

## MAIN FEATURES OF CORPUS LINGUISTICS

**Annotation:** This article discusses programs developed in computer technologies for corpus linguistics, their differences, as well as the views of a group of scientists on corpus linguistics.

**Key words:** features of corpus linguistics, models of corpus linguistics, corpus linguistics.

## KORPUS LINGVISTIKASINING ASOSIY XUSUSIYATLARI

**Аннотация:** Ushbu maqolada korpus lingvistikasi bo'yicha kompyuter texnologiyalari orqali ishlab chiqilgan dasturlar haqida, ularni farqi va bir guruh olimlarni korpus lingvistikasi bo'yicha fikri keltirib o'tilgan.

**Калит so'zlar:** korpus lingvistikasining xususiyatlari, korpus lingvistikasi modellari, korpus tilshunoslik.

## ОСНОВНЫЕ ОСОБЕННОСТИ КОРПУСНОЙ ЛИНГВИСТИКИ

**Аннотация** В данной статье рассматриваются программы, разработанные в компьютерных технологиях для корпусной лингвистики, их отличия, а также взгляды группы ученых на корпусную лингвистику.

**Ключевые слова:** особенности корпусной лингвистики, модели корпусной лингвистики, корпусная лингвистика.

Kompyuter texnologiyalarining rivoji natijasida zamonaviy tilshunoslikda korpus va kompyuter lingvistikasi terminlarini semantik tahlil, korpus lingvistikasi, kritik diskurs tahlil, shuningdek, deskriptiv, qiyosiy va komponent tahlil metodlari asosida tadqiq etish tobora dolzarb ahamiyat kasb etib bormoqda. Tilshunoslik nuqtai nazaridan yondashsak, kompyuter lingvistikasining paydo bo'lishi bugungi kunda fanning bir qancha muammolariga echim topish imkoniyatini beradi. Kompyuter texnologiyalarini

tilshunoslikning ilmiy-nazariy, falsafiy muammolari: til va nutq, sistema va struktura, semiotika, sintagmatika va paradigmatika, tipologiya, lug'atshunoslik kabi muhim hamda murakkab masalalari bo'yicha olib borilayotgan ilmiy tadqiqotlarda qo'llashning ijobiy tomonlariga tayangan holda korpus lingvistikasi sohasi yaratildi. Kompyuter va korpus lingvistikasi bir-biriga o'xshash va ba'zi hollarda bir xil degan noto'g'ri qarashlar uchrab turadi. XX asrning 90-yillari birinchi yarmida korpus tilshunosligi til to'g'risidagi fanning alohida qismi sifatida shakllandi. Shu bilan birga, u kompyuter tilshunosligi bilan yaqindan hamkorlik qiladi, uning yutuqlaridan foydalanadi va, o'z navbatida, uni boyitadi.

Korpus lingvistikasining xususiyatlarini o'rganishda muhim ilmiy va amaliy ahamiyatga ega. Shuningdek, olib borilgan va borilayotgan izlanishlar soha muammolarining mavjudligi hamda ularga yechim topishdagi yondashuvning o'ziga xosligini ko'rsatish bilan bir qatorda, uning boshqa fanlardan mustaqil ekanligini ko'rsatib bermoqda. Qolaversa, tilshunoslikka yo'naltirilgan ko'plab oliy o'quv yurtlarida korpus lingvistikasiga ixtisoslashtirilgan kafedra va yillik kesimida o'tkaziladigan xalqaro korferentsiyalarning tashkil etib kelinishi, qolaversa, korpus tahlillarini qo'llashga doir maqolalarni chop etib kelayotgan qator ilmiy jurnallarning mavjudligi ushbu yangi tarmoqning bugungi kundagi ahamiyatini yaqqol ko'rsatib turibdi.

Kompyuter lingvistikasi ham o'z muammo hamda yutuqlari nuqtai nazaridan o'rganib chiqilgan. A.N.Baranov o'z ishlarida sun'iy ong muammosini ochib berishga to'xtalgan, B.Orexov semantik analiz, texnik tarjima va avtomatik morfologiya tomonidan fanni o'rganib chiqqan bo'lsa I.A.Chesebiyev kompyuter lingvistikasining metodlari rivojlanish tarixini o'rganib chiqqan. Shu o'rinda, N. Xomskiyning sintaktik strukturalar va universal grammatika nazariyasi fan rivojiga katta hissa qo'shgan. Ch.Fillmor esa ushbu nazariyaga qo'shimcha sifatida tillardagi predlog kabi

noo'xshashliklar va ularning til strukturasi boshqa tushunchalar orqali ifoda etilishi nuqtai nazardan yondashgan. U. Vuds (Woods) fanga protsessual semantika g'oyasini kiritdi. Yuqorida nomlari keltirilgan olimlar o'z ishlarida til va ong tushunchalarini texnikada jamlashni o'z oldilariga maqsad qilib olishgan. Zero, kompyuter lingvistikasining eng dolzarb vazifalaridan biri – bu inson va kompyuter o'rtasida avtomatik muloqot interfeysini yaratish. Bundan kelib chiqadiki, ushbu sohada asosiy maqsad til o'rganuvchilari va izlanuvchilarini maxsus texnik qurilma yoki programmalar bilan ta'minlash.

V.Mamontova va N.Xomskiy kabi bir guruh olimlar korpus lingvistikasini kompyuter tilshunosligining tarkibiy qismi sifatida baholashadi. Ularning ko'pchiligi quyidagicha ta'rifni o'rtaga surishadi: korpus tilshunosligi – kompyuter texnikasi yordamida tilshunoslik korpusini qurish va ulardan foydalanishning umumiy tamoyillarini ishlab chiqadigan birliklarni yaratiladigan kompyuter tilshunosligi. Shu bilan birga, odatda, kompyuter asbob-uskunalari, dasturlari, korpusning kompyuter texnologiyalari va ma'lumotlarini qayta ishlashning keng doirada ishlash mexanizmini modellashtirish tarzida tushuniladi.

Kompyuter tilshunosligi korpus tilshunosligi uchun vositalarni (ya'ni dasturlarni) yaratadi. Shu ma'noda ular bir-birini to'ldiradilar. Masalan, korpus tilshunoslari korpusda so'zlarni avtomatik tarzda belgilash uchun maxsus vositalarga muhtoj. Agar yuz milliondan ortiq foydalanadigan matn birligi mavjud bo'lib, har bir so'z uchun nutqning bir qismini belgilash kerak bo'lsa, bu jarayonni to'laligicha qo'lda amalga oshirib bo'lmaydi, shu bois tizimli dasturiy ta'minot talab qilinadi. Unga ko'ra so'zlarni tasnifiy belgilariga ko'ra belgilash avtomatik ravishda amalga oshiriladi.

I.A.Chesebiyev o'z izlanishlarida kompyuter lingvistikasidan foydalanishning ayrim jihatlariga to'xtalib, xususan, quyidagilarni eslatib o'tadi:

- avtomatik tarjima;

- autentik va yarim-autentik matnlardan avtomatlashtirilgan ma'lumotlarni olish;
- inson va mashina o'rtasida qulay interfeyslarni yaratish, tabiiy tillarda aloqaning kvantitativ (muallifdan: Quantitative – (ing. – songa asoslangan) ixtiyoriy tanlangan manba xususidagi xulosani aniq raqamlar asosida chiqarish) tavsifi.

Ilk korpusning yaratilishi 1812 yilga borib taqaladi, bunda nemis olimi Kadling o'zining nemis so'zlaridagi undosh harflar distibutsiyasini tahlil qilgan. Vaholanki, u davrda hali kompyuter termining o'zi ham bo'lmagan. Keyinchalik, zamonaviy ingliz tili korpusi namunalaridan biri sifatida Brovn korpusi 1960–1961 yillarda yaratildi va u ilk bor bosma holda chop etildi. Oradan bir yil o'tibgina, ushbu korpus elektronlashtirildi. Ko'rinib turibdiki, ilk korpuslar kompyuter texnologiyalarisiz ham mavjud bo'lgan va izlanishlar olib borilgan. Keyinchalik fan rivoji qidiruv metodi(konkordans)ni elektron formatdagi matnlarda o'tkazishni taqozo etdi va kompyuter lingvistikasi bilan sohalararo munosabatga ehtiyoj tug'ildi.

Shu bilan birga, kompyuter tilshunosligi “odatda, kompyuter vositalari (dasturlari, ma'lumotlarni tashkil qilish va qayta ishlash uchun kompyuter texnologiyalari)ni muayyan sharoitlarda, vaziyatlarda, muammoli sohalarda va tillardagi modellarning ko'lamini nafaqat tilshunoslikda, balki boshqa fanlarga ham qo'llashni nazarda tutadi”. A.N.Baranov “tilni kompyuter modellashtirish tilshunoslikdagi dasturlash nazariyasi (kompyuter modellarini qurish va kompyuter grafikasi fani) ni qo'llash sohasi sifatida qaralishi mumkin. Chunki ushbu muammo keyingi yillarda faqatgina tilshunoslik masalasi sifatida e'tirof etilmoqda”, deb ta'kidlagan. Korpus tilshunosligi kompyuterni aniq vosita sifatida ishlatadi va ularsiz, albatta, u bajaradigan vazifalarni bajarolmaydi. Biroq bu zamonaviy bilimlarning deyarli har bir sohasiga taalluqli bo'lib, ularni kompyuter fanining ajralmas qismlariga aylantirmaydi.

Korpus tilshunosligi o'z materiallarini yaratadi, anirog'i, uni mustaqil ravishda tuzadi. Bu esa, mustaqil tilshunoslikni o'z ichiga oladi – u ishlatilgan og'zaki materiallarning o'ziga xos xususiyati va vositalari, ya'ni korpuslar tahlili uchun dasturlarga egadir. Ilm-fanning mustaqilligi tahlil materiallari yoki tadqiqot usullari mavjudligi bilan belgilanadi. Korpus tilshunosligi unisiga ham, bunisiga ham ega.

Korpus lingvistikasiga tilshunoslikning lisoniy korpuslar yaratish va ulardan foydalanishni o'rganadigan bo'limi sifatida qaraymiz va bu tanlab olingan matnlar asosida ifodalangan tillarni o'rganish demakdir. A.N.Baranov kompyuternaya lingvistikasi terminiga quyidagi ta'rifni beradi: “Kompyuter lingvistikasi – kompyuter texnologiyalardan til faoliyatlarini modellashtirishda ma'lumotlar yig'ish, saralash va keng ko'lamda foydalanishdir. Ba'zi muammoli vaziyatlarga ko'ra, tilning kompyuter modeli nafaqat tilshunoslikda, balki u bilan chambarchas bog'liq bo'lgan fanlarda ham qo'llash nazarda tutiladi”. Olim o'z fikrining izchilligini saqlab qolgan holda, boshqa bir asarida shunday uni asoslaydi, “kompyuter lingvistikasida tilga asosiy urg'u berilmasdan, balki eng so'nggi o'rinlarda e'tibor qaratilib, bunda tilning kompyuter modellashtiruvi tilshunoslik sohalaridagi programmalashtirishga oid texnik ta'minot nazariyalarigagina ahamiyat beriladi”. Korpus lingvistikasi esa kompyuterlardan vosita sifatida foydalanadi. Tan olish kerakki, agar kompyuterlar bo'lmaganida u hozirdagi faoliyatini amalga oshira olmas edi. Fanlarning taraqqiyoti natijasida korpus lingvistikasi barcha fanlarda tatbiq etiladi, lekin bu orqali ushbu fanlar kompyuter sohasi (computer science)ning bir qismi bo'lib qolmaydi.

Zamonaviy tilshunoslikda har ikkala fanning o'rganilish darajasi va mazkur sohalar bo'yicha dunyo olimlarining izlanishlari va ilmiy qarashlariga tayangan holda kompyuter lingvistikasi korpus lingvistikasidan quyidagi xususiyatlariga ko'ra keskin farqlanishi bo'yicha xulosalar chiqarildi:

1. Nomlanishidagi farq;
2. Mazkur ikki sohaning belgilangan maqsadlari jihatidan:

a) kompyuter lingvistikasi tilga oid programmalarni ishlab chiqishni o'z oldiga asosiy maqsad qilib qo'yadi. Ta'kidlash joizki, bu bevosita tilshunoslik oldidagi muamolarni hal etishga qaratilmagan;

b) korpus lingvistikasi esa, fanning maqsadi sifatida mavjud tilning korpusini yaratish va uning yordamida tildagi so'zlar, iboralar hamda boshqa lisoniy birliklarning qo'llanilish chastotasini aniqlashni belgilab qo'ygan. Bu esa korpus lingvistikasini aynan tilshunoslikka oid ekanligini ko'rsatadi.

3. Kompyuterning barcha fanlarda foydalanish o'рни va ahamiyati tomonidan:

a) kompyuter lingvistikasida kompyuter qurilmalari asos bo'lib, ushbu fandagi barcha izlanishlar kompyuterlar va ularning afzalliklarini oshirishga qaratiladi;

b) korpus lingvistikasi esa kompyuter qurilmalaridan faqatgina vosita sifatida foydalanadi, bunda matnlarni yig'ish va saralash amaliyotlarini bevosita ular yordamida amalga oshiradi.

Har qanday yuksalayotgan fan o'z amaliyotida kompyuterlar taraqqiyotining yutuqlaridan foydalanadi, lekin ushbu fanlar (korpus lingvistikasi, kompyuter lingvistikasi, kompyuter matematikasi, kompyuter fizikasi) kompyutershunoslikning bir qismi sifatida qabul qilinishi joiz emas. Mazkur fanlar tahlil ob'ektini tadqiq etishda kompyuterdan vosita sifatida foydalanadi, xolos.

Shuni aytish mumkinki, korpus lingvistikasi kompyuter lingvistikasining alohida bir qismi yoki bo'limi emas. Ular yo'nalish hamda tadqiqot ob'ekti, maqsad va muammolari turlicha bo'lgan mustaqil soha hisoblanadi.

#### Foydalanilgan adabiyotlar ro'yxati

1. O'zbekiston Respublikasi Prezidentining "O'zbekiston Respublikasida xorijiy tillarni o'rganishni ommalashtirish faoliyatini sifat jihatidan yangi bosqichga olib chiqish chora-tadbirlari to'g'risida" 2021 yil 19 maydagi PQ-5117-son Qarori, [www.lex.uz](http://www.lex.uz).

2. O'zbekiston Respublikasi Vazirlar Mahkamasining "Xorijiy tillarni o'rganishni ommalashtirishni samarali tashkil etish chora-tadbirlari to'g'risida" 2021 yil 19 maydagi 312-son Qarori, [www.lex.uz](http://www.lex.uz).

3. Ataboev N.B. Chomsky's Critics to Corpus and the Problem of Expert Speakers' Linguistic Intuition // Филология масалалари. – Тошкент. (4)/2015. – Б. 108-111.

4. Атабоев Н.Б. Корпус тилшунослиги шаклланишининг диахроник тадқиқи // Бердақ номидаги Қорақалпоқ давлат университетининг Ахборот-номаси. – Нукус. №4/ 2018. – Б. 147-150.

5. Juraev E. S. Foreign experience in conducting financial policies for small business development. *Economy and Society*. – 2018.

6. Razzakov S. J., Xolboev Z. X., Juraev E. S. Investigation of the Stress-Strain State of Single-Story Residential Buildings and an Experimental/Theoretical Approach to Determining the Physicomechanical Characteristics of Wall Materials // *Solid State Technology*. – 2020. – Т. 63. – №. 4. – С. 523-540.