

MATNLARNI LINGVISTIK TAHLIL QILISHDA OMONIMIYANI HAL ETISHNING KOMPYUTER USULLARI

Isroilov Jasur Baxodir o'g'li,

Namangan davlat universiteti, Mustaqil tadqiqotchi

e-mail: jasurbek9109@gmail.com

Annotatsiya: *Ushbu maqolada tabiiy tillarni qayta ishlashda omonimiya bo'yicha olib borilgan tadqiqot natijalari e'lon qilingan.*

Kalit so'zlar: *omonimiya, NLP, O'zbek korpusi, tabiiy tillarni qayta ishlash, token, lemma, model, lingvistik tahlil.*

КОМПЬЮТЕРНЫЕ МЕТОДЫ РАЗРЕШЕНИЯ ОМОНИМИИ ПРИ ЛИНГВИСТИЧЕСКОМ АНАЛИЗЕ ТЕКСТОВ

Исроилов Жасур Баходир угли,

Наманганский государственный университет, Исследователь

e-mail: jasurbek9109@gmail.com

Аннотация: *В данной статье представлены результаты исследования омонимии при обработке естественного языка.*

Ключевые слова: *омонимия, NLP, узбекский корпус, обработка естественного языка, токен, лемма, модель, лингвистический анализ.*

COMPUTER METHODS OF RESOLVING HOMONYMY IN THE LINGUISTIC ANALYSIS OF TEXTS

Isroilov Jasur Bakhodir o'g'li,

Namangan state university, Researcher

e-mail: jasurbek9109@gmail.com

Abstract: *This article presents the results of the research on homonymy in natural language processing.*

Keywords: *homonymy, NLP, Uzbek corpus, natural language processing, token, lemma, model, linguistic analysis.*

KIRISH

Har qanday tilning shakllanishida tilshunos olimlarning tilga, uni o'rganish, tadqiq qilishga bo'lgan munosabati muhim ahamiyat kasb etadi. Har qanday

millatning millat sifatida o'zligini namoyon etishida aynan til muhim omil bo'lib xizmat qiladi. Globallashuv davrida har bir millat, mustaqil davlat o'z milliy manfaatlarini ta'minlash, bu borada avvalo o'z madaniyatini, azaliy qadriyatlarini, ona tilini asrab-avaylash va rivojlantirish masalasiga ustuvor ahamiyat qaratishi tabiiydir[1].

Hukumatimiz tomonidan ham o'zbek tilini rivojlantirish bo'yicha bir qator tizimli ishlar olib borilmoqda. Prezidentimiz Sh. Mirziyoyev 2020-yil 20-oktabrda "Mamlakatimizda o'zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to'g'risidagi" farmonni [2] imzoladi. Ushbu Farmonda o'zbek tili jonkuyarlari, tilshunos olimlar va mas'ul tashkilotlarga uzluksiz ta'limning barcha turlari uchun imlo lug'ati, sinonimlar lug'ati, frazeologizmlarning katta izohli lug'ati, omonimlar lug'atlarini yatarish, o'zbek tiliga oid barcha ilmiy, nazariy va amaliy ma'lumotlarni o'zida jamlagan elektron ko'rinishdagi o'zbek tili milliy korpusini yaratish vazifalari yuklatilgan. Bugungi kunda yuqorida topshiriqlarni ijrosi bo'yicha bir qator tizimli ishlar olib borilmoqda. O'zida 50 milliondan ortiq so'zni jamlagan, professor N.Abdurahmonova boshchiligida O'zbek milliy korpusi yaratildi. Unda tabiiy tillarni qayta ishlash (NLP) bilan bog'liq bir qator muammolar o'z yechimini topib kelmoqda. Ushbu ilmiy maqolada O'zbek tili korpusi uchun omonimlik hodisasi muammosi va uni yechimlari ishlab chiqilgan.

ADABIYOTLAR TAHLILI:

Tabiiy tillarni qayta ishlash jarayonida omonimiya masalasi bugungi kunga qadar muammoli holatlardan biri hisoblanadi. Bizga ma'lumki mashina tarjimasida, qidiruv tizimlarida, psixolingvistikada, morfoanalizda, sintaktik va semantik analizlarda omonimlar bilan ishlash jarayonida bir qator muammolarga duch kelinadi. Shu jihatdan, omonimiyani hal etish bilan bog'liq tadqiqotlarning rivojlanishiga turtki bo'lmoqda.

Omonimiya umumiy semantik hodisa bo'lib, ayrim holatda mavhumlikka olib keladi, chunki omonim ikki yoki undan ortiq ma'noga ega bo'lishi mumkin va shuning uchun jumla konteksti mavjudligiga qaramay, bir ma'no boshqasiga

almashtirilishi ehtimoli katta. Yuqoridagilardan xulosa qilib aytish mumkinki, omonimlar, nafaqat tarjimada, balki, fanning turli sohalari va zamonaviy axborot almashinuv tizimida ham jiddiy muammolarni keltirib chiqaradi. Jahon tilshunosligida omonimlarga tadqiqotchilar tomonidan turli ta'riflar berilgan. Xususan, S. Ulmann [4] omonimiyani — lingvistik universal deb atagan. S. Ulmanning fikricha, omonimiya statistik universal, ya'ni nisbiy.

Omonimiya nafaqat yevropa olimlari, balki, turkolog olimlarning ham tadqiqot mavzusi bo'ldi. Xususan, rus, qozoq, qirg'iz, tatar, turk, turkman, boshqird, xakas, o'zbek olimlari tomonidan qator tadqiqotlar amalga oshirildi. R.Bekdjanova(qirg'iz tili), T.Kerim (qirg'iz tili), Z.Sattarova (qrim-tatar tili), V.Abayeva, O.S.Axmanova, A.I.Smirniskiy, V.V.Vinogradov, K.Axanov(qozoq tili), M.X.Axtyamov(boshqird tili), A.A.Gasanov (ozarbayjon tili), A.Geldimuradov(turkman tili) kabi turkolog olimlarning tadqiqotlari kompyuter lingvistikasida omonimiyani o'rganish uchun asos vazifasini o'tadi. Korpus lingvistikasida omonimlik muammosini yechish, omonim birliklarni teglash va matnni avtomatik o'qish jarayonida omonimiyani bartaraf etish masalasiga oid qator tadqiqotlar vujudga kelgan. Jumladan, G.I.Kustova, O.N.Lyashevskaya, Ye.V.Paducheva, Ye.V.Raxilina, B.P.Kobrisov, T.I.Reznikova, V.V.Kukanova, A.A.Kretovlar kompyuter lingvistikasida omonimiyani bartaraf etishga qaratilgan qator tadqiqot ishlarini olib borishgan. Omonimiyani bartaraf etish kompyuter tilshunosligining ko'plab dasturlarida, xususan, qidiruv tizimlarida ahamiyatli. Chunki u muayyan so'rov sinflarini qayta ishlash aniqligini oshirishi yoki saqlangan axborot miqdorini kamaytirishi mumkin.

O'zbek tili kompyuter lingvistikasi yo'nalishida amalga oshirilgan sanoqli tadqiqotlarda kompyuter xotirasi omonim birliklarni "tanishi" va "o'qiy olishi"ga mo'ljallangan tahlil dasturlarini yaratish borasida harakatlar mavjud. O'zbek tilidagi omonimlarni teglash muammolari bo'yicha ayrim mulohazalar, omonimiyani aniqlash algoritmini tuzish bo'yicha dastlabki harakatlar amalga oshirilgan. Tadqiqotchilar M.A'lamova, N.Abduraxmonova, Sh.Gulyamova,

X.Axmedova, G'.Abduvahobov kabi qator tadqiqotchilarning izlanishlari bunga yaqqol misoldir.

TADQIQOT METODOLOGIYASI

Jahon korpus lingvistikasida omonim birliklarni teglash va matnni avtomatik o'qish jarayonida omonimiyani bartaraf etish masalasi alohida tadqiq qilingan. Ammo o'zbek korpus lingvistikasi yosh soha bo'lgani bois omonimlarni semantik teglash va farqlash masalasiga bag'ishlangan tadqiqotlar juda kam. Shu boisdan leksik omonimlarni farqlash uchun olib borilgan tadqiqotlarni o'rganishda turkiy tillar va rus tilshunosligidagi manbalarga tayanamiz.

Omonimlarni o'rganish natijasida korpusda omonimlarni izlash va tokenlarga ajratishda quyidagi holatlar asosiy muammolarni keltirib chiqarishi aniqlandi.

a) Eski o'zbek tilidagi hozirda iste'moldan chiqqan so'zlar.

Sh.Rahmatullayevning ta'kidlashicha, omonim, avvalo, til doirasida belgilanadi. Masalan, hozirgi o'zbek tilida ishlatilmaydigan (eski o'zbek tiliga mansub) leksema bilan hozirgi o'zbek tilidagi leksemani o'zaro omonim deyish o'rinli emas. D.B.Axmedovaning fikriga ko'ra hozirgi o'zbek tilida yozilgan asarlar matnida eski o'zbek tiliga mansub leksemalar anchagina uchraydi, shuni hisobga olib, hozirgi o'zbek tilidagi leksemaning omonimi bo'lgan ba'zi istorizm leksemalar lug'atga kiritilgan. Masalan:

dol I ot. Bolalarning o'yinida ganga (qubbaga) birinchi bo'lib urish huquqini beradigan holat, soqqaning ganga yaqinroq ekanligi. "Kim dol?" – "Men dol".

dol II ot; eski. Arab alifbosidagi sakkizinchi harfning nomi. Bu so'z dol bilan emas, zol bilan yoziladi.

dol III ravish; eskirgan; so'zlash. Xuddi, naq, to'ppa-to'g'ri. Nishonning dol o'rtasiga tekizmoq.

b) Ayrim atoqli va turdosh otlar.

Matnlar tarkibida ayrim atoqli otlar va sifat so'z turkumidagi so'zlarni farqlashda noaniqliklar kelib chiqadi. Bu holat, asosan, sifat so'z turkumidan otga ko'chgan so'zlarda kuzatiladi. Masalan, *go'zal*, *botir*, *jasur* kabi sifat so'z

turkumidagi soʻzlar atoqli otga koʻchgan va ot soʻz turkumida hisoblanadi. Atoqli otlar doimo bosh harflar bilan yoziladi. Bunday omonimlarni modellash tirishda

$$[HW \rightarrow W.cl \rightarrow Noun] \quad (2)$$

modelidan foydalanishimiz mumkin. Bunda $HW \rightarrow$ omonim soʻz, agar soʻz bosh harf bilan yozilsa ($W.cl$), u holda bu soʻz ot ($Noun$) turkumida. $W.cl$ – capital letter, yaʼni bosh harf. Ammo ayrim atoqli otga koʻchgan sifatlar gap boshida kelib, bosh harf bilan yozilsa, ushbu modelga koʻra dastur sifat soʻz turkumidagi soʻzni ham ot deb oʻqishi mumkin. Bunday omonim soʻzlarni faqatgina kontekstdan aniqlash mumkin.

c) Urgʻu bilan farqlanadigan soʻzlar

Tilimizga chet tilidan oʻzlashma soʻzlar kirib kelishi natijasida oʻzbek tili lugʻatida urgʻu bilan maʼnosi farqlanadigan omonim soʻzlar yuzaga keldi. Masalan, soya soʻzi urgʻu olishiga koʻra maʼnolari farqlanadi. Urgʻu I. soya soʻzining oxirgi boʻgʻiniga tushsa, salqin joy, oftob tushmaydigan degan maʼnolarda qoʻllanadi. Urgʻu II. Soya soʻzining birinchi boʻgʻiniga tushsa, isteʼmol mahsuloti maʼnosida qoʻllanadi. Matnlardagi bunday urgʻu bilan farqlanadigan soʻzlarni analiz qilishda ular oʻzidan oldin va keyin keladigan birikmalar bazasini yigʻish maqsadga muvofiq.

d) Qoʻshimchalar qoʻshilishi natijasida yuzaga keladigan ayrim soʻz shakllar.

Soʻz asosiga ayrim yasovchi va grammatik qoʻshimchalar qoʻshilishi natijasida omonimlik hosil boʻladi. Bunga bir nechta misollar va ularning modellarini koʻramiz.

changi

I. Changi: Noun+i (egalik qoʻsh). $[Noun+N.gsh=HW1.N]$

II. Changi: Noun+I (feʼl yas). $[Noun+i (yas)= HW2.V]$.

terim

I. Terim: Noun+im(egalik qoʻsh). $[Noun+N.gsh=HW1.N];$

II. Terim: Noun+m(egalik qoʻsh). $[Noun+N.N.gsh=HW2.N];$

III. Terim: Verb+im(yasovchi). $[Verb+im=HW3.N]$.

ko‘zlar

I. Ko‘zlar: Noun+lar(ko‘plik). [Noun+N.gsh=HW1.N]

II. Ko‘zlar: Noun+la (yasovchi)+r (sh.yas). [Noun+la+V.gsh=HW2.V]

Sana

I. Sana: Noun+a (fe‘l yas). [Noun+a (yas)=HW1.V]

II. Sana: Noun. [HW2.N]

ko‘tara

I. Ko‘tara: Verb+a (sifat yas). [Verb+a (yas)= HW1.Adj]

II. Ko‘tara: Verb + (fe‘l lug‘. sh). [Verb+V.gsh= HW2.V].

yuvindi

I. Yuvindi: Verb + in (fe‘l lug‘. sh.) +di (fe‘l zamon). [Verb+V.ns+V.zm=HW1.V]

II. Yuvindi: Verb+ indi (ot yas.). [Verb+indi(yas). [Verb+indi(yas)= HW2.N]

Suvchi

I. Suvchi: Noun+chi (ot yas). Turdosh otdan shaxs oti yasalyapti.

[Noun+chi=HW1.N]

II. Suv-chi: Noun+ -chi (kuchaytiruv-ta‘kid yuklamasi). [Noun+ chi.part=HW2.N]

Suvchi so‘zini farqlash uchun yuqoridagi modellardan foydalanamiz, ya‘ni agar -chi qo‘shimchasi so‘zga qo‘shib yozilsa, yasovchi affiks. Agar otdan keyin -chi affiksi chiziqcha bilan yozilsa, yuklama hisoblanadi. Part qisqartma shakli yuklamani ifodalaydi.

Senla

I. Senla: Pron+la (fe‘l yas.) = Verb. [Pron+la(yas)=HW1.V]

II. Sen-la: Pron+ -la (bilan ko‘makchisining qisqartma shakli). Sen bilan birga ma‘nosida. [Pron+ -la (Pr.)= HW1.Pron.Pr]

Fe‘l so‘z turkumidagi senlamoq va sen-la shaklini farqlashda yuqoridagi modeldan foydalanamiz. Bu yerda agar –la so‘zga qoshib yozilsa, yasovchi affiks va fe‘l so‘z turkumidagi leksemani hosil qiladi. Agar –la so‘zda chiziqcha bilan ajratib yozilsa, ko‘makchining qisqartma shaklini ifodalaydi.

e) Sifat va ot so‘z turkumidagi so‘zlar

Sifat ot soʻz turkumidagi soʻzlar bir xil grammatik shakllarni qabul qilganligi sababli analiz qilish jarayonida turkumlarga ajratishda xatoliklar kelib chiqadi. Sifatlar otlashganda otga xos boʻlgan egalik, kelishik va koʻplik qoʻshimchalarini qabul qiladi. Bu turkumdagi soʻzlarni farqlash uchun ular birikib keladigan birikmalar bazasini yigʻish maqsadga muvofiq. Ammo shunda ham turkumlarga ajratishda bu modellar xatoliklarni butunlay bartaraf etmaydi. Faqat gap konteksi orqaligina soʻzlar turkumini aniqlash mumkin.

Yuqoridagi omonimlarning filologik modellari algoritmlashtirilib, token boʻyicha soʻzlarni qidirilganda quyidagi xatoliklar aniqlandi. (1-rasm) Istisno tariqasida aniqlangan xato soʻzlar uchun alohida filologik model ishlab chiqildi.

soʻz	soʻz turkumi	forma	izohi	model
ari	Noun		verb chiqyapti	ari+N.gsh/V.gsh
ot	Noun	ot+a+l+adigan	noun chiqyapti	ot+a+V.gsh
aya	Noun		teg topilmadi	aya+N.gsh/aya+V.gsh
bit	Verb	bit+gan+da	Noun chiqyapti, teg	bit+V.gsh
bor	M.V	... bor	noun+P3+SG	bor+WN.Ø/edi, ekan, emish
burun	Adv		noun+P3+SG	burun+N.gsh
daydi	Adj		Verb+imp.+P2+SG	daydi+Adj.L/N.gsh
yon	Verb		ADV chiqyapti	yon+N.gsh
yor	Verb	yor+il+ib	Noun chiqyapti	yor+V.gsh
yot	Verb	yot+gan+i+ni	Adj chiqyapti	yot+V.gsh
yoz	Verb	yoz+ib	Noun chiqyapti	yoz+V.gsh
yoy	Verb	yoy+gan+da	Noun xato	yoy+V.gsh
yosh	Adj		Noun	yosh+Adj.gsh/N.gsh
yogʻ	Verb		P	yogʻ+ V.gsh
kechik	Verb	kechik+kan+i+ga	Noun xato	kechik+V.gsh
kuy	Verb	kuy+guncha	Noun xato	kuy+V.gsh
kul	Verb	kul+ib	Noun xato	kul+V.gsh
kurash	Verb	kurash+a+man	Noun xato	kurash+V.gsh
koʻk	Adj		Lemma topilmadi	koʻk+N.gsh/Adj.L
koʻr	Verb		Lemma topilmadi	koʻr+ V.gsh
koʻch	Verb		Lemma topilmadi	koʻch+ V.gsh

1-rasm: Istisno soʻzlarni filologik modellari

XULOSA

Olib borilgan tadqiqotlar shuni koʻrsatmoqdaki, butun dunyo hamjamiyatida mukammal semantik analizator yaratilmagan. Lekin semantik analizator yaratish uchun dastlabki ishlar amalga oshirilgan. Omonimlik hodisasi boʻyicha ilmiy-tadqiqot ishlarimizni davom ettirib, kelgusida oʻzbek korpusida omonimlik hodisasini aniqlay oladigan mukammal dastur tuzish ishlarini olib boramiz.

FOYDALANILGAN ADABIYOTLAR

1. Oʻzbekiston Respublikasi Prezidenti Shavkat Mirziyoyevning 2019- yil 21-oktabrdagi “Oʻzbek tilining davlat tili sifatidagi nufuzi va mavqeyini tubdan

oshirish chora-tadbirlari to'g'risida"gi PF-5850-son farmoni. <https://lex.uz/docs/-4561730>

2. O'zbekiston Respublikasi Prezidenti Shavkat Mirziyoyevning 2020-yil 20-oktabrdagi "Mamlakatimizda o'zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to'g'risida"gi PF-6084-son farmoni. <https://lex.uz/docs/5058351>.

3. Кустова Г.И., Ляшевская О.Н., Падучева Е.В., Рахилина Е.В. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. М.: Индрик, 2005. – С. 155-174

4. Ulman. S. Ruination of our mother tongue GAnG GAnG. The Canberra Times.1970. –P. 61

5. Бекджанова Р. Омонимы в киргизском языке (на кирг. яз.): Дисс. канд. филол. наук. Фрунзе, 1967. 182 с

6. Abdurahmonova N. O'zbek tili elektron korpusining kompyuter modellari: Filol. Fanlari Doctory (DSc) ... diss. mavhum – Toshkent, 2021 yil

7. Gulyamova Sh.K. Semantik analizator uchun omonimlikni farqlashning ayrim masalalari xususida // "O'zbekistonning umidli yoshlari" mavzusidagi Respublika ilmiy talabalar, magistral yosh tadqiqotchilar va mustaqil izlanuvchilar uchun onlayn konferensiya materiallari to'plami. -Toshkent:

8. Abduvahobov G'. Yangi avlod omonimlar o'quv lug'ati mobil ilovasini yaratishning lingvistik asoslari: filol. Fan bo'yicha falsafa dok. (PhD) ...diss. avtoref. –Farg'ona, 2023. –B.25