi, masalan, ochiq manbali parallel korpus (Open source parallel corpus/ OPUS). Qiyosiy korpus turli sohalar va muassasalarga moslashtirilgan korpuslarini o'z ichiga olgan onlayn korpus, shu bilan birga solishtirish mumkin bo'lgan janr, mavzu va vaqt oralig'i kabi xususiyatlarni o'zida jamlagan ikki yoki undan ortiq matnlar to'plamini o'z ichiga olgan va ularni solishtirish mumkin bo'lgan korpuslar. Bunga misol sifatida Ingliz tilining xalqaro korpusi (ICE) va Ingliz tilini o'rganuvchi xalqaro korpusi (International Corpus of Learner English/ICLE) keltirsa bo'ladi.

Yuqorida keltirilgan korpus turlaridan tashqari birinchi jadvalda ham ko'plab til korpuslarining mavjudligini bir qator xususiyatlarga ko'ra tasniflashga imkon beradi. Taqdim etilgan tasnif V. V. Rykov [Рыков В.В.] va V. P. Zaxarov va S.Yu. Bogdanova [Захаров В.П., Богданова С.Ю.] tasnifi asosida tuzilgan.

Belgisi	Korpus turlari
Til ma'lumotlarining turi	Og'zaki, yozma, multimedia, aralash
Matnlarni taqdim etish tili	Bir tili, ikki tili, ko'p tili, aralash
Til normasi	Adabiy, dialekt, aralash
Matnlarning diskursiv-stilistik mansubligi	Adabiy, folklor, so'zlashuv, jurnalistik, ilmiy (terminologik), aralash
Matnlarning janr mansubligi	Nasriy, she'riy, dramaturgiya, aralash
Foydalanish maqsadi	Ko'p maqsadli, ixtisoslashgan
Kirish usuli	Erkin foydalanish uchun, tijorat, yopiq, aralash
Maqsad	Tadqiqot, tasviriy, aralash
Xronologik xususiyat	Sinxron, diaxronik
Korpusning mavjudligi usuli	Dinamik (to'ldiriladigan), statik
Indeksatsiya	Belgilangan (izohlangan), belgilanmagan
Belgilashning tabiati	Morfologik, sintaktik, semantik, prosodik, aralash
Matnlar hajmi	To'liq matnli, fragment matnli, aralash

Jadval 1. Til korpuslarining tasnifi (V.V.Rykov)

Ushbu kontekstda ona tilida so'zlashuvchilar tomonidan emas, balki ushbu tilni o'rganuvchilar tomonidan yozilgan matnlardan iborat o'quv korpuslari (learner corpus) alohida e'tiborga loyiqdir (insholar, imtihon materiallari va boshqalar). O'quv binolari ikkinchi tilni

o'zlashtirish jarayonini o'rganishda, shuningdek, ona tilida so'zlashuvchilarning nutqini ushbu tilni o'rganuvchilarning nutqi bilan taqqoslashda kontrastli tadqiqotlarda qo'llaniladi [Baker P., Hardie A., McEnery T.A.].

Korpus turlari bo'yicha xitoylik olim D.Liu ham korpus kategoriyalarining asosiy to'rt turini va ular ta'lim maqsadi va mazmun mohiyatiga ko'ra bo'linishini farqlab bergan. Tilni o'qitish va o'rganishda muhim hisoblangan umumlashgan, maxsus, o'rganuvchi va parallel korpuslarga ajratgan. [D.Liu]. Yuqorida keltirilgan korpus kategoriyalar faqat til o'rganish uchun eng muhimlari hisoblanib, ular yana korpus maqsadlariga qarab boshqa korpus turlariga ham bo'linadi.

Xitoylik olim tomonidan berilgan korpus turlaridan farqli o'laroq G.R.Bennet ham korpus kategoriyalarini umumiy jihatdan 8 turga, ya'ni umumiy, maxsus, o'rganuvchi, pedagogik, tarixiy, parallel, qiyosiy va monitor turlariga bo'linadi deb hisoblagan. Va bulardan tilni o'rganish va o'qitish maqsadiga ko'ra umumiy, maxsus, o'rganuvchi va pedagogik korpuslarni ajratib keltirgan [G.R.Bennet, 13-14].

Korpuslarni maqsadiga va xususiyatlariga ko'ra eng ommabop 8 turga bo'linishi borasida E.Friginal ham ma'lumotlar berib, uning fikricha ular: umumiy, maxsus, yozma, og'zaki, qiyosiy, izohli, parallel, va monitorga ajratgan. [E.Friginal].

Til o'rganish va o'qitishda ushbu maqolada keltirilgan korpus turlari til o'rganish jarayonini yanada samarali qilish uchun keng qo'llaniladi. Bunda korpuslar til o'rganuvchilarga haqiqiy til materiallarini taqdim etish orqali ularga tilni o'zlashtirishda yordam beradi. Korpuslarning pedagogik qo'llanilishiga oid misollarni keltirib o'tadigan bo'lsak. Til o'rganishda yozish ko'nikmalarni rivojlantirish uchun yozma korpuslardan foydalanish ularning tilni o'rganishda haqiqiy matnlardan foydalanishga imkon beradi. Ingliz tilini o'rganuvchilar uchun Britaniya Milliy Korpusidan foydalanish orqali ular zamonaviy ingliz tilidagi yozma asarlar, maqolalar va hikoyalarni o'qishadi. Bu orqali til o'rganuvchilar tilning boyligini va turli mualliflar va janrlar o'rtasidagi uslubiy farqlarni tahlil qilish imkoniyatiga ega bo'lishadi.

Og'zaki korpuslardan foydalanish til o'rganuvchilarga haqiqiy suhbatlar va muloqotlarni tinglash imkonini beradi. Og'zaki korpuslarga misol sifatida Kembridj ingliz tili korpusidan foydalanib, ular turli suhbatlar, interviyular va podkastlarni eshitishlari mumkin. Bu jarayon natijasida til o'rganuvchilar o'zlarining og'zaki nutq va intonatsiyalarni rivojlantirishadi va turli kommunikativ vaziyatlarda qanday qilib samarali muloqot qilishni o'rganadilar.

Huquq va tibbiyot sohasidagi talabalar uchun maxsus korpuslar yaratilishi mumkin. Misol uchun Medline korpusi tibbiyot sohasidagi ilmiy maqolalar va tadqiqotlarni o'z ichiga oladi. Talabalar o'zlarining sohasidagi maxsus terminlarni o'rganishadi. Ushbu terminlarni qanday qilib amaliyotda qo'llashni tushunishadi.

Tarjimonlik kurslaridagi talabalar parallel korpuslardan foydalanish talabalarga bir xil matnlarning bir necha tildagi tarjimalarini taqqoslash imkoniyatini beradi. Parallel korpuslarni foydalanish natijasida talabalar turli tillardagi leksik va grammatik strukturalarini taqqoslash orqali tilni yanada chuqurroq tushunishadi, talabalar tarjima jarayonida qanday usullarni qo'llashni o'rganadilar.

Xulosa

Korpuslar turli turlarini o'z ichiga oladi va ularni o'rganish, tahlil qilish va qo'llash jarayonida turli mutaxassislar ishtirok etadi. Bu korpuslar yordamida til va muloqot qilish jarayonlarini chuqurroq o'rganish, tilni o'qitish va tarjima jarayonlarini yaxshilash mumkindir. Korpuslar tilshunoslik va boshqa sohalarida yangi imkoniyatlar yaratadi va ilmiy tadqiqotlar va amaliyotlarda muhim ahamiyatga ega.

Foydalanilgan adabiyotlar:

1. Bennet G. Using corpora in the language learning classroom. Michigan, University of Michigan Press, 2010.-12-14 p.
2. Gatto M. The Web as Corpus: Theory and practice. Cambridge: Cambridge University Press, 2013.

3. Friginal E. Corpus linguistics for English teachers new tools, online resources and classroom activities. -New-York.:Taylor and Francis., 2018.-p.12.
4. Liu D., & Lei L. Using corpora for language learning and teaching.-Annapolis Junction, MD.: TESOL Press. 2017.p. 140.
5. McEnery T., &Hardie A. Corpus linguistics.-Cambridge,UK.:Cambridge University Press., 2012.-321 p.
6. Sinclair J. Corpus, concordance, collocation.-oxford, UK: Oxford University Press., 1991- p.13-18.
7. Захаров В.П. Богданова С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов. - Иркутск: ИГЛУ, 2011.-с. 7-10.
8. Рыков В.В. Корпус текстов как онтология речевой деятельности и Труды Международной семинара Диалог. - 2004.- М.: Наука, 2004 – 56-61 с.
9. Аноркулов, С. (2021). Хронотоп как отличительная черта романного универсума ФМ Достоевского. Просвещение и познание, (5 (5)), 16-23.
10. Аноркулов, С. ПРОСВЕЩЕНИЕ И ПОЗНАНИЕ. ПРОСВЕЩЕНИЕ И ПОЗНАНИЕ Учредители: Прудник Ксения Евгеньевна, (5), 16-23.