



Nilufar NORPO'LOTOVA,
Samarqand davlat chet tillar instituti tayanch doktoranti
E-mail mirjahon776@gmail.com

Professor B.Mengliyev taqrizi asosida

KORPUS LINGVISTIK TA'MINOTINING NAZARIY ASOSLARI

Annotatsiya

Ushbu maqolada kompyuter texnologiyalari bilan tilshunoslikning o'zaro bog'liqligi natijasida kompyuter lingvistikasi degan yangi yo'nalishning vujudga kelishi sohada juda katta ko'tarilish bo'lgani, tilshunoslikning alohida sohasi hisoblangan kompyuter lingvistikasi fanining paydo bo'lishi, korpus lingvistik ta'minotining nazariy asoslari haqida xorij va o'zbekistonlik olimlarning fikrlariga tayangan holda atroflicha fikr yuritilgan.

Kalit so'zlar: Kompyuter texnologiyalari, pragmatolingvistika, lingvokulturologiya, sotsiolingvistika, kompyuter lingvistikasi, kompyuter dasturlari, matematik lingvistika.

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ЯЗЫКОВОГО ОБЕСПЕЧЕНИЯ КОРПУСА

Аннотация

В данной статье подробно рассматривается возникновение нового направления под названием компьютерная лингвистика в результате взаимодействия компьютерных технологий и лингвистики, возникновение дисциплины компьютерная лингвистика как отдельной области языкознания, а также теоретические основы корпусной лингвистики, основанные на мнениях зарубежных и узбекских ученых.

Ключевые слова: Компьютерные технологии, прагматика, лингвокультурология, социолингвистика, компьютерная лингвистика, компьютерные программы, математическая лингвистика.

THEORETICAL FOUNDATIONS OF CORPUS LINGUISTIC PROVISION

Annotation

This article discusses in detail the emergence of a new direction called computer linguistics as a result of the interaction of computer technologies with linguistics, the emergence of the science of computer linguistics, which is considered a separate branch of linguistics, and the theoretical foundations of corpus linguistics, based on the opinions of foreign and Uzbek scientists.

Key words: Computer technology, pragmatolinguistics, linguoculturology, sociolinguistics, computer linguistics, computer programs, mathematical linguistics.

Kirish. XXI asrga kelib hayotimizning hech bir sohasini kompyuter texnologiyalarisiz tasavvur etib bo'lmay qoldi. Butun dunyoni axborot texnologiyalari qamrab oldi va bu, o'z navbatida, ilm-fanga ham sezilarli ta'sir ko'rsatdi. Bu tilshunoslik sohasiga ham bevosita aloqadordir. Ayni shu sohada bir qancha zamonaviy yo'nalishlar paydo bo'ldi. Masalan, pragmatolingvistika, lingvokulturologiya, sotsiolingvistika kabilar. Ayniqsa, kompyuter texnologiyalari bilan tilshunoslikning o'zaro bog'liqligi natijasida kompyuter lingvistikasi degan yangi yo'nalishning vujudga kelishi sohada juda katta ko'tarilish bo'ldi. Bu mavzuda xorij [1] va yurtimiz olimlari [2] bir qancha izlanishlar olib borishgan va tadqiqotlar o'z samarasini bermoqda.

Ayniqsa, oxirgi o'n yillikda [1] kompyuterlar odatiy ishchi tilshunoslarning professional hayotini tubdan o'zgartirdi. Bu o'zgarishlar asta-sekin turli xil uzoq tajribalar asosida yuzaga keldi. Qo'l yozuvi allaqachon ekranda paydo bo'ldi va o'zgarish tezligi kundan kunga kuchaymoqda. Hisoblash va tilshunoslik o'rtasidagi munosabatlar texnologik o'zgarishlarni yanada chuqurroq va qiziqarli qiladi. Shu o'rinda kompyuter lingvistikasining o'zi nima hamda u o'z oldiga qanday maqsad va vazifalarni qo'yadi degan savol yuzaga keladi. Kompyuter lingvistikasining [3] asosiy maqsadi tilshunoslik muammolarini yechishning kompyuter dasturlarini ishlab chiqishdir. Barchamizga [2] ma'lumki, lingvistika fani mustaqil fan sifatida XIX asrda, aniqrog'i, 1816-yilda shakllana boshladi. Mana shu davrdan keyin tilshunoslik sohasida turli tadqiqotlar qilinmoqda. Keyingi

yillarda barcha fanlar singari tilshunoslik fani ham ikki fanning o'zaro birikuvi asosida qiyosan o'rganiladigan bo'ldi. Bunday fanlar sirasiga sotsiolingvistika (sotsiologiya va tilshunoslik), psixolingvistika (psixologiya va tilshunoslik), etnolingvistika (etnografiya va tilshunoslik), neyrolingvistika (nevrologiya va tilshunoslik), matematik lingvistika va kompyuter lingvistikasi fanlarini misol tariqasida keltirish mumkin. Buni fanlarning o'zaro aloqadorligi yoki hamkorligi deyish mumkin.

Kompyuter lingvistikasi [2] eng mashhur va tilshunos olimlarning diqqat markazida bo'lgan, fanlarning aralashuvi asosida yuzaga kelgan, zamonaviy tilshunoslik yo'nalishlaridan biridir. Bu fan matematik lingvistika yo'nalishining mantiqiy davomi bo'lib, amaliy tilshunoslikning eng muhim qirrasini hisoblanadi. U 1954-yil AQShda Jorjtaun universitetida mashina tarjimai bo'yicha dunyoda o'tkazilgan birinchi tajriba asosida yo'nalish sifatida shakllandi va 1960-yilga kelib mustaqil fan sifatida o'rganila boshlandi. Kompyuter lingvistikasi inglizcha "computational linguistics" so'zining o'zbekcha muqobilidir. Shu o'rinda turli xil savollar yuzaga keladi. Xo'sh, kompyuter lingvistikasining asosiy vazifasi nimalardan iborat va uning tilshunoslikdagi ahamiyati qanday?

Kompyuter lingvistikasining asosiy vazifalariga quyidagilar kiradi:

avtomatik o'qitish tizimini ishlab chiqish;
bilimlarni tekshirish;
matnlarni turli jihatdan avtomatik tahrirlash;

matnlarning avtomatik tarzda morfologik, sintaktik va semantik tahlilini ta'minlovchi tizimlar yaratish;
 mashina tarjimasini uchun mo'ljallangan dasturlarni ishlab chiqish;
 matn strukturasi gipertekst texnologiyasini yaratish;

elektron lug'atlar va tražeurslar yaratish;
 sun'iy intellekt tizimini ishlab chiqish;
 matnlar korpusini yaratish;
 matnning avtomatik referatlash;
 sujet strukturasi modellash;
 nutqni avtomatik tushunish tizimini ishlab chiqish;
 matndan faktlarni ajratib olish;

Yuqoridagilardan kelib chiqib aytish mumkinki, bugungi kunda kompyuter lingvistikasi barcha fanlar uchun nihoyatda zarur. Kompyuter lingvistikasi fanining eng istiqbolli yo'nalishlaridan biri sun'iy intellektidir. Ijtimoiy tarmoqlarda unga quyidagicha ta'rif berilgan: "Sun'iy intellekt – bu kompyuter yoki robotning shunday xususiyatiki, asosan, insonning aqliy faoliyati bilan bo'g'liq bo'lgan masalalarni, masalan, fikrlash, nutq mazmunini anglash, ma'lumotlarni umumlashtirish kabilarni yechishga yo'naltirilgan bo'ladi. Bu atama mazkur xususiyatlarga ega sistemalarni ishlab chiqish bilan bog'liq kompyuter texnologiyalarining bir bo'limiga nisbatan ishlatilmoqda".

Mana shu o'rinda korpus lingvistikasining obyekti, predmeti, asosiy maqsad va vazifalari nimalardan iborat ekanligi haqida alohida to'xtalib o'tish kerak. Korpus lingvistikasining obyekti bu tabiiy tildagi matnlardir. Korpus lingvistikasining predmeti esa til korpuslari sanaladi. Korpus lingvistikasining o'z predmeti keng doiradagi foydalanuvchilarning manfaati uchun tadqiqotlarga mo'ljallangan til ma'lumotlarining yirik majmualarini yaratish va ulardan foydalanishning nazariy asoslari va amaliy mexanizmlaridir. Korpus lingvistikasining asosiy maqsadi til tizimining lingvistik tavsifi sanaladi, bunda muloqotning nutqiy jarayondan olingan ma'lumotlari korpusda o'z aksini topadi. Manzura Abjalovaning keltirishicha, korpus lingvistikasining asosiy vazifalari quyidagilarni o'z ichiga oladi:

ushbu sohaning nazariy asoslarini ishlab chiqish;
 har xil turdagi korpuslarni yaratish va ulardan foydalanish tajribasini tahlil qilish;
 korpusga qo'yilgan umumiy talablarni shakllantirish;
 turli xil tadqiqot va o'quv vazifalari uchun korpuslarni yaratish;

tilshunoslikning turli sohalarida matn korpuslarini qo'llashning samarali usullarini shakllantirish.

Amerika Milliy Korpusi (AMK) ingliz tilida bo'lib, 1990-yildan beri ishlab chiqarilgan 22 million so'z, yozma va og'zaki ma'lumotlarni o'z ichiga oladi. Hozirda AMK tarkibiga elektron pochta, tvitlar va veb-ma'lumotlar kabi bir qancha janrlar kiradi.

Britaniya Milliy Korpusi (BMK) XX asrdagi Britaniya ingliz tilining yozma va og'zaki qismining 100 million so'zdan iborat to'plamidir. Buning asosiysi yozma va qolgan kichik hajmdagi qismi esa og'zakidir. Britaniya og'zaki qismining audio formati ham mavjud bo'lib, undan ko'pchilik samarali foydalanadi.

Butun dunyo bo'yicha mana shu korpusning paydo bo'lish davrini [7] XX asr deb baholaydilar va bu yangilik butun fan olamida ulkan kashfiyot sifatida qaraladigan bo'ldi. O'zbek tili korpusini yaratish yo'lida qilingan va hali bajarilishi lozim bo'lgan ishlar juda ko'p. Bu sohada faqatgina kompyuter lingvistikasining ma'lum bir qismi sifatida qaralib kelinmoqda va hozirgi kunga kelib ushbu sohada ko'plab ishlar amalga oshirilmoqda. Bu sohaning to'liq shakllanishi uchun esa ma'lum bir katta hajmdagi baza kerak bo'ladi va shu baza orqali o'zbek tilidagi barcha ma'lumotlar matn shaklida kiritilib boriladi. Bu yo'nalishda esa yaqin yillardan boshlab ma'lum ishlar amalga oshirib kelinmoqda, o'tgan asrlarda elektron tarzda mavjud bo'lmagan bo'lsa-da, o'zbek tili tarixida yaratilgan va undan ma'lum davrda amalda foydalanilgan. Bunga misol tariqasida bir nechta mashhur asarlarni kiritishimiz mumkin.

Ushbu kichik hajmdagi tadqiqotlardan shuni ko'rish mumkinki, "korpus lingvistikasi" atamasi bilan bog'liq ba'zi qiyinchiliklar mavjud. Avvalo, bu atama kutilganidan ancha kam qo'llanilgan. Ayniqsa, CLRA korpusidagi natijalar hayratlanarli, chunki undagi maqolalar "korpus lingvistikasi" atamasi kalit so'zlarda, sarlavha yoki annotatsiyada ishlatilgani sababli tanlab olingan edi. Albatta, bu kutilmagan kam uchrash holatini talqin qilishning ko'plab mumkin bo'lgan usullari mavjud va bu yerda faqat bir necha taxminlar ilgari surilgan. Bugungi kunda Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universitetida bir qancha fidoyi va jonkuyar olimlar tashabbusi bilan o'zbek tili milliy korpusi yaratilishiga erishildi. Darhaqiqat, u vaqt va mehnatni tejaydigan qurol, ona tilimizning zamonaviy shakliga xos axborot tizimi sifatida texnik jarayonni tezlashtirishi shubhasiz.

ADABIYOTLAR

1. A Practical Guide Using computers in linguistics Edited by John M. Lawler and Helen Aristar Dry. London and New York. First published 1998 by Routledge 11 New Fetter Lane, London EC4P 4EE This edition published in the Taylor & Francis e-Library, 2003 page 239
2. Rahimov A. Kompyuter lingvistikasi asoslari / A. Rahimov; mas'ul muharrir A. Nurmonov; O'zR Oliy va o'rta maxsus ta'lim vazirligi, Andijon davlat universiteti. – T.: Akademnashr, 2011. – 160-b.
3. Po'latov A, Muhammedova S. Kompyuter lingvistikasi (qayta ishlanib, to'ldirilgan nashri) O'zbekiston Respublikasi Oliy va o'rta maxsus ta'lim vazirligi oliy o'quv yurtlari uchun o'quv qo'llanma sifatida nashrga tavsiya etgan, – Toshkent, 2009, 90-bet.
4. Xolmanova Z. Kompyuter lingvistikasi. – Toshkent, 2009, "Asian book house", 196-bet.
5. Abjalova M. Korpus lingvistikasi. (Uslubiy qo'llanma) "BOOKMANY PRINT", – Toshkent, 2022, 103-bet.