

O'ZBEK TILI MATNLARIDA SO'Z BIRIKMASINI AVTOMATIK ANIQLASHDA OMONIMIYA MASALASI

Abdullayeva Oqila Xolmo'minovna,
Filologiya fanlari falsafa doktori (PhD)
abdullayevaoqila@gmail.com
ToshDO'TAU doktoranti DSc

Nizomova Fotima Bo'rixo'ja qizi,
Kompyuter lingvistikasi yo'nalishi talabasi
nizomovaf44@gmail.com
ToshDO'TAU

Annotatsiya. Ushbu maqolada o'zbek tili matnlarida so'z birikmasini avtomatik aniqlashning zarurligi va uning sintaktik tahlildagi ahamiyati o'rganib chiqilib xulosa va fikrlar berildi. Bunda o'zbek tili matnlarida so'z birikmasini avtomatik aniqlashda omonimiya masalasi haqida fikr yuritildi. Maqolada boshqaruvli so'z birikmasining nazariyasi o'rganilib chiqildi va undagi omonimiya masalasi ko'rsatildi. Boshqaruvli so'z birikmasining teg va modellari ishlab chiqildi va uning ibora (turg'un bog'lanma) bilan omonimligi va farqlari haqida ham fikr va xulosalar berildi.

Annotation. In this article, the necessity of automatic identification of word combinations in Uzbek text and its importance in syntactic analysis were examined, along with conclusions and insights provided. The issue of ambiguity (homonymy) in automatic identification of word combinations in Uzbek texts was discussed. The theoretical framework of governed word combinations was studied, and the issue of homonymy within it was highlighted. Tags and models for governed word combinations were developed, and conclusions and insights were also provided regarding their ambiguity and differences compared to idiomatic expressions (fixed phrases).

Аннотация. В этой статье была рассмотрена необходимость автоматического распознавания словосочетаний в текстах на узбекском языке и его важность для синтаксического анализа, а также приведены выводы и наблюдения. Обсуждался вопрос омонимии при автоматическом распознавании словосочетаний в узбекских текстах. Была изучена теоретическая основа управляемых словосочетаний, и выделена проблема омонимии в их рамках. Были разработаны теги и модели для управляемых словосочетаний, а также приведены выводы и наблюдения касаясь их омонимии и различий по сравнению с идиоматическими выражениями (фиксированными фразами).

Kalit soʻzlar: Soʻz birikmasi, boshqaruvli soʻz birikmasi, turgʻun birikma, erkin birikma, teg, POS teglash, omonimiya.

Kirish. XX asrning ikkinchi yarmida struktur sintaksis, semantik sintaksis yoʻnalishlari paydo boʻlishi bilan gapning sintaktik tuzilishiga, gap boʻlaklariga nisbatan yondashuv ham oʻzgardi. Formal grammatika talqiniga koʻra, faqat kesimgina gapning grammatik yadrosi sifatida eʼtirof etildi. Gapning markaziy boʻlagini belgilashga doir qarashlar dunyo tilshunosligida turlicha boʻlib kelgan. Fransuz tilshunosi Lyusen Tenyerning verbosentrizm (lotincha “verbum” – “feʼl”, “kesim” maʼnolarini anglatadi, “centrum” – “markaz”) konseptsiyasi feʼlning gap markazi boʻlgan kesim vazifasida kelishini ifoda etadi.¹

Sintaksisni tekshirish orfografiyani tekshirishdan koʻra murakkabroq. Chunki orfografiyani tekshiruvchi dasturlar yaratilgan, lekin norasmiy matnlarning sintaktik jihatdan avtomatik tahriri masalasi haligacha hal etilgani yoʻq. Rasmiy matnlarning avtomatik tahriri boʻyicha muhim natijalarga erishilgan [Xolmanova, 2020. 176]

Oʻzbek tilida matnlarni avtomatik sintaktik tahlil qilish uchun bugunga qadar kompyuter lingvistikasida amaliy natijalarga erishilgani yoʻq. Ammo, bu sohada sintaktik analizator yaratish uchun dastlabki vazifalar amalga oshirilmoqda. Sintaktik analizator yaratishda soʻz birikmasining tobe va hokim qismlarni bir-biriga bogʻlanish aloqalarini oʻrganib chiqish va ularning teg va modellarini ishlab chiqish oldimizda turgan asosiy vazifalardan biri hisoblanadi.

Oʻzbek tili matnlarida soʻz birikmasini avtomatik aniqlashning ahamiyati va dolzarbligi

Ahamiyati

- Soʻz birikmalari – bir yoki bir nechta soʻzlardan tashkil topgan, oʻzaro bogʻlanib, bir maʼno ifodalovchi sintaktik birliklardir.
- Oʻzbek tilidagi soʻz birikmalari sintaktik analizatorni yaratishda juda muhim ahamiyatga ega, chunki ular matnning toʻliq va aniq tahlilini amalga oshirishda asosiy rol oʻynaydi.

Dolzarbligi

- Soʻz birikmalarini toʻgʻri aniqlash tilning sintaktik tahlilini toʻliq amalga oshirishga yordam beradi.
- Sintaktik analizatorlar soʻz birikmalarini toʻgʻri aniqlashga asoslanib, matnning toʻgʻri tahlilini amalga oshiradi.

Boshqaruv aloqasi. Bu aloqada tobe soʻz hokim soʻzning talabi bilan maʼlum forma – shaklga kiradi: tobe soʻz yo biror kelishik qoʻshimchasini oladi, yo biror koʻmakchi bilan birga keladi. Shunga qarab, boshqaruv aloqasi ikki turga ajraladi:

- Kelishikli boshqaruvda tobe soʻz hokim soʻzning talabi bilan tushum, joʻnalish, oʻrin-payt va chiqish kelishiklaridan birida qoʻllanadi. Bu uning tobeligini koʻrsatuvchi belgidir: *Iloji boʻlsa, sizni bu saroydan hovliga koʻchiramiz* (A. Qodiriy).

- Koʻmakchili boshqaruvda tobe soʻz hokim soʻzning talabiga binoan biror koʻmakchi bilan birga qoʻllanadi. Bu uning tobeligini koʻrsatuvchi belgidir: 1. *U daraxtlarning qurigan shoxlarini tokqaychi bilan kesdi.* 2. *Chang salomatlik uchun zararlidir.*

Boshqaruv aloqasi orqali obyektli va holli munosabatlar ifodalanadi: bogʻni tomosha qilmoq (obyektli munosabat), dalaga bormoq (holli munosabat), kitobga chizmoq (obyektli munosabat).

Boshqaruv aloqasida tobe soʻz ot yoki ot oʻrnida qoʻllanadigan soʻzlardan, shuningdek, infinitiv va sifatdoshdan boʻladi. Hokim soʻz funksiyasida esa koʻpincha feʼl qoʻllanadi. Biroq bu vazifada ot, sifat, ravish va boshqa soʻzlar ham kelishi mumkin. Bunda hokim soʻzning oʻzgarishi tobe soʻzga taʼsir qilmaydi.

Misollar: *Oyni etak bilan yopib boʻlmas* (Maqol).

Yigʻlagani uchun choʻchidi. (“Mushtum”).

Paxtaga kon Oʻzbekiston (Qoʻshiq).

Uyat oʻlimdan ogʻir (Maqol).

Qoʻrqqanga qoʻsh koʻrinar (Maqol).

Koʻkaldosh ham Alpomishdan kam emas (F.Yoʻldosh).

Oʻzingda yoʻq - olamda yoʻq (Maqol).

Barchin qizim Qorajonga lozimdir (F.Yoʻldosh). [Abdullayev, 1976. 14].

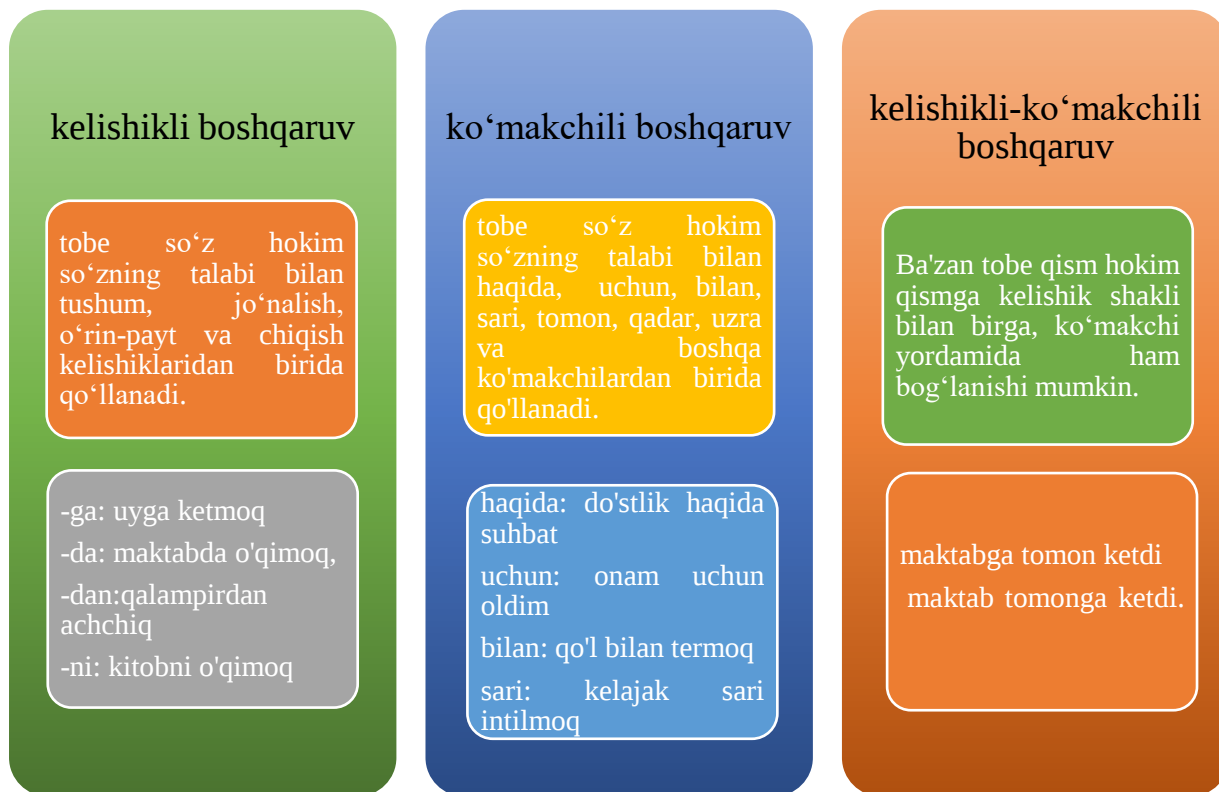
Tobe qismining hokim qismga koʻmakchilar yordamida bogʻlanishi koʻmakchili boshqaruv hisoblanadi: *maktab tomon ketdi, ariq boʻylab yurdi.*

Baʼzan tobe qism hokim qismga kelishik shakli bilan birga, koʻmakchi yordamida ham bogʻlanishi mumkin. Bunday ikkita vositali boshqaruv kelishikli-koʻmakchili boshqaruv sanaladi: *maktabga tomon ketdi - maktab tomonga ketdi.* Shunga koʻra, bunday aloqani variantli va variantli boʻlmagan aloqaga boʻlish mumkin.

Variantli boʻlmagan aloqada tobe qism hokim qism talabi bilan faqat bir shaklga ega boʻladi: *uyda chiqmoq, Karimni koʻrmoq.* Variantli aloqada esa tobe qism hokim qismga bir necha shaklda bogʻlanadi: *uzum yemoq – uzumdan yemoq, maktabga ketmoq – maktab tomon ketmoq – maktab tomonga ketmoq* variantlarining har biri oʻzaro umumiylik-xususiylik, aniqlik-noaniqlik maʼnolari bilan farq qiladi. [Mahmudov, Nurmonov, 1995. 14]

Ushbu nazariyalardan xulosa qilib quyidagi sxemani keltirish mumkin.

So‘z birikmasi o‘z ichida 3 turga bo‘linadi:



So‘z birikmalari komponentlarining ma’no mustaqilligi va ularning birikish darajasiga ko‘ra *turg‘un birikma* va *erkin birikmaga* bo‘linadi.

Turg‘un birikma. Tilda bir qolipda, bo‘linmaydigan holda qo‘llanadigan so‘z birikmalari bor. Bunday iboralar so‘zlovchining xotirasida bir butun holatda saqlanadi va nutqda ham xuddi shu holatda qo‘llanadi. So‘zlovchi turg‘un iboralarni so‘zni ishlatgani kabi tanlab ishlatadi. Bunday iboralar gapda bitta so‘roqqa javob bo‘lib, bir bo‘lak vazifasida keladi. Masalan, G‘alaba shodiyonasi millionlar qatori Dildorning ham boshini ko‘klarga ko‘tardi (Y.Sh.) gapida boshini ko‘klarga ko‘tardi – nima qildi? – bitta bo‘lak – kesim vazifasida qo‘llangan. [Abdurahmonov, Sulaymonov, Xoliyorov, Omonturdiyev, 1979. 9]

Turg‘un birikma tarkibiga kiruvchi komponentlar tashqi tomondan grammatik ko‘rsatkichlarga ega bo‘lib, ular ma’lum sintaktik aloqa turlarini hosil qiladilar (boshini ko‘tarmoq — boshqaruvli birikma shakli kabi). Biroq birikma qismlari orasidagi sintaktik aloqa jonli emas, birikma yaxlit holda boshqa biror bo‘lakni boshqarib keladi yoki shu holicha boshqa bo‘lak tomonidan boshqariladi:

Men sevinchimdan terimga sig'may Xadraga yugurdim kabi turg'un birikma bir leksik ma'no anglatuvchi bo'lib qolganligi – leksikalizatsiyalashganligi uchun leksik birlik hisoblanadi. [Abdurahmonov, Sulaymonov, Xoliyorov, Omonturdiyev, 1979. 10].

Turg'un birikmalar, asosan, ikki yirik guruhga bo'linadi [Abdurahmonov, Sulaymonov, Xoliyorov, Omonturdiyev, 1979. 10]:

murakkab terminlar:

- Sovet Ittifoqi Kommunistik partiyasi, Termiz Davlat pedagogika instituti, Janubiy Surxon suv ombori, O'rta Rus pasttekisligi, "O'qish kitobi", ishlab chiqarish boshqarmasi kabi;

frazeologik iboralar:

- Og'zing qani desa, qulog'ini ko'rsatmoq; Tegirmonga tushsa butun chiqmoq; Yulduzni beneravon urmoq; Qamchisidan qon tommoq; Orasidan qora mushuk o'tmoq kabi

So'z birikmasi va ibora (turg'un bog'lanma)ning farqi [Rasulova, 2018. 15]

<i>So'z birikmasi</i>	<i>Ibora (turg'un bog'lanma)</i>
• <i>So'z birikmasida so'zlar o'zaro erkin bog'langan bo'ladi</i>	• <i>Iboralar turg'un bog'langan, so'zlar ma'noviy butunlik uchun birlashgan-yaxlitlangan bo'ladi</i>
• <i>Har ikki qism alohida tushuncha anglatadi: ko'z yummoq: ko'z-a'zo, yummoq - harakat</i>	• <i>Ibora bir leksik ma'no anglatadi: ko'z yummoq-vafot etmoq, ko'z yummoq-e'tibor bermaslik</i>
• <i>So'z birikmasi lug'atlarda berilmaydi</i>	• <i>Turg'un bog'lamaga bir so'z sifatida qaraladi, lug'atlarda so'zlar qatorida beriladi, chunki ular so'zlar kabi ma'no ifodalaydi: yog' tushsa yalagudek-toza</i>
• <i>So'z birikmasidagi so'zlar erkin bog'langani uchun almashtirish mumkin: baxtdan kuylab-kuylab baxtdan.</i>	• <i>Turg'un bog'lamadagi so'zlarni almashtirib bo'lmaydi: oq yo'l. Iboralar bir ma'no ifodalash qonuniyati asosida turg'unlashgan, yaxlitlangan bo'ladi</i>
• <i>So'z birikmasi tarkibidagi so'zlar hokim va tobe qismga ajratiladi</i>	• <i>Iboralar bir so'zga teng, ajralmas holga kelib qolgan, uning tarkibidagi qismlarni hokim va tobe so'zga ajratish mumkin emas.</i>
• <i>So'z birikmasi tarkibidagi so'zlarga alohida-alohida savol beriladi: aravani (nimani) olib qochmoq (nima qilmoq)</i>	• <i>Bir leksik ma'no, ya'ni bir so'z (tushuncha)ni ifodalagani uchun iboralarga bitta so'roq beriladi: aravani quruq olib qochmoq iborasiga nima qilmoq so'rog'i beriladi.</i>

Teg – Kompyuter yordamida matn tahlilini amalga oshirish jarayonini tezlashtirish va osonlashtirishga xizmat qiluvchi shartli belgi yoki maxsus kod. Teglar bir necha turlarga bo'linadi: semantik teg, sintaktik teg va grammatik teg. Grammatik teg, shuningdek, PoS (Part of speech) tegging nomi bilan ham mashhur.

POS teglash – bu berilgan gapdagi har bir so'zshaklga uning turkum (ot, fe'l, sifat, son, ravish yoki olmosh)ga mansubligini belgilash (teglash) vazifasidir. POS teglash tabiiy tilni qayta ishlash (Natural Language Processing, NLP)ning asosiy vazifalaridan biri bo'lib, pipeline konveyerining muhim bosqichi hisoblanadi (1-rasm).



1-rasm. Matnga boshlang'ich ishlov berish bosqichlari

O'zbek tili matnlarida so'z birikmasini avtomatik aniqlashda dastur dastlab matnni tokenizatsiya qiladi. Tokenlarni esa POS tagging jarayonidan o'tkazadi, lemmasini aniqlaydi va stemlaydi. So'z birikmasini avtomatik aniqlashda POS tagginglarni quyidagi ro'yxat asosida amalga oshiriladi.

№	POS teg	Ingliz	Belgilanishi
1.	Ot	Noun	N
2.	Sifat	Adjective	JJ
3.	Fe'l	Verb	VB
4.	Son	Number	NUM
5.	Ravish	Adverb	RR
6.	Olmosh	Pronoun	P
7.	Bog'lovchi	Conjunction	C
8.	Ko'makchi	preposition	II
9.	Yuklama	particle	Prt
10.	Modal so'z	modal words	MD
11.	Undov so'z	exclamatory words	UH
12.	Taqlid so'zlar	imitation words	IM

2-rasm. O'zbek tilidagi POS teglar ro'yxati (qisqa variant)

So'z birikmasi va ibora (turg'un birikma)ning shakliy o'xshashligi POS tegging jarayonida omonimlikni keltirib chiqaradi. Bu esa dasturiy ta'minotda murakkablikni keltirib chiqaradi. Tokenizatsiya jarayoni gapni, yoki matnni tahlil qilganda uni qismlarga bo'lib tashlab, POS teglash jarayoniga tayyorlab beradi. Bunda o'zbek tili matnlarida so'z birikmasini avtomatik aniqlashda tokenizatsiya

soʻz birikmasini alohida tokenlarga ajratib berishi, ibora (turgʻun birikma)ni esa yaxlit bir holatda holatda POS teglash jarayoniga oʻtkazib berishi zarur. Buni quyidagi jadval orqali aniq koʻrishimiz mumkin.

Yigit oqimga qarshi suzib ariqning boshiga yetdi-soʻz birikmasi

№	Soʻz	Oʻzak	POS teg	Izoh
1.	Yigit	yigit	Ot	“yigit” alohida leskema, tarkibiy qismlarga boʻlinmaydi
2.	Oqimga	oq	Ot	“oqim” oq+im; -im soʻz yasovchi qoʻshimcha; leksema+grammatik shakl
3.	Qarshi	qarshi	sifat	“qarshi” alohida leskema, tarkibiy qismlarga boʻlinmaydi
4.	Suzib	suz	feʼl	suz+ib lugʻaviy shakl yasovchi; leksema+grammatik shakl
5.	Ariqning	ariq	ot	ariq+ning sintaktik shakl yasovchi; leksema+grammatik shakl
6.	Boshiga	bosh	ot	bosh+i+ga sintaktik shakl yasovchi; leksema+grammatik shakl+grammatik shakl
7.	Yetdi	yet	feʼl	Yet+di sintaktik shakl yasovchi; leksema+grammatik shakl

Qoʻrsligi, beʼmani gaplari axiri uning boshiga yetdi-ibora

№	Soʻz	Oʻzak	POS teg	Izoh
1.	Qoʻrsligi	Qoʻrs	ot	Qoʻrs+lik+i; soʻz yasovchi+egalik qoʻshimchasi; leksema+grammatik shakl yasovchi+grammatik shakl yasovchi
2.	Bemaʼni	maʼni	sifat	Shakl yasovchi+oʻzak; yasama soʻz
3.	Gaplari	gap	ot	Gap+lar+i; leksema+grammatik shakl yasovchi+grammatik shakl yasovchi
4.	Axiri	axiri	yuklama	Axiri; yuklama
5.	Uning	u	olmosh	U+ning; leksema+grammatik shakl
6.	Boshiga yetdi	Boshiga yetdi	ibora	Bir leksik maʼnoga ega boʻlgan ibora.

Xulosa. O'zbek tili matnlarida so'z birikmasini avtomatik aniqlash dasturi kompyuter lingvistikasi sohasida muhim ahamiyat kasb etadi. Ushbu jarayonni amalga oshirishda omonimiya masalasi eng dolzarb muammolardan biri hisoblanadi. So'z birikmasi va turg'un iboralarning shakliy o'xshashligi POS tegging jarayonida murakkablik tug'diradi. Shuning uchun dasturiy ta'minotni ishlab chiqishda turg'un iboralarni yaxlit birlik sifatida aniqlash va ularni erkin so'z birikmalaridan farqlash alohida algoritmik yondashuvni talab qiladi. Boshqaruvli so'z birikmalari nazariyasini chuqur o'rganish va ularning teg hamda modellari asosida ishlov berish dasturiy yechimning asosiy poydevorini tashkil etadi. Ushbu yo'nalishda olib borilgan izlanishlar kelajakda o'zbek tili matnlarini avtomatik tahlil qilishga mo'ljallangan ilg'or tizimlarni yaratish imkonini beradi. Shu bilan birga, o'zbek tili kabi morfologik jihatdan boy tillarni qayta ishlashga oid tadqiqotlar xalqaro lingvistik va texnologik hamjamiyatga ulkan hissa qo'shadi.

Foydalanilgan adabiyotlar:

1. Теньер Л. Основы структурного синтаксиса. - М.: Прогресс, 1988.
2. Elov B.B. Xudayberganov N.U. O'zbek tili korpusi matnlarini pos teglash usullari: file:///C:/Users/User/Downloads/elov-b.b.-xudayberganov-n.u.-ozbek-tili-korpusi-matnlarini-pos-teglash.pdf
3. Xolmanova Z. Kompyuter lingvistikasi. -Toshkent, 2020.
4. Mahmudov N., Nurmonov A. O'zbek tilining nazariy grammatikasi. – Toshkent, 1995.
5. Abdurahmonov G'., Sulaymonov A., Xoliyorov X., Omonturdiyev J., Hozirgi o'zbek adabiy tili; Sintaksis – Toshkent, 1979.
6. Rasulova N. Ona tilidan ma'ruzalar. – Toshkent, 2018.
7. Abdullayev F.A. O'zbek tili grammatikasi. II jild. – Toshkent, 1976.