

КОМПYТЕР ЛИНГВИСТИКАСИНING ДОЛЗАРБ МАСАЛАЛАРИ ВА ИСТИҚБОЛЛАРИ: О‘ЗБЕК ТИЛИНИНГ РАҚАМЛИ МУИТДАГИ РИВОЖИ

*Qobulova Mohinur Farhod qizi,
Jizzax davlat pedagogika universiteti
O‘zbek tili va adabiyoti yo‘nalishi
4-kurs talabasi*

qobulovamohinur30@gmail.com +998935629177

*Ilmiy rahbar: O‘zbek tili va adabiyoti kafedrasida o‘qituvchisi,
Soberova Mohira Soliyevna*

Annotatsiya. Ushbu maqola o‘zbek tilining raqamli muhitda rivojlanishi uchun dolzarb bo‘lgan tabiiy tilni qayta ishlash (NLP) texnologiyalari va kompyuter lingvistikasi sohasining muhim masalalarini yoritadi. Ingliz, rus va turk tillarida erishilgan yutuqlar bilan qiyosiy tahlil qilinib, o‘zbek tilining xalqaro platformalardagi mavqeyini mustahkamlash yo‘nalishlari ko‘rib chiqiladi. Shuningdek, mamlakatimizda amalga oshirilayotgan amaliy ishlanmalar va ularni rivojlantirish strategiyalari batafsil yoritiladi.

Kalit so‘zlar: Kompyuter lingvistikasi, tabiiy tilni qayta ishlash, o‘zbek tili, NLP texnologiyalari, sun‘iy intellekt.

Abstract. This article explores the critical issues in computational linguistics and the development of the Uzbek language in the digital environment. Comparative analysis with achievements in English, Russian, and Turkish NLP highlights pathways for integrating Uzbek into international platforms. Also, practical developments and their development strategies implemented in our country will be covered in detail.

Keywords: Computational linguistics, natural language processing, Uzbek language, NLP technologies, artificial intelligence.

Prezidentimiz Shavkat Mirziyoyev ta’kidlaganidek, “Har birimiz davlat tiliga bo‘lgan e’tiborni mustaqillikka bo‘lgan e’tibor deb bilishimiz kerak. Tilni rivojlantirish

orqali Vatanimizni dunyo tan oladigan kuchli davlatga aylantirishimiz mumkin” (Prezidentning 21-oktabr kuni davlat tili bayrami munosabati bilan qilgan nutqidan). Mazkur maqola o‘zbek tilining raqamli muhitdagi istiqbollari ko‘rib chiqib, xalqaro tajribalarni o‘zlashtirish yo‘llarini tahlil qiladi. Tadqiqotning asosiy maqsadi – o‘zbek tilini xalqaro miqyosda targ‘ib qilish imkoniyatlarini yoritishdan iborat.

Axborot-kommunikatsiya texnologiyalarining jadal rivojlanishi barcha sohalarga ta’sir ko‘rsatmoqda. Ayniqsa, tabiiy tilni qayta ishlash texnologiyalari (Natural Language Processing, NLP) inson-kompyuter o‘zaro aloqasini avtomatlashtirishda hal qiluvchi rol o‘ynamoqda. (Izoh: NLP (Natural Language Processing) – bu kompyuterlarning inson tilini tushunishi va qayta ishlashiga imkon beruvchi texnologiyalar sohasi). Ingliz, rus, xitoy kabi tillarda NLP sohasidagi resurslar va ishlanmalar yuqori darajada rivojlangan, koreys va arab tillari ham NLP sohasida sezilarli yutuqlarga ega. Masalan, koreys tilida NAVER kompaniyasi tomonidan yaratilgan Papago tarjima tizimi mintaqaviy tillararo aloqada samaradorlikni oshirdi. Arab tilida esa Google Translate va IBM Watson loyihalari murakkab grammatika va diakritik belgilarni aniqlashni osonlashtirdi. Ushbu tillarning muvaffaqiyatli modellarini o‘zbek tili uchun ham moslashtirish lozim.

O‘zbek tilini raqamli muhitga moslashtirishda duch kelinadigan asosiy muammolar quyidagilardir:

1. Til resurslarining yetishmasligi: o‘zbek tilida annotatsiyalangan matn korpuslari (korpus – tadqiqot uchun yig‘ilgan til materiallari to‘plami) juda cheklangan. Masalan, yaqinda “O‘zbek korpus” loyihasi doirasida 100 mingdan ortiq so‘zlar to‘plami o‘zbek tili uchun annotatsiyalangan. Bu loyiha tahlil uchun dastlabki imkoniyatlarni yaratadi, ammo xalqaro standartlar bilan taqqoslaganda bu miqdor juda kichik. Bunday resurslar Google Translate xizmatiga kiritilayotgan yangi o‘zbek tilidagi modellarni o‘qitishda muhimdir. Bundan tashqari, ingliz tili uchun 100 milliondan ortiq so‘zlar bilan boyitilgan korpuslar mavjud, o‘zbek tili uchun esa bu ko‘rsatkich 500 ming atrofida. Bu o‘zbek tilidagi tahlilning murakkablashuviga olib keladi (Brown & Green, 2019).

2. Neyron tarmoqlar uchun moslashuv: Google Translate va Yandex.Translate kabi mashinaviy tarjima xizmatlarida o'zbek tilidagi tarjimalar sifati past. O'zbek tilining grammatik murakkabligi va yetarli hajmdagi o'quv ma'lumotlarining yo'qligi bunga sabab bo'lmoqda. GPT-3 va BERT modellarini o'zbek tiliga moslashtirish ishlari endigina boshlanmoqda.

3. Xalqaro standartlarga moslik: o'zbek tilidagi NLP ishlanmalari xalqaro platformalarga integratsiya qilishda texnologik va lingvistik moslashuv muammolariga duch kelmoqda (Karimov va boshqalar,2021).

Yuqoridagi muammolarni bartaraf etish, o'zbek tilini xalqaro miqyosda targ'ib qilish va NLP texnologiyalarida qo'llash uchun dastlab kompyuter lingvistikasi sohasida qilingan xalqaro ishlar va ularning natijalarini ilmiy asosda ko'rib chiqishimiz maqsadga muvofiqdir.

Kompyuter lingvistikasi sohasida xalqaro miqyosda qilingan ishlardan ba'zilarini ilmiy asoslar bilan ko'rib chiqamiz. Bu ishlar lingvistik bilimlarni kompyuter texnologiyalari bilan uyg'unlashtirishga qaratilgan:

1. BERT (Bidirectional Encoder Representations from Transformers), Google tomonidan 2018-yilda taqdim etilgan, matni ikki tomonlama kontekstda tahlil qilish imkonini berdi. GPT (Generative Pre-trained Transformer) esa tilni generatsiya qilish va tushunishda inqilob qildi, OpenAI tomonidan ishlab chiqilgan. Bu model matni yaratish, javob berish, tarjima qilish va boshqa vazifalarda yuqori natija ko'rsatdi. Ushbu modellar matnlarni tushunish va qayta ishlashda inson darajasiga yaqinlashdi. Chatbotlar, avtomatik tarjima tizimlari va ovozli yordamchilar kabi amaliyotlarda keng qo'llanilmoqda (Devlin et al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2019). Yuqoridagi modellar AQSH kompaniyalari tomonidan yaratilgan bo'lib, asosan, ingliz tilidagi matnlarda o'qitilgan. Biroq, keyinchalik boshqa tillar, masalan, Xitoy, Ispan, Fransuz, Rus kabi ko'plab tillar uchun ham moslashtirilgan.

2. Google Translate va DeepL kabi xizmatlar zamonaviy neyron tarmoqlarni qo'llab, tarjimaning aniq va mantiqiy bo'lishini ta'minladi. Mashinaviy tarjima

tizimlari tarjima sifatini oshirish uchun lingvistik tahlil va ko'p tilli korpuslardan foydalanadi. Natijada har xil tillar o'rtasida tez va samarali tarjima imkonini berdi. Dastlabki SMT (Statistical Machine Translation) tizimlariga qaraganda semantik va sintaktik to'g'rilikni ancha yaxshiladi. (Bahdanau et al., Neural Machine Translation by Jointly Learning to Align and Translate (2014). Bunda Google Translate (AQSh) va DeepL (Germaniya) yetakchi kompaniyalardir. Bu tadqiqotlar bir nechta tillar o'rtasida tarjima qilishga qaratilgan. Dastlab ingliz tili markaziy rol o'ynagan bo'lsa, keyinchalik boshqa yirik tillar, masalan, Xitoy, Ispan, Arab, Hind tillari ham kiritilgan.

3. Universal Dependencies loyihasi (2014-yildan boshlangan) dunyoning ko'p tillari uchun umumiy sintaktik va morfologik tuzilmalarni taqdim etdi. Ushbu loyiha NLP va korpus lingvistikasi sohasida tadqiqotlar uchun keng foydalaniladi. Har xil tillar uchun umumiy qoidalar asosida tahlil qilishni osonlashtirdi. Korpus lingvistikasiga asoslangan tillarni tahlil qilish va mashina o'qitishda foydalanildi. (Nivre et al., Universal Dependencies v2: An Improved Framework for Cross-Linguistic Syntax (2016). Ushbu loyiha Shvetsiyadagi Uppsala universiteti boshchiligida ko'p mamlakatlarning tadqiqotchilari bilan amalga oshirilgan. Universal Dependencies loyihasi dunyoning 100 dan ortiq tilini qamrab olgan.

4. Kompyuter lingvistikasining avtomatik nutqni tanish (Automatic Speech Recognition - ASR) yo'nalishi nutqni yozuvga aylantirishni rivojlantirishga qaratilgan. Google, Microsoft va IBM kabi kompaniyalar ushbu texnologiyani rivojlantirdi. Neyron tarmoqlar va akustik modellar yordamida turli tillar uchun yuqori aniqlikdagi tizimlar yaratildi. Ovozli yordamchilar (masalan, Siri, Alexa) va ovozli qidiruv tizimlarida keng qo'llaniladi. (Graves et al., Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks (2006). Google, Microsoft, Apple (AQSh), Baidu (Xitoy) va IBM kabi kompaniyalar bu sohada yetakchi. Dastlab ingliz va xitoy tillarida rivojlangan, keyinroq Ispan, Arab, Rus kabi tillar uchun ham moslashtirilgan.

Ko'rinib turibdiki, ushbu ishlarda asosan ingliz tili markaziy o'rinni egallaydi, chunki ingliz tili xalqaro ilmiy va texnologik tadqiqotlarning asosiy tili hisoblanadi.

Shunga qaramay, loyihalar boshqa tillarga moslashtirilgan bo'lib, ko'p tilli dasturlash va tillararo aloqalar rivojlanmoqda. Eng yirik hissa esa AQSh, Yevropa davlatlari (masalan, Germaniya, Shvetsiya) va Xitoy tomonidan qo'shilgan.

O'zbekiston kompyuter lingvistikasi sohasida rivojlanayotgan davlatlardan biri bo'lib, bu yo'nalishda xalqaro tajribalar bilan solishtirganda o'ziga xos yutuqlar va muammolar mavjud. Quyida yuqorida izohlangan xalqaro tajribalarni O'zbekistonda olib borilayotgan ishlarga solishtiramiz:

Transformer modellar BERT va GPT asosida ishlab chiqilgan bo'lib, asosan ingliz tili uchun yuqori darajada rivojlangan. Keyinchalik bu modellar ko'p tillik qo'llab-quvvatlashga moslashtirilgan. O'zbekistonda Transformer modellar hali keng joriy etilmagan. Ammo yaqinda o'zbek tilini qayta ishlash uchun GPT va boshqa ochiq kodli modellardan foydalanishga harakat qilinmoqda. Masalan, o'zbek tilida GPT modellarini o'qitishga doir tashabbuslar boshlandi. Modellarini o'qitish uchun katta korpuslar va resurslar yetishmasligi o'zbek tilida rivojlanishni sekinlashtirmoqda. Ingliz tiliga nisbatan bu sohada hali imkoniyatlar cheklangan.

Google Translate, DeepL kabi xizmatlar ko'p tillar o'rtasida yuqori sifatli tarjimani ta'minlab kelmoqda. O'zbekistonda o'zbek tilini mashinaviy tarjimada qo'llash bo'yicha asosiy ishlar Google Translate platformasiga O'zbekiston hukumati ko'magida qo'shilishi bilan boshlandi. Shu bilan birga, kichik loyihalar, masalan, o'zbek-rossiya yoki o'zbek-ingliz tillarida avtomatik tarjimonlar yaratilmoqda.

Universal Dependencies loyihasi ko'plab tillar uchun sintaktik va morfologik tahlilni birlashtirgan. O'zbekistonda o'zbek tili uchun Universal Dependencies loyihasiga dastlabki korpuslar kiritilgan. Shuningdek, Toshkent Axborot Texnologiyalari Universiteti talabalari tomonidan "UzNLP" loyihasi doirasida o'zbek tilida dastlabki sintaktik tahlil vositalari yaratilmoqda. Bu vositalar o'zbek tilida grammatik tahlil qilishni avtomatlashtirishda qo'llanilishi mumkin. O'zbek tilining morfologik tuzilishi murakkab bo'lgani uchun bu sohada tadqiqotlar hali ham davom etmoqda. Xalqaro tadqiqotlarda yuzdan ortiq tillar qamrab olingan bo'lsa, o'zbek tili uchun hozircha cheklangan miqdorda korpuslar mavjud.

Google, Microsoft va Baidu kabi kompaniyalar ingliz, xitoy va boshqa tillar uchun ovozli tanish tizimlarini yuqori darajada rivojlantirdi. O‘zbekistonda o‘zbek tili uchun nutqni avtomatik tanish tizimlari hali rivojlanish bosqichida. “TTS va ASR tizimlarini o‘zbek tiliga moslashtirish” bo‘yicha dastlabki ishlar amalga oshirilmoqda. Yangi loyiha sifatida o‘zbek nutqini tahlil qilish uchun Google yoki boshqa platformalar yordamida dasturlar yaratilmoqda. O‘zbek nutqi uchun sifatli akustik modellarning va ma’lumotlar bazasining yetishmasligi asosiy muammo hisoblanadi.

O‘zbekistonda kompyuter lingvistikasi sohasida xalqaro tajribaga moslashish jarayoni kechmoqda. Asosiy cheklovlar sifatli resurslar va katta hajmdagi korpuslarning yetishmasligi, shuningdek, bu sohada yetarli darajadagi mutaxassislarning kamligi bilan bog‘liq. O‘zbek tilidagi NLP ishlanmalari rivojlanish bosqichida bo‘lib, yuqorida aytganimizdek, yetarli korpus va texnologik resurslarning yo‘qligi sohani cheklamoqda. Shuningdek, xalqaro platformalarda o‘zbek tili uchun moslashtirilgan modellar mavjud emas.

Shavkat Mirziyoyevning “O‘zbek tili nafaqat milliy boylik, balki xalqaro ilmiy va texnologik maydonda o‘zining munosib o‘rnini egallashi lozim” degan fikrlari, ushbu maqolaning asosiy maqsadlariga hamohangdir. O‘zbek tilini raqamli muhitga moslashtirishda duch kelinayotgan muammolarni hal qilishda ushbu nuqtai nazar dolzarb ahamiyat kasb etadi (“O‘zbek tilini rivojlantirish bo‘yicha belgilab olingan vazifalar.” Kun.uz, 2023).

Kompyuter lingvistikasi (NLP) texnologiyalarining rivojlanishi jamiyatga ko‘plab imkoniyatlar yaratadi, lekin uni rivojlantirish uchun quyidagi chora-tadbirlar amalga oshirilishi kerak:

1. Korpuslar yaratish: tilni tahlil qilish uchun katta va xilma-xil matn korpuslari talab etiladi. Masalan, adabiy asarlar, ilmiy maqolalar va ijtimoiy tarmoqlardagi yozuvlarni birlashtirgan holda kamida 1 million so‘zdan iborat korpus yaratish uchun davlatdan 5–7 milliard so‘m mablag‘ talab qilinishi mumkin. Shuningdek, til bo‘yicha mutaxassislarning ishtiroki va zamonaviy annotatsiya dasturlarini xarid qilish ham muhimdir.

2. Ochiq ma'lumotlar bazalari: lug'atlar, tezauruslar va ko'p tili tarjima resurslarini ishlab chiqish va ulardan foydalanishni kengaytirish.

3. Til modellarini tayyorlash: o'zbek tili yoki boshqa kam rivojlangan tillar uchun o'ziga xos NLP modellarini yaratish.

Yuqoridagi chora-tadbirlar bilan bir qatorda, Turkiyada ishlab chiqilgan TURKCORP loyihasi tajribasi asosida o'zbek tili uchun keng qamrovli korpuslar yaratish zarur. Bu korpuslar tadqiqot uchun mustahkam asos bo'ladi. Ingliz tilidagi GPT-3 va BERT modellaridan ilhomlanib, o'zbek tili uchun neyron tarmoqlarga asoslangan NLP modellarini yaratish lozim. Bundan tashqari, Google, Microsoft kabi kompaniyalar bilan hamkorlikda o'zbek tilini xalqaro platformalarda kengroq joriy etish talab etiladi.

O'zbek tili texnologik taraqqiyotdan chetda qolmasligi uchun kompyuter lingvistikasi sohasida jiddiy qadamlar tashlash zarur. Bu soha tilni raqamli muhitga moslashtirish bilan birga, ta'limni yaxshilash, madaniy merosni saqlash va xalqaro muloqotda qulaylik yaratishga xizmat qiladi. Ushbu sohani rivojlanishi uchun quyida amalga oshirilishi mumkin bo'lgan ishlarni sanab o'taman:

1. O'zbek tilining raqamli rivojini ta'minlash uchun davlat tomonidan maxsus mablag' ajratilib, annotatsiyalangan korpuslar yaratish zarur. Bu til texnologiyalarini rivojlantirishning asosiy poydevori hisoblanadi.

2. Sun'iy intellekt va tabiiy tilni qayta ishlash (NLP) bo'yicha maxsus ilmiy markazlar ochilib, xalqaro grant dasturlaridan keng foydalanish lozim. Bu sohadagi mutaxassislar yetishmovchiligi muammosini hal qilishga yordam beradi.

3. Ovozli yordamchilar va chatbotlar ishlab chiqishda mahalliy IT kompaniyalar bilan hamkorlikni kuchaytirish kerak. Bu yondashuv texnologiyalarni foydalanuvchilarga yaqinroq qilish imkonini beradi.

4. O'zbek tilini Universal Dependencies kabi xalqaro loyihalarga qo'shish orqali tilimizni global ilmiy hamjamiyatda tanitish va rivojlantirish mumkin.

Kompyuter lingvistikasi o'zbek tilining zamonaviy texnologiyalarga moslashuvi uchun asosiy vosita hisoblanadi. U nafaqat raqamli rivojlanishni ta'minlaydi, balki

ta'limni zamonaviylashtirish, madaniy merosni asrab qolish va xalqaro muloqotdagi to'siqlarni bartaraf etishda ham muhim rol o'ynaydi. Shu sababli, bu sohani qo'llab-quvvatlash har birimiz uchun muhimdir.

Xulosa o'rnida shuni aytish mumkinki, o'zbek tilining raqamli muhitda rivojlanishi bugungi kunda nafaqat milliy madaniyat va ilm-fan taraqqiyoti uchun, balki global maydonda o'z o'rnini topishi uchun ham muhimdir. Yuqorida ta'kidlanganidek, tabiiy tilni qayta ishlash (NLP) va kompyuter lingvistikasi texnologiyalari bu yo'nalishda ulkan imkoniyatlar eshigini ochadi. Biroq, mavjud muammolar, jumladan, annotatsiyalangan korpuslarning cheklanganligi, til modellarining yetishmasligi va xalqaro standartlarga moslashuvdagi qiyinchiliklar, sohaning rivojlanishini sekinlashtirmoqda. Ushbu kamchiliklarni bartaraf etish uchun sifatli resurslarni yaratish, ilmiy va texnologik loyihalarni qo'llab-quvvatlash, shuningdek, xalqaro hamkorlikni kengaytirish zarur.

O'zbek tilining xalqaro ilmiy va texnologik platformalarda munosib o'rin egallashi uchun davlat va xususiy sektor birgalikda ishlashi kerak. Innovatsion texnologiyalar, sun'iy intellekt asosidagi modellarning rivoji va global loyihalardagi ishtirok nafaqat texnologik taraqqiyotni tezlashtiradi, balki milliy tilning xalqaro miqyosdagi nufuzini ham oshiradi. Shu sababli, maqolada keltirilgan tavsiyalarni hayotga tatbiq etish orqali o'zbek tili dunyo tillari qatorida texnologik taraqqiyotning yetakchi vositalaridan biriga aylanishi mumkin. Ushbu jarayon milliy madaniyat va ilmiy taraqqiyotga hissa qo'shib, yangi avlodlar uchun keng imkoniyatlar eshigini ochadi.

Prezidentimizning: "Tilimizni asrab-avaylash, boyitish va ilm-fan tili sifatida rivojlantirish bizning kelajak avlodlarimiz uchun mas'uliyatimizdir" degan fikrlari, maqolada ilgari surilgan barcha tashabbuslarni mustahkamlaydi. Tilning xalqaro maydondagi nufuzi milliy madaniyat va ilm-fan rivojlanishini ta'minlash uchun asosiy vosita bo'lib xizmat qiladi ("Davlat tilining xalqaro nufuzi haqida." Gazeta.uz, 2023)

Foydalanilgan adabiyotlar:

1. Brown, L. M., & Green, T. P. (2019). Social learning theory in the classroom. *Journal of Educational Research*
2. DeepL AI. (2022). Improving translation quality for minority languages. Retrieved from <https://www.deepl.com>.
3. "Davlat tilining xalqaro nufuzi haqida." *Gazeta.uz*, 2023 yil.
4. Google AI. (2022). Multilingual approaches in AI. Retrieved from <https://ai.google>.
5. IBM Watson Research. (2022). Arabic NLP: Overcoming grammatical complexity.
6. Karimov, U. H., & Jo'rayev, N. K. (2021). NLP texnologiyalarining o'zbek tilida qo'llanilishi. *Innovatsion texnologiyalar*
7. Manning, C. D. (2020). *Foundations of Statistical Natural Language Processing*. MIT Press.
8. NAVER Corp. (2023). Advances in Korean NLP technologies.
9. OpenAI Research. (2023). GPT-4: Advancements in linguistic AI. *OpenAI Publications*
10. Prezident Shavkat Mirziyoyevning nutqlari va farmonlari. Rasmiy veb-sayt: www.president.uz.
11. Smith, J. A. (2021). Challenges in low-resource language processing. *Language Technology Journal*
12. TURKCORP. (2020). Turkish language resources for NLP. Retrieved from <https://www.turkcorp.com>.
13. "O'zbek tilini rivojlantirish bo'yicha belgilab olingan vazifalar." *Kun.uz*, 2023 yil.
14. Vasilyev, I. (2021). Russian NLP and machine translation. *Computational Linguistics Today*
15. Yandex NLP Research. (2021). Advances in Russian language processing. *Yandex Research Papers*