



ISSN  
2181-1784  
SJIF: 5.423



# ORIENTAL RENAISSANCE:

*Innovative, educational,  
natural and social sciences*

[www.oriens.uz](http://www.oriens.uz)

*Exact sciences*  
*Natural sciences*  
*Engineering sciences*  
*Pedagogical sciences*  
*Social and Human sciences*  
*Philological sciences*  
*Philosophy sciences*  
*Economic sciences*

**No10  
2021**



## **OMONIMIYA VA LINGVISTIK TIZIMLARDA OMONIMLARNI ANIQLASH USULLARI**

**Abjalova Manzura Abdurashetovna**

Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti  
dotsent v.b., filologiya fanlari bo'yicha falsafa doktori ( PhD).

[abjalova.manzura@gmail.com](mailto:abjalova.manzura@gmail.com)

### **ANNOTATSIYA**

*Omonimlikni aniqlash tabiiy tilni qayta ishlash (NLP, Natural Language Processing)da dolzarb masalalardan biri hisoblanadi. Mazkur masalada o'zbek tilidagi matnlarda uchraydigan omonimshakllarni aniqlash, tahlil qilishning bir necha usullari xususida so'z yuritildi.*

***Kalit so'zlar:** omonimshakl, usul, yashirin Markov modeli, N-gramma, so'z birikmalari modeli.*

### **ABSTRACT**

*Determining homonymy is one of the most pressing issues in Natural Language Processing (NLP). There are several ways to identify and analyze homonymous forms in Uzbek texts.*

***Keywords:** homonym, method, hidden Markov model, N-gram, phrase model.*

### **АННОТАЦИЯ**

*Определение омонимии - одна из самых актуальных проблем в обработке естественного языка (NLP). Существует несколько способов выявления и анализа омонимических форм в узбекских текстах.*

***Ключевые слова:** омоним, метод, скрытая марковская модель, N-грамма, фразовая модель.*

### **KIRISH**

Matnlarni avtomatik qayta ishlash bir necha bosqichga bo'linadi va ulardan biri morfologik tahlil bosqichi hisoblanadi. Mazkur bosqichda har bir so'zga morfologik tavsif beriladi: lemma [1]si (asosi), kelishigi, soni, darajasi, nisbati, shaxsi va hk. Morfoanalizning so'zlarni morfologik teglash vazifasi omonimshakllar bilan murakkablashadi.

Ma'lumki, omonim so'zlar shakli bir xil, ammo semantikasi turfa xil bo'lgan leksik birliklar hisoblanadi. Avtomatik qayta ishlashda omonimlik hodisasi quyidagi birliklarda mavjud:

1. so'z omonimligi – shakldoshlik so'z asosida bo'ladi, ya'ni muayyan so'z bir so'z turkumi yoki bir necha turkumga mansub ma'no beradi. Masalan:

ot	modal
Avval birliklarni, keyin oʻnliklarni qoʻshamiz.	Suv bor joyda hayot bor. Ruchkang bormi?
ot	hisob soʻz
Kuch – birlikda.	Bir necha bor taklif yubordim.
ot	ot
Ogʻirlik birliklari. Til birliklari	<i>Bor – kimyoviy element.</i>
	feʼl
	Ishga bormoq. Maktabga bormoq.

Birinchi ustunda “birlik” soʻzi bir soʻz turkumi doirasida omonim hisoblanadi: 1) oʻngacha boʻlgan butun son; grammatik koʻplik aksi (ot); 2) birlashish, hamjihatlik (ot); 3) bir turdagi miqdorlarni oʻzaro baholash uchun qabul qilingan oʻlchov; til qurilishiga xos termin (ot).

Ikkinchi ustunda “bor” soʻzi ikki xil turkumga mansub shakldoshlikni yuzaga keltirgan: 1) mavjud (modal); 2) marta, dafʼa, bora (hisob soʻz)

2. Qoʻshimcha omonimligi – muayyan qoʻshimcha vazifasiga koʻra qoʻshimchalarning turli guruhiga mansub boʻladi. Masalan:

-ki	<i>koʻchki, tepki, turtki ustki, ichki, kechki</i>	(ot yasaydi) (sifat yasaydi)
-	<i>uying, kitobing, ishing</i>	(shakl yasaydi: sintaktik mun.shakli – egalik qoʻshimchasi)
(i)ng	<i>koʻring, boring, tayyyorlang</i>	(shakl yasaydi: sintaktik mun.shakli – shaxs-son qoʻshimchasi)

3. Iboralar omonimligi – shakli, yaʼni tuzilishi bir xil, ammo semantikasi turlicha boʻlgan frazeologik birliklar.

qattiq shovqin soldi	yuksak darajada izzat-hurmat qildi
<i>uyini boshiga koʻtarmoq</i>	<i>onasini boshiga koʻtarmoq</i>

4. Gap omonimligi – muayyan gap ifoda maqsadi yoki mazmuniga koʻra farqlanadi.

darak gap	soʻroq gap
Ishni bajarmadim.	Ishni bajarmadim?

## MANBALAR TAHLILI

Rus tilida omonimlik turlarida soʻzlarning turkumligi boʻyicha omonimlik, morfologik omonimlik va leksik omonimlik farqlanadi [2]. Eʼtiborli jihati shundaki soʻzlarni morfologik va leksik omonimligi boʻyicha guruhlanishi tabiiy tilni qayta ishlashda muhim ahamiyat kasb etadi.

Morfologik omonimlikda bir turkumga mansub boʻlgan shakldosh soʻzlar lemma (asosi)si turlicha, ammo muayyan shakllaridagina omonimlikni yuzaga keltiruvchi soʻzshakllar eʼtiborga olinadi. Masalan:

lemmasi *ter*

***terim*** – mening terim

(-im egalik qoʻshimchasi: I shaxs, birlik)

lemmasi *terim*

***terim*** – hosil

Leksik omonimlikda bir lemmaga mansub soʻz turli maʼnolarni beradi:

ot

**bogʻ** – toʻdalab bogʻlangan holat

ot

**bogʻ** – oʻsimlik va daraxtlar  
koʻp ekilgan joy

Bogʻlamoq, bogʻlam. Bir bogʻ  
piyoz

bogʻ-rogʻ, bogʻ-boʻston;  
Uzumini ye, bogʻ ini surishtirma.

Omonimshakllarning morfologik va leksik guruhlanishi matnlarni qayta ishlovchi dasturiy taʼminot va tizimlarda lemmatizatsiya va stemming jarayonlari uchun muhim sanaladi [3]. Har ikki texnologiya soʻz yoki soʻzshaklning asosini topishga yoʻnaltirilgan boʻlib,

Taʼkidlash oʻrinliki, omonimlik hodisasi avtomatik qayta ishlash jarayonida eng dolzarb masala hisoblanadi. Shu bois NLPda omonimshakllarni aniqlash va ularni tahlil qilish maxsus oʻrganiladi, hatto bir necha usullar ham ishlab chiqilgan.

## METODLAR

Omonimlikni aniqlash metodlarining barchasi ikki guruhga boʻlinadi:

1. Qoidalarga asoslangan usullar. Oʻz navbatida, ular quyidagilarga boʻlinadi:

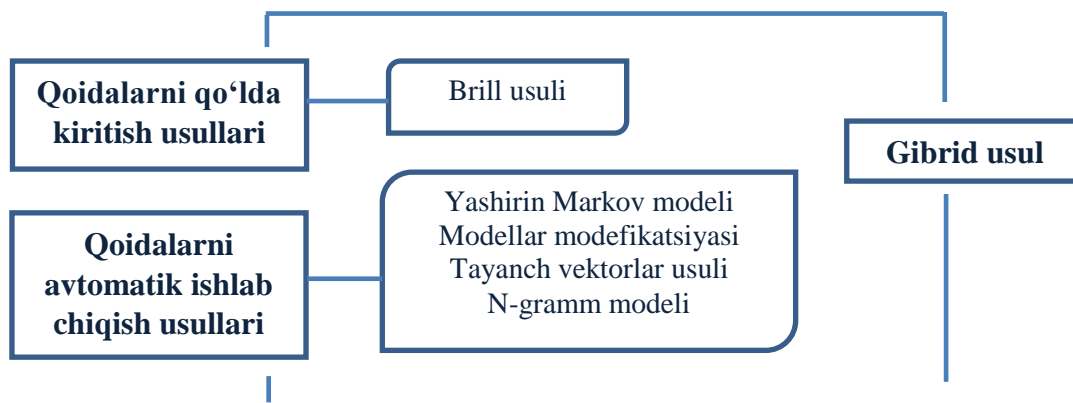
a) Qoidalarni qoʻlda kiritish usullari.

b) Qoidalarni avtomatik ishlab chiqarish usullari.

2. Statistika asoslangan usullar.

Ushbu guruhlarning har birining oʻziga xos afzalliklari va kamchiliklari mavjud. Bunday vaziyatlarda tez-tez sodir boʻladiganidek, ikkala guruhning xususiyatlarini

(va afzalliklarini) bir usulda birlashtirish avval erishilgan natijalarga qaraganda yaxshiroq natijani ko'rsatishi mumkin. Bunday usul gibril usuli deb nomlanadi.



1-sxema. Omonimlikni aniqlash usullari.

Mazkur usullarga tayanuvchi tizimlar o'z navbatida quyidagi guruhni tashkil etadi:

1. Qo'lda yaratilgan qoidalarga asoslangan tizimlar.
2. Ehtimoliy modellar asosida yaratilgan va tavsiflangan korpuslarga tayanadigan tizimlar.
3. Ehtimollik modellari va qoidalarga asoslangan gibril tizimlar.

Omonimiyani aniqlash uchun har bir so'zshaklni "tasniflash" kerak, ya'ni uning lemmasi, so'z turkumi va bir tegga birlashuvchi morfologik xususiyatlar to'plami bilan bog'lab qo'yiladi.

Yashirin Markov modeli Baum L.E. va uning hamkasblari tomonidan ishlab chiqilgan [6] mazkur model omonimlikni aniqlashning statistik metodi statistik jarayonda yuzaga keladigan barcha variantlar ehtimoligini hisobga olishga yordam beradi. Masalan, ma'lum bir matnda ot turkumiga oid so'zlar bog'lovchiga nisbatan tez-tez va ko'p uchrasa unda ayni kontekstda mavjud omonim katta ehtimollik bilan bog'lovchi emas, ot turkumiga oid so'z bo'ladi, keyingi ehtimollikda bog'lovchi sifatida hisobga olinadi. Kontekstni tavsiflash uchun N-grammadan foydalaniladi. N-gramma – matnlarga avtomatik ishlov berishda keng qo'llaniladigan matematik hisob vositasidir. O'zbek kompyuter lingvistikasida S.Rizayev harf birikmalarini bigramm, trigramm terminlari bilan ifodalagan [5].

N-gramma – so'zlar yoki teglar kabi N-identifikator elementlarning ketma-ketligini ifodalaydi. Ikki element ketma-ketligi – bigramma, uch element ketma-ketligi esa trigramma, deyiladi. Masalan, *old qo'shimcha+ot* holati bigrammaga misol bo'ladi.

Omonimlikni aniqlashning oddiy statistik metodi va boshqa shu kabi metodlarning tavsifini keltirish uchun quyidagi usullar ishlatiladi:

–  $w_i$  – jumladagi  $i$ -o‘rinda joylashgan so‘z,  $t_i$  – ushbu so‘zning identifikatori (tegi).

–  $D_{(w)} = \{t_1^w, t_2^w, \dots, t_k^w\}$   $w$  so‘zining barcha mumkin bo‘lgan belgilar majmui. Ushbu ma’lumotlarni morfologik lug‘at yordamida olish mumkin. Agar so‘z lug‘atda bo‘lmasa uni Brill usulida bajarilganidek, ot so‘z turkumi sifatida hisoblash mumkin, ammo lingvistik ta’minot ishonchli bo‘lishi uchun barcha mumkin bo‘lgan teglarni qo‘yib chiqish kerak.

–  $C$  – korpusdagi muayyan holatlar soni ( $n$ -gramm). Bunda  $C(t)$  –  $t$  teglar soni; va  $C(t_1, t_2)$  – bigrammalar soni ( $t_1, t_2$ ).

–  $C_t(w, t)$  –  $t$  tegli  $w$  so‘zlar soni.

–  $F(w, t)$  –  $w$  so‘zida  $t$  tegi mavjudligi ehtimoli. Tavsiflar quyidagi formula bo‘yicha hisoblanadi:

$$F(w, t) = \frac{C_t(w, t)}{C(t)}$$

–  $P(t_i|t_{i-1})$  – bu  $t_{i-1}$  tegidan keyin  $t_i$  tegining kelish ehtimollik holati. Bunda  $i = 1$  bo‘lganda  $t_i$  tegi gapda birinchi teg hisoblanadi. Hisoblash formulasi quyidagicha:

$$P(t_i|t_{i-1}) = \frac{C(t_{i-1}, t_i)}{C(t_{i-1})}$$

Yashirin Markov modeliga asoslanib, omonimlikni aniqlashning statistik metodi yordamida ishlash natijasida  $n$  uzunlikdagi jumlada  $T_i \in D(w_i)$  bo‘lganda  $T = \{T_1, T_2, \dots, T_n\}$  teglarning ehtimoliy ketma-ketligi topiladi [7].

## XULOSA

Xulosa qilib aytganda, dunyo kompyuter lingvistikasida omonimlikni bartaraf etish usullari o‘rganilganida, bu xususdagi tajribadan foydalanib o‘zbekcha matnlardagi so‘zshakllarning tegishli tekshirish formulasi yaratildi. Omonimlikni bartaraf etish uchun har bir so‘zni “tasniflash” kerak, ya’ni uni lemma – gap bo‘lagi va morfologik xususiyatlar majmui bilan taqqoslash mumkin, ular qulaylik uchun bir tegga qo‘shiladi. Barcha mumkin bo‘lgan teglarni o‘rganish uchun morfologik lug‘atdagi so‘zlarga tegishli havolalarni topish yoki MyStem kabi morfologik analizatorni ishlatish yetarli bo‘lib, u so‘z teglarini topishda yordam beradi. Shundan so‘ng bir nechta teglar orasidan faqat tegishli tegni tanlash kerak bo‘ladi.

Omonim soʻzshakllarni tahlil qilishda qoʻllanilgan optimal lingvistik usul matnlarni tahrir va tahlil qilish, mashina tarjimasini, matnlarni qayta ishlash jarayonlarida muhim omil boʻladi.

### REFERENCES

1. Большакова Е.И. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. Учебное пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. – Москва: МИЭМ, 2011. – 272 с.
2. Порохин А.А. Анализ статистических методов снятия омонимии в текстах на русском языке. Вестник АГТУ. Сер.: Управление, вычислительная техника и информатика. – 2013. № 2. – С. 168-174.
3. Rahmatullayev Sh. Oʻzbek tili omonimlarining izohli lugʻati. – Toshkent: Oʻqituvchi, 1984. – B.5.
4. Abjalova M. Tahrir va tahlil dasturlarining lingvistik modullari: Monografiya. – Toshkent, 2020. – B. 25-27.
5. Rizayev S. Oʻzbek tilshunosligida lingvostatistika asoslari. – Toshkent: Fan, 2006. – B . 18.
6. Baum, L. E.; Sell, G. R. [Growth transformations for functions on manifolds](#). Pacific Journal of Mathematics. 27 (2) 1968. – P. 211–227.; [https://en.wikipedia.org/wiki/Hidden\\_Markov\\_model](https://en.wikipedia.org/wiki/Hidden_Markov_model).
7. [http://www.academia.edu/15517740/Анализ\\_статистических\\_алгоритмов\\_снятия\\_морфологической\\_омонимии\\_в\\_русском\\_языке](http://www.academia.edu/15517740/Анализ_статистических_алгоритмов_снятия_морфологической_омонимии_в_русском_языке).