

TABIY TILNI QAYTA ISHLASHDA TO'LDIRUVCHINING MODELLARI

Nuriddinova Gulira'no Baxtiyor qizi,
1-kurs magistrant
guliranonuriddinova@gmail.com
ToshDO'TAU

Annotatsiya. Ushbu maqolada o'zbek tilida mavjud gap bo'laklaridan ikkinchi darajali bo'lak hisoblanmish to'ldiruvchining modellari haqida so'z yuritiladi. To'ldiruvchining lingvistik xususiyatlari kompleks tarzda tahlil qilinadi va mazkur bo'lakning sintaktik va semantik funksiyalari aniqlanib, umumiy modellar ishlab chiqiladi. Taklif etilgan modellar gap tuzilmasida ikkinchi darajali bo'laklarning o'rni va o'zaro munosabatlarini aniqlashga xizmat qiladi. Tadqiqot natijalari o'zbek tilining sintaktik tizimini chuqurroq anglashga, shuningdek, avtomatik sintaktik tahlil, morfologik analiz va tabiiy tilni qayta ishlash tizimlarini yaratishda va til strukturasi umumiy modellarga tayangan holda baholashda muhim ahamiyat kasb etadi.

Kalit so'zlar: *NLP, lingvistik model, to'ldiruvchi, valentlik, sintaktik tahlil, BIO tagging, korpus*

Abstract. This article examines the models of secondary sentence constituents in the Uzbek language within the framework of Natural Language Processing (NLP). The study provides a comprehensive analysis of the linguistic features of objects, attributes, and adverbial modifiers. It identifies their syntactic and semantic functions and develops generalized models for these constituents. The proposed models serve to clarify the roles and interrelations of secondary sentence elements within sentence structure. The results of the study contribute to a deeper understanding of the syntactic system of the Uzbek language, as well as play an important role in the development of automatic syntactic analysis, morphological analysis, and Natural Language Processing (NLP) systems, and in evaluating language structure based on generalized models.



Keywords: *Natural Language Processing (NLP), linguistic model, object, valency, syntactic analysis, BIO tagging, corpus*

O'zbek tilining sintaksis bo'limida gap bo'laklari muhim strukturaviy va semantik birliklar sifatida namoyon bo'ladi. Gap bo'laklari mustaqil so'zlar va so'z birikmalari bilan ifodalanadi. Har qaysi turkum so'zlari gapda ma'lum bir tipik-sintaktik funksiyada qo'llanadi[2:39]. Gapning mazmunini to'liq va aniq ifodalashda ikkinchi darajali bo'laklar to'ldiruvchi, aniqlovchi va hol alohida ahamiyat kasb etadi. Ushbu bo'laklar asosiy predikativ bo'lak ma'nosini kengaytirib, harakatning obyektini, predmetning belgilarini hamda harakat va holatning turli sharoitlarini ifodalash orqali gapning mazmun ko'lamini oshiradi.

An'anaviy tilshunoslikda ikkinchi darajali bo'laklar ko'proq ularning grammatik shakllari va sintaktik vazifalari nuqtayi nazaridan tavsiflab kelingan. Biroq zamonaviy lingvistika, xususan, tabiiy tilni qayta ishlash doirasida ushbu birliklarni formal modellar asosida ifodalash zarurati ortib bormoqda. Bu esa ikkinchi darajali bo'laklarning nafaqat lingvistik xususiyatlarini, balki ularning strukturaviy va funksional munosabatlarini ham tizimli ravishda o'rganishni va shunga muvofiq gap bo'laklarini modellashtirishni talab etadi. Gap bo'laklarini modellashtirish – tilni qayta ishlash jarayonidagi eng asosiy va murakkab masalalardan biridir. Umuman olganda, modellashtirishda tadqiqotchi obyektning o'zini emas, balki uning modelini o'rganadi [7:21]. O'zbek tilida gap bo'laklarini avtomatik tahlil qilishda so'z birikmalarining qoliplarini aniqlash muhim bosqichdir. Tadqiqotchi O. Xidirov ta'kidlaganidek, so'z birikmalarini lisoniy-sintaktik qoliplar asosida teglash tizimning aniqligini ta'minlaydi [8:22]. Ushbu maqolada to'ldiruvchining lingvistik modellari ishlab chiqilib, ularning sintaktik va semantik xususiyatlari tahlil qilinadi. Shuningdek, ushbu bo'laklarning gap tarkibidagi o'рни va o'zaro aloqalari modellashtirish nuqtayi nazaridan ko'rib chiqiladi.



Ikkinchi darajali bo'laklar tavsifi. O'zbek tilida gapni tashkil etuvchi bo'laklar ya'ni bosh bo'laklar ega va kesim hisoblanib, ikkinchi darajali bo'laklarning ishtiroki gap strukturasi yangi bosqichga olib chiqadi. Ya'ni endi gap faqatgina harakat va harakat bajaruvchisini emas, balki harakat sodir bo'lgan joy, vaqt, tarzni, harakatning obyektini, harakat bajaruvchisining sifati kabi bir qancha qo'shimcha ma'nolarni anglatuvchi tugal lisoniy hodisa sifatida voqealanadi. Tabiiy tilni qayta ishlash sohasida ham ikkinchi darajali bo'laklarning modellari ML hamda DeepL sohalarida muhim ahamiyat kasb etadi. Boisi, mashinaviy o'qitish jarayonida gapning semantik va sintaktik tuzilmasini to'liq ifodalash model tomonidan kontekstni chuqurroq anglash, to'g'ri klassifikatsiya va aniq natijalar olish imkonini yaratadi.

To'ldiruvchining ifodalanishi. Gapning kesimi bilan bog'lanib predmetlik ma'nosini ifodalovchi va kimni?, nimani?, kimga?, nimaga?, kimdan?, nimadan?, kim (bilan, orqali, uchun)?, nima (bilan, orqali, uchun)? so'roqlariga javob bo'lgan ikkinchi darajali bo'lak to'ldiruvchidir. To'ldiruvchi kesimdan anglashilgan harakat-holat, belgi-xususiyat bilan bog'langan predmetni bildirganligi tufayli ot, olmosh va otlashgan so'zlar bilan ifodalanadi [6:62]. Bundan tashqari, gapning fe'l bilan ifodalangan biror bo'lagiga bog'lanib, uning ma'nosini to'ldirib keladi. Odatda to'ldiruvchi kesim bilan bog'lanib, u to'g'rida qo'shimcha ma'lumot beradi (4:127).

Sintaktik modelda to'ldiruvchi gap tuzilishida kesimga bog'langan holda aniqlanadi. U odatda fe'l bilan semantik va grammatik aloqada bo'ladi. Masalan, “*Men kitobni o'qidim*” gapida “kitobni” so'zi kesimga bog'lanib, harakat obyektini bildiradi. Shuningdek, “*U do'stiga yordam berdi*” gapida “do'stiga” bilvosita to'ldiruvchi vazifasida kelgan.

Morfologik modelda to'ldiruvchi so'zning shakliy ko'rsatkichlari orqali aniqlanadi. O'zbek tilida bu asosan kelishik qo'shimchalari yordamida ifodalanadi: tushum (-ni), jo'nalish (-ga), chiqish (-dan). Masalan, “*Talaba maqolani yozdi*”



(maqolani – tushum), “*Men ukamga sovg‘a berdim*” (ukamga jo‘nalish), “*U suvdan ichdi*” (suvdan – chiqish). Bundan tashqari, ko‘makchilar yordamida ham ifodalanadi: “*Men do‘stim bilan bordim*”, “*Bu kitob sen uchun*”, “*U internet orqali ishlaydi*”.

Bundan tashqari, to‘ldiruvchining semantik ma‘nosi ham mavjud bo‘lib, u harakatning obyektini bildiradi. Bu yondashuvda shakldan ko‘ra ma‘no ustun turadi. Masalan, “*Men kitob o‘qidim*” gapida tushum kelishigi qo‘shimchasi ishlatilmagan bo‘lsa ham, “kitob” semantik jihatdan to‘ldiruvchi hisoblanadi. Demak, ayrim hollarda to‘ldiruvchi yashirin holda ifodalanishi mumkin.

To‘ldiruvchining valentligi fe‘lning talab qiladigan argumentlaridan biri sifatida qaraladi. Har bir fe‘l ma‘lum miqdordagi obyektlarni talab qiladi. Masalan, “*bermoq*” fe‘li kamida ikki argumentni talab qiladi: kimga? nimani? (“*Men ukamga kitob berdim*”). Bu yerda “ukamga” va “kitob” fe‘lning ma‘no to‘ldiruvchilari hisoblanadi.

To‘ldiruvchining mazmuniy sintagmatik munosabati deyilganda esa to‘ldiruvchi vaziyatidagi sintaktik shakl ifodalagan obyekt bilan kesim vaziyatidagi sintaktik shakl ifodalagan predikat o‘rtasidagi munosabat nazarda tutiladi[5:50].

To‘ldiruvchining qoidalarga asoslangan modellari. Tabiiy tilni qayta ishlashda gapdagi munosabatlarni kompyuterga tushuntirish uchun har bir so‘z shaklining sintaktik qolipi (pattern) yaratilishi shart. Bu borada to‘ldiruvchining fe‘l bilan bog‘lanish imkoniyatlarini matematik-lingvistik modellashtirish asosiy vazifa hisoblanadi [1:56]. Qoidalarga asoslangan yondashuvda to‘ldiruvchi kelishik qo‘shimchalari va ko‘makchilar orqali aniqlansa, zamonaviy neyron modellar kontekst asosida aniqlaydi. Xususan, BERT va mBERT kabi transformer modellar yordamida to‘ldiruvchi yashirin semantik bog‘lanishlar orqali ham aniqlanishi mumkin. Mashinaviy o‘qitishda dastlab til imkoniyatlarini ochib berish uchun qoidalarga asoslangan yondashuvdan foydalanishimiz mumkin. Bunda har bir gap



bo‘lagi qaysi so‘z turkum bilan va qanday morfologik ko‘rsatkichlar bilan gapda kela olishini aniqlaymiz. Tajriba uchun 500 ta gapdan iborat annotatsiyalangan sintaktik korpus tuzdik. To‘ldiruvchini BIO tagging usuli yordamida aniqladik va umumiy xulosalar chiqardik. Qoidalarga asoslangan yondashuv yordamida to‘ldiruvchini aniqlash o‘zbek tilida morfologik ko‘rsatkichlarga tayanib samarali amalga oshirildi. Xususan, tushum (-ni), jo‘nalish (-ga), chiqish (-dan) kelishik qo‘shimchalari hamda “bilan”, “uchun”, “orqali” kabi ko‘makchilar to‘ldiruvchini aniqlashda muhim indikatorlar sifatida xizmat qiladi. Korpusning tuzilishini 1-jadvalda ko‘rishingiz mumkin:

1-jadval. Korpusning tuzilishi

sent_id	token_id	token	gap_chunk	gap_bo‘lagi
120	1	Eshik-derazani	B-O	To‘ldiruvchi
120	2	ochmang	B-P	Kesim
120	3	,	_	_
120	4	iloji	B-P	Kesim
120	5	bo‘lsa	I-P	Kesim
120	6	niqob	B-O	To‘ldiruvchi
120	7	taqing	B-P	Kesim
120	8	.	_	_

Korpusni annotatsiyalash uchun dastlab yig‘ilgan matnga ishlov berildi va buning uchun to‘ldiruvchining qoidalarga asoslangan modellari tuzib chiqildi:

```
import re
CASE_SUFFIXES = ["ni", "ga", "dan"]
POSTPOSITIONS = ["bilan", "uchun", "orqali"]
def clean_word(word):
return re.sub(r"[\w'"]", "", word.lower())
def find_objects(sentence):
words = sentence.split()
objects = []
```

Mazkur kod fragmenti qoidalarga asoslangan modelning boshlang‘ich qismi bo‘lib, unda:



1-rasm. Qoidalarga asoslangan yondashuvning boshlang'ich bosqichi

Keyingi bosqichlarda aynan shu ma'lumotlar asosida to'ldiruvchi elementlar aniqlanadi va natijalar ko'rib chiqiladi:

```
{'object': 'ukamga', 'type': 'kelishik (ga)', 'position': 1}  
{'object': 'kitobni', 'type': 'kelishik (ni)', 'position': 2}  
{'object': 'do'stim bilan', 'type': 'ko'makchili', 'position': 5}  
{'object': 'shahardan', 'type': 'kelishik (dan)', 'position': 6}  
{'object': 'u uchun', 'type': 'ko'makchili', 'position': 9}  
{'object': 'internet orqali', 'type': 'ko'makchili', 'position': 12}
```

Mazkur yondashuvning afzalligi uning sodda, tushunarli va tezkor ishlashidir. U kichik hajmdagi ma'lumotlar bilan ham samarali ishlay oladi va qo'shimcha o'qitishni talab qilmaydi. Shu bilan birga, modelni kengaytirish va yangi qoidalar qo'shish orqali aniqlikni oshirish mumkin. Biroq, ushbu yondashuv kontekstni chuqur hisobga olmasligi sababli, erkin so'z tartibi yoki qo'shimchalarning tushirib qoldirilishi holatlarida xatoliklar yuzaga kelishi mumkin. Shunga qaramay, qoidalarga asoslangan model to'ldiruvchini aniqlashda boshlang'ich va asosiy yechim sifatida muhim ahamiyat kasb etadi.

Ushbu maqolada tabiiy tilni qayta ishlash jarayonida to'ldiruvchini modellashtirish masalasi qoidalarga asoslangan yondashuv bilan tahlil qilindi. To'ldiruvchi gapning muhim sintaktik komponenti sifatida harakatning obyektini ifodalaydi va uning to'g'ri aniqlanishi matnni chuqur tushunishda muhim ahamiyat kasb etadi. Tadqiqot davomida to'ldiruvchini aniqlashning qoidalarga asoslangan yondashuvi asosiy usul sifatida ko'rib chiqildi. Ushbu yondashuvda tushum kelishik qo'shimchalari hamda ko'makchilar to'ldiruvchini aniqlovchi asosiy lingvistik indikatorlar sifatida belgilandi. Amaliy tajriba sifatida ishlab chiqilgan model ushbu belgilar asosida gap tarkibidan to'ldiruvchilarni ajratib olish imkonini berdi.



Natijalar shuni ko'rsatdiki, qoidalarga asoslangan model sodda tuzilishga ega bo'lib, tezkor ishlashi va tushunarli implementatsiyasi bilan ajralib turadi. U kichik hajmdagi ma'lumotlar bilan ham samarali ishlay oladi va murakkab o'qitish jarayonini talab qilmaydi. Shu bilan birga, modelni kengaytirish orqali yangi qoidalar qo'shish va aniqlikni oshirish imkoniyati mavjud. Biroq, mazkur yondashuvning ayrim cheklovlari ham mavjud. Xususan, o'zbek tilida erkin so'z tartibi, ayrim hollarda kelishik qo'shimchalarining tushirib qoldirilishi hamda kontekstga bog'liq ma'no farqlari model aniqligiga salbiy ta'sir ko'rsatishi mumkin. Shu sababli, qoidalarga asoslangan model ko'proq boshlang'ich va bazaviy yechim sifatida qaraladi.

O'zbek tilida gap bo'laklarini avtomatik tahlil qilishda so'z birikmalarining qoliplarini aniqlash muhim bosqichdir. Biz o'z tadqiqotimizni davomi sifatida ushbu yondashuvni BERT modeli bilan integratsiya qilishni maqsad qildik [2:5] Umuman olganda, to'ldiruvchini modellashtirishda lingvistik qoidalar muhim asos bo'lib xizmat qiladi. Kelgusida ushbu yondashuvni yanada takomillashtirish, uni statistik va neyron modellar bilan kuchaytirish hamda o'zbek tiliga mos korpuslar asosida sinovdan o'tkazish ushbu yo'nalishdagi tadqiqotlarni rivojlantirishga xizmat qiladi.

Foydalanilgan adabiyotlar ro'yxati:

1. Abdullayeva O., Xudayarova S. O'zbek tilida so'z birikmalarining lisoniy sintaktik qoliplari va ularni modellashtirish masalasi. – Toshkent: O'zbekiston til va madaniyat, 2023.
2. Abdurahmonov G'. Hozirgi o'zbek adabiy tili. – Toshkent: “O'qituvchi”, 1979.
3. Devlin J. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. – arXiv preprint, 2019.
4. Irisqulov M. Tilshunoslikka kirish. – Toshkent: “O'qituvchi”, 1992.



5. Mahmudov N., Nurmonov A. O‘zbek tilining nazariy grammatikasi (Sintaksis). – Toshkent: “O‘qituvchi”, 1995.
6. Nurmonov A., Mahmudov N., Sobirov A. Hozirgi o‘zbek adabiy tili. – Toshkent: “Ilm-ziyo”, 2004.
7. Po‘latov A. Kompyuter lingvistikasi. – Toshkent, 2009.
8. Xidirov O. So‘z birikmalarini lisoniy-sintaktik qoliplar asosida sintaktik teglash. – Toshkent: “O‘zbek amaliy filologiyasi istiqbollari”, 2022