

O'ZBEK TILIDAGI SO'Z BIRLIKLARINI OLMOSH SO'Z TURKUMI BO'YICHA POS TEGGLASHDAGI MUAMMOLAR VA YECHIMLAR

Karimova Zilola Dilmurod qizi,

Kompyuter lingvistikasi mutaxassisligi magistranti

karimovazilola791@gmail.com

ToshDO'TAU

Annotatsiya. Ushbu maqolada o'zbek tilidagi so'z birliklarini Olmosh so'z turkumi doirasida avtomatik teglash (POS tagging) orqali aniqlashdagi asosiy muammolar va ularning yechimlari tahlil qilinadi. Xususan, olmoshlarning morfologik shakllari, kontekstual ko'rinishlari va leksik-semantik xususiyatlari tufayli yuzaga keladigan xatolar ko'rib chiqiladi. Muammolarning texnik va lingvistik jihatlari o'rganilib, tegishli qismlarni takomillashtirish bo'yicha takliflar bildiriladi.

Abstract. This paper analyzes the core challenges of automatic part-of-speech (POS) tagging of pronouns in the Uzbek language. It focuses on morphological forms, contextual usage, and lexical-semantic characteristics of pronouns that often cause tagging errors. Statistical data and diagrams based on annotated Uzbek corpora are presented, and several practical solutions are proposed to enhance model accuracy.

Аннотация. В данной статье рассматриваются основные проблемы автоматической части речи (POS tagging) для местоимений в узбекском языке. Особое внимание уделяется морфологическим формам местоимений, их контекстуальным проявлениям и лексико-семантическим особенностям, вызывающим ошибки в системах автоматической обработки. На основе аннотированных корпусов узбекского языка проведён статистический анализ и предложены возможные решения для повышения точности.

Kalit so'zlar: *Olmosh so'z turkumi, statistik tahlil, POS tagging, morfologik analiz, avtomatik belgilash, korpus tilshunosligi*

KIRISH

Zamonaviy kompyuter lingvistikasi sohasida tabiiy tilni avtomatik qayta ishlash masalalari tobora dolzarb bo'lib bormoqda. Xususan, so'z turkumlarini avtomatik aniqlash – Part-of-Speech (POS) tagging texnologiyasi tilshunoslik tadqiqotlarida asosiy vosita sifatida xizmat qilmoqda. Ko'plab rivojlangan tillarda (ingliz, rus, fransuz va boshqalar) bu yo'nalishda mustahkam korpuslar, model va

teglovchilar mavjud bo'lsa-da, o'zbek tilida POS tagging texnologiyasining to'liq shakllanmaganligi ilmiy-amaliy izlanishlar zaruratini yuzaga keltirmoqda[1].

Hozirgi kunda tabiiy tilni qayta ishlash (NLP) sohasida matndagi so'zlarning kategoriyalarini aniqlash muammosi dolzarb hisoblanadi. So'zlarning ot, olmosh, fe'l, ravish kabi kategoriyalarga mansubligini aniqlash POS (Parts of Speech) tagging deb ataladi. Bu jarayon matnni tahlil qilishda va mashinali o'qitishda muhim ahamiyatga ega[2].

O'zbek tilidagi olmoshlar semantik va sintaktik jihatdan murakkab birliklar hisoblanadi. Ularda kontekstga bog'liq holda o'zgaruvchanlik mavjudligi tufayli, avtomatik tizimlar tomonidan noto'g'ri belgilanish hollari ko'p uchraydi. Xususan, shaxs olmoshlari (“men”, “sen”, “u”), ko'rsatish olmoshlari (“bu”, “shu”, “o'sha”) orasidagi formal yaqinlik, yoki so'roq (“kim”, “nima”) va gumon olmoshlari (“kimdir”, “qandaydir”) orasidagi semantik noaniqlik natijasida POS tagging tizimlarida xatoliklar kuzatiladi.

Olmoshlarning turlari – kishilik, ko'rsatish, so'roq, o'zlik, belgilash, bo'lishsizlik va gumon olmoshlaridir. Ularning har biri morfologik va semantik jihatdan farqlanadi. Ular o'z funksional yuklamasiga ko'ra gapda turli sintaktik rol ni bajaradi. Masalan, kishilik olmoshlari ko'pincha ega vazifasida ishlatilsa, ko'rsatish olmoshlari aniqlik kirituvchi yoki ko'rsatkich bo'lib xizmat qiladi.

Bundan tashqari, olmoshlar tuzilishiga ko'ra quyidagi ko'rinishlarda uchraydi:

1. *Sodda olmoshlar*: sen, u, biz, bu, shu
2. *Qo'shma olmoshlar*: har kim, har nima, hech kim
3. *Juft olmoshlar*: kim-kim, nima-nima
4. *Takroriy olmoshlar*: o'sha-o'sha, shu-shu

1-jadval. Olmoshlarning gapdagi grammatik roli

№	Olmosh turi	Namuna	Grammatik rol
	Kishilik	men, sen, u	Subyekt, obyekt
	Ko'rsatish	bu, shu, o'sha	Aniqlik kirituvchi
	So'roq	kim, nima, qaysi	Savol bildiruvchi
	Gumon	kimdir, nimadir	Noaniqlik bildiruvchi
	O'zlik	o'zim, o'zing	Refleksiv holat
	Belgilash	har kim, har nima	Universal
	Bo'lishsizlik	hech narsa, hech kim	Inkor holat

Bu jadvaldagi olmoshlar POS teglovchi modellarni tuzishda har bir olmosh turini alohida yondashuv bilan ko'rib chiqishni talab etadi. Shuningdek, olmoshlar sintaktik rolga qarab turlicha turlarga kiritilishi, ular asosida avtomatik teg qo'yuvchi algoritmlarni sinovdan o'tkazish zaruratini vujudga keltiradi.

Asosiy qism. Nutqning bir qismi (POS) – soʻzning grammatik funksiyasini tavsiflash usuli. POS teglari NLP ning yuqori darajadagi funksiyalarida qoʻllanilsa-da, ularni oʻz-oʻzidan tushunish muhim va ularni matn tahlilida foydali maqsadlarda ishlatish mumkin[3].

Olmosh soʻz turkumi ikki xil koʻrinishga ega, yaʼni *sof olmosh tabiatli soʻzlar* hamda *olmosh vazifasida qoʻllaniluvchi soʻzlar*. Sof olmosh soʻzlar oʻzida olmoshning barcha jihatlarini aks ettiradi, uning tub mohiyati ham olmosh maʼnosini ifodalaydi, ularga misol sifatida men, sen, ular, kimdir, nimadir va hokazolarni keltirish mumkin. Olmosh vazifasida qoʻllaniluvchi istilohlar esa tabiatan olmosh xarakteriga ega boʻlsa-da, biroq ularning kelib chiqishi aslan boshqa soʻz turkumlariga yaʼni ot, sifat soʻz turkumlariga oid boʻladi. Nomlari qayd etilgan har ikki toifa bir-biridan baʼzi jihatlariga koʻra farqlanadi hamda ularning ayrim oʻxshash jihatlarini ham mavjud[4]. Uni quyidagi 2-jadval asosida koʻrib chiqamiz.

2-jadval. Olmosh soʻz turkumining qoʻllanish oʻrinlari

№	Sof olmosh tabiatli soʻzlar	Olmosh vazifasida keluvchi soʻzlar
	Olmoshning barcha xususiyatlarini oʻzida aks ettiradi	Olmoshning barcha xususiyatlarini oʻzida aks ettiradi
	Kelib chiqishiga koʻra ham olmosh turkumiga mansub boʻladi	Kelib chiqishiga koʻra boshqa soʻz turkumlariga oid boʻlishi mumkin
		Ayrim hollarda asl soʻz turkumi bajargan vazifasini bajaradi (masalan, sifat soʻz turkumiga oid boʻlsa, sifat soʻz turkumining xususiyatini aks ettirishi mumkin)

Oʻzbek tilida olmoshlarning POS teglashdagi asosiy muammolaridan biri ularning koʻp funksiyali va kontekstsiz ajratishning murakkabligidir. Masalan, “u” soʻzi kishilik olmoshi sifatida ham (“*U maktabga bordi*”), koʻrsatish olmoshi sifatida ham (“*U bola sinfda qolibdi*”) ishlatilishi mumkin. Bu holat avtomatik teglashda ikki xil tegni talqin qilishga olib keladi. Shuningdek, soʻroq va gumon olmoshlari ham koʻp hollarda bir-biriga yaqin strukturada boʻlib, ayniqsa matn korpuslarida notekis ishlatilgan. Masalan, “kim” (soʻroq) va “kimdir” (gumon) olmoshlari korpuslarda koʻpincha yagona teg bilan belgilangan, bu esa statistik modellar uchun chalqituvchi hisoblanadi[5].

Olmoshlar tuzilmasiga koʻra, ularni sodda, qoʻshma, juft va takroriy turlarga ajratish mumkin. Sodda olmoshlar bir morfemali boʻlsa, qoʻshma olmoshlar ikki yoki undan ortiq soʻzning birikmasidan tashkil topadi (“*har kim*”, “*biror narsa*”). Juft olmoshlar esa ikki marta takrorlangan yoki oʻzaro juftlashgan olmoshlardir (“*kim-kim*”, “*nima-nima*”). Takroriy olmoshlar esa ayni bir olmoshning ikki marta takrorlanishi orqali hosil boʻladi (“*shu-shu*”, “*oʻsha-oʻsha*”). Bu strukturalarning

ko'pligi va kontekstdagi rolini hisobga olgan holda, POS teglashda morfologik analizatorlarning aniqligini oshirish uchun chuqur semantik va sintaktik tahlil zarur. Masalan, mashinali o'qitish asosidagi modellar uchun olmoshlarni kontekstual

Misollar bilan boyitish, yoki BERT modeli asosida maxsus mexanizmni joriy qilish taklif etiladi. Yuqoridagi misollar shuni ko'rsatadiki, POS teglashda nafaqat sirtqi ko'rinish, balki olmoshlarning gapdagi vazifasi, atrofdagi birliklar bilan munosabati ham muhimdir. Bu esa olmoshlarning alohida korpuslarda semantik belgilanishi, qo'shimcha atributlar bilan boyitilishi, teglashda rule-based va statistik yondashuvlarning kombinatsiyasini talab qiladi[6].

Olmoshlar bilan bog'liq POS teglashdagi muammolarni quyidagicha tahlil qilishimiz mumkin:

- Olmoshlarning mazmuni va ularning grammatik funksiyalari kontekstga bog'liq bo'ladi. Masalan, “u” so'zi “u” olmoshi bo'lishi mumkin, ammo u shaxsni bildirgan holda ishlatilganida “u”ning o'zi haqida gapirsa, boshqacha xususiyatga ega bo'lishi mumkin. Bu kabi muammolarni to'g'ri tahlil qilishda POS tagging tizimlari uchun qiyinchilik tug'diradi.

- Kishilik va o'zlik olmoshlarini ajratish: Ba'zi olmoshlar bir nechta shartli vaziyatlarga ega bo'lishi mumkin. Misol uchun, “men” olmoshi shaxsni bildiradi, ammo “o'zim” olmoshi o'zlikni bildiradi. Bu ikki turdagi olmoshlarni to'g'ri ajratib olish tizim uchun qiyin bo'lishi mumkin.

- Ko'rsatish olmoshlarining aniq tahlili: “Bu” va “shu” kabi ko'rsatish olmoshlari, odatda, predmetni yoki shaxsni aniq belgilash uchun ishlatiladi, ammo kontekstga qarab ular farqlanishi mumkin. Bu holatni aniqlash uchun kontekstni tahlil qilish talab etiladi.

- Bo'lishsizlik olmoshlari: Bo'lishsizlik olmoshlari (“hech kim”, “hech narsa”) juda ko'p holatlarda, ayniqsa so'roq gaplarida yoki inkor javoblarda ishlatiladi. Ularni to'g'ri tegishli kategoriyaga o'tkazish uchun muayyan algoritmlar ishlab chiqilishi zarur.

3-jadval Olmoshlar bilan bog'liq POS teglashdagi muammolar

№	Muammo turi	Tavsif	Misollar
	Morfologik	Olmoshlarning boshqa so'z turkumlari bilan shaklan o'xshashligi, qo'shimchalarga ega bo'lishi	“Uning kitobi”, “O'zini yo'qotdi”
	Kontekstual	Gapdagi o'rni va vazifasiga qarab olmoshlarning noto'g'ri turkumlanishi.	“Bu yaxshi fikr”, “Bu esa boshqa masala”
	Leksik-semantik	Ma'no va referentga bog'liq holda olmoshlarning chalkash teglanishi	“Hech kim kelmadi”, “Kim bo'ldi u?”

POS tagging tizimlarida olmoshlarni tahlil qilishda quyidagicha yechimlarni berishimiz mumkin:

1. *Statistik va ma'lumotlarga asoslangan modellar*: HMM (Hidden Markov Model) va boshqa statistik modellar, masalan, Viterbi algoritmi yordamida olmoshlarni tahlil qilishda kontekstga asoslanib to'g'ri POS tegini aniqlash mumkin. Bu modellarda har bir so'zning ehtimoliy pozitsiyasini aniqlash orqali olmoshlarning to'g'ri kategoriyasini belgilash mumkin[7].

2. *Deep Learning metodlari*: O'qitilgan neyron tarmoqlari yordamida olmoshlar turini aniqlashda, masalan, LSTM (Long Short-Term Memory) yoki BERT (Bidirectional Encoder Representations from Transformers) modellarida foydalanish mumkin. Ushbu metodlar kontekstni chuqur o'rganib, olmoshlarning to'g'ri tahlilini amalga oshiradi[8].

3. *Qayta ishlash va morfologik tahlil*: Olmoshlarni to'g'ri ajratish uchun morfologik tahlil qilish va qayta ishlash texnikalari ishlatiladi. Bu jarayonlarda so'zning tarkibiy qismlari, affikslar va sintaktik o'zgarishlar tahlil qilinadi.

XULOSA.

Olmosh so'z turkumi o'zbek tilining murakkab va ko'p qatlamli leksik-morfologik qatlamlaridan biri bo'lib, ularni avtomatlashtirilgan tizimlar orqali aniqlash va teg belgilash (POS tagging) amaliyotida bir qator muammolar yuzaga keladi. Ushbu maqolada biz aynan morfologik, kontekstual va leksik-semantik muammolarni tahlil qildik va ularning har biri POS tagging tizimlariga qanday ta'sir ko'rsatishini ko'rsatib berdik. Morfologik muammolar asosan olmoshlarning boshqa so'z turkumlari bilan shaklan o'xshash bo'lishi va morfemik jihatdan bir nechta variantlarga ega bo'lishi natijasida yuzaga keladi. Kontekstual muammolar esa, ayniqsa ko'rsatish, so'roq va belgilash olmoshlariga xos bo'lib, bir xil ko'rinishdagi birliklarning har xil kontekstlarda turli sintaktik vazifalarni bajarishi bilan bog'liq. Leksik-semantik muammolar esa referentga bog'liq holda olmoshning asl ma'nosini to'g'ri aniqlashdagi murakkabliklar bilan bog'liq bo'lib, ular faqat sintaktik va semantik bog'liqlik orqali hal qilinishi mumkin. Zamonaviy NLP (Natural Language Processing) texnologiyalarining rivojlanishi ushbu muammolarni hal qilish imkonini bermoqda. Xususan, gibrid yondashuvlar – ya'ni morfologik tahlil, kontekstual model (masalan, HMM, CRF, BERT asosidagi modellari) va semantik tahlil birgalikda qo'llanilganda olmoshlarni to'g'ri teglash darajasi sezilarli darajada oshadi. Bundan tashqari, o'zbek tiliga moslashtirilgan yorliqlash tizimlari, annotatsiyalangan korpuslar, va qo'llanmalarining mavjudligi ilmiy-amaliy izlanishlar uchun mustahkam zamin yaratmoqda. Shu bois, olmoshlarning POS teglashdagi aniqligi tilshunoslik va kompyuter lingvistikasi uchun nihoyatda muhim bo'lib, bu boradagi tadqiqotlar yanada chuqurlashtirilmog'i zarur. O'zbek tilining turkiy tillar orasida o'ziga xos morfologik xususiyatlarga ega ekanini inobatga olgan holda, mahalliy tilshunos va dasturchilar hamkorligida ilg'or va aniq ishlovchi tizimlar ishlab chiqilishi dolzarb masala sanaladi.

Foydalanilgan adabiyotlar:

1. Marie Novotná, Markéta Masopustová. Using Syntax Analyser SET as a GrammarChecker for Czech.2018.9.
2. Elov B., Hamroyeva Sh., Xusainova Z., Xudayberganov N., Yodgorov U., Yuldashev A. POS tagging of uzbek texts using Hidden Markov Models (HMM) and Viterbi algorithm.2023.105-bet.
3. Toshpo'latov Z. Hozirgi o'zbek adabiy tili. –T.: Fan (2017.45-bet)
4. Brayan Roepke. A Quick Guide to Part of Speech Tagging. 2021.14.
5. Ne'matov H., Rasulov R..O'zbek tili system leksikologiyasi asoslari.- T.:Fan, 1975.36-bet
6. Mirtojiyev M. O'zbek tilida polisemiya.-T.:Fan,1975.48.
7. Jurafsky and Martin. Speech and Language Processing, 3rd edition, Draft chapters (Stanford NLP)2018.56.
8. Karel Pala. Comments onCzechMorphological Tagset (2018.67-69)
9. <https://part of speech and taggings, natural language processing>
10. Rada Mihalcea, Raymond Mooney, Dan Jurafsky. Word Classes and Part of Speech Tagging. (2018.10-12)