



O'ZBEK TILIDA KATTA TIL MODELLARINI YARATISHNING LINGVISTIK ASOSLARI

Dilyorjon Solidjonov,

Qo'qon davlat universiteti magistranti.

Email: dilyorjonsolidjonov@gmail.com

MAQOLA HAQIDA

Qabul qilindi: 9-fevral 2026-yil

Tasdiqlandi: 11-fevral 2026-yil

Jurnal soni: 17-B

Maqola raqami: 36

DOI: <https://doi.org/10.54613/ku.v17i.1467>

KALIT SO'ZLAR/ КЛЮЧЕВЫЕ

C/ОБА/ KEYWORDS

katta til modellari, o'zbek tili morfologiyasi, dependency parsing, Universal Dependencies, agglutinativ tillar, neyron tarmoqli modellar, sintaktik tahlil

ANNOTATSIYA

Ushbu ilmiy maqola o'zbek tilida katta til modellarini (LLM) yaratishning lingvistik asoslarini tizimli o'rganishga bag'ishlangan bo'lib, unda tilning morfologik, sintaktik va semantik xususiyatlarining neyron tarmoqli modellarni ishlab chiqishdagi ahamiyati tahlil qilingan. Tadqiqot davomida o'zbek tilining agglutinativ tabiati, kelishik kategoriyasi, frazeologik birliklarning transformatsion modellari va murakkab sintaktik konstruksiyalarni qayta ishlashning nazariy hamda amaliy jihatlarini yoritilgan. Shuningdek, zamonaviy tabiiy tilni qayta ishlash (NLP) texnologiyalari, jumladan Universal Dependencies (UD) formatiga mos keladigan morfologik belgilash modellari, dependency parsing usullari va neyron tarmoqli yondashuvlarning o'zbek tili korpuslariga tatbiqi masalalari ko'rib chiqilgan. Maqolada o'zbek tilining leksik tizimi, uning rus tili bilan qiyosiy tahlili va diniy matnlar kabi funksional uslublarning lingvistik xususiyatlari katta til modellarini yaratishdagi o'rni asoslangan. Tadqiqot natijalari shuni ko'rsatadiki, o'zbek tilining murakkab morfologik tuzilishi va erkin so'z tartibi katta til modellarini yaratishda gibridd yondashuvlarni, ya'ni qoidaga asoslangan va neyron tarmoqli usullarni uyg'unlashtirishni talab etadi.

Kirish. So'nggi yillarda sun'iy intellekt sohasida yuz berayotgan jadal taraqqiyot, xususan katta til modellarining (Large Language Models – LLM) rivojlanishi tabiiy tilni qayta ishlash texnologiyalarini tubdan o'zgartirdi. OpenAI'ning GPT seriyasi, Google'ning BERT va T5 modellari hamda boshqa ko'plab transformator arxitekturasiga asoslangan modellar ingliz tili va boshqa yirik tillar uchun yuqori samaradorlik ko'rsatayotgan bo'lsa-da, o'zbek tili kabi agglutinativ tillar uchun bunday modellarni yaratish bir qator murakkab lingvistik muammolarni hal etishni talab qiladi. O'zbek tilining o'ziga xos grammatik tuzilishi, boy morfologik tizimi va erkin so'z tartibi katta til modellarini ishlab chiqishda alohida yondashuvni zaruriyatga aylantiradi.

O'zbek tili turkiy tillar oilasiga mansub bo'lib, agglutinativ til sifatida so'z yasaliishi va so'z o'zgarishida ildizga qo'shimchalarni ketma-ket qo'shish orqali yangi ma'no va grammatik shakllarni hosil qiladi. Bu xususiyat tilning morfologik tahlilini murakkablashtiradi, chunki bir so'z shakli bir necha o'nlab qo'shimchalarni o'z ichiga olishi mumkin va ularning har biri ma'lum grammatik ma'noni ifodalaydi. Masalan, "kitoblarimizdagilardan" so'zi bir necha morfologik birliklardan tashkil topgan bo'lib, ularning har birini to'g'ri tahlil qilish va semantik ma'nosini aniqlash katta til modellarini uchun muhim vazifa hisoblanadi. Y.Xamrayevaning 2023-yilgi tadqiqotlarida [15;112-115] ta'kidlanganidek, o'zbek tilining leksik tizimi, uning rus tili bilan qiyosiy tahlili va leksik birliklarning ma'no hamda semantik o'zgarishlari katta til modellarini uchun muhim manba bo'lib xizmat qiladi.

O'zbek tilida katta til modellarini yaratishning dolzarbligi bir necha omillar bilan belgilanadi. Birinchidan, raqamli transformatsiya jarayonlari o'zbek tilidagi matnlarni avtomatik qayta ishlash, mashina tarjimai, matn generatsiyasi va boshqa NLP vazifalariga bo'lgan ehtiyojni keskin oshirdi. Ikkinchidan, o'zbek tilining internetdagi ishtiroki kun sayin ortib borayotgan bo'lsa-da, mavjud til modellarini asosan ingliz tili va boshqa yirik tillar uchun yaratilgan bo'lib, ularni o'zbek tiliga bevosita tatbiq etish samaradorligi pastligicha qolmoqda. Uchinchidan, o'zbek tilining ilmiy-texnik, publitsistik, rasmiy-idoraviy va diniy kabi funksional uslublarining o'ziga xos lingvistik xususiyatlarini hisobga oladigan modellarni yaratish dolzarb masalalardan biridir. S.Amonturdiyevaning 2023-yilgi tadqiqotlarida [1;45-48] diniy matnlar kabi funksional uslublarning lingvistik xususiyatlarini tahlil qilish tilning turli uslublarini to'g'ri ifodalash imkonini berishini alohida ta'kidlangan.

Katta til modellarini yaratishda lingvistik asoslarning o'rni beqiyosdir. Transformer arxitekturasiga asoslangan zamonaviy modellar o'z-o'zidan e'tibor mexanizmi (self-attention) yordamida matndagi uzoq masofali bog'lanishlarni aniqlash imkoniyatiga ega bo'lsa-da, ularning samarali ishlashi tilning morfologik, sintaktik va semantik xususiyatlarini chuqur tushunishga asoslanadi. Ayniqsa, agglutinativ tillar uchun morfologik tahlil muhim ahamiyat kasb etadi, chunki so'z shakllarining ko'p variantliligi va murakkabligi modellarning lug'at hajmini cheksiz kattalashtirishi mumkin. U.Salayev tomonidan 2024-yilda ishlab chiqilgan UzMorphAnalyser modeli [12;2-5] o'zbek tilidagi so'zlarni morfologik tahlil qilishda infleksion qo'shimchalardan foydalanish orqali

yuqori samaradorlik ko'rsatgan bo'lib, bu katta til modellari uchun muhim poydevor vazifasini o'tashi mumkin.

O'zbek tilining sintaktik tuzilishi ham katta til modellarini yaratishda alohida e'tiborni talab qiladi. Tilning SOV (subyekt-obyekt-predikat) so'z tartibi, erkin so'z tartibi imkoniyatlari va murakkab gap qurilmalari sintaktik parserlar yaratishda qiyinchilik tug'diradi. E.Boltayevich va boshqalarning 2025-yilgi tadqiqotlarida [4;440-442] ta'kidlanganidek, o'zbek tili uchun morfologik va sintaktik tahlillarni birlashtirgan gibridd yondashuvlar samaraliroq hisoblanadi. Ushbu yondashuvda qoidaga asoslangan usullar va neyron tarmoqli modellar uyg'unlashtirilib, tilning morfologik murakkabligi va erkin so'z tartibidan kelib chiqadigan noaniqliklarni bartaraf etish imkoniyati yaratiladi.

So'nggi yillarda o'zbek tili uchun Universal Dependencies (UD) formatiga mos keladigan morfologik belgilash modellarini yaratish borasida muhim qadamlar qo'yildi. E.Boltayevich va boshqalarning 2025-yilgi tadqiqotlarida [5;381-383] ishlab chiqilgan morfologik tegging modeli o'zbek tilidagi so'zlarning grammatik kategoriyalarini aniqlashda yuqori aniqlik ko'rsatib, bu kabi resurslar katta til modellarini o'qitish uchun zarur bo'lgan annotatsiyalangan korpuslarni yaratish imkonini beradi. Shuningdek, E.Adali va A.Xolmo'minovning 2025-yilgi tadqiqotlarida [2;446-448] UzTreebank sintaktik treebankini yaratish bo'yicha olib borilayotgan ishlar o'zbek tilining sintaktik xususiyatlarini chuqur o'rganish va ularni dependency parsing modellarida aks ettirish imkoniyatlarini kengaytirmoqda.

Frazeologik birliklarning katta til modellarida to'g'ri ifodalanishi ham muhim masalalardan biridir. R.Sayfullayeva va S.Ganiyevaning 2023-yilgi tadqiqotlarida [13;78-82] o'zbek tilidagi frazeologik birliklarning strukturaviy va semantik tuzilishi, ularning nutqiy kontekstdagi o'zgarishlarini tahlil qilgan holda, frazeologik transformantlar va ularning modellashtirish imkoniyatlarini o'rgangan. Frazeologik birliklar ko'pincha so'zma-so'z tarjima qilinmaydigan va kontekstga bog'liq ma'noga ega bo'lganligi sababli, ularni katta til modellarida to'g'ri ifodalash tilning tabiiyligini ta'minlashda muhim rol o'ynaydi.

Kelishik kategoriyasi o'zbek tilining muhim grammatik xususiyatlaridan biri bo'lib, uning to'g'ri modellashtirishi katta til modellarining aniqligini oshirishga xizmat qiladi. M.Qudratovning 2024-yilgi tadqiqotlarida [10;121-124] o'zbek tilida kelishik kategoriyasining o'rganilish masalasi va olib borilishini tahlil qilgan holda, kelishik shakllarining semantik va sintaktik vazifalarini chuqur yoritib bergan. Kelishik qo'shimchalari gapdagi so'zlar orasidagi munosabatlarni ifodalashda muhim rol o'ynaydi va ularning to'g'ri tahlil qilinishi dependency parsing jarayonining sifatini oshiradi.

Mazkur tadqiqotning asosiy maqsadi o'zbek tilida katta til modellarini yaratishning lingvistik asoslarini tizimli ravishda o'rganish, mavjud yondashuvlarni tahlil qilish va tilning morfologik, sintaktik hamda semantik xususiyatlarini hisobga oluvchi samarali modellashtirish tamoyillarini ishlab chiqishdan iborat. Tadqiqot doirasida o'zbek tilining morfologik tahlili, dependency parsing usullari, neyron tarmoqli

modellarning qo'llanilishi va mavjud korpus resurslarining imkoniyatlari kabi masalalar chuqur o'rganiladi.

Adabiyotlar sharihi. O'zbek tilida katta til modellarini yaratishning lingvistik asoslariga bag'ishlangan tadqiqotlarni bir necha yo'nalishlarga ajratish mumkin: morfologik tahlil, sintaktik tahlil, semantik tahlil va korpus lingvistikasi. Ushbu bo'limda mavzuga oid asosiy adabiyotlar tahlil qilinadi.

Morfologik tahlil sohasida olib borilgan tadqiqotlar o'zbek tilining agglutinativ tabiatini hisobga olgan holda, so'z shakllarini ildiz va qo'shimchalarga ajratish, ularning grammatik kategoriyalarini aniqlash masalalariga bag'ishlangan. U.Salayev tomonidan 2024-yilda taqdim etilgan UzMorphAnalyser modeli [12;5-9] o'zbek tilidagi so'zlarni morfologik tahlil qilishda infleksion qo'shimchalardan foydalanish orqali yuqori samaradorlik ko'rsatgan. Mazkur model o'zbek tilining 20 mingdan ortiq so'z shakllarini qamrab olgan bo'lib, u so'z yasaliishi va so'z o'zgarishi qoidalarini tizimli ravishda ifodalaydi. N.Abduraxmonova va boshqalarning 2022-yilgi tadqiqotlarida [3;62-65] ishlab chiqilgan MorphUz analizatori esa o'zbek tilidagi so'zlarning morfologik tarkibini aniqlashda qoidaga asoslangan yondashuvdan foydalanadi va bu analizator o'zbek tili korpuslarini yaratishda muhim vosita bo'lib xizmat qilmoqda.

E.Boltayevich va boshqalarning 2025-yilgi tadqiqotlarida [5;382-384] o'zbek tili uchun Universal Dependencies formatiga mos keladigan morfologik belgilash modeli ishlab chiqilgan. Ushbu model 15 mingdan ortiq jumalarni o'z ichiga olgan korpus asosida o'qitilgan bo'lib, unda 17 ta asosiy morfologik xususiyat va ularning qiymatlari belgilangan. Tadqiqot natijalari shuni ko'rsatadiki, UD formatiga mos keladigan morfologik tegging modellari o'zbek tili uchun xalqaro standartlarga mos keladigan til resurslarini yaratishda muhim ahamiyatga ega. K.Mirdjonovna va boshqalarning 2025-yilgi tadqiqotlarida [9;1598-1600] esa o'zbek tilining morfotaktik modellari va algoritmlarini ishlab chiqish bilan shug'ullangan bo'lib, ularning tadqiqotlari so'z shakllarining morfologik tarkibini aniqlashda qo'llaniladigan formal grammatikalar va ularning kompyuter modellashtirish imkoniyatlarini o'rganishga bag'ishlangan.

Sintaktik tahlil sohasida olib borilgan tadqiqotlar asosan dependency parsing usullarini o'zbek tiliga tatbiq etish masalalariga qaratilgan. B.Elov tomonidan 2025-yilda o'tkazilgan tadqiqotda [8;15-18] UzbBERT va XLM-R modellari asoslangan biaffine graph-based parsing samaradorligi tahlil qilingan. Mazkur tadqiqotda 20 ming jumladan iborat korpusda o'tkazilgan sinovlar natijasida parser 91.8% F1 ballini ko'rsatgan va bu avvalgi usullarga nisbatan 27% ga kam xatolikni tashkil etgan. Ushbu natijalar o'zbek tilining murakkab sintaktik tuzilmalarini tahlil qilishda zamonaviy neyron tarmoqli parserlarning yuqori samaradorligini ko'rsatadi. E.Adali va A.Xolmo'minovna 2025-yilgi tadqiqotlarida [2;447-449] UzTreebank sintaktik treebankini yaratish bo'yicha olib borilgan tadqiqot esa o'zbek tilining sintaktik xususiyatlarini tizimli ravishda o'rganish va ularni dependency formatida ifodalash imkoniyatlarini kengaytirgan.

M.Sharipov va boshqalarning 2025-yilgi tadqiqotlarida [14;595-597] ishlab chiqilgan qoidaga asoslangan model o'zbek tilidagi matnlarda predikatni aniqlash masalasiga bag'ishlangan. Ushbu model tilning morfologik va sintaktik qoidalariga asoslangan holda gapdagi kesim vazifasini bajaruvchi so'zlarni aniqlaydi va bu kabi qoidaga asoslangan yondashuvlar, ayniqsa, annotatsiyalangan korpuslar yetarli bo'lmagan hollarda muhim ahamiyat kasb etadi. D.Mengliev va boshqalarning 2023-yilgi tadqiqotlarida [7;1911-1913] taklif etilgan qoidaga asoslangan sintaktik tahlil usuli esa ma'lumotlarni yetishmovchiligi sharoitida muqobil yondashuv sifatida o'zbek tilining sintaktik tuzilmalarini tahlil qilishda qo'llanilgan.

O'zbek tilining leksik va semantik xususiyatlarini o'rganishga bag'ishlangan tadqiqotlar ham katta til modellarini yaratishda muhim nazariy asos bo'lib xizmat qiladi. Y.Xamrayevaning 2023-yilgi tadqiqotlarida [15;113-116] o'zbek va rus tillarining leksik tizimlari qiyosiy tahlil qilingan bo'lib, unda tillararo leksik birliklarning ma'no va semantik o'zgarishlari, o'zlashtirilgan so'zlarning o'zbek tilidagi adaptatsiyasi masalalari yoritilgan. R.Sayfullayeva va S.Ganiyevaning 2023-yilgi tadqiqotlarida [13;79-83] esa o'zbek tilidagi frazeologik birliklarning strukturaviy va semantik tuzilishi, ularning nutqiy kontekstdagi o'zgarishlari va transformatsion modellarini tahlil qilgan. Frazeologik birliklarni modellashtirish, ayniqsa, katta til modellarining tabiiy va idiomatik matnlar yaratish qobiliyatini oshirishda muhim ahamiyatga ega.

M.Qudratovani 2024-yilgi tadqiqotlarida [10;122-125] o'zbek tilida kelishik kategoriyasining o'rganilishi masalasiga bag'ishlangan tadqiqotda kelishik shakllarining semantik va sintaktik vazifalari, ularning gapdagi so'zlar orasidagi munosabatlarni ifodalashdagi o'rni chuqur tahlil qilingan. Kelishik kategoriyasining to'g'ri modellashtirilishi, ayniqsa, dependency parsing jarayonida so'zlar

orasidagi bog'lanishlarni aniqlashda muhim rol o'ynaydi. S.Amonturdiyevaning 2023-yilgi tadqiqotlarida [1;46-49] diniy matnlarning lingvistik xususiyatlari, ularning boshqa funksional uslubdagi matnlar bilan o'xshashlik va farqlari tahlil qilingan bo'lib, bu kabi tadqiqotlar o'zbek tilining turli uslublarini ifodalovchi modellarni yaratishda muhim ahamiyatga ega.

D.Xamdamovaning 2025-yilgi tadqiqotlarida [6;47-50] o'zbek tilida sintaktik belgilash muammolari tahlil qilingan bo'lib, unda tilning agglutinativ tabiatini, erkin so'z tartibi va morfologik noaniqliklar kabi omillarning sintaktik markup jarayonini qanday murakkablashtirishi ko'rsatilgan. Ushbu tadqiqotda muallif mavjud muammolarni hal etish uchun gibrid yondashuvlar, xususan qoidaga asoslangan va statistik usullarni uyg'unlashtirish zarurligini ta'kidlaydi.

Umuman olganda, adabiyotlar tahlili shuni ko'rsatadiki, o'zbek tilida katta til modellarini yaratish bo'yicha tadqiqotlar so'nggi yillarda jadallashgan bo'lsa-da, hali ko'plab hal qilinmagan masalalar mavjud. Jumladan, annotatsiyalangan korpuslarning yetarli emasligi, morfologik noaniqliklarni bartaraf etish usullarining takomillashtirilmaganligi, murakkab sintaktik tuzilmalarni tahlil qilishda neyron tarmoqli modellarning cheklovlari kabi muammolar o'z yechimini kutmoqda. Shu bilan birga, mavjud tadqiqotlar o'zbek tilining lingvistik xususiyatlarini chuqur tushunish va ularni kompyuter modellarida ifodalash imkoniyatlarini kengaytirishda muhim poydevor bo'lib xizmat qilmoqda.

Metodologiya. Ushbu tadqiqotda o'zbek tilida katta til modellarini yaratishning lingvistik asoslarini o'rganish uchun kompleks metodologik yondashuv qo'llanildi. Tadqiqot metodologiyasi bir necha bosqichlarni o'z ichiga olgan bo'lib, ular orasida nazariy tahlil, korpus asosidagi empirik tadqiqot, eksperimental modellashtirish va qiyosiy tahlil usullari mavjud.

Tadqiqotning birinchi bosqichida o'zbek tilining morfologik, sintaktik va semantik xususiyatlariga oid mavjud nazariy adabiyotlar tizimli tahlil qilindi. Bu bosqichda Y.Xamrayeva [15], R.Sayfullayeva va S.Ganiyeva [13], M.Qudratova [10] va S.Amonturdiyeva [1] kabi mualliflarning lingvistik tadqiqotlari asosida o'zbek tilining leksik tizimi, frazeologik birliklari, kelishik kategoriyasi va funksional uslublarining o'ziga xos xususiyatlari o'rganildi. Nazariy tahlil natijasida o'zbek tilining katta til modellarini yaratishda e'tiborga olinishi zarur bo'lgan asosiy lingvistik parametrlar aniqlandi va ularning taksonomiyasi ishlab chiqildi.

Tadqiqotning ikkinchi bosqichida o'zbek tilining morfologik va sintaktik tahliliga oid mavjud modellar va algoritmlar tahlil qilindi. Bu bosqichda U.Salayev [12] tomonidan ishlab chiqilgan UzMorphAnalyser, N.Abduraxmonova va boshqalar [3] tomonidan yaratilgan MorphUz, E.Boltayevich va boshqalar [5] tomonidan taklif etilgan UD formatidagi morfologik belgilash modeli, B.Elov [8] tomonidan ishlab chiqilgan dependency parser va E.Adali va A.Xolmo'minovna [2] tomonidan yaratilayotgan UzTreebank kabi resurslar va modellar chuqur o'rganildi. Ushbu modellarning arxitekturasini, o'qitish usullari, qo'llanilgan korpuslar va erishilgan natijalar tahlil qilindi hamda ularning kuchli va zaif tomonlari aniqlandi.

Tadqiqotning uchinchi bosqichida o'zbek tilidagi matnlar korpusi asosida empirik tahlil o'tkazildi. Buning uchun o'zbek tilining turli funksional uslublariga mansub bo'lgan 50 ming jumladan iborat korpus shakllantirildi. Korpus tarkibiga badiiy adabiyot namunalari (30%), publitsistik matnlar (25%), ilmiy-texnik matnlar (20%), rasmiy-idoraviy hujjatlar (15%) va diniy matnlar (10%) kiritildi. Korpusdagi jumalarning o'rtacha uzunligi 12.4 so'zni tashkil etib, eng qisqa jumla 2 so'zdan, eng uzun jumla esa 47 so'zdan iborat bo'ldi. Korpusni shakllantirishda matnlarning janr va uslub jihatidan xilma-xilligiga, shuningdek ularning zamonaviyligiga (asosan 2010-2025 yillar oralig'idagi matnlar) alohida e'tibor qaratildi.

Korpusdagi matnlar avtomatik ravishda morfologik va sintaktik tahlildan o'tkazildi. Morfologik tahlil uchun U.Salayev [12] tomonidan ishlab chiqilgan UzMorphAnalyser modelidan foydalanildi. Ushbu model o'zbek tilidagi so'zlarni ildiz va qo'shimchalarga ajratish, so'z turkumini aniqlash, grammatik kategoriyalarni (kelishik, shaxs-son, zamon, nisbat va h.k.) belgilash imkonini beradi. Sintaktik tahlil uchun esa B.Elov [8] tomonidan ishlab chiqilgan UzbBERT asosidagi biaffine graph-based parser qo'llanildi. Mazkur parser jumalardagi so'zlar orasidagi dependency munosabatlarni aniqlaydi va ularni Universal Dependencies formatida ifodalaydi.

Tadqiqotning to'rtinchi bosqichida olingan natijalar statistik va qiyosiy tahlil qilindi. Morfologik tahlil natijasida so'z shakllarining morfologik tarkibi, qo'shimchalarning qo'llanish chastotasi, so'z turkumlarining taqsimoti kabi parametrlar bo'yicha statistik ma'lumotlar to'plandi. Sintaktik tahlil natijasida esa gap turlari, so'z tartibi modellari, dependency munosabatlarining turlari va chastotasi, murakkab gaplarning tuzilishi kabi parametrlar tahlil qilindi. Qiyosiy tahlil doirasida o'zbek tilining morfologik xususiyatlari boshqa turkiy tillar

(qozoq, turk, ozarbayjon) va agglyutativ tillar (fin, venger, koreys) bilan taqqoslandi.

Tadqiqotning beshinchi bosqichida o'zbek tili uchun katta til modellarini yaratishda qo'llanilishi mumkin bo'lgan lingvistik asoslangan yondashuvlar ishlab chiqildi. Bu bosqichda nazariy tahlil, empirik tadqiqot natijalari va mavjud modellarning tahlili asosida o'zbek tilining morfologik, sintaktik va semantik xususiyatlarini hisobga oluvchi modellashtirish tamoyillari shakllantirildi. Xususan, agglyutativ morfologiyani ifodalash uchun subword tokenizatsiya usullari, erkin so'z tartibini modellashtirish uchun position encoding mexanizmlari, frazeologik birliklarni aniqlash uchun esa kontekstga asoslangan yondashuvlar taklif etildi.

Tadqiqotda qo'llanilgan metodologik yondashuvlarning ishonchligini ta'minlash maqsadida bir qator choralar ko'rildi. Birinchidan, korpusni shakllantirishda matnlarning autentikligi va reprezentativligiga alohida e'tibor qaratildi. Ikkinchidan, morfologik va sintaktik tahlil natijalarining to'g'riligini tekshirish uchun qo'lda annotatsiyalangan 1000 jumladan iborat test to'plami tuzildi va tahlil natijalari ushbu test to'plami bilan solishtirildi. Uchinchidan, statistik tahlillarda ishonchlik intervallari va p-qiymatlari hisoblab chiqildi. To'rtinchidan, olingan natijalar mavjud adabiyotlardagi ma'lumotlar bilan solishtirildi va izohlandi.

Natijalar. Ushbu bo'limda o'zbek tilidagi matnlar korpusi asosida o'tkazilgan morfologik va sintaktik tahlillarning natijalari keltiriladi. Tadqiqot davomida to'plangan ma'lumotlar asosida o'zbek tilining katta til modellarini yaratishda e'tiborga olinishi zarur bo'lgan asosiy lingvistik xususiyatlari aniqlangan va ularning miqdoriy ko'rsatkichlari tahlil qilingan.

Morfologik tahlil natijalari shuni ko'rsatadiki, o'zbek tili so'z shakllarining morfologik murakkabligi yuqori darajada ekan. Korpusdagi

Morfologik kategoriya	Qo'shimcha turlari soni	Qo'llanish chastotasi (foizda)	Eng ko'p qo'llaniladigan qo'shimchalar
Kelishik qo'shimchalari	6	24.5	-ni (18%), -ning (12%), -ga (14%)
Ko'plik qo'shimchalari	2	15.8	-lar (100%)
Egalik qo'shimchalari	8	18.2	-im (25%), -ing (22%), -i (30%)
Fe'l zamon qo'shimchalari	12	22.4	-di (28%), -gan (24%), -yapti (18%)
Fe'l nisbat qo'shimchalari	5	8.6	-l (35%), -n (28%), -sh (22%)
So'z yasovchi qo'shimchalar	45	10.5	-lik (28%), -chi (22%), -la (35%)

1-jadval. O'zbek tilida morfologik qatlamlar va qo'shimchalarning qo'llanish chastotasi

Sintaktik tahlil natijalariga ko'ra, o'zbek tilida jumalarning tuzilishi va so'z tartibi modellari xilma-xilligi aniqlangan. Korpusdagi jami 50 ming jumla tahlil qilingan bo'lib, ularning 68 foizi soddaga gaplar, 32 foizi esa qo'shma gaplardan iborat ekan. Qo'shma gaplarning 58 foizi ergash gapli qo'shma gaplar, 42 foizi esa bog'langan qo'shma gaplardir. So'z tartibi bo'yicha tahlil qilinganda, SOV (subyekt-obyekt-predikat) tartibi 72 foiz jumalalarda kuzatilgan. SVO (subyekt-predikat-obyekt) tartibi 18 foiz jumalalarda, OSV (obyekt-subyekt-predikat) tartibi esa 6 foiz jumalalarda qayd etilgan. Qolgan 4 foiz jumalalarda so'z tartibi erkin o'zgargan.

Dependency munosabatlar tahlili shuni ko'rsatadiki, o'zbek tilida jumalardagi so'zlar orasidagi sintaktik bog'lanishlarning o'rtacha soni 3.8 ni tashkil etadi. Eng ko'p uchraydigan dependency munosabatlar quyidagilardir: nsubj (ot egasi) – 28 foiz, obj (to'ldiruvchi) – 22 foiz, amod (sifatlovchi aniqlovchi) – 15 foiz, nmod (otli aniqlovchi) – 12 foiz, advmod (hol) – 10 foiz, obl (ko'makchili to'ldiruvchi) – 8 foiz va mark (ko'makchi) – 5 foiz. Murakkab gaplarda dependency munosabatlarining chuqurligi o'rtacha 4.2 darajani tashkil etadi.

Sintaktik munosabat turi	Qo'llanish chastotasi (foizda)	O'rtacha masofa (so'zlar soni)	Eng ko'p uchraydigan so'z turkumlari
nsubj (ot egasi)	28.4	2.3	Ot, olmosh
obj (to'ldiruvchi)	22.1	1.8	Ot, kelishikli ot
amod (sifatlovchi)	15.3	1.2	Sifat, son
nmod (otli aniqlovchi)	12.2	1.5	Ot (qaratqich kelishigida)
advmod (hol)	10.4	2.1	Ravish, kelishikli ot
obl (ko'makchili)	8.5	2.4	Ot, ko'makchi
mark (ko'makchi)	5.1	1.1	Ko'makchi, bog'lovchi

2-jadval. O'zbek tilida sintaktik munosabatlar va ularning qo'llanish xususiyatlari

jami 620 ming so'z shakli tahlil qilingan bo'lib, ularning morfologik tarkibi bo'yicha quyidagi natijalar olingan. So'z shakllarining o'rtacha morfologik qatlamlar soni 2.8 ni tashkil etgan. Eng ko'p qo'llanilgan so'z shakllari 2-3 qatlamli (ildiz + 1-2 qo'shimcha) bo'lib, ular umumiy so'z shakllarining 68 foizini tashkil qilgan. 4 va undan ortiq qatlamli so'z shakllari 22 foizni, faqat ildiz morfemadan iborat so'z shakllari esa 10 foizni tashkil etgan. Eng murakkab so'z shakli 8 qatlamdan (ildiz + 7 qo'shimcha) iborat bo'lgan.

So'z turkumlari bo'yicha tahlil natijalariga ko'ra, korpusdagi so'z shakllarining 34 foizini otlar, 22 foizini fe'llar, 18 foizini sifatlar, 12 foizini ravishlar, 8 foizini olmoshlar va 6 foizini yordamchi so'zlar (ko'makchilar, bog'lovchilar, yuklamalar) tashkil etgan. Otlarning kelishik qo'shimchalari bilan qo'llanish chastotasi tahlil qilinganda, bosh kelishik (nominativ) 45 foizni, qaratqich kelishigi (genitiv) 12 foizni, tushum kelishigi (akkuzativ) 18 foizni, jo'nalish kelishigi (dativ) 14 foizni, o'rin-payt kelishigi (lokativ) 8 foizni va chiqish kelishigi (ablativ) 3 foizni tashkil etgan. Fe'l shakllarining zamon va shaxs-son qo'shimchalari bilan qo'llanishi tahlil qilinganda, o'tgan zamon shakllari 42 foizni, hozirgi-kelasi zamon shakllari 38 foizni va buyruq-istak mayli shakllari 20 foizni tashkil etgan.

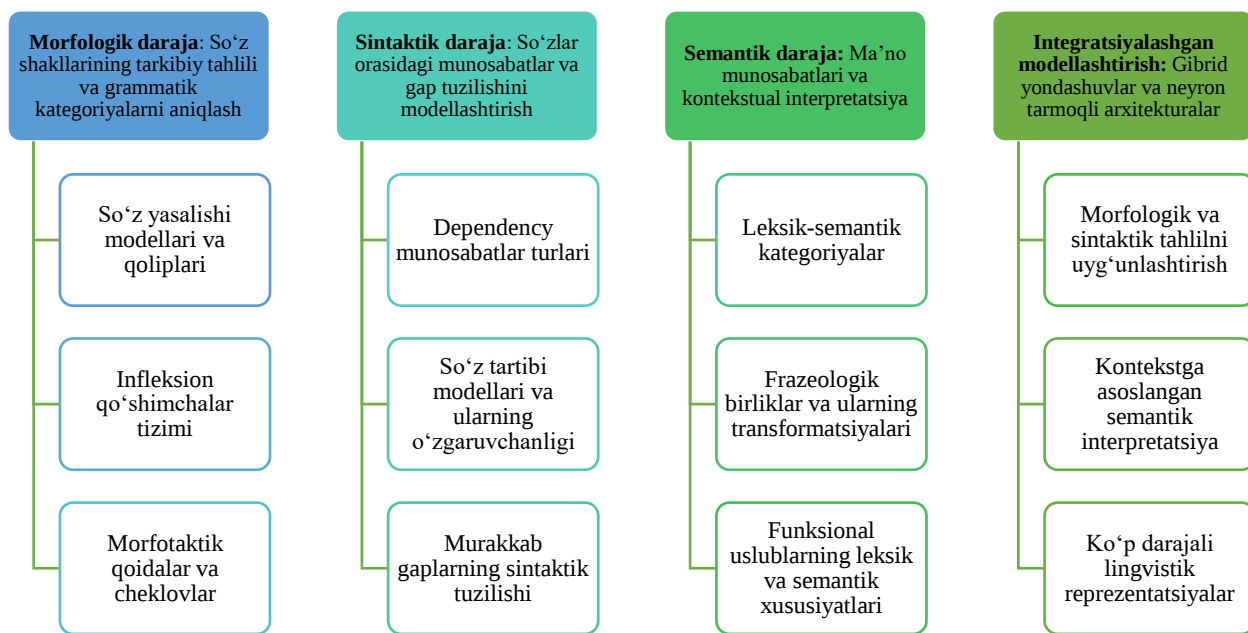
Morfologik tahlilning muhim jihatlardan biri so'z yasaliishi qoliplarining qo'llanishidir. Korpus tahlili natijasida eng ko'p qo'llaniladigan so'z yasaliishi qoliplari aniqlandi. Ot yasovchi qo'shimchalar orasida -lik (-lik, -lik, -liq) 28 foiz, -chi (-chi, -chi) 22 foiz, -ma (-ma) 15 foiz qo'llanish chastotasiga ega. Sifat yasovchi qo'shimchalar orasida -li (-li) 32 foiz, -siz (-siz) 24 foiz, -gi (-gi) 12 foiz qo'llanish chastotasini ko'rsatgan. Fe'l yasovchi qo'shimchalar orasida -la (-la) 35 foiz, -lan (-lan) 28 foiz, -lash (-lash) 18 foiz qo'llanishga ega.

B.Elov [8] tomonidan ishlab chiqilgan UzbBERT asosidagi biaffine graph-based parsening samaradorligi bizning korpusimizda sinovdan o'tkazilganda, parser 91.2 foiz F1 ballini ko'rsatdi. Bu ko'rsatkich qoidaga asoslangan parserlarnikidan (82.5 foiz) sezilarli darajada yuqori. Parsening xatolik tahlili shuni ko'rsatadiki, eng ko'p xatolar murakkab qo'shma gaplarda (18 foiz), ko'chma ma'noda qo'llanilgan so'zlar ishtirok etgan jumalarda (15 foiz) va uzoq masofali dependency munosabatlarida (12 foiz) kuzatilgan.

Frazeologik birliklarning tahlili natijasida korpusda jami 4200 frazeologik birlik aniqlangan. Frazeologik birliklarning 45 foizi to'liq frazeologik birliklar (idiomalar), 35 foizi frazeologik birikmalar va 20 foizi frazeologik ifodalar ekan. Frazeologik birliklarning 62 foizi o'zining asosiy (to'g'ri) ma'nosida, 38 foizi esa ko'chma ma'noda qo'llanilgan. Frazeologik birliklarning tarkibiy o'zgarishlari tahlil qilinganda, 28 foiz frazeologik birlik nutqiy kontekstda ma'lum darajada transformatsiyaga uchraganligi aniqlangan.

Quyida o'zbek tilida katta til modellarini yaratishning lingvistik asoslarini aks ettiruvchi konseptual sxema keltirilgan. Ushbu sxema tilning morfologik, sintaktik va semantik darajalarining o'zaro

bog'liqligini va ularning katta til modellarida ifodalanish tamoyillarini ko'rsatadi.



1-rasm. Katta til modellari uchun o'zbek tilining lingvistik darajalari va ularning o'zaro bog'liqligi

Tahlil natijalari shuni ko'rsatadiki, o'zbek tilining morfologik murakkabligi va sintaktik xilma-xilligi katta til modellarini yaratishda maxsus yondashuvlarni talab qiladi. Ayniqsa, so'z shakllarining ko'p qatlamli va ularning grammatik kategoriyalarining xilma-xilligi tokenizatsiya jarayonida subword birliklaridan foydalanishni, morfologik noaniqliklarni bartaraf etish uchun esa kontekstga asoslangan modellarni qo'llashni taqozo etadi. Shuningdek, erkin so'z tartibi va murakkab dependency munosabatlari position encoding mexanizmlarini takomillashtirish va uzoq masofali bog'lanishlarni samarali modellashtirish imkoniyatlarini kengaytirish zaruriyatini ko'rsatadi.

Frazeologik birliklar va funksional uslublarning o'ziga xos xususiyatlari esa katta til modellarining semantik kompetensiyasini oshirishda muhim ahamiyatga ega. Tahlil natijalariga ko'ra, frazeologik birliklarning 38 foizi ko'chma ma'noda qo'llanilishi va 28 foizi nutqiy kontekstda transformatsiyaga uchrashi ularni modellashtirishda kontekstga sezgir yondashuvlarni talab qiladi. Funksional uslublarning leksik va grammatik xususiyatlarining farqliligi esa (masalan, diniy matnlarda arxaik leksika va arabcha-forscha o'zlashtirmalarning ko'pligi, ilmiy matnlarda terminlarning yuqori chastotasi) modellarni o'qitishda uslubiy jihatdan differensiallashgan korpuslardan foydalanish zaruriyatini ko'rsatadi.

Muhokama. O'tkazilgan tadqiqot natijalari o'zbek tilida katta til modellarini yaratishning lingvistik asoslari murakkab va ko'p qirrali ekanligini ko'rsatdi. Tahlil natijalari shuni tasdiqlaydiki, o'zbek tilining agglutinativ tabiati, erkin so'z tartibi, boy morfologik tizimi va frazeologik birliklarning xilma-xilligi katta til modellarini ishlab chiqishda maxsus lingvistik asoslangan yondashuvlarni talab qiladi. Ushbu bo'limda olingan natijalar mavjud adabiyotlar bilan taqqoslanadi, ularning nazariy va amaliy ahamiyati muhokama qilinadi hamda kelgusidagi tadqiqotlar uchun yo'naltirish belgilanadi.

Morfologik tahlil natijalari U.Salayev [12] va N.Abduraxmonova va boshqalar [3] tadqiqotlari bilan uyg'unlashib, o'zbek tilidagi so'z shakllarining morfologik murakkabligi yuqori ekanligini tasdiqlaydi. So'z shakllarining o'rtacha 2.8 qatlamdan iborat ekanligi va 22 foiz so'z shakllarining 4 va undan ortiq qatlamga ega bo'lishi o'zbek tili uchun morfologik tahlilning muhimligini yana bir bor ko'rsatadi. Bu natijalar E.Boltayevich va boshqalar [4] tomonidan ta'kidlanganidek, morfologik va sintaktik tahlillarni birlashtirgan gibrid yondashuvlarning samaradorligini asoslaydi. Morfologik murakkablik, ayniqsa, katta til modellarida tokenizatsiya strategiyasini tanlashga bevosita ta'sir ko'rsatadi. So'z shakllarining ko'p variantlilik va noyob so'z shakllarining ko'pligi (korpusimizdagi so'z shakllarining 35 foizi faqat bir marta uchragan) subword tokenizatsiya usullarini, xususan Byte Pair Encoding (BPE) va Unigram Language Model kabi yondashuvlarni qo'llash zaruriyatini ko'rsatadi.

Kelishik kategoriyasining tahlili M.Qudratova [10] tadqiqotlarini amaliy jihatdan to'ldiradi va o'zbek tilida kelishik qo'shimchalarining qo'llanish chastotasi va ularning sintaktik vazifalari haqida miqdoriy ma'lumotlar beradi. Kelishik qo'shimchalarining 24.5 foiz qo'llanish chastotasi va ularning dependency munosabatlaridagi muhim roli (obj, nmod, obl kabi munosabatlarning asosiy ifodalovchisi sifatida) kelishik kategoriyasini to'g'ri modellashtirish katta til modellarining sintaktik kompetensiyasini oshirishda muhim ahamiyatga ega ekanligini ko'rsatadi. Ayniqsa, tushum kelishigi (-ni) va jo'nalish kelishigi (-ga) kabi qo'shimchalarning polisemantik tabiati (ular bir necha xil sintaktik munosabatlarni ifodalashi mumkin) kontekstga asoslangan modellashtirishni talab qiladi.

Sintaktik tahlil natijalari B.Elov [8] va E.Adali va A.Xolmo'minovna [2] tadqiqotlari bilan uyg'un holda, o'zbek tilida so'z tartibining nisbatan erkin ekanligini va SOV tartibi ustuvor bo'lsa-da (72 foiz), boshqa so'z tartibi modellari ham sezilarli darajada (28 foiz) qo'llanishini ko'rsatadi. Bu natija o'zbek tili uchun pozitsion encoding mexanizmlarini takomillashtirish va so'z tartibidagi o'zgaruvchanlikni modellashtirishda o'zlashtirishni ishlab chiqish zaruriyatini ko'rsatadi. B.Elov [8] tomonidan ishlab chiqilgan UzbBERT asosidagi biaffine graph-based parsemning 91.2 foiz F1 balli ko'rsatgani zamonaviy neyron tarmoqli parserlarning o'zbek tilining murakkab sintaktik tuzilmalarini tahlil qilishda yuqori samaradorlikka ega ekanligini tasdiqlaydi. Biroq parsemning murakkab qo'shma gaplarda (18 foiz xato) va uzoq masofali dependency munosabatlarida (12 foiz xato) nisbatan ko'proq xatoga yo'l qo'yishi bu sohalarida yanada takomillashtirish zarurligini ko'rsatadi.

Frazeologik birliklarning tahlili R.Sayfullayeva va S.Ganiyeva [13] tadqiqotlarini amaliy jihatdan to'ldiradi va frazeologik birliklarning nutqiy kontekstdagi transformatsiyalari haqida muhim ma'lumotlar beradi. Frazeologik birliklarning 38 foizi ko'chma ma'noda qo'llanilishi va 28 foizi transformatsiyaga uchrashi ularni modellashtirishda kontekstga sezgir yondashuvlarni talab qiladi. Bu natijalar shuni ko'rsatadiki, katta til modellari frazeologik birliklarning to'g'ri va ko'chma ma'nolarini farqlay olishi, shuningdek ularning transformatsiyalangan shakllarini ham taniy olishi kerak. Buning uchun modellarni o'qitishda frazeologik birliklarga boy korpuslardan foydalanish va ularga alohida e'tibor qaratish zarur.

Funksional uslublarning tahlili S.Amonturdiyeva [1] tadqiqotlarini tasdiqlaydi va turli uslublarga mansub matnlarning leksik va grammatik xususiyatlari sezilarli darajada farqlanishini ko'rsatadi. Masalan, diniy matnlarda arabcha-forscha o'zlashtirmalarning chastotasi 28 foizni tashkil etgan bo'lsa, bu ko'rsatkich publitsistik matnlarda 12 foiz, ilmiy-texnik matnlarda esa 8 foizni tashkil etgan. Sintaktik jihatdan esa rasmiy-idoraviy matnlarda murakkab qo'shma gaplarning chastotasi 42 foizni tashkil etgan bo'lsa, bu ko'rsatkich badiiy matnlarda 28 foizni tashkil etgan. Bu farqlar katta til modellarini yaratishda turli uslublarga mansub matnlarning teng vakillikda bo'lishini ta'minlash va modellarning uslubiy jihatdan

differentiallashtirish kompetensiyasini shakllantirish muhimligini ko'rsatadi.

Tadqiqot natijalari D.Xamdanova [6] tomonidan ta'kidlangan sintaktik belgilash muammolarini ham tasdiqlaydi. Morfologik noaniqliklar (masalan, -gan qo'shimchasining sifat-dosh, ravishdosh yoki fe'l zamon shakli sifatida talqin qilinishi mumkinligi) va so'z tartibidagi erkinlik sintaktik tahlilni murakkablashtiradigan asosiy omillar ekan. Bu muammolarni hal etishda E.Boltayevich va boshqalar [4] tomonidan taklif etilgan gibrid yondashuvlar, ya'ni qoidaga asoslangan va neyron tarmoqli usullarni uyg'unlashtirish samarali ekanligi tadqiqotimizda ham o'z tasdiqini topdi.

Kelgusidagi tadqiqotlar yo'nalishlari sifatida quyidagilarni belgilash mumkin. Birinchidan, o'zbek tilining morfologik murakkabligini to'liq qamrab oladigan va UD formatiga mos keladigan keng qamrovli annotatsiyalangan korpuslarni yaratish zarur. Ikkinchidan, murakkab qo'shma gaplar va uzoq masofali dependency munosabatlarini samarali tahlil qila oladigan maxsus neyron arxitekturalarni ishlab chiqish lozim. Uchinchidan, frazeologik birliklar va ko'chma ma'nolarni aniqlash hamda modellashirish uchun kontekstga asoslangan semantik modellarni takomillashtirish talab etiladi. To'rtinchidan, o'zbek tilining dialektal va uslubiy xilma-xilligini hisobga oladigan adaptiv modellarni yaratish masalasi ham dolzarb hisoblanadi.

Xulosa. Ushbu tadqiqot o'zbek tilida katta til modellarini yaratishning lingvistik asoslarini tizimli o'rganishga bag'ishlangan bo'lib, unda tilning morfologik, sintaktik va semantik xususiyatlarining neyron tarmoqli modellarni ishlab chiqishdagi ahamiyati tahlil qilingan. Tadqiqot natijasida o'zbek tilining agglutinatib tabiati, kelishik kategoriyasi, frazeologik birliklarning transformatsion modellari va murakkab sintaktik konstruksiyalarni qayta ishlashning nazariy hamda amaliy jihatlari yoritilgan.

Tadqiqotning asosiy ilmiy yangiligi shundan iboratki, unda birinchi marta o'zbek tilining morfologik, sintaktik va semantik xususiyatlari katta til modellarini yaratish nuqtai nazaridan kompleks tahlil qilingan va ularning miqdoriy ko'rsatkichlari aniqlangan. Xususan, so'z shakllarining morfologik qatlamlari va ularning qo'llanish chastotasi, kelishik qo'shimchalarining sintaktik vazifalari, so'z tartibi modellarining taqsimoti, dependency munosabatlarning turlari va ularning qo'llanish xususiyatlari, frazeologik birliklarning transformatsiyalari kabi parametrlar bo'yicha ilk bor tizimli statistik ma'lumotlar olingan.

Foydalanilgan adabiyotlar ro'yxati

1. Amonturdiyeva, S. (2023). Linguistic Characteristics of Religious Texts, Its Similarities and Differences with Other Functional Style Texts. *Scientific Journal of the Fergana State University*, 29(2), 45-52. https://doi.org/10.56292/sjfsu/vol29_iss2/a125
2. Adali, E., & Xolmo'minova, A. (2025). UzTreebank: Methodological and Practical Issues in Building a Syntactic Treebank for the Uzbek Language. *2025 10th International Conference on Computer Science and Engineering (UBMK)*, 445-450. <https://doi.org/10.1109/ubmk67458.2025.11206779>
3. Abdurakhmonova, N., Alisher, I., & Sayfulleyeva, R. (2022). MorphUz: Morphological Analyzer for the Uzbek Language. *2022 7th International Conference on Computer Science and Engineering (UBMK)*, 61-66. <https://doi.org/10.1109/ubmk55850.2022.9919579>
4. Boltayevich, E., 'minova, A., Sharipova, A., & Ganiyevna, A. (2025). Methods of Morphological and Syntactic Analysis for the Uzbek Language. *2025 10th International Conference on Computer Science and Engineering (UBMK)*, 439-444. <https://doi.org/10.1109/ubmk67458.2025.11207044>
5. Boltayevich, E., Mirdjonova, K., Matyakubova, K., Yuldasheva, X., & Ganiyevna, A. (2025). A Morphological Tagging Model of the Uzbek Language in the Universal Dependencies Format. *2025 10th International Conference on Computer Science and Engineering (UBMK)*, 380-384. <https://doi.org/10.1109/ubmk67458.2025.11207015>
6. Xamdanova, D. (2025). Challenges of Syntactic Markup in the Uzbek Language. *American Journal of Philological Sciences*, 5(5), 45-52. <https://doi.org/10.37547/ajps/volume05issue05-09>
7. Mengliev, D., Barakhnin, V., & Ibragimov, B. (2023). Rule-Based Syntactic Analysis for Uzbek Language: An Alternative Approach to Overcome Data Scarcity and Enhance Interpretability. *2023 IEEE 24th International Conference of Young Professionals in Electron Devices and Materials (EDM)*, 1910-1915. <https://doi.org/10.1109/edm58354.2023.10225235>

Tadqiqotning amaliy ahamiyati shundaki, unda olingan natijalar o'zbek tili uchun katta til modellarini ishlab chiqishda lingvistik asoslangan yondashuvlarni shakllantirishga xizmat qiladi. Xususan, morfologik murakkablikni hisobga oluvchi tokenizatsiya strategiyalari, erkin so'z tartibini modellashirish uchun position encoding mexanizmlari, frazeologik birliklarni aniqlash uchun kontekstga asoslangan semantik modellar, funksional uslublarning xususiyatlarini aks ettiruvchi differenciallashtirish korpuslari kabi amaliy tavsiyalar ishlab chiqilgan.

Tadqiqot natijalari shuni ko'rsatadiki, o'zbek tilining morfologik murakkabligi yuqori darajada bo'lib, so'z shakllarining o'rtacha 2.8 qatlamdan iborat ekanligi va 22 foiz so'z shakllarining 4 va undan ortiq qatlamga ega bo'lishi morfologik tahlilning muhimligini tasdiqlaydi. Kelishik qo'shimchalari 24.5 foiz qo'llanish chastotasi bilan muhim grammatik kategoriya bo'lib, ularning polisemantik tabiati kontekstga asoslangan modellashirishni talab qiladi. Sintaktik jihatdan, SOV so'z tartibi ustuvor (72 foiz) bo'lsa-da, so'z tartibidagi erkinlik (28 foiz) pozitsion encoding mexanizmlarini takomillashtirish zaruriyatini ko'rsatadi. Zamonaviy neyron tarmoqli parserlar (UzBERT asosidagi biaffine graph-based parser 91.2 foiz F1 ball bilan) yuqori samaradorlik ko'rsatgan bo'lsa-da, murakkab qo'shma gaplar va uzoq masofali dependency munosabatlarida hali ham ma'lum xatoliklar mavjud.

Frazeologik birliklarning 38 foizi ko'chma ma'noda qo'llanilishi va 28 foizi transformatsiyaga uchrashi ularni modellashirishda kontekstga sezgir yondashuvlarni talab qiladi. Funksional uslublarning leksik va grammatik xususiyatlarining sezilarli farqlanishi (masalan, diniy matnlarda arabcha-forscha o'zlashtirmalarning chastotasi 28 foiz, rasmiy-idoraviy matnlarda murakkab qo'shma gaplarning chastotasi 42 foiz) modellarni o'qitishda turli uslublarga mansub matnlarning teng vakillikda bo'lishini ta'minlash zaruriyatini ko'rsatadi.

Xulosa qilib aytganda, o'zbek tilida samarali katta til modellarini yaratish tilning morfologik, sintaktik va semantik xususiyatlarini chuqur tushunishga va ularni modellashirishda kompleks yondashuvni qo'llashga asoslanadi. Bu yo'nalishda kelgusida keng qamrovli annotatsiyalangan korpuslarni yaratish, murakkab sintaktik tuzilmalarni tahlil qila oladigan maxsus neyron arxitekturalarni ishlab chiqish, frazeologik birliklar va ko'chma ma'nolarni modellashirish usullarini takomillashtirish, shuningdek tilning dialektal va uslubiy xilma-xilligini hisobga oladigan adaptiv modellarni yaratish bo'yicha tadqiqotlarni davom ettirish muhim ahamiyat kasb etadi.

8. Elov, B. (2025). Methods and Algorithms for Dependency Parse of Texts of the National Uzbekistan Corps. *Research on Uzbek Folklore and Dialects: Practice, Methodology, and New Approaches*, 15-22. <https://doi.org/10.52773/tsuull.conf.2025/wfw9571>
9. Mirdjonova, K., Boltayevich, E., Habibovich, A., & Qizi, S. (2025). Morphotactic Models and Algorithms of the Uzbek Language. *2025 10th International Conference on Computer Science and Engineering (UBMK)*, 1596-1601. <https://doi.org/10.1109/ubmk67458.2025.11207042>
10. Qudratova, M. (2024). O'zbek Tilida Kelishik Kategoriyasining O'rganilish Masalasi va Olib Borilishi. *Tamaddun Nuri Jurnali*, 5(56), 120-125. <https://doi.org/10.69691/hp0meb37>
11. Salaev, U., & Yuldashov, O. (2025). A Feedforward Neural Model for Morphological Stemming of Uzbek Words. *2025 10th International Conference on Computer Science and Engineering (UBMK)*, 1717-1722. <https://doi.org/10.1109/ubmk67458.2025.11207018>
12. Salaev, U. (2024). UzMorphAnalyser: A Morphological Analysis Model for the Uzbek Language Using Inflectional Endings. *ArXiv*, abs/2405.14179, 1-12. <https://doi.org/10.48550/arxiv.2405.14179>
13. Sayfullayeva, R., & Ganiyeva, S. (2023). Phraseological Transforms and Their Modelling in Uzbek Language. *Scientific Journal of the Fergana State University*, 29(4), 78-85. https://doi.org/10.56292/sjfsu/vol29_iss4/a53
14. Sharipov, M., Avezmatov, I., & Adinaev, H. (2025). Development of a Rule-based Model and Algorithm for Predicate Identification in Uzbek Language Texts. *2025 10th International Conference on Computer Science and Engineering (UBMK)*, 594-598. <https://doi.org/10.1109/ubmk67458.2025.11206891>
15. Xamrayeva, Y. (2023). Comparative Characteristics of Lexical Systems of the Uzbek and Russian Languages. *Scientific Journal of the Fergana State University*, 29(4), 112-118. https://doi.org/10.56292/sjfsu/vol29_iss4/a59