

KORPUS LINGVISTIKASINING SHAKLLANISH VA RIVOJLANISH TARIXIDAN LAVHALAR

Dildor Baxtiyarovna Otajonova

Chirchiq davlat pedagogika universiteti

d.otajonova@cspi.uz

ABSTRACT

This article deals with the problem of corpus linguistics and concordance, and also its origin and development. Corpus linguistics nowadays is perhaps most readily associated in the minds of linguists with searching through screen after screen of concordance lines and wordlists generated by computer software, in an attempt to make sense of phenomena in big texts or big collections of smaller texts. This method of exegesis based on detailed searches for words and phrases in multiple contexts across large amounts of text that can be traced back to the thirteenth century, when biblical scholars and their teams of minions pored over page after page of the Christian Bible and manually indexed its words, line by line, page by page. Concordancing arose out of a practical need to specify for other biblical scholars, in alphabetical arrangement, the words contained in the Bible, along with citations of where and in what passages they occurred. Technology has been the major enabling factor in the growth of corpus linguistics but has both shaped and been shaped by it. The ability to store masses of data on relatively small computer drives and servers meant that corpora could be as big as one wanted. In this regard, lexicographers led the way. Their aim has always been to collect the maximum amount of data possible, to capture even the rare events in a language.

Keywords: corpus linguistics, concordance, corpora, history, development of concordance.

KIRISH

Xorijiy tillarga bo'lgan qiziqish, yuqori darajada bilishga bo'lgan talab tarjima sohasining birmunchayin ilgarilashiga turtki bo'ldi desak, mubolag'a bo'lmaydi. Kundan kunga taraqqiy etayotgan tarjimaning turli jabhalari, nafaqat tilga xos bilimlar, balki, IT sohasiga oid ko'nikmalarni ham chuqur o'zlashtirish kabi masalalarning dolzarbligidan dalolat beradi So'nggi yillarda tilshunoslik jabhalarida, jumladan, leksikografiya, terminologiya, kompyuter lingvistikasi, korpus lingvistikasi sohalarida bir qator tadqiqotlar olib borilayotganligiga qaramasdan, nutqiy faoliyat va unda



ifodalangan soʻz maʼnolari bilan bogʻliq masalalar toʻlalgicha oʻz yechimini topgani yoʻq. Taʼkidlash joizki, leksemalar va matnlar korpuslarini yaratish hamda ulardan foydalanishning oʻziga xos jihatlarini oʻz obʼekti sifatida belgilab olgan soha – korpus lingvistikasi bugungi kunda koʻplab lingvistik tadqiqotlarga moʻljallangan til maʼlumotlarining yirik majmualarini, ulardan foydalanishning nazariy asoslari va amaliy mexanizmlarining shakllanishi uchun zamin yaratadi.

ADABIYOTLAR TAHLILI VA METODOLOGIYA

Korpus lingvistikasi tilshunoslarning takbiri boʻyicha, “muvofiqlik asosida hosil qilingan satrlar va soʻz roʻyhatini ekrandan qidirish bilan kom'yuter dasturlari orqali katta matnlar yoki kichik matnlarning katta toʻplamdagi hodisalarni tushunishga urunish”ning eng oson usuli sifatida ularning ongida shakllangandir. Bu kabi matn izohlash usuli katta hajmga ega matnlardagi bir nechta kontekstda ifodalangan soʻzlar va iboralarni toʻliq qidirishga asoslangan boʻlib, ushbu usul yordamida XIII asrda injilshunos olimlar va ularning izdoshlari tomonidan injil sahifalari satrlarga va soʻzlarga koʻchirilgan. Shuningdek, konkordans yohud muvofiqlik asosida injildagi soʻzlarning alifbo tartibida, iqtiboslari hamda qaysi matnlarda qoʻllanilganligini koʻrsatishning amaliy ehtiyojidan kelib chiqqan holda boshqa olimlar tomonidan amalga oshirilgan.

Soʻzning etimologik talqiniga nazar tashlaydigan boʻlsak, *concordantia* lotincha *cum*, yaʼni, “*bilan*” va *cor* esa “*yurak*” maʼnolarini anglatib, injilning shunchaki matnlardan tashkil topgan kitob emas, balki ushbu mashaqqatli urunishning asl gʻoyaviy asosi uning ilohiyatga uygʻunlashgan xabar ekanligi haqidagi daʼvoni taʼkidlashdan iborat boʻlgan. *Korpus* leksemasiga keltirilgan bir qator tavsiflar mavjud boʻlib, uning asosiy vazifasini yoritib berishga qaratilgandir.

NATIJALAR

E.Finegan oʻzining qoʻllanmasida: “korpus – matn yaratilgan vaziyat va unga tegishli maʼlumotlar, jumladan, soʻzlovchi, muallif, adresat yoki auditoriya haqidagi maʼlumotlarni oʻz ichiga olgan, mashina oʻqiydigan formatga ega reprezentativlangan matnlar majmuasi,” deya eʼtirof etadi. Vikipediya elektron lugʻatida esa, “lingvistik, til yoki matnlar korpusi – annotatsiyalangan va strukturlangan, mashina oʻqiydigan formatda taqdim etilgan, muayyan lingvistik masalalarni hal etishga moʻljallangan til maʼlumotlarining filologik kompetentli massivi,”¹ deb taʼriflanadi. T.Mak-Eneri va E.Vilsonlar korpusni “til modeli sifatida foydalanish uchun aniq til mezonlari boʻyicha

¹https://ru.wikipedia.org/wiki/Корпусная_лингвистика

tanlangan til fragment (qism / parcha)lari to‘plami,” deb hisoblashadi. Shuningdek, J.Sinklarning fikriga ko‘ra, “korpus lingvistikasining asoschilaridan biri, korpusni “tilni yoki uning o‘zgarishini to‘liq aks ettirish uchun tashqi mezonlarga muvofiq tanlangan elektron shakldagi matnlarning parchalari to‘plami”. Keltirilgan tavsiflardan xulosa qilish mumkinki, korpusning muayyan maqsadga muvofiq yaratilishi, formatga ega bo‘lishi, reprezentativlik va metalingvistik jihatlarni o‘zida namoyon etishi zaruratini ta’kidlaydi.

Antonio Padua injilning birinchi anonim konkordansi *Concordantiae Morales* bilan bog‘liq *Vulgate* – injilning V asrga tegishli bo‘lgan, lotin tilidagi muqobiliga asoslangan. Shu davrning o‘zida anchayin mukammallashtirilgan nusxasi 1230 yil, Avliyo James monastirida, Karolik Kardinal *Hugo* tashabbusi bilan tashkil etilgan 500 kishidan iborat Dominikan rohiblari jamoasi tomonidan “so‘zlar indeksi” tuzilganligini kuzatish mumkin. Shundan buyon, injil konkordanslarining ko‘plab boshqa tubdan o‘zgargan muqobillari, jumladan, 1737 yil *Crudenning* “Muqaddas yozuvlarga to‘liq muvofiqlik” va 1890 yil Strongning “Injilning to‘liq muvofiqlik” kabi konkordanslari yaratilgan.

MUHOKAMA

Hozirgi kunga kelib, komp'yuter muvofiqlashtiruvchi programmalari 500 ta rohibning ishini bir necha soniyada amalga oshiradi. Shuningdek, Shekspirning asarlari ham konkordansning asosiy vositasi sifatida ushbu sohaning rivojlanishiga hissa qo‘shgan. Misol tariqasida keltirib o‘tish joizki, Becketning² 1787 yilda Shekspirning asarlari asnosida yaratgan konkordansida so‘z va uning lingvistik konteksti hamda joylashuvi berilgan bo‘lib, adabiyot shinavandalari uchun qimmatli ma'lumotlarni taqdim etgan. Ilgari davrda, qiyinchilik bilan terilgan konkordanslarning o‘zagi va maqsadlari bizga ma'lum bo‘lgan komp'yuter programmlarida saqlab kelinmoqda. Elektron korpus yaratish tashabbusiga injil hamda badiiy asarlarning muvofiqlashtirish va indekslash jarayonidagi mashaqqatli mehnat sabab bo‘la olmagan.

Binobarin, Xristian dini rohibi Roberto Busa (Jesuit priest Roberto Busa) tomonidan 1950-yillarda boshlangan va 1970-yillarning oxirlarida yakunlangan, avliyo Tomas Akvinas (St Thomas Aquinas)ning to‘liq asarlarini o‘z ichiga qamrab oluvchi “Index Thomisticus / Tomas indeksi” nomli qisqa elektronik raqamlashtirilgan bazasining yaratilishi konkordansning elektron usulining vujudga kelishiga turtki bo‘ldi. Shu o‘rinda, yana ikkita asosiy manbani sanab o‘tish kerak, masalan, leksikograflar va Chomskiydan oldin

²Becket (1787) A Concordance to Shakespeare



faoliyat yuritgan struktur tilshunoslarning tadqiqotlari ham zamin bo'la olgan. Har ikkala usul ham tadqiqotchilar uchun ma'lumotlarni to'plashda muhim ahamiyat kasb etgan. Shuningdek, doktor Samuel Jonson (Dr Samuel Johnson)ning 1755 yilda chop etilgan birinchi "Ingliz tilining izohli lug'ati" ko'p yillar, jumladan, 1560-1660 yillar mobaynida qo'llanilgan qog'ozlarga tushirilgan misollarni yig'ish asnosida qog'ozda korpus yaratishning mahsuli bo'lib xizmat qildi. Ehtimol, uch milliondan oshiq qog'oz bo'laklari korpusida mujassamlashgan so'zlarning qo'llanishini amalda ta'minlovchi lug'at, ya'ni, 1880-yillarda Oxford English Dictionary tarkibiga kiritilgan. Dunyoning mashhur lug'atlaridan birini yaratishda ushbu millionlab qog'oz parchalaridagi so'zlar tahrirlangan holda matnning mazmunga ega bir qismiga aylantirildi.

Leech (1992)ning ta'kidlashicha, 1950-yillarda amerikalik strukturalistlar, jumladan, Harris, Fries and Hilllarning tadqiqotlarida "ma'lumotlarni yig'ish" tushunchasi keng qo'llanilganligini kuzatish mumkin. Shuningdek, ilk injilshunos olimlar va yozuvchilarning so'zlarni qidirish va indekslash borasidagi olib borilgan chora-tadbirlari nafaqat tilga oid ma'lumotlarni to'plash, balki, ona tiliga xos bilimlarni tadqiq etish nuqtai nazaridan strukturalistlarning konkordans ilmi rivojidadagi peshqadamligidan dalolat beradi. Ona tilini o'zlashtirishga bo'lgan qiziqish transkripsiyalangan ma'lumotlar – 1960-yillarda yaratilgan CHILDES³ til bazasiga asoslangan bo'lsada, ushbu loyiha 1984 yildagina rasmiy tasdiqlangan.

XULOSA

Yuqorida keltirilgan tahlillarga tayangan holda, xulosa qilish mumkinki, korpus lingvistikasi uzoq taraixga ega bo'lib, didniy asarlar yaratilgan davrlarda shakllanib ulgurganligining guvohi bo'lamiz. Shuningdek, ingliz yozuvchilarining asarlari to'plamining yaratilishida ham korpusning o'rni biqiyos bo'lgan. Natijada, hozirgi kunga kelib uning anchayin mukammallashgan muqobili yaratilgan. Demak, korpus asosida boshqa sohalarni ham taraqqiy ettirish mumkin.

REFERENCES

1. Aarts, J. and Meijs, W. (eds) (1984) *Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research*. Amsterdam: Rodopi.
2. Bernardi, S. (2000) *Competence, Capacity, Corpus*. Bologna: CLUEB.
3. Biber, D., Johansson, S., Leech, G., Conrad, S. and Finnegan, E. (1999) *The Longman Grammar of Spoken and Written English*. London: Longman.

³childes.psy.cmu.edu



