

MATNNI AVTOMATIK TAHLIL QILISHDAGI MUAMMOLAR

Suyunova Malika Odil qizi

Alisher Navoiy nomidagi Toshkent davlat
o'zbek tili va adabiyoti universiteti
Kompyuter lingvistikasi mutaxassisligi
2-bosqich magistranti,
malikasuyunova0@gmail.com

Annotatsiya. Bugungi kunda korpus lingvistikasi sohasida ko'plab ishlar amalga oshirilmoqda. Ushbu maqolada morfologik analizator yuzasidan qilingan ishlar, matnni avtomatik tahlil qilish, tahlil jarayonidagi muammolar, ko'p ma'noli va omonim so'zlarni korpusda farqlash masalasi, grammatik omonimlar bilan bog'liq muammolar va ularning yechimlari haqida so'z yuritiladi.

Kalit so'zlar: *korpus, kompyuterli grammatika, kompyuterli semantika, morfologik analizator, omonimlar, grammatik omonimlar.*

Abstract. Today, there are many work is being done in the field of corpus linguistics. This article discusses the work done on the morphological analyzer, automatic text analysis, problems in the analysis process, the issue of distinguishing multi-meaning and homonymous words in the corpus, problems related to grammatical homonyms and their solutions.

Keywords: *corpus, computer grammar, computer semantics, morphological analyzer, homonyms, grammatical homonyms.*

Абстрактный. На сегодняшний день ведется большая работа в области корпусной лингвистики. В данной статье рассматриваются проделанная работа по морфологическому анализатору, автоматический анализ текста, проблемы в процессе анализа, проблема различения многозначных и омонимичных слов в корпусе, проблемы, связанные с грамматическими омонимами и пути их решения.

Ключевые слова: *корпус, компьютерная грамматика, компьютерная семантика, морфологический анализатор, омонимы, грамматические омонимы.*

Kompyuter texnologiyalarida matn asosiy o'rinni egallaydi. Amaliy tilshunoslik sohasida tadqiqotchilar matnni bir tildan boshqa bir yoki bir necha tillarga avtomatik ravishda tarjima qiluvchi, matnni morfologik, sintaktik, semantik jihatdan avtomatik tahlil qiluvchi, matnda ishtirok etgan so'zlarning chastotasini ko'rsatuvchi va boshqa shu kabi dasturlarni yaratish ustida amaliy ishlarni olib borishmoqda. Xorijda bu kabi ishlar keng miqyosda amalga oshirilib, katta-katta hujjatlar matematik metodlar asosida tahlil qilinmoqda. 1950-yillardan boshlab matnni avtomatik tahlil qilish bo'yicha ko'plab tajribalar o'tkazildi. Ularning ko'pchiligi ko'p ishlatiladigan atamalar yoki iboralardan iborat matnni tahlil qilish

ustida olib borildi. Bu vaqtning tejalishiga, inson tomonidan qo'yiladigan kamchiliklarning kamayishiga olib keladi. Avallari ko'p vaqt, aqliy mehnat va kuch sarflab qilingan ishlar, endi tez fursatlarda inson aqli yordamida yuzaga kelgan dasturlar yordamida samarali natijalarga erishilmoqda. Shu bilan bir qatorda, axborot-qidiruv tizimida kiritmalarni to'g'ridan-to'g'ri tahlil va tahrir qilish imkoniyati kengaymoqda. Bugungi kunda matnni texnik tahlil qilish borasida yaratilayotgan ishlar asosan quyidagi yo'nalishda olib borilmoqda [Abjalova M., 2020:56]:

1. Kompyuterli grammatika
 - a) kompyuterli morfologiya
 - b) kompyuterli sintaksis
2. Kompyuterli semantika

Matnni avtomatik tahlil qilishda, eng avvalo, o'sha tildagi barcha so'zlarni o'zida jamlagan lug'at fondi muhim ahamiyatga ega. Chunki lug'at xalqning madaniyati, qadriyatlari va fikrlarini bilish uchun eshikdir. Matndagi so'zlar tahlil qilinayotganda bazaga kiritilgan lug'atga murojaat qilinadi. Bu usul bilan tahlil qilish eng keng tarqalgan va yaxshi samara bergan usullardan biridir. Lug'at matni tahliliga misol sifatida avtomatlashtirilgan kontent tahlilini keltirish mumkin. Bu tahlilda tez-tez ishlatiladigan so'zlar chiqarib tashlanadi. So'zlar asos shaklga ya'ni lug'atdagi ko'rinishga qadar kesib tashlanadi va matndagi so'zning chastotasi ko'rsatiladi.

Morfologik tahlilning vazifasi matndagi har bir so'zning qaysi turkumga mansubligini avtomatik tarzda muayyanlashtirish; so'zlarning qaysi leksik-grammatik sinfga tegishlilikini aniqlashdan iborat. Rus tilida formal morfologiya rivojlanganligi sababli buni amalga oshirishning keng imkoniyatlari mavjud. Ingliz tilida ham algoritm nisbatan oddiy: leksik ko'p ma'nolilik mavjud bo'lsa-da, morfoanaliz dasturlari (tegger) matndagi so'zning turkumini 90 foizda to'g'ri aniqlay oladi. Rus tili matnlarining morfologik analizini amalga oshirish uchun A. Zaliznyak grammatik lug'atining kompyuter versiyasidan foydalanilgan, ingliz tilining morfoanalizi Myuller grammatik lug'ati asosida amalga oshirilgan. Rus va ingliz morfoanalizini yaratish tajribasi shuni ko'rsatadiki, morfologik analizni amalga oshirish uchun formal grammatika yaratilgan bo'lishi kerak [Xampoeva III., 2020:37].

Matnni avtomatik tahlil qilishda ko'chma ma'noda qo'llangan so'zlar ya'ni ko'p ma'noli so'zlar, omonim so'zlarning qaysi ma'noda ishlatilayotganini kompyuterda aniqlash eng dolzarb masalalardan biridir. Chunki inson aqligina so'zning qaysi ma'noda kelayotganini matn mazmunidan kelib chiqib aniqlay oladi. Buni kompyuter dasturlari orqali amalga oshirish ko'pchilikni o'ylantirgan va har xil yondashuvlar kelib chiqqan. O'zbek tilshunosligida yangi soha hisoblangan korpus tilshunosligida matnni avtomatik tahlil qilishda ko'p ma'noli so'zlarni aniqlash, omonimiya muammosini yechish, ularni teglash kabi masalalar kun tartibiga qo'yilgan. Bu borada qator tadqiqotlar amalga oshirilgan. Jumladan, Sh.Gulyamova “O'zbek tili semantik analizatorining lingvistik asoslari” monografiyasida

polifunksional va omonim soʻzlarni korpusda farqlash uchun semantik filtr ishlab chiqqan. Bu nafaqat morfoanalizatorida, balki semantik analizatorida eng qiyin muammoni bartaraf etishga xizmat qiladi. Omonimiya hodisasi qadimdanoq dolzarb, keng muhokamalarga sabab boʻluvchi muammolardan biridir. Maʼlumki, soʻz har qanday tilning leksik boyligini tashkil etadi, bir-biri bilan oʻzaro muloqotda boʻladi, turli qadriyat hamda vazifalarni bajaradi. Soʻzlashuv nutqida soʻzlar maʼno jihatdan bir-biriga yaqin yoki aksincha, zid maʼnoga ega boʻladi. Baʼzi soʻzlar maʼno jihatdan yaqin boʻlmasa ham, koʻrinishi, shakli, imlo, talaffuzi bilan bir-biriga oʻxshaydi. Bunday soʻzlar guruhi tilning umumiy sistemasida omonimlar deyiladi. Omonim – grek tilidan olingan murakkab soʻz, *homos* “bir xil” va *onoma* “ism” degan maʼnolarni bildiradi. Shu bilan birga, omonimlar *bir xil nomlar* deb tarjima qilinadi [Гулямова III., 2021:53]. Sh. Rahmatullayevning “Oʻzbek tili omonimlarining izohli lugʻati” da omonimlar yozilishi bir xil fonemalardan iborat shakllar boʻlib, omonimiyada ifoda plani yetakchi turadi deb taʼrif berilgan[Rahmatullayev Sh., 1984:5].

Omonim hodisada yasama leksema va tub leksema, shuningdek, ularga grammatik maʼno ifodalovchi affiks qoʻshilganda voqe boʻladigan shakliy teng kelish ham qamrab olinadi. Masalan, *terim* soʻzshakli *ter* – feʼl leksemasiga *–im* leksema yasovchisining birikishi natijasida hosil qilingan leksemaning birlik, bosh kelishik formasiga teng.

Aniqroq boʻlishi uchun, bir necha turdagi omonimlarni ajratib koʻrsatish mumkin:

1. Nutqning bir qismi yoki lugʻatdagi shakli (lemma) farqlanadi;
2. Baʼzi morfologik xususiyatlar, masalan, kelishik, egalik va sonda (boy morfologiyali tillarda) farqlanadi;
3. Bir-biridan faqat mazmun jihatdan farqlanadi (bu holat semantik farqlanishda uchraydi) [Abjalova M., 2020:71].

Bugungi kunda lingvistik jihatdan tadqiqotlar olib borishda til korpuslarining ahamiyati katta. Chunki til korpuslari turli lingvistik topshiriqlarni bajarishga xizmat qiladi. Tilshunoslikning ilk va oʻrta bosqichlarida omonimiya hodisasining paydo boʻlishi, taraqqiyoti va yondosh hodisalarga munosabatlari oʻrganilgan boʻlsa, zamonaviy lingvistikada uning korpusda berilishi masalasi kun tartibiga qoʻyilgan. Sh. Hamroyeva taʼkidlaganidek, til korpusi elektron kutubxonadan farqli ravishda muayyan tilni tadqiq qilish, oʻrganish va oʻrgatish uchun zaruriy, foydali va qiziqarli matnlar toʻplanishini nazarda tutadi [Hamroyeva Sh., 2018:53].

Matnlarni avtomatik qayta ishlash bir necha bosqichga boʻlinadi va ulardan biri morfologik tahlil bosqichi hisoblanadi. Mazkur bosqichda har bir soʻzga morfologik tavsif beriladi: lemmasi (asosi), kelishigi, soni, darajasi, nisbati, shaxsi va hk. Morfoanalizning soʻzlarni morfologik teglash vazifasi omonim shakllar bilan murakkablashadi. Omonimiyani aniqlashning bir qator usullari aniqlangan. Mazkur usullarga tayanuvchi tizimlar oʻz navbatida quyidagi guruhni tashkil etadi:

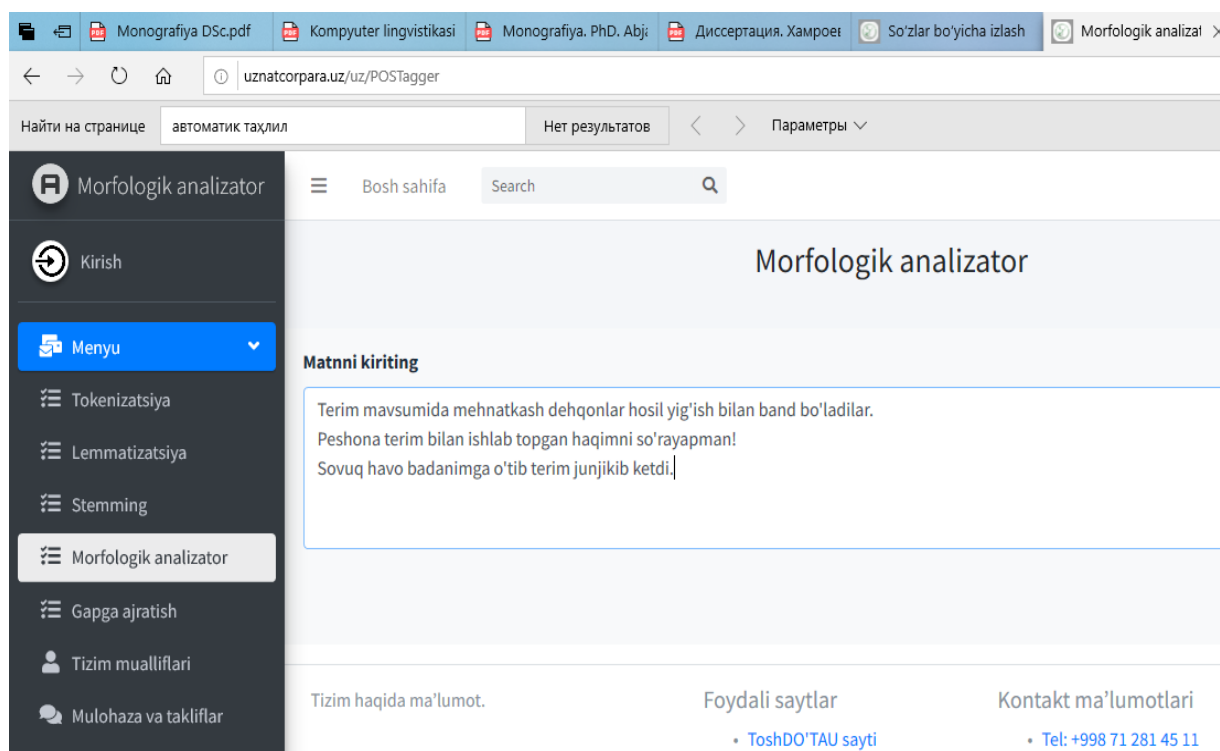
1. Qoʻlda yaratilgan qoidalarga asoslangan tizimlar.

2. Ehtimoliy modellar asosida yaratilgan va tavsiflangan korpuslarga tayanadigan tizimlar.

3. Ehtimollik modellari va qoidalarga asoslangan gibrid tizimlar[Abjalova M.,2021].

Biz tilimiz tabiati va xususiyatini inobatga olib, ulardan foydalanamiz. Omonimlarning bir qancha turlari mavjud bo'lib, ulardan leksik omonimlar lug'ati Sh.Rahmatullayev tomonidan tuzilgan. Leksik omonimlarni Sh.Gulyamova semantik analizatorida aniqlash uchun filtr ishlab chiqib muammoning yechimini topgan. Omonimlarning yana bir turi grammatik omonimlar bo'lib, ular so'zlarga grammatik shakllar qo'shilganda yuzaga keluvchi nutqiy jarayondir. Ayrim manbalarda leksik omonimlardan farqli ravishda grammatik omonimlar ham borligi haqida to'xtalib o'tilgan.

Morfologik analizator jarayonida so'zlar turkumga ajratilganda grammatik omonimlar shaklan bir xil bo'lganligi, lekin leksik omonimlar safiga kiritilmaganligi sababli qaysi turkumda va qaysi ma'noda kelayotganligini aniqlash juda qiyin. Quyida grammatik omonimlarning morfoanalizator jarayonida tahlil qilinishini ko'ramiz [<http://uznatcorpara.uz/uz/POSTag>]:



1-rasm. Morfologik analizatorga grammatik omonimlar ishtirokidagi gaplarning kiritilishi

Morfologik analizatorga grammatik omonimlar ishtirok etgan uchta gapni kiritdik(1-rasm) va tahlil qildik:

Morfologik analizator				
Kirish				
Menyu				
Tokenizatsiya				
Lemmatizatsiya				
Stemming				
Morfologik analizator				
Gapga ajratish				
Tizim mualliflari				
Mulohaza va takliflar				

№	So'z	Lemma	So'z turkumi	O'zak va qo'shimchalar
1	Terim	Terim	Ot ...	{Terim}
2	mavsumida	mavsum	Ot ...	{mavsum}-i-da
3	mehnatkash	mehnat	Ot ...	{mehnat}-kash
4	dehqonlar	dehqon	Ot ...	{dehqon}-lar
5	hosil	hosil	Ot ...	{hosil}
6	yig'ish	yig'ishmoq	Fe'l ...	{yig'ish}
7	bilan	bilan	Bog'lovchi ...	{bilan}
8	band bo'ladilar	band bo'lmoq	Fe'l ...	{band bo'li}-a-dilar

2-rasm. Morfoanalizatorga kiritilgan birinchi gapning tahlil natijasi

Birinchi gapda *terim* so'zi ot so'z turkumi sifatida to'g'ri tahlil qilingan(2-rasm). Ot so'z turkumining chuqurroq tahlilini tekshirib ko'ramiz(3-rasm)

Morfologik analizator													
terim													
Ot													
Tuzilishiga ko'ra (Ot)				Yasalishiga ko'ra (Ot)		Yasalish usuli (Ot)		Predmetning yakka/to'dasini bildirishiga ko'ra		Bevosita kuzatish mumkinligiga ko'ra		Turdosh ot	
Sodda ot	Qo'shma ot	Juft ot	Takror ot	Tub ot	Yasama ot	Kompozitsiya (ot)	Affiksatsiya (ot)	Yakka ot	Jamlovchi ot	Mavhum ot	Aniq ot	Narsa/buyum otlari	Shaxs otlari nomla
+					+		+	+			+		

3-rasm. Grammatik omonim so'zning tahlil qilinishi

terim													
Bosita kuzatish mumkinligiga ko'ra		Turdosh ot											
Mavhum	Aniq ot	Narsa/buyum otlari	Shaxs otlari nomlari	Inson tana a'zolari nomlari	Hayvon va uning a'zolari otlari	Oziq-ovqat nomlari	O'simlik va uning qismlari otlari	Makon/maskan otlari	Zamon otlari	Diniy-ilohiy otlar	Miqdor otlari	Faoliyat/jarayon otlari	
	+												+

4-rasm. Grammatik omonim so'zning tahlil qilinishi (davomi)

Terim so'zi ot:

1. Sodda ot
2. Yasama ot
3. Affiksatsiya
4. Yakka ot
5. Aniq ot
6. Faoliyat/jarayon otlari

Bunda morfoanalizatorida *terim* so'zi barcha jihatdan to'g'ri tahlil qilingan.
Keyingi gapning tahlil qilinishini ko'ramiz(5-rasm):

Kirish				
Menyu				
Tokenizatsiya				
Lemmatizatsiya				
Stemming				
Morfologik analizator				
Gapga ajratish				
Tizim mualliflari				
Mulohaza va takliflar				
9	Peshona	Peshona	Ot ...	{Peshona}
10	terim	terim	Ot ...	{terim}
11	bilan	bilan	Bog'lovchi ...	{bilan}
12	ishlab	ishlamoq	Fe'l ...	{ishla}-b
13	topgan	top	Fe'l, Ot, Sifat ...	{top}-gan
14	haqimni	haq		{haq}-im-ni
15	so'rayapman	so'ramoq	Fe'l ...	{so'ra}-yap-man

5-rasm. Grammatik omonim ishtirokidagi ikkinchi gapning tahlil qilinishi

terim														
Ot														
Tuzilishiga ko'ra (Ot)				Yasalishiga ko'ra (Ot)		Yasalish usuli (Ot)		Predmetning yakka/to'dasini bildirishiga ko'ra		Bevosita kuzatish mumkinligiga ko'ra		Turdosh ot		
Sodda ot	Qo'shma ot	Juft ot	Takror ot	Tub ot	Yasama ot	Kompozitsiya (ot)	Affiksatsiya (ot)	Yakka ot	Jamlovchi ot	Mavhum ot	Aniq ot	Narsa/buyum otlari	Shaxs otlari	nomla
+					+		+	+			+			

6-rasm. Grammatik omonim so'zning tahlil qilinishi

terim												
vosita tashish imkinligiga ra		Turdosh ot										
vhum	Aniq ot	Narsa/buyum otlari	Shaxs otlari nomlari	Inson tana a'zolari nomlari	Hayvon va uning a'zolari otlari	Oziq- ovqat nomlari	O'simlik va uning qismlari otlari	Makon/maskan otlari	Zamon otlari	Diniy- ilohiy otlar	Miqdor otlari	Faoliyat/jarayon otlari
	+											+

7-rasm. Grammatik omonim soʻzning tahlil qilinishi (davomi)

Ikkinchi gapda ham yuqoridagidek tahlil qilindi(6-7-rasm). Aslida *terim* yasama emas tub ot, faoliyat-jarayon oti emas narsa oti deb tahlil qilinishi kerak.

Morfoloqik analizator				
Kirish				
Menyu				
Tokenizatsiya				
Lemmatizatsiya				
Stemming				
Morfoloqik analizator				
Gapga ajratish				
Tizim mualliflari				
Mulohaza va takliflar				

16	Sovuq	Sovuq	Ot, Sifat ...	{Sovuq}
17	havo	havo	Ot ...	{havo}
18	badanimga	badan	Ot ...	{badan}-im-ga
19	o'tib	o't	Fe'l, Ot ...	{o't}-ib
20	terim	terim	Ot ...	{terim}
21	junjikib	junjikmoq	Fe'l ...	{junjik}-ib
22	ketdi			ketdi

8-rasm. Grammatik omonim ishtirokidagi uchinchi gapning tahlil qilinishi

Uchinchi gapda ham *terim* soʻzi yasama, faoliyat-jarayon oti sifatida tahlil qilindi. Aslida tub, inson ta'na a'zosi sifatida tahlil qilinishi kerak edi(8-rasm).

Ko'rinib turibdiki, avtomatik tahlil jarayonida grammatik omonimlarni aniqlab, ularni to'g'ri tahlil qilish nihoyatda kerakli masalalardan biridir. Buning uchun esa grammatik omonimlarni umumiy va xususiy birikuvchilarini aniqlab olish va shu asosda ularning semantik maydonini shakllantirib olish, barcha grammatik omonimlar uchun model ishlab chiqish kabi vazifalar kun tartibida turibdi.

Foydalanilgan adabiyotlar:

1. Хамроева Ш. Ўзбек тили морфологик анализаторининг лингвистик таъминоти. – GlobeEdit, 2020: 268.
2. Гулямова Ш. Ўзбек тили семантик анализаторининг лингвистик асослари. Монография – Тошкент: 2021: 277.



3. Rahmatullayev Sh. O'zbek tili omonimlarining izohli lug'ati. – Toshkent: O'qituvchi, 1984.
4. Hamroyeva Sh. O'zbek tili mualliflik korpusini tuzishning lingvistik asoslari: Fil.fan. bo'yicha falsafa dokt. (PhD) diss. avtoref. – Qarshi, 2018.
5. Elov B., Hamroyeva Sh., Xusainova Z. NLP (tabiiy tilga ishlov berish)ning vazifalari va zamonaviy yondashuvlar. TerDU, Filologik tadqiqotlar: til, adabiyot, ta'lim. 2022, 5-6.
6. Elov B., Hamroyeva Sh., Elova D. Morfologik analizatorni yaratish usullari. O'zbekiston: til va madaniyat (Amaliy filologiya), 2022, 5(1). 67-87.
7. <http://uznatcorpara.uz/>