

**ЎЗБЕКИСТОН РЕСПУБЛИКАСИ ВАЗИРЛАР МАҲКАМАСИ ҲУЗУРИДАГИ ДАВЛАТ  
ТИЛИНИ РИВОЖЛАНТИРИШ ДЕПАРТАМЕНТИ**

**ЎЗБЕКИСТОН РЕСПУБЛИКАСИ ОЛИЙ ВА ЎРТА МАХСУС ТАЪЛИМ ВАЗИРЛИГИ**

**ЎЗБЕКИСТОН РЕСПУБЛИКАСИ ХАЛҚ ТАЪЛИМИ ВАЗИРЛИГИ**

**ЎЗБЕКИСТОН РЕСПУБЛИКАСИ ИННОВАЦИОН РИВОЖЛАНИШ ВАЗИРЛИГИ**

**АЛИШЕР НАВОИЙ НОМИДАГИ  
ТОШКЕНТ ДАВЛАТ ЎЗБЕК ТИЛИ ВА АДАБИЁТИ УНИВЕРСИТЕТИ**

**“ЎЗБЕК МИЛЛИЙ ВА ТАЪЛИМИЙ  
КОРПУСЛАРИНИ ЯРАТИШНИНГ НАЗАРИЙ  
ВА АМАЛИЙ МАСАЛАЛАРИ”**

**мавзусидаги  
халқаро илмий-амалий конференция  
материаллари  
(2021 йил 7 май)**

## OMONIM SO‘ZLARNI FARQLASH USULLARI

### HOW TO DISTINCE HUMAN WORDS

Elov Botir Boltayevich\*  
Axmedova Xolisa Ilxomovna\*  
Abjalova Manzura Abdurashetovna\*

**Annotatsiya:** Tabiiy tillarni qayta ishlash jarayonida qator muammolar o‘z yechimini kutmoqda. Shunday muammolardan biri so‘zlarni semantik tahlil qilish muammosidir. Bu muammoni yechish uchun esa so‘z turkumlarini teglash usullari va modellari bizga eng muhim va samarali yechimdir. Qoidalarga asoslangan teglash usullari yordamida ikkita so‘z turkumi doirasidagi omonim so‘zlarni farqlash modellari haqida ushbu maqolada so‘z boradi.

**Kalit so‘zlar:** Tabiiy tillarni qayta ishlash, qoidalarga asoslangan teglash usuli, omonim so‘zlar, to‘plamlar, matematik model.

**Annotation:** A number of processes in the process of processing natural languages are awaiting resolution. One such problem is the problem of semantic analysis of words. To solve this problem, the methods and models of tagging are the most important and effective solution for us. This article discusses models for distinguishing homonymous words within two word groups using rule-based tagging methods.

**Keywords:** Natural Languages Processing, rule-based tagging method, homonymy words, sets, math model.

Tabiiy tillarni qayta ishlash jarayoni (NLP – Natural Languages Processing) ning muammolarini ketma-ket modellashtirish zamon talabidir. Ayniqsa so‘z turkumlarini teglash qadimiy va eng mashhur muammolardan hisoblanadi. So‘z turkumlarini teglashtirishdagi asosiy maqsadimiz gap tuzilishini modellashtirishdan iborat. Masalan, gapni o‘qish va qanday so‘zlar ot, olmosh, fe‘l, ravish va hokazo vazifasini bajarishini aniqlash nutq teglarining bir qismi hisoblanadi. Korpus tilshunosligida so‘z turkumlarini teglash grammatik teglash yoki so‘z turkumlariga ajratish deb ham ataladi, bu matn (korpus) dagi so‘zni nutqning ma‘lum bir qismiga mos keladigan tarzda belgilash, uning ta‘rifi va kontekstdan kelib chiqqan holda, ya‘ni o‘zidan oldin va keyin kelgan va o‘zaro bog‘liq ibora, jumla yoki paragrafdagi so‘zlar.

O‘zbek tilida turkumlarni teglash jarayonida [Michael Collins, 2011] keltirilgan teglash usulining ikkalasidan ham foydalanamiz. Masalan, o‘zbek tilidagi ikki so‘z turkumi doirasidagi omonim so‘zlarning ma‘nolarini kontekstdan aniqlash jarayonida qoidalarga asoslangan teglash usuli hamda stoxastik teglash usuli ham muhim ahamiyat kasb etadi. Ushbu maqolada qoidalarga asoslangan teglash usuli yordamida omonim so‘zlarni teglash jarayonini ko‘rib chiqamiz.

#### **Qoidalarga asoslangan teglash**

So‘z turkumlarini avtomatik teglash bu tabiiy tilni qayta ishlash jarayoni bo‘lib, unda statistik metodlar bilan qoidalarga asoslangan usullardan ko‘ra ko‘proq muvaffaqiyatga erishilgan. Odatda qoidalarga asoslangan yondashuvlar noma‘lum yoki noaniq so‘zlarga teglar belgilash uchun kontekstli ma‘lumotlardan foydalanadi. So‘zning lisoniy xususiyatlarini ajratish, oldingi, keyingi so‘zini va boshqa jihatlarini tahlil qilish orqali amalga oshiriladi. Masalan, oldingi o‘rindagi so‘z *kitob* bo‘lsa, u holda bu so‘z ot bo‘lishi kerak. Shu kabi boshqa so‘z turkumlari uchun ham qoidalar keltirish mumkin. Agar noma‘lum X so‘zidan oldin aniqlovchi va undan keyin ot kelsa, u holda bu so‘z sifat deb belgilanadi.

Yuqoridagi keltirilgan qoidalar to‘plami orqali aniqlash juda murakkab jarayon bo‘lib, u doimiy emas. Shunga qaramasdan, o‘zbek tilida ba‘zi jarayonlar aynan shu qoidalar to‘plamiga bo‘ysunadi.

\* Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti texnika fanlari bo‘yicha PhD.  
[ebb@mail.ru](mailto:ebb@mail.ru)

<sup>20</sup>\* Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti tayanch doktoranti.  
[xolisa9029@mail.ru](mailto:xolisa9029@mail.ru)

<sup>21</sup>\* Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti filologiya fanlari bo‘yicha PhD.  
[manzura\\_ok@mail.ru](mailto:manzura_ok@mail.ru)

## Omonim soʻzlar

Tilda soʻzlar shakl va maʼno munosabatiga koʻra omonim, sinonim, antonim va paronim soʻzlarga boʻlinishi barchaga maʼlum. Bu esa nutqning chiroyli, taʼsirchan ifodalanishi va leksemalardagi nozik maʼno qirralarini farqlashga yordam beradi. Shakli bir xil, maʼnosi har xil soʻzlar omonim soʻzlardir. Shu jumladan, omonimlar ham shakli va qoʻllanilish uslubi jihatdan alohida oʻringa ega soʻzlardir. Omonimlarning yuzaga kelish sabablari har xildir:

1. Tilda azaldan mavjud boʻlgan ayrim soʻzlarning shakllari tasodifan teng boʻlib qoladi: **baqa**-qurbaqa, **baqa**-suv tegirmoni parrakdan harakat olib, tegirmon toshini aylantiruvchi metall qismi.
2. Bir maʼnoli soʻz leksik maʼnoning koʻchishi oqibatida koʻp maʼnoli soʻzga aylanadi, keyinroq bosh maʼno va hosila maʼno oʻrtasidagi bogʻlanish unitilib, bir soʻz negizida ikki boshqa-boshqa leksema paydo boʻladi: **kun**-quyosh; **kun**- sutka, kun chiqqandan yana kun chiqqungacha boʻlgan vaqt.
3. Boshqa tillardan oʻzlashtirilgan ayrim leksemalar oʻzbek tilidagi u yoki bu leksemaga shaklan teng boʻlib qoladi: **toy** (fors-tojik)-katta toʻp qilib taxlab yoki bosib bogʻlangan mol va shu tarzda mol oʻlchovi.
4. Boshqa tillardan oʻzlashtirilgan leksemalar orasida shaklan teng boʻlgan soʻzlarning uchrashi omonimiyaga olib keladi: **surat** (arabcha)-rasm, **surat** (arabcha)-urf-odat.
5. Leksemalarning yasalihi ham baʼzan omonimlarni keltirib chiqaradi: **qoʻnoq**-tariq, **qoʻnoq** (qoʻn+oq)-mehmon kabi.

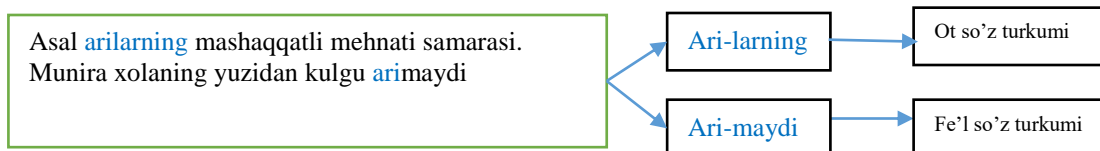
Omonim soʻzlarning paydo boʻlishi sabablaridan koʻra ularning qaysi soʻz turkumlariga mansubligi, qanday qoidalar asosida maʼnolarini farqlanishi muhim.

Soʻz turkumlarini teglashning qoidalarga asoslangan usulidan biz bir turkum doirasidagi omonimlikni farqlashda foydalanamiz. Quyidagi misollarga qaraymiz.

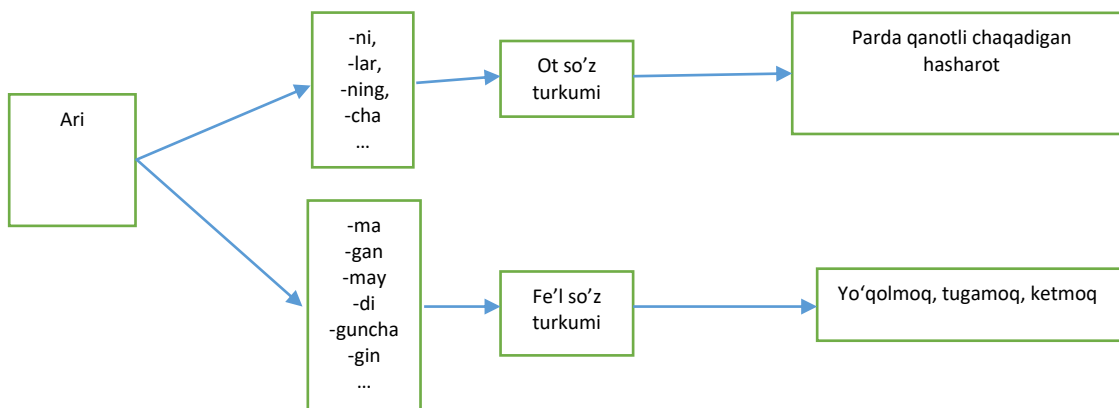
Asal-arilarning mashaqqatli mehnati mahsulidir.

Munira xolaning yuzidan kulgu hech arimaydi.

Ushbu jumalarda keltirilgan ari soʻzining qaysi soʻz turkumiga tegishliligi va gapda qanday maʼnolarda kelishini aniqlashimiz lozim.



Qoidalarga asoslangan teglash usulining mazmunidan kelib chiqib, quyidagi koʻrinishni hosil qilamiz, yaʼni ot soʻz turkumi va feʼl soʻz turkumi orasidagi omonim soʻzlarning maʼnolari qoidalarga asosan teglash usuli yordamida farqlanadi. Yuqorida keltirilgan ikkita jumla tarkibida berilgan omonim soʻzlarni oʻzak va qoʻshimchalarga ajratilishi va maʼnolari farqlanishini quyidagi grafikda koʻramiz:



Shunga oʻxshash:

*Pahlovonlar oʻrtasidan yoydan oʻq otish musobaqasi oʻtkazildi.*

*Uydirma gaplarni yoyib gap soʻz qilish gʻiybatchilarning ishi.*

jumlalarda ham yuqorida keltirilgan grafiklarni takrorlash mumkin. Bu ma'lumotlar asosida ot va fe'l so'z turkumlari orasidagi omonimlikni farqlash uchun kerak, bunday holda quyidagi matematik modelni keltiramiz:

$$H_{N,V} = \begin{cases} N + x, & x \in \{aff_{s,l_i}^N\}, i = 1..n \\ V + x, & x \in \{aff_{s,l_j}^V\}, j = 1..m \end{cases}$$

$H_{N,V}$ -Ot va fe'l so'z turkumlari orasidagi omonimlik,  $N$ -Ot,  $x$ -lug'aviy va sintaktik shakl yasovchi qo'shimchalar kombinatsiyasi,  $V$ -fe'l,  $aff_{s,l_i}^N$ - OT so'z turkumidagi omonim so'zga qo'shiluvchi sintaktik va lug'aviy shakl yasovchi qo'shimchalar va ularning kombinatsiyalaridan iborat to'plam,  $aff_{s,l_j}^V$ -Fe'l so'z turkumidagi omonim so'zga qo'shiluvchi sintaktik va lug'aviy shakl yasovchi qo'shimchalar va ularning kombinatsiyalaridan iborat to'plam.

Birinci namunadagi jumlalardagi omonim so'zlar uchun ushbu keltirilgan matematik formulani tadbiq etamiz.  $H_{N,V} = \{ari\}$  – ot va fe'l so'z turkumlari doirasidagi omonim so'zlar to'plamining bir elementi:

$aff_{s,l_i}^N = \{-ni, -ning, -lar, -da, -ga, -larga, -larni, -larning, -larda, -cha, -lardagi, -chada, -chaga, -chani, -channing, -chalarini, -chalarga, -chalarning, -chalarda, -larniki, -niki, -dir, -nikidir\}$

$aff_{s,l_j}^V = \{-di, -maydi, -gan, -maydigan, -masdi, -masdan, -magan, -masin, -sin, -ydi\}$

$x = \{ff_{s,l_j}^V, aff_{s,l_i}^N, aff_{s,l_i}^{Adj}, aff_{s,l_i}^{Adv}\}$ -lug'aviy va sintaktik shakl yasovchi qo'shimchalar va ularning kombinatsiyalaridan iborat to'plam. Endi ushbu matematik model asosida algoritm yaratamiz. Algoritm uchun quyidagi belgilashlarni kiritamiz:

$Z$ -berilgan matn;

$S_i$  (inglizcha Sentences so'zining bosh harfidan olingan)-matndagi gaplardan iborat to'plam  $i = 1..n$ ;

$W_j$  –  $S_i$  ning tarkibidagi gaplardan iborat to'plam,  $j = 1..m$ ;

$H_l$  – o'zbek tilida mavjud bo'lgan ot va fe'l so'z turkumlari doirasidagi omonim so'zlardan iborat to'plam,  $l = 1..q$ ;

$P_{l1} = \{., !, ?, !?, "", ..., !!!, <<>>\}$ -gapni yakunlovchi belgilar,  $l1 = 1..8$ ;

$D_k$  – o'zbek tilida mavjud omonim so'zlarning izohlaridan iborat to'plam,  $k = 1..t$ .

1.  $Z$  matn  $P_{l1}$  belgilar yordamida  $S_i$  ga ajratib chiqiladi.
2. Har bir  $S_i$  gapni  $W_j$  ga o'zlashtirib chiqiladi.
3. Endi esa har bir  $W_j$  so'zni o'zak va qo'shimchalarga ajratib chiqiladi.
4. Ajratilgan o'zak so'zni  $H_i$  so'zlar bazasidagi so'zlar bilan solishtirib chiqiladi.
5. Agar ajratilgan o'zak so'z  $H_i$  to'plamda mavjud bo'lsa, uni ma'lumotlar bazadagi ID raqami aniqlanadi va o'zak so'zdan keyin kelgan qo'shimcha yoki qo'shimchalar kombinatsiyadan iborat keltma-ketlik ajratib olinadi va  $q$ -o'zgaruvchisiga o'zlashtiriladi aks holda keyingi so'zga o'tiladi.



6. Endigi qadamda oldingi qadamda aniqlangan  $q$  ni  $x$  to‘plam ichidagi to‘plamlar elementlari bilan solishtirib chiqiladi, agar  $q \text{ aff}_{s-l_i}^N$  -to‘plam elementlari orasida mavjud bo‘lsa 7.
7. 5-qadamda aniqlangan ID bo‘yicha  $D_k$  to‘plamdan Omonim so‘zning izohi va qaysi so‘z turkumida tegishli ekanligi aniqlanadi.
8. Natijalar ekranga chiqariladi.
9. Shu tariqa kiritilgan matn gap va so‘zlarga ajratiladi, so‘zlar esa o‘z navbatida o‘zak va qo‘shimchalarga ajratiladi, ajratilgan o‘zak so‘z omonim so‘zlar bazasidagi so‘zlar bilan solishtiriladi, uning qo‘shimchalari esa qo‘shimchalar bazasi bilan solishtirilib omonim so‘zlar farqlanadi. Biz yuqorida ko‘rib chiqqan model qoidalarga asoslangan teglash usulini o‘zbek tili semantikasida qo‘llashga misol bo‘la oladi.

#### ADABIYOTLAR:

1. Michael Collins. Tagging with Hidden Markov Models //2011.
2. Divya Godayal, An introduction to part-of-speech tagging and the Hidden Markov Model// 2018.
3. Кутузов А.Б. Корпусная лингвистика. – (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс)- // lab314.brsu.bAy/kmp-lite/kmp-video/CL/CorporeLingva.pdf;
4. Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – Санкт-Петербург. –2006. – 26 с.
5. Mengliyev B. va b. O‘zbek tilining milliy korpusi // Ma’rifat. – 26/04/2018.
6. Raxmatullayev Sh. Omonimlarning izohli lug‘ati. – Toshkent: O‘qituvchi, 1984. – 214 b.