**JDPI o’qituvchi**

**Mamirov Ozodjon**

**KOMPYUTER LINGVISTIKASIDA O‘ZBEK KORPUSINING TAKOMILLASHUVI.**

Kompyuter lingvistikasi fani, tarixi va zamonaviy holati.Tilshunoslik tarixida kompyuter lingvistikasining tutgan o‘rni. An’anaviy tilshunoslikdan kompyuter lingvistikasining farqlari. Evropadagi kompyuter lingvistikasiga oid tadqiqotlar. Rossiyadagi kompyuter lingvistikasiga oid tadqiqotlar. Turkologiya va o‘zbek tilshunosligidagi kompyuter lingvistikasiga oid tadqiqotlar.

Kompyuter lingvistikasining asosiy yo‘nalishlari.Avtomatik tarjima. Avtomatik tarjima yo‘nalishining tarixi. Tarjimon dasturlar. Avtomatik tarjima yo‘nalishining zamonaviy holati va istiqboli. Avtomatik tarjimaning ma’lumotlar bazasini yaratish. Bu jarayon asosan, XX asrning ikkinchi yarmiga to‘g‘ri kelsa-da, o‘zbek tilshunosligida ushbu sohalar (neyrolingvistika, kompyuter lingvistikasi, psixolingvistika, geografik tilshunoslik kabi) rivoji mustaqillik davridan so‘ng kirib kela boshladi va bir qancha ilmiy ishlarning tadqiqot obyektiga aylana boshladi. Shu jumladan, “XXI asr – kompyuter texnologiyalari asri” ning o‘zbek tilshunosligiga olib kirgan yangiligi bo‘lmish kompyuter lingvistikasi sohasining rivoji ham. Kompyuter lingvistikasi fani 1960-yilda AQShda shakllandi.[[1]](#footnote-2) Mazkur fan XX asrning o`rtalarida paydo bo`lgan matematik lingvistika fani asoslaridan kelib chiqqan. Tabiiy tillar grammatikasi formal modeli kompyuterda tarjima qilish, tillarga o`rgatish, bilimlarni diagnostika qilish, matnlarni tahrir etish kabi amaliy dasturlarni ishlab chiqishga asos bo`lgani holda kompyuter lingvistikasi yo`nalishiga zamin yaratdi.

Kompyuter leksikografiyasi.Matematik statistika haqida tushuncha. Kompyuter leksikografiyasiga oid asosiy tushuncha va terminlar. Tabiiy tillarning kompyuter lug`atlari. O`zbek tilining tub so`zlar lug`atlarini kompyuterda yaratish muammolari. Bunday lug’atlarda yasama so`zlarni berish. Atoqli otlar va geografik nomlar lug`atlarini kompyuterda yaratish muammolar ochib beriladi.

Matematik statistika va uning tilshunoslikdagi ahamiyati**.**Matematik statistika. Matematik statistikaning ma’lumotlar bazasini yaratishdan iborat.

Til o‘rgatish jarayonini avtomatlashtirish. Til o‘rgatuvchi avtomatlar uchun nutqiy materiallarni tayyorlash, o‘quv dasturlarini yaratish. Bilimlarni baholash dasturlarini tuzish. Til o‘rgatish jarayoni uchun ma’lumotlar bazasini yaratish.

Kompyuter lingvistikasining avtomatik tahrir yo‘nalishi**.**Avtomatik tahrir yo‘nalishining vujudga kelishi va uning dunyo tilshunosligidagi rivoji.Avtomatik tahrir dasturlarini yaratish tamoyillari. Avtomatik tahrir uchun ma’lumotlar bazasini yaratish.

Kompyuter analizi va rus (ingliz) tilidagi gaplarni o‘zbek tiliga tarjima qilish algoritmi.Rus (ingliz) tilidagi gaplarni o‘zbek tiliga tarjima qilish dasturining algoritmi. Rus (ingliz) tilidagi gaplarni o‘zbek tiliga tarjima qilish dasturining ishlash tamoyili.

O‘zbek tilining axborot- kompyuter uslubini yaratish**.**Tilning axborot- kompyuter uslubi nazariyasi. O‘zbek tilining axborot- kompyuter uslubini yaratish mezonlari. Axborot-kompyuter uslubida sinonimlar, frazeologizmlar va polisemantik so‘zlarning mavqeyi.

“Kompyuter lingvistikasi” fanining rivojlanish istiqbollari**.** Lingvistik kompyuter dasturlarini takomillashtirish yo‘llari. Kompyuter tilshunosligining zamonaviy yo‘nalishlaridan biri hisoblanadi. Hozirgi kunda kompyuter lingvistikadi korpus, o‘zbek tili korpusi, o‘zbek milliy korpusi kabi terminlar uchraydi. O‘zbek tili korpusi bo‘yicha fiolologiya fanlari doktori Nilufar Abdurahmonova tomonidan ishlab chiqilgan “o‘zbekcorpus.uz” saytini misol qilishimiz mumkin.[[2]](#footnote-3)

  Fan oldiga yangi vazifalarning qo‘yilishi bilan tilshunoslik fani EHMlar yordamida tarjima jarayonini takomillashtirish bilan XX asrning 30-yillaridan boshlab shug’ullana boshladi. Keyingi yillarda avtomatik EHM larning yaratilishi munosabati bilan ilmiy va badiiy tarjima jarayonini mexanizasiyalash imkoniyatlari yanada kengaydi. Mashina yordamida biror matnni bir tildan ikkinchi bir tilga tarjima qilish uchun qo‘yidagilar zarurdir:

**1.**Maxsus tanlangan so‘zlar bilan birga matnni mexanik ravishda tarjima qilish imkoniyatini beruvchi qoidalar tizimini yaratish. Bu qoidalarni va so‘zlarni “mashina tili” da yozish, ya’ni usha qoidalarni reallashtiruvchi dastur ishlab chiqish.

  Tarjima qilish qoidalari faqat ikki til uchun zarur bo‘lib, bir umumiy yo‘nalishdagi rasmiy va ilmiy matnlarni, ba’zan badiiy matnlarni ham tarjima jarao‘nini EHMda amalga oshirish uchun quyidagi lingvistik ta’milot ishlarini amalga oshirish kerak:

  a) har ikki tildagi negizlar lug’atini tuzish;

  b) oborotlar (o‘ramlar), turg’un sintaktik qurilmalar lug’atini tuzish;

  d) old ko‘makchilar, so‘ng ko‘makchilarni tarjima qilish jadvallarini tuzish;

  e) omonimlarni bir-biridan farqlash qoidalarini ishlab chiqish;

  f) har ikki tildagi affikslar jadvalini ishlab chiqish;

  g) har ikki tildagi grammatik qoidalarni analiz va sintez qiluvchi qoidalar tizinini yaratish;

  h) mavjud tahlil qilingan qoidalarni sintezlovchi qoidalar tizimini ishlab chiqish va boshqalar.

         Bunday ma’lumotlar mashinaning xotirasiga kiritiladi, ana shundan keyin mashinada tarjima qilish jarayoni boshlanadi. Bu jarayon quyidalarni o‘z ichiga oladi:

         a) lug’atdan kerakli so‘zni qidiruvchi, uning turini ko‘rsatuchi belgisi bilan mashina xotirasining ishlovchi qismiga kiritiladi;

         b) aniqlangan oborot (o‘ram)larning bir-biriga muvofiq keltiruvchi usullarini ishlab chiqish;

         d) aniqlangan omonimlarni bir-biridan farqlash va tarjima qilish usullarini ishlab chiqish;

         e) analiz qilish, analiz qiluvchi qoidalar yordamida ma’lum izchillik asosida har bir so‘z turkumi doirasida tarjima qilinuvchi so‘zlar ro‘yxatini tuzish;

         f) olingan natijalar asosida sintezlash, ya’ni sintezlovchi qoidalar yordamida ruscha gap tuzish, inglizcha va o‘zbekcha gap tuzish dasturlarini mashina xotirasiga kiritish va hokazo.[[3]](#footnote-4)

         Tarjima qilish jarayoni amalga oshirilganda dastlabki ish **lug’atdan so‘z qidirishdan**boshlanadi, ya’ni **EHM**tarjima qilinadigan matndan so‘zlarni birin-ketin olib, ularning har biriga mos keladigan negizlarni, negizlar lug’atidan qidirib topadi.

Keyin ularni taqqoslaydi, harflar soni jihatidan maksimal darajada mos keluvchi so‘zra sig’adigan negizni qidirib topari va uni tarjima qiladi.

         Analiz qiluvchi qoidalar **EHM**da so‘z turkumlari bo‘yicha guruhlashadi. Har bir tilda analiz qilish fe’ldan boshlanadi, chunki so‘zlar orasida fe’l har qanday gapning ma’no kaliti sanaladi. Shuning uchun ko‘pgina tillarda fe’l ham analitik, ham sintetik xususiyatga ega bo‘ladi. Bunday ikki tomonlama, murakkab xususiyat o‘zbek tilidagi fe’llarga ham xosdir:

**Analitik xususiyat                       Sintetik xususiyat**

*Kulib yubordi                               kuldi, kulimsiradi*

*Kulib qo‘ydi                                   jilmaydi*

*Yig’latib qo‘ydi                                       yig’latdi*

*Topib keldi                                     topdi*

         Fellar o‘z shakllari orqali shaxs, son, zamon, mayl va nisbat xususiyatlarini ifodalaydi. Ularni tarjima jarayonida aniq ifodasi **EHM**ga kiritilayotgan lingvistik ta’minotning aniq bo‘lishiga bog’liq.

Xulosa o’rnida shuni aytish mumkinki hozirgi zamon tilshunosligi asosida tarjima jarayonini avtomatlashtirish bilan bog’liq muammolar kundan-kunga ko‘payib bormoqda. Kompyuter lingvistikasini korpussiz tasavvur qilish qiyin. Shuning uchun korpus, milliy korpus ishlab chiqilyapti. Yaqin yillarda bu amaliyotda qo’llaniladi.

**Foydalanilgan adabiyotlar ro‘yxati**

1. Nilufar Abduraxmonova «"O‘zbek tili elektron korpusining kompyuter modellari"» 2021.

2.Nilufar Abduraxmonova «Til pedagogikasida o‘zbek tili ta’limiy korpusining ahamiyati» 2020

3. Yo‘ldoshev B. O‘zbekkompyuter lingvistikasi muammolari haqida // O‘zbek tilshunosligi masalalari (ilmiy maqolalar to‘plami). – Samarqand: SamDU nashri, 2004. – B. 17-22.

4. NilufarAbduraxmonova «Kompyuterlingvistikasi» 2021

5.AbduraxmonovaN. Mashina tarjimasining lingvistik ta’minoti (monografiya) / Toshkent: Muharrir, 2018, 165 b.

1. Nilufar Abduraxmonova «"O‘zbek tili elektron korpusining kompyuter modellari"» 2021 [↑](#footnote-ref-2)
2. NilufarAbduraxmonova «Kompyuterlingvistikasi» 2021 [↑](#footnote-ref-3)
3. Nilufar Abduraxmonova «Til pedagogikasida o‘zbek tili ta’limiy korpusining ahamiyati» 2020 [↑](#footnote-ref-4)