

O‘ZBEK TILI BIRLIKLARI OT SO‘Z TURKUMINI TUZILISHIGA KO‘RA POS TEGLASH

Xudayberganov Nizomaddin Uktambay o‘g‘li

Katta o‘qituvchi

nizomaddin@navoiy-uni.uz

ToshDO‘TAU

Annotatsiya. Ushbu maqolada o‘zbek tili birliklarini ot so‘z turkumini tuzilishiga ko‘ra avtomatik belgilash (POS tagging) masalasi chuqur ko‘rib chiqiladi. O‘zbek tilining agglutinativ morfologik tuzilishi kontekstida ot turkumining yasaliş yo‘llari, grammatik kategoriyalari – kelishik (6 ta), egalik (shaxs-son), ko‘plik (-lar), aniqlik/noaniqlik – hamda xos ot (PROPN) bilan farqlash muammolari tahlil qilinadi. Ot so‘z turkumini teglashda yuzaga keladigan ko‘p ma‘nolilik – xususan “olma” (NOUN: meva) va “ol-ma” (VERB: inkor) kabi holatlar – va ularni yechish yo‘llari muhokama etiladi. Kengaytirilgan ot tagset (NOUN + kelishik + egalik + ko‘plik + yasaliş turi) taklif etiladi. Tajriba natijalariga ko‘ra, BERT-asoslangan modellar ot turkumini 96.8% aniqlikda teglash imkonini beradi.

Kalit so‘zlar: *POS teglash, ot so‘z turkumi, o‘zbek tili, kelishik, egalik affiksi, ko‘plik, xos ot (PROPN), UzbekPOS, BBPOS, UD Treebank, agglutinativ morfologiya, kengaytirilgan tagset*

Abstract. This article comprehensively examines Part-of-Speech (POS) tagging of Uzbek language units according to the structural typology of nouns. Within the context of Uzbek's agglutinative morphology, the formation methods of the noun class, grammatical categories – case (6 forms), possessive (person-number), plural (-lar), definiteness – and challenges in distinguishing nouns (NOUN) from proper nouns (PROPN) are analyzed. The homonymy challenge – particularly “olma” (NOUN: apple) and “ol-ma” (VERB: negative imperative) – and solutions are discussed. An extended noun tagset (NOUN + case + possessive +

plural + formation type) is proposed. Experimental results indicate BERT-based models can tag nouns with up to 96.8% accuracy.

Keywords: *POS tagging, noun part-of-speech, Uzbek language, case, possessive affix, plural, proper noun (PROPN), UzbekPOS, BBPOS, UD Treebank, agglutinative morphology, extended tagset*

KIRISH

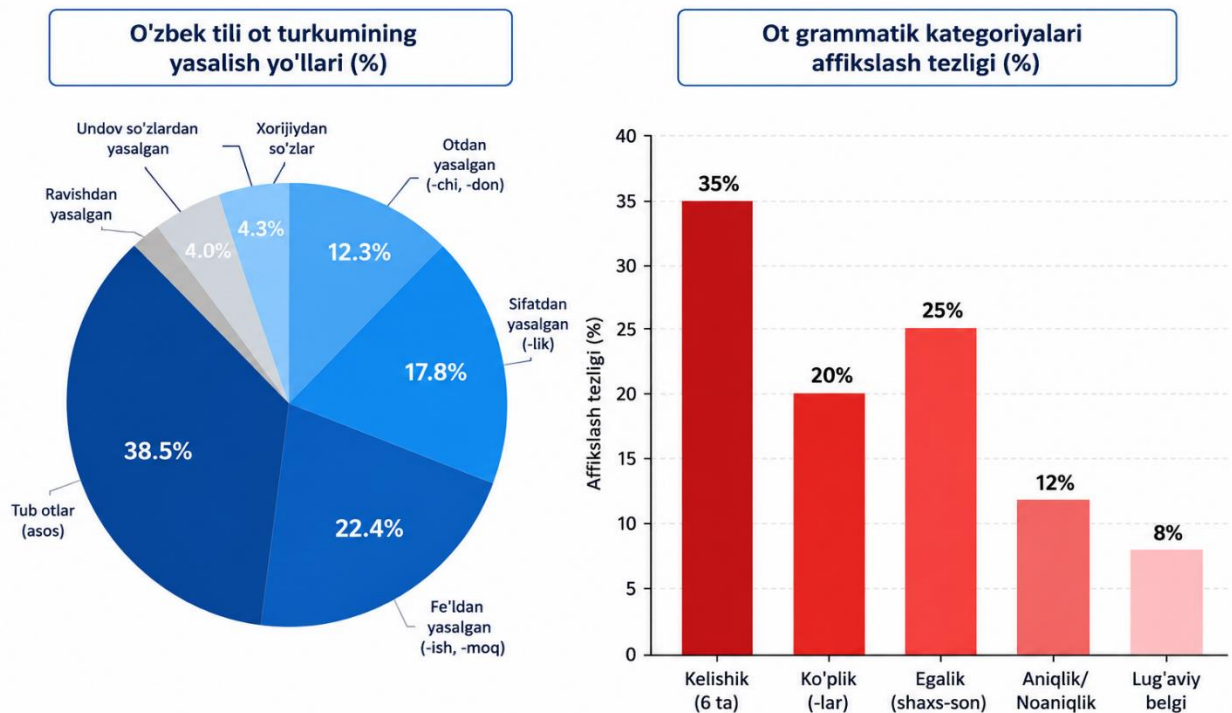
Tabiiy tillarni qayta ishlash (NLP) sohasida soʻz turkumlarini avtomatik belgilash – POS teglash – matn tahlilining eng muhim bosqichi hisoblanadi. Oʻzbek tili – agglutinativ morfologiyaga ega til[1] sifatida POS teglashda ingliz va boshqa analitik tillardan farqli bir qator murakkabliklarni taqdim etadi. Ularning eng koʻp uchraydigan va ayni paytda muammoli turkumi – **ot (NOUN)** soʻz turkumidir.

Oʻzbek tili grammatikasiga koʻra[2] ot soʻz turkumi quyidagi asosiy grammatik kategoriyalarga ega:

- Kelishik (Case) – 6 ta: bosh kelishik (-Ø), qaratqich (-ning), joʻnalish (-ga/-a/-ka), tushum (-ni), oʻrin-payt (-da), chiqish (-dan)
- Egalik (Possessive/Person-number) – I shaxs (-m/-im/-imiz), II shaxs (-ng/-ing/-ingiz), III shaxs (-si/-i)
- Koʻplik (Plural) – (-lar): kitoblar, shaharlar, oʻquvchilar
- Aniqlik/Noaniqlik – oʻzbek tilida articles yoʻq, lekin 'bir' (noaniqlik) va koʻrsatish olmoshi (aniqlik) ishlatiladi
- Lugʻaviy belgi – tub ot (ildiz), yasama ot (affikslangan), murakkab ot (qoʻshma soʻz)

Karimova (2025)[3] taʼkidlashicha, kengaytirilgan POS teglashda ot uchun nafaqat NOUN tegi, balki grammatik atributlar ham berilishi morfosintaktik tahlil sifatini sezilarli oshiradi. Masalan: “**maktabda**” soʻzi → **[NOUN] [yasama emas] [oʻrin-payt kelishik] [III shaxs egalik yoʻq] [birlik]**

O‘zbek tilida otlar bir nechta asosiy yo‘l bilan yasaladi. 1-rasmda yasalish yo‘llari foizlar bilan keltirilgan. Tub otlar (38.5%) eng ko‘p, fe‘ldan yasalgan otlar (22.4%) esa ikkinchi o‘rinda turadi.



1-rasm. O‘zbek tili ot turkumining yasalish yo‘llari va grammatik kategoriyalari

Fe‘ldan ot yasash – eng faol yo‘l: “-ish” (o‘qish, yozish), “-moq” (o‘qimoq, yurmoq), “-uvchi” (o‘quvchi, ishlovchi). Sifatdan ot yasash: “-lik” (kattalar → kattalik, yaxshi → yaxshilik). Otdan ot yasash: “-chi” (osh → oshpaz, kitob → kitobxon).

1-jadval. O‘zbek tili ot turkumining grammatik kategoriyalari va POS teglash uchun ahamiyati

Kategoriya	Affiks/shakl	Misol	POS teglashdagi ahamiyati
Bosh kelishik	-Ø (affikssiz)	kitob, maktab	NP boshi; ega, to‘ldiruvchi funksiyasi



Qaratqich kelishik	-ning	kitobning, maktabning	Qaratuvchi; izofali birikma hosil qiladi
Jo'nalish kelishik	-ga / -a / -ka	kitobga, uyga, ko'chaga	Mo'ljal va yo'nalish ma'nosi
Tushum kelishik	-ni	kitobni, maktabni	To'g'ri to'ldiruvchi; o'timli fe'l bilan
O'rin-payt kelishik	-da	maktabda, uyda	Joylashuv yoki vaqt ma'nosi
Chiqish kelishik	-dan	maktabdan, uydan	Manba, chiqish nuqtasi
Ko'plik	-lar	kitoblar, maktablar	Birlik va ko'plik farqi
I shaxs egalik	-m / -im / -imiz	kitobim, maktabim	Birinchi shaxs egalik
II shaxs egalik	-ng / -ing / -ingiz	kitobing, maktabing	Ikkinchi shaxs egalik
III shaxs egalik	-si / -i / -lari	kitobi, maktabi	Uchinchi shaxs egalik
Murakkab kelishik	-dagi, -dagi, -niki	maktabdagi, uniki	Kelishik + egalik kombinatsiyasi
Xos ot (PROPN)	Bosh harf belgisi	Toshkent, Alisher	NER uchun alohida; NOUN dan farqli

O'zbek tilidagi eng mashhur ko'p ma'nolilik misoli – “olma” so'zi. Bu misol Elov et al. (2023)[4] maqolasida batafsil ko'rsatilgan: “olma” – meva nomi (NOUN), lekin “ol-ma” – “olmoq” fe'lining inkor shakli (VERB). Kontekstsiz teglash bu so'zni noto'g'ri turkumga kiritishi mumkin.

Bu muammo nafaqat “olma” da, balki yuzlab bunday juft shakllarda mavjud. Shuningdek, Xolmo'minovna et al. (2023)[5] turk, uyg'ur va o'zbek tillari solishtirmasida ko'p ma'noli so'zlar agglutinatив tillarda qoidabazali tizimlar uchun eng og'ir muammo ekanligini ko'rsatadi.

Ba'zi o'zbek so'zlari kontekstga qarab NOUN yoki ADJ bo'lib kelishi mumkin. Masalan, “katta” so'zi sifat (ADJ) bo'lishi mumkin (“katta uy”) yoki mustaqil ravishda ot vazifasini bajarishi ham mumkin (“kattalar”). Bunga qo'shimcha, “-lik” affiksi bilan sifatdan hosil qilingan otlar ham chalkashlik yaratishi mumkin.

2-jadval. Ot teglashdagi asosiy ko'p ma'nolilik holatlari va ularni yechish usullari

So'z	Holat-1	Holat-2	Teg-1	Yechish usuli
olma	“olma” – meva (NOUN)	“ol-ma” – inkor fe'l (VERB)	NOUN vs VERB	Kontekst tahlili; qo'shni fe'l yo'qligi → NOUN
yozuv	Ot: “xat” (NOUN)	Fe'l ildizi: “yozuv” → yozmoq	NOUN vs VERB	Lug'at qidirish + kontekst
katta	Sifat: “katta uy” (ADJ)	Ot: “kattalar” → katta+lar	ADJ vs NOUN	“-lar” affiksi → NOUN; kontekstsiz ADJ
O'zbekiston	Xos ot (PROPN)	Sifat: O'zbek+iston (NOUN?)	PROPN vs NOUN	Bosh harf + NER lug'ati
yangi	Sifat: “yangi kitob” (ADJ)	Ot: “yangilar” (NOUN+lar)	ADJ vs NOUN	“-lar” affiksi tekshirish
yaxshilik	Ot: “-lik” affiksi (NOUN)	Sifat: “yaxshi” asos (ADJ)	NOUN vs ADJ	“-lik” → NOUN; asos sifat bo'lsa ham
baxt	Ot: “baxt” (NOUN)	Sifat: “baxtli” (ADJ+li)	NOUN vs ADJ	“-li” affiksi tekshirish → ADJ

Hozirda o'zbek tili POS teglashida uch xil yondashuv qo'llaniladi: 12 taglik (UzbekTagger, pos.uznlp.uz [6]), 16 taglik (UzbekPOS, 2026), va 17 UPOS taglik

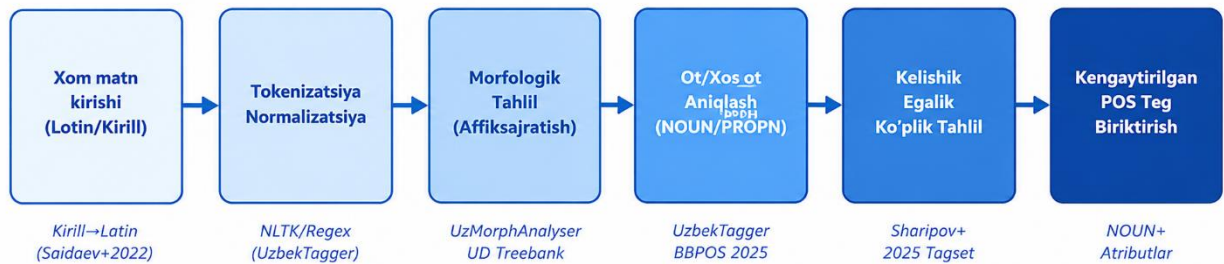
UD standarti (BBPOS, UD Treebank). Biroq ularning barchasi ot uchun faqat bir teg (NOUN yoki N) beradi va grammatik atributlarni akslantirmaydi.

Abdullayeva et al. (2022)[7] 100+ morfologik teg taklifi qilgan – lekin bu hali annotatsiyalangan datasetga tatbiq etilmagan. Ushbu maqolada **amaliy qo'llanishi mumkin bo'lgan kengaytirilgan ot tagset** taklif etiladi:

3-jadval. POS teglash – ot turkumi bo'yicha amaliy misollar

So'z shakli	Asosiy teg	Kengaytirilgan teglar	Tahlil: morfemalar
kitob	NOUN	NOUN.NOM + NOUN.ROOT	kitob = ildiz; bosh kelishik; birlik
kitoblarida	NOUN	NOUN.PLU + NOUN.P3 + NOUN.LOC	kitob + lar + i + da = ko'plik+III shaxs egalik+o'rin
maktabning	NOUN	NOUN.GEN + NOUN.ROOT	maktab + ning = qaratqich kelishik
o'qishga	NOUN	NOUN.DER.V + NOUN.DAT	o'qi + sh + ga = fe'ldan yasalgan+jo'nalish
yaxshilik	NOUN	NOUN.DER.A + NOUN.NOM	yaxshi + lik = sifatdan yasalgan; bosh kelishik
o'quvchi	NOUN	NOUN.DER.V + NOUN.ROOT	o'qi + uvchi = fe'ldan yasalgan ot
Toshkentga	PROPN	PROPN + NOUN.DAT	Toshkent + ga = xos ot + jo'nalish kelishik
kitobimni	NOUN	NOUN.P1 + NOUN.ACC	kitob + im + ni = I shaxs egalik + tushum
shaharlar	NOUN	NOUN.PLU + NOUN.NOM	shahar + lar = ko'plik; bosh kelishik
o'qituvchi	NOUN	NOUN.DER.V + NOUN.ROOT	o'qi + t + uvchi = orttirma+fe'ldan yasalgan

2-rasmda ot soʻz turkumini tuzilishiga koʻra teglashning toʻliq pipeline sxemasi koʻrsatilgan. Sxema UzbekTagger (2023) va UzbekPOS (2026) metodologiyasi asosida ishlab chiqilgan.



2-rasm. Ot soʻz turkumini tuzilishiga koʻra POS teglash pipeline sxemasi

Oʻzbek tili ot soʻz turkumini tuzilishiga koʻra POS teglash – bu faqat NLP uchun emas, balki mashina tarjimasi, lugʻat yaratish, til taʼlim tizimlari va axborot qidirish uchun ham fundamental ahamiyatga ega vazifadir. UzbekPOS, BBPOS va UD Treebank kabi yangi resurslar ushbu sohani jadal rivojlantirish imkonini beradi.

Foydalanilgan adabiyotlar roʻyxati

1. Agglutinativ til grammatik maʼnolar soʻz oʻzagiga affikslar qoʻshish orqali ifodalanadi. Johanson, L. & Csató, É.A. (2015). The Turkic Languages. Routledge. – Oʻzbek, turk, qozoq, fin, venger tillari bunga misol.
2. Oʻzbek tili grammatikasi. Oʻzbekiston Fanlar Akademiyasi, Toshkent: Fan, 2009.
3. Karimova, Z.D. (2025). POS tagging of lexical units in the Uzbek language. GIIRJ, 13(1), pp. 94–96.
4. Elov, B., Sirojiddinov, S., Khamroeva, Sh. et al. (2023). POS tagging of Uzbek text using hidden Markov model. UBMK-2023, pp. 63–68.



5. Xolmo‘minovna, A.O., Yuldashevna, K.Z. & Uktambov o‘g‘li, X.N. (2023). The problem of POS tagging and stemming for agglutinative languages. CEUR-WS, 2022, pp. 93–98.
6. pos.uznlp.uz O‘zbek POS Tagger onlayn tizimi.
7. Abdullayeva, O., Uzokova, M. & Elov, B. (2022). O‘zbek tilida POS tagging masalasi: muammo va takliflar. Uzbekistan: Language and Culture, 2(5), 51–68.