O‘ZBEK TILI KORPUS LINGVISTIKASINI TAKOMILLASHTIRISHNING NAZARIY VA AMALIY MUAMMOLARI

Shermamatova Madina Obidjon qizi
“Behbudiy vorislari” xususiy maktabi
o‘qituvchisi, tadqiqotchi
https://doi.org/10.5281/zenodo.22938246

Annotatsiya. Mazkur maqolada o‘zbek tili korpus lingvistikasini takomillashtirishning nazariy va amaliy muammolari tahlil qilinadi. Unda o‘zbek tilining milliy elektron korpusini shakllantirish, korpus tarkibining reprezentativligi va muvozanatliligini ta’minlash, matnlarni elektron shaklda yig‘ish, qayta ishlash hamda lingvistik belgilash masalalariga alohida e’tibor qaratiladi. Shuningdek, o‘zbek tilining fonetik, leksik, morfologik, sintaktik va semantik xususiyatlarini korpusda aks ettirish bilan bog‘liq nazariy masalalar hamda korpusni avtomatik tahlil qilishda zamonaviy axborot texnologiyalari, tabiiy tilni qayta ishlash (NLP) va sun’iy intellekt imkoniyatlaridan foydalanish istiqbollari yoritiladi.

Kalit so‘zlar: o‘zbek tili, korpus lingvistikasi, milliy korpus, elektron korpus, lingvistik korpus, korpus ma’lumotlari, morfologik tahlil, sintaktik tahlil, NLP.

Hozirgi zamonaviy jamiyatda muloqot va axborot almashinuvi alohida ahamiyat kasb etib, inson hayotining deyarli barcha sohalari axborot muhitining bevosita ta’siri ostida rivojlanmoqda. Shu bilan birga, kompyuter texnologiyalarining jadal taraqqiyoti yangi davr talablariga moslashish va murakkab vazifalarni hal etishda muhim vositaga aylanib bormoqda. Ular kundalik hayotimizning barcha jabhalariga chuqur kirib borib, turli jarayonlarni avtomatlashtirish va samaradorlikni oshirish imkonini bermoqda. Globallashuv sharoitida inson tili bilan mashina tili o‘rtasidagi tafovutlar bir qator muammolarni yuzaga keltirmoqda. Inson tilining murakkab tuzilishi, ko‘p ma’noliligi va kontekstga bog‘liqligi ushbu masalani yanada dolzarb etadi. Shu sababli hisoblash lingvistikasi hamda tarjima lingvistikasi zamonaviy ilm-fanning muhim yo‘nalishlaridan biri sifatida e’tirof etiladi.

Hisoblash lingvistikasi tilshunoslik, kompyuter texnologiyalari va matematika fanlarining o‘zaro integratsiyasi asosida shakllangan fanlararo soha bo‘lib, u tilni avtomatik qayta ishlashning nazariy va amaliy jihatlarini o‘rganadi. Ushbu yo‘nalishda matematik modellar, algoritmlar va dasturiy vositalar muhim rol o‘ynaydi, chunki kompyuter tizimlarida til bilan ishlash aynan shunday vositalar orqali amalga oshiriladi. Mazkur fan nutq va matn shaklida namoyon bo‘ladigan tilning ikki tomonlama - og‘zaki va yozma ko‘rinishlarini o‘rganadi hamda ularni qayta ishlashga qaratilgan. Uning asosiy yo‘nalishlari qatoriga til muhandisligi, sun’iy intellekt, mashinaviy tarjima, axborot va kommunikatsiya texnologiyalari, qidiruv tizimlarini yaratish, katta hajmdagi matnlardan ma’lumot ajratib olish kabi sohalar kiradi.

Bundan tashqari, hisoblash lingvistikasi yordamida yirik ma’lumotlar bazalari va internet tarmoqlarida turli savollarga javob topish, matn va nutqni avtomatik qayta ishlash, ovozli buyruqlarni qabul qilish va bajarish, matnni nutqqa va nutqni matnga aylantirish, matn mazmunini tahlil qilish hamda samarali ma’lumotlar bazalarini yaratish imkoniyati yuzaga keladi. Shu jihatdan, ushbu soha zamonaviy axborot jamiyatining muhim ilmiy va amaliy asoslaridan biri hisoblanadi.

Tegishli fanlar majmui inson tili, inson tafakkuri va psixologiyasi, shuningdek, sun’iy kompyuter intellekti bilan bog‘liq murakkab masalalarni qamrab oladi. Shu sababli mazkur yo‘nalish turli ta’lim tizimlarida turlicha talqin qilinadi: ayrim hollarda u tilshunoslik tarkibida

qaralsa, boshqa hollarda kompyuter texnologiyalari va matematika fanlarining tarkibiy qismi sifatida o‘rganiladi. Zamonaviy texnologiyalar rivoji natijasida til bilan bog‘liq tadqiqotlar doirasi yanada kengaydi. Xususan, telefon, radio va boshqa elektron qurilmalar orqali nutqni yozib olish va uzatish imkoniyati paydo bo‘lishi bilan til aloqa muhandisligi sohasining ham muhim tadqiqot obyektiga aylandi. Bu esa nutq signalini qayta ishlash, uzatish va qabul qilish jarayonlarini ilmiy asosda o‘rganishga olib keldi.

Kompyuter texnologiyalarining jadal taraqqiyoti tabiiy tilni o‘rganish va qayta ishlash usullariga sezilarli ta’sir ko‘rsatdi. Bugungi kunda imlo va grammatik xatolarni aniqlash, matnlarni avtomatik tahrirlash, hujjatlar va ma’lumotlar bazalaridan kerakli axborotni izlab topish, shuningdek, bir tildan boshqasiga tarjima qilish kabi funksiyalarni bajaruvchi dasturiy vositalar keng tarqalgan. Bunday tizimlar global miqyosda keng qo‘llanilib, millionlab foydalanuvchilar tomonidan ishlatilmoqda. Shunga qaramay, hozirgi dasturiy vositalarning aksariyati hali to‘liq darajada mukammal sun’iy intellektga ega emas. Ular ko‘pincha oldindan belgilangan qoidalar va modellarga asoslanib ishlaydi, natijada ayrim murakkab lingvistik holatlarda xatolarga yo‘l qo‘yishi mumkin. Ayniqsa, tabiiy tillarning nozik semantik va pragmatik jihatlarini hisobga olishda, shuningdek, bir tilni boshqasiga to‘liq va aniq tarjima qilishda mavjud tizimlar hali ham cheklangan imkoniyatlarga ega.

Hisoblash tilshunosligining muhim yo‘nalishlaridan biri - tabiiy tilni to‘g‘ri aniqlash, avtomatik tarjima qilish hamda inson va mashina o‘rtasidagi muloqotni mukammallashtirishga xizmat qiluvchi dasturiy vositalarni ishlab chiqishdir. Ushbu yo‘nalish tilni chuqur tahlil qilish, uni kompyuter uchun tushunarli shaklga keltirish va samarali kommunikatsiya tizimlarini yaratishni o‘z ichiga oladi. Shu sababli kompyuter lingvistikasini o‘qitishda idrok xaritalari (mind map) kabi vizual usullardan foydalanish samarali natija beradi, chunki ular murakkab tushunchalarni tizimli ravishda anglashga yordam beradi. Kompyuter lingvistikasi sohasining shakllanganiga qariyb yarim asr vaqt o‘tgan bo‘lsa-da, bu yo‘nalishda barcha muammolar to‘liq hal etilgan deb bo‘lmaydi. Ayniqsa, tabiiy tilning murakkabligi, ko‘p ma’noliligi va kontekstga bog‘liqligi hali ham dolzarb ilmiy muammolarni keltirib chiqarmoqda.

Aslida, turli xalqlar o‘rtasida samarali muloqotni ta’minlash masalasi tarixan ham muhim bo‘lgan. Insoniyat doimo umumiy tushuniladigan universal til yaratishga intilgan. Masalan, o‘rta asrlarda xalqaro ilmiy va siyosiy muloqotlarda asosan lotin tili qo‘llanilgan va u o‘ziga xos universal aloqa vositasi vazifasini bajargan. Bugungi kunda esa texnologiyalar rivoji tufayli til to‘siqlarini bartaraf etishga qaratilgan ko‘plab dasturiy yechimlar yaratilgan. Hozirgi vaqtda aloqa muammolarini hal qilishga xizmat qiluvchi 600 dan ortiq turli dasturiy tizimlar mavjud bo‘lib, ular avtomatik tarjima, nutqni aniqlash va inson-mashina muloqotini rivojlantirishga xizmat qilmoqda. Shunga qaramay, universal va mukammal til tizimini yaratish masalasi hanuzgacha dolzarbligicha qolmoqda.

Tarixiy tilshunoslik nuqtai nazaridan ayrim tillarning keng hududlarda tarqalganini kuzatish mumkin. Jumladan, ayrim davrlarda muayyan til butun bir mintaqada hukmron mavqega ega bo‘lib, davlat boshqaruvi va rasmiy muloqot vositasi sifatida xizmat qilgan. Masalan, mo‘g‘ullar hukmronligi davrida va undan oldin ham ayrim tillar Hindiston hududida rasmiy aloqa vositasi sifatida qo‘llanilgan. Ushbu til boy lug‘at zaxirasi va ifoda imkoniyatlari bilan ajralib turib, nafaqat hukmdorlar, balki keng xalq ommasi orasida ham mashhurlikka erishgan. Vaqt o‘tishi bilan u o‘zining nutqiy jozibadorligi va ifoda qudrati tufayli millionlab insonlar tomonidan qabul qilinib, Tojikiston, Afg‘oniston hamda Markaziy Osiyoning ayrim hududlarida keng tarqalgan.

Bugungi kunda esa o‘zbek tilida hisoblash tilshunosligi imkoniyatlarini o‘rganish alohida ahamiyat kasb etmoqda. Bu jarayon nafaqat o‘zbek tilida so‘zlashuvchi jamiyatning kognitiv imkoniyatlarini kengaytiradi, balki boshqa tillarda yaratilgan ilmiy va amaliy bilimlarni o‘zbek

tiliga samarali tarzda o‘tkazish zaruratini ham yuzaga keltiradi. Hozirgi davrda ilmiy ma’lumotlarning katta qismi ingliz va rus tillarida yaratilishi hamda internet orqali tarqatilishi sababli, ularni o‘zbek tiliga moslashtirish muayyan qiyinchiliklar bilan bog‘liq.

Shu nuqtai nazardan, kompyuter lingvistikasi asosida yaratilgan tizimlar matn va so‘zlarni tarjima qilish, audio va video ma’lumotlarni qayta ishlash hamda turli tillar o‘rtasida axborot almashinuvini ta’minlashda muhim rol o‘ynaydi. Bu esa o‘zbek tilining zamonaviy axborot makonida o‘z o‘rnini mustahkamlashiga xizmat qiladi. Boshqa tomondan, elektron platformalarning kengayishi jamiyat hayotida yangi raqamli muhitni shakllantirdi. Til esa ijtimoiy institut sifatida ushbu muhitda asosiy kommunikativ vositaga aylandi. Ayniqsa, mobil qurilmalar, veb-saytlar va kompyuterlar orqali amalga oshirilayotgan matnga asoslangan muloqot shakllari keng ommalashdi. Bu jarayon natijasida lingvistik kompyuter ilovalari, jumladan, elektron ta’lim tizimlari (e-learning) rivojlanib, ta’lim sohasida yangi imkoniyatlar yaratdi.

Shu bilan birga, o‘zbek tilini elektron platformalar orqali o‘rgatish jarayonida muayyan muammolar ham mavjud. Tilning yozuv tizimi va grammatik xususiyatlari ba’zan raqamli muhitga to‘liq moslashishni qiyinlashtiradi. Tarixan til o‘rgatishda audio va vizual vositalardan – audiokassetalar, kompakt disklar va raqamli xotira qurilmalaridan keng foydalanilgan bo‘lsa-da, zamonaviy elektron ta’lim tizimlarida o‘zbek tilining to‘liq integratsiyasi hali yetarli darajada emas. Ilmiy manbalarga ko‘ra, XX asrning 50-yillaridan boshlab tabiiy tillarni kompyuter yordamida qayta ishlash yo‘nalishida jiddiy tadqiqotlar olib borila boshlagan. Dastlabki izlanishlar asosan mashinaviy tarjima muammosiga qaratilgan bo‘lib, bu holat o‘sha davrning siyosiy sharoiti bilan ham bevosita bog‘liq edi. Xususan, Sovuq urush davrida yirik davlatlar o‘rtasidagi raqobat sababli rus tilidan ingliz tiliga avtomatik tarjima tizimlarini yaratishga alohida e’tibor qaratilgan.

Hisoblash tilshunosligining shakllanishi va rivojlanishi matematika, muhandislik, axborot nazariyasi hamda radioaloqa kabi sohalarning taraqqiyoti bilan chambarchas bog‘liqdir. Ushbu fanlar tilni modellashtirish, algoritmlashtirish va avtomatik qayta ishlash imkoniyatlarini kengaytirib, mashinaviy tarjima tizimlarini yaratishda muhim nazariy va amaliy asos bo‘lib xizmat qilgan.

Hozirgi kunda ushbu yo‘nalishlar izchil rivojlanib, yanada takomillashib bormoqda hamda katta istiqbolga ega soha sifatida qaralmoqda. Zamonaviy kompyuter texnologiyalari til bilan bog‘liq ko‘plab murakkab masalalarni avtomatlashtirish va samarali hal etish imkonini bermoqda. Shu jihatdan, hisoblash tilshunosligi kompyuterlashtirish jarayonining muhim tarkibiy qismi bo‘lib, tilshunoslikdagi ko‘plab nazariy va amaliy muammolarga yechim topishda yangi imkoniyatlar yaratmoqda.

So‘nggi yillarda tilshunoslik va axborot texnologiyalari kesishgan nuqtada ilmiy adabiyotlarni avtomatik tarjima qilish, ma’lumotlarni saqlash, kodlash va uzatish bilan bog‘liq ko‘plab yangi ilmiy-texnik muammolar yuzaga kelmoqda. Bu jarayon hisoblash lingvistikasining dolzarbligini yanada oshirib, tilni avtomatik qayta ishlash sohasida chuqur tadqiqotlar olib borishni taqozo etmoqda.

Tadqiqotchilar ta’kidlashicha, mashinaviy tarjima g‘oyasi ilk bor sovet muhandisi va ixtirochisi P.P. Smirnov-Troyanskiy tomonidan ilgari surilgan bo‘lib, u 1933-yilda dastlabki mexanizatsiyalashgan lug‘atni yaratishga muvaffaq bo‘lgan. Keyinchalik, 1946-yilda amerikalik tilshunos A. Uiver ushbu g‘oyani yanada rivojlantirib, mashina tarjimasining nazariy asoslarini shakllantirgan.

Amaliy tajribalar esa 1950-yillardan boshlab olib borila boshladi. Xususan, 1954-yilda AQShda IBM-701 elektron hisoblash mashinasi yordamida rus tilidan ingliz tiliga tarjima qilish bo‘yicha dastlabki tajribalar o‘tkazildi. Ushbu tajribalarda taxminan 250 ta so‘zdan iborat lug‘at asosida texnik matnlar tarjima qilinib, mashinaviy tarjimaning amaliy imkoniyatlari namoyon etildi.

O‘zbek tili korpus lingvistikasini takomillashtirish zamonaviy tilshunoslikning muhim yo‘nalishlaridan biridir. Milliy elektron korpusni shakllantirish, uning tarkibini boyitish, matnlarni lingvistik annotatsiyalash va avtomatik tahlil qilish o‘zbek tilini chuqur hamda empirik o‘rganish imkonini beradi. Shu bilan birga, korpusning reprezentativligi, ma’lumotlarning aniqligi va qidiruv tizimlarini takomillashtirish muhim vazifalardan hisoblanadi. Zamonaviy axborot texnologiyalari, NLP va sun’iy intellekt imkoniyatlaridan samarali foydalanish o‘zbek tili korpus lingvistikasining rivojlanishiga xizmat qiladi.

FOYDALANILGAN ADABIYOTLAR

1. Muhamedova S. O‘zbek kompyuter lingvistikasi: rivojlanish tarixi, buguni va kelajagi // Til va adabiyot ta’limi, 2004. – 18 b.
2. Abduraxmonova N.Z. O‘zbek tili elektron korpusining kompyuter modellari (monografiya). – Toshkent, 2021. 202 b. (B. 45–67)
3. Raupova L., Elov B., Abjalova M., Alayev R. O‘zbek tilining ta'limiy korpusi
4. Sh.Hamroyeva. O‘zbek tili mualliflik korpusini tuzishning lingvistik asoslari. F.f.f.d…diss.av. – Qarshi, 2018. – 27 b.
5. B.Mengliyev. O‘zbek tili korpusshunosligining dolzarb masalalari. Toshkent, 2022.12 b.