

ISSN 2181-922X

LANGUAGE & CULTURE

UZBEKISTAN O'ZBEKISTON

UZBEKISTAN

TIL VA MADANIYAT

KOMPYUTER
LINGVISTIKASI

2024 Vol. 3 (6)

www.compling.tsuull.uz

ISSN 2181-922X

O‘ZBEKISTON

TIL VA MADANIYAT

KOMPYUTER LINGVISTIKASI

2024 Vol. 3 (6)

compling.tsuull.uz

Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti

Bosh muharrir:

Botir Elov

Bosh muharrir o‘rinbosari:

Shahlo Hamroyeva

Mas‘ul kotib:

Oqila Abdullayeva

Tahrir kengashi

Shuhrat Sirojiddinov (O‘zbekiston), Eshref Adali (Turkiya), [Viktor Zaxarov] (Rossiya), Vladimir Benko (Slovakiya), Ayrat Gatiatullin (Tataristan), Rinat Gilmullin (Tataristan), Murat O‘rxun (Turkiya), Suyun Karimov (O‘zbekiston), Abduvali Qarshiyev (O‘zbekiston), Muxammadjon Musayev (O‘zbekiston), Kamoliddin Shukurov (O‘zbekiston), O‘tkir Hamdamov (O‘zbekiston), Tal‘at Zuparov (O‘zbekiston), Bahodir Mo‘minov (O‘zbekiston), Faxriddin Nurullayev (O‘zbekiston), Zulxumor Xolmanova (O‘zbekiston), Muqaddas Abdurahmonova (O‘zbekiston), Elova Dilrabo (O‘zbekiston), Ruhillo Alayev (O‘zbekiston), Rasuljon Atamuratov (O‘zbekiston), Malika Abdullayeva (O‘zbekiston), Mannon Ochilov (O‘zbekiston), Xolisa Axmedova (O‘zbekiston), Zilola Xusainova (O‘zbekiston), Uldona Abdurahmonova (O‘zbekiston).

Jurnal haqida ma‘lumot

“O‘zbekiston: til va madaniyat. Kompyuter lingvistikasi” seriyasi – Oliy attestatsiya komissiyasi ilmiy nashrlar ro‘yxatidagi “O‘zbekiston: til va madaniyat” akademik jurnalining ilovasi hisoblanib, unda professor-o‘qituvchilar, doktorantlar, stajor-tadqiqotchilar, mustaqil izlanuvchilar, magistrantlarning kompyuter lingvistikasi, jumladan, tabiiy tilga ishlov berish (NLP), o‘zbek tilining formal grammatikasi, korpus lingvistikasi, mashina tarjimasi, nutqni qayta ishlash tizimlari, intellektual tizimlar, kompyuter leksikografiyasi hamda lingvistik ontologiyalar kabi sohalarga oid tadqiqotlari nashr qilinadi.

Jurnal ilovasi bir yilda to‘rt marta chop etiladi.

O‘zbek, turk, rus va ingliz tillarida yozilgan maqolalar qabul qilinadi.

Jurnalda kitoblarga yozilgan taqrizlar, adabiyotlar sharhi, konferensiyalar hisobotlari va tadqiqot loyihalari natijalari ham e‘lon qilinadi.

Mualliflar fikri tahririyat nuqtayi nazaridan farq qilishi mumkin.

“O‘zbekiston: til va madaniyat. Kompyuter lingvistikasi” seriyasi 2023-yildan chiqa boshlagan.

Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti. O‘zbekiston, Toshkent, Yakkasaroy tumani, Yusuf Xos Hojib ko‘chasi, 103-uy.

E-mail: kompling@navoiy-uni.uz

Website: kompling.tsuull.uz

Alisher Navo'i Tashkent State University of the Uzbek Language and Literature

Chief editor: Botir Elov
Deputy editor-in-chief: Shahlo Hamroyeva
Responsible secretary: Oqila Abdullayeva

Editorial board

Shukhrat Sirojiddinov (Uzbekiston), Eshref Adali (Turkiye), [Viktor Zakharov] (Russia), Vladimir Benko (Slovakia), Ayrat Gatiatullin (Tataristan), Rinat Gilmullin (Tataristan), Murat Orhun (Turkey), Suyun Karimov (Uzbekistan), Abduvali Karshiyev (Uzbekistan), Mukhammadjon Musayev (Uzbekistan), Kamoliddin Shukurov (Uzbekistan), O'tkir Hamdamov (Uzbekistan), Tal'at Zuparov (Uzbekistan), Bahadir Mo'minov (Uzbekistan), Fakhridin Nurullayev (Uzbekistan), Zulkhumor Kholmanova (Uzbekistan), Muqaddas Abdurakhmonova (Uzbekistan), Elova Dilrabo (Uzbekistan), Ruhillo Alayev (Uzbekistan), Rasuljon Atamuratov (Uzbekistan), Malika Abdullayeva (Uzbekistan), Mannon Ochilov (Uzbekistan), Kholisa Akhmedova (Uzbekistan), Zilola Khusainova (Uzbekistan), Uldona Abdurakhmonova (Uzbekistan).

Information about the magazine

"Uzbekistan: language and culture. "Computer Linguistics" series is an appendix of the academic journal "Uzbekistan: Language and Culture" in the list of scientific publications of the Higher Attestation Commission, in which computer linguistics, including natural language processing (NLP) of professors-teachers, doctoral students, intern-researchers, independent researchers, master's students, researches related to formal grammar of the Uzbek language, corpus linguistics, machine translation, speech processing systems, intelligent systems, computer lexicography and linguistic ontologies are published.

The magazine supplement is published four times a year.

Articles written in Uzbek, Turkish, Russian and English languages are accepted.

The journal also publishes book reviews, literature reviews, conference reports, and research project results.

The opinion of the authors may differ from the editorial point of view.

"Uzbekistan: language and culture. "Computer Linguistics" series has been published since 2023.

Tashkent State University of Uzbek Language and Literature named after Alisher Navoi. Yusuf Khos Hajib street, 103, Yakkasaray district, Tashkent, Uzbekistan.

E-mail: kompling@navoiy-uni.uz

Website: kompling.tsuull.uz

MUNDARIJA

Talat Zuparov

Word2vec metodi orqali matnlarni raqamlashtirish va mashinali o'qitish usullari orqali qayta ishlash.....6

Umidjon Yodgorov

O'zbek tili frazemalarining morfologik shakli va variantlari tadqiqiga an'anaviy va korpus tahlili yondashuvi.....29

Go'zal Erkinjonova

Fe'l so'z turkumiga oid birliklarning leksik-grammatik xususiyati.....49

Botir Elov, Maftuna Baratova

Pos (part of speech) teglash usullari.....62

Zebo Qodirova

Tibbiy atamalarni tur va sinflarga ajratishda ontologik tamoyillar.....78

Gulira'no Nuriddinova

Tabiiy tilni qayta ishlashda eganing modellari.....90

Botir Elov, Maftuna Ahmedova

N-gramlar asosida imloni tuzatish tizimini ishlab chiqish.....101

Mavluda Urazaliyeva

Audiomatnlarni korpusga kiritish muammolari tahlili.....115

CONTENT

Talat Zuparov

Digitization of texts using the word2vec method and processing through machine learning techniques.....27

Umidjon Yodgorov

The traditional and corpus-based approach to the study of morphological forms and variations of uzbek phraseology.....46

Go'zal Erkinjonova

Lexical and grammatical characteristics of verbal word class units.....60

Botir Elov, Maftuna Baratova

Pos (part of speech) tagging methods.....76

Zebo Kodirova

Ontological principles in the division of medical terms into types and classes.....88

Gulira'no Nuriddinova

Models of subject in natural language processing.....100

Botir Elov, Maftuna Ahmedova

Development of a spell correction system based on n-grams.....113

Mavluda Urazaliyeva

Analysis of problems in incorporating audio texts into a corpus.....124

TABIY TILNI QAYTA ISHLASHDA EGANING MODELLARI

Gulira'no Nuriddinova¹

Annotatsiya. Ushbu maqolada gap bo'laklaridan bosh bo'lak hisoblanmish eganing modellari so'z turkumlari (ot, fe'l, sifat, son, ravish, olmosh, ibora, taqlid, modal va undov so'zlar) kesimida aniqlangan, eganing har bir so'z turkum doirasida ifodalanishi misollar yordamida izohlangan hamda shu asosida mashina uchun o'qitishga mo'ljallangan modellari tuzilgan. Maqolaning ikkinchi qismida tabiiy tilni qayta ishlash jarayonida dependency parsing modellari ahamiyati, afzalliklari va ular yordamida eganing modellarini yaratish haqida ma'lumotlar berilgan. Eganing so'z turkumlar doirasida hamda dependency parsing asosida bunday modellarini yaratish o'zbek tili uchun sintaktik analizatorning ishlab chiqilishidagi dastlabki qadamlardan biri bo'lishi maqsad qilingan. Tadqiqot uchun besh yuzta gap saralab olingan va ularning tahlili asosida eganing modellari yaratilgan. Maqola tilshunoslik va kompyuter lingvistikasidagi yangi tadqiqotlar uchun ilmiy asos bo'lib xizmat qiladi.

Kalit so'zlar: *ega, tilni modellashtirish, NLP, sintaktik tahlil, amaliy tilshunoslik, tobelik tahlili (dependency parsing), token.*

Kirish

Zamonaviy axborot-texnologiyalar rivoji boshqa sohalar singari tilshunoslikka ham o'z ta'sirini o'tkazmasdan qolmadi. Tabiiy tillarni mashina uchun tushunarli bo'lgan sun'iy tilga o'tkazish masalasi tilshunoslik oldidagi dolzarb masalaga aylandi, shu boisdan, XX asrning 40-yillaridan boshlab rivojlana boshlagan ushbu ta'limotlar natijasida kompyuter lingvistikasi, korpus lingvistikasi, mashina tarjimai kabi bir qancha ilmiy yo'nalishlar vujudga keldi (Xonnazarov, 2023. 3). Jahon tilshunosligida ushbu sohalarda olib borilgan tadqiqotlar diqqatga sazovor. Tilshunoslar tomonidan

¹Nuriddinova Gulira'no Baxtiyor qizi - Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti Kompyuter lingvistikasi mutaxassisligi 1-kurs magistranti.

E-pochta: guliranonuriddinova@gmail.com

gap bo'laklarini modellashtirish, sintaktik tahlil bo'yicha bir qator nazariy yondashuvlar ishlab chiqilgan. Bulardan Chomskiyning Transformational Grammar (Transformatsion grammatika), Dependency Grammar (bog'lanishli grammatika), Lexical Functional Grammar (Leksik-funksional grammatika), Cognitive Grammar (Kognitiv grammatika), Probabilistic Parsing (Ehtimolli bog'lanish), Combinatory Categorical Grammar (Kombinatorik kategoriyali grammatika) kabilar amaliy tilshunoslik uchun asos bo'lib xizmat qilmoqda. Bundan tashqari BERT, GPT kabi modellar, spaCy, AllenNLP kutubxonalari, MaltParser, UDPipe, Berkeley Parser analizatorlari, Google Translate va DeepL kabi mashina tarjima xizmatlarida gap bo'laklarini modellashtirish va analiz qilish uchun kuchli NLP algoritmlari ishlab chiqilgan. Siri, Alexa, Google Assistant kabi virtual yordamchilar esa gap bo'laklarini tushunib, ma'noli javob qaytarish uchun NLP modellaridan foydalanadi. Bu tizimlar gapdagi subyekt, predikat va boshqa grammatik bo'laklarni ajratish imkoniyatiga ega. 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 (Po'latov, 2009. 21). Inson va nutqning tabiiylik darajasi, so'zlar va gap bo'laklari o'rtasidagi murakkab munosabatlar kompyuter tizimlariga bu bog'lanishlarni aniqlashda ko'plab qiyinchiliklarni keltirib chiqaradi. Shunga qaramay, zamonaviy til texnologiyalari (avtomatik tarjima tizimlari, matn tahlili, ovozli yordamchilar va boshqalar) aniq va samarali ishlashi uchun gap bo'laklarini to'g'ri modellashtirish muhimdir. Til o'z tabiatiga muvofiq doim o'zgarishda va rivojlanishda bo'ladi. Bugungi kunda qaysidir tilning keng ommalashishi, foydalanuvchilari sonining ortishi hamda umrboqiyliги o'z-o'zidan uning zamonaviy texnologiyalar orqali qayta ishlanishiga, nazariy tilshunoslikdagi masalalarning amaliy tilshunoslikda ifodalanishiga bog'liq bo'lib qolmoqda. Shu jumladan, har bir til uchun til modellarini yaratish tilshunoslik, kompyuter lingvistikasi sohasi uchun, shu va shu kabi boshqa sohalarda amalga oshiriladigan tadqiqotlar uchun muhim ahamiyat kasb etadi. Til modellari tabiiy tilni qayta ishlashda (Natural Language Processing – NLP), matnlardagi gap bo'laklari va so'zlarning umumiy strukturasi aniqlashda muhimdir. Ushbu maqolada gap bo'laklaridan bosh bo'lak hisoblanmish eganing modellari so'z turkumlari doirasida aniqlanadi va dependency parsing orqali eganing modellarini tuzishning ahamiyati tahlil qilinadi.

Ega va kesim gapning bosh bo'laklari hisoblanadi va kesimdan anglashilgan harakat va holat ega orqali shaxs va sonda aniqlashadi. Eganing umumiy strukturasi to'xtaladigan bo'lsak, u hamisha bosh kelishik shaklida keladi hamda o'ziga mos egalik va ko'plik qo'shimchalarini olishi mumkin. Tabiiy tilni qayta ishlashda eganing modellarini tuzish sintaktik analizator yaratishda asos bo'lib xizmat qiladi. O'zbek tili lug'at tarkibi boy hamda grammatik qoidalari nisbatan murakkab til bo'lganligi sababli unda til modellarini tuzish ham bir qadar murakkablik kasb etadi. Shu boisdan qoidalarga asoslangan til modellarini tuzish uchun tilshunosdan puxta bilim va to'g'ri yondashuv talab qilinadi. Maqolada eganing so'z turkumlari bo'yicha ifodalanishi hamda gapda boshqa bo'laklar bilan bo'lgan aloqasi tahlil qilinib, shu yondashuvlar asosida eganing modellari tuziladi.

Asosiy qism

1.1. Eganing so'z turkumlar doirasida ifodalanishi

Ega (S-subject) gapda turli xil so'z turkumlari va so'zlar bilan ifodalanib kelishi mumkin. Ularga ot, olmosh, sifat, son, ravish, fe'lning harakat nomi, sifatdosh shakllari, taqlid, modal, undov so'zlar, ibora kiradi. Ular gapda ega bo'la olishlari uchun bosh kelishik shaklida kelishi, egalik qo'shimchasini olishi talab qilinadi (keyingi o'rinlarda eganing modellarini ifodalashda bosh kelishik uchun - b.k, egalik qo'shimchasi uchun - e.g, ko'plik shakli uchun - k belgilaridan foydalaniladi). Shu jihatdan eganing so'z turkumlar doirasidagi modellari quyidagicha ifodalanishi mumkin:

Eganing ot bilan ifodalanishi (N). Ega gapda eng ko'p ot so'z turkumi bilan ifodalanadi. Buning bir nechta sabablari bor, albatta. Birinchidan otga xos egalik, bosh kelishik qo'shimchalari yordamida aynan ega shakllanadi. Morfologik hamda sintaktik tahlil jarayonida beriladigan so'roq ham har ikkisida bir xil, ya'ni "kim?, nima?, qayer?". Keyingi omil sifatida o'zbek tilida so'z turkumlari doirasida otlarning ko'pligini olishimiz mumkin. Demak, ega gapda ot so'z turkumi bilan ifodalanganda bosh kelishik shaklida hamda egalik, ko'plik shakllarini yo olib, yo olmasdan kelish huquqiga ega. Ushbu so'z turkum doirasida eganing ifodalanishini turli misollar yordamida ko'rib chiqdik va eganing ushbu modellarini yaratdik:

1-jadval

S=N+b.k	<i>Oila</i> -muqaddas dargoh.
S=N+b.k, k	<i>Imtixonlar</i> yil yakuniga kelganda boshlanadi.

S=N+b.k, e.g	<i>Dardim</i> senikidan-da kattaroq.
S=N+b.k, k, e.g	<i>Xayollarim</i> har tomonga tarqab ketdi.

Eganning ot soʻz turkumi doirasidagi shakllari shu holatda ifodalanadi va tabiiy tilni qayta ishlash jarayonida ushbu modellar mashina uchun egani aniqlash jarayonini osonlashtiradi.

Eganning olmosh bilan ifodalanishi (P). Olmosh soʻz turkumi boshqa soʻz turkumlari hamda soʻzlar bilan almashib kela oladi. Shuning uchun bu soʻz turkumi bir nechta gap boʻlaklarining yasalişiga xizmat qiladi. Olmosh bilan ifodalanganda egani ushbu modellar yordamida aniqlaymiz.

2-jadval

S=P+b.k	<i>Sen</i> ertaga yetuk kadr boʻlib yetishishing kerak.
S=P+b.k, e.g	Uning bu dunyoda otasidan boshqa <i>hech kimi</i> yoʻq.
S=P+b.k, k	<i>Ular</i> allaqachon yetib kelishgan.
S=P+b.k, k, e.g	<i>Bularing</i> haligacha kelmadi.
S=P+k, niki; S=P+niki	<i>Ular</i> niki ham, <i>seniki</i> ham tayyor.

Eganning son bilan ifodalanishi (NUM). Son soʻz turkumi gapda ega vazifasida kelishi uchun otlashgan holatda boʻlishi kerak. Ot va olmosh uchun egalik, kelishik, koʻplik shakllari ularning turkumlik xususiyatiga xos hisoblansa, boshqa bir soʻz turkumlari va soʻzlar ham borki, ularning ushbu qoʻshimchalarni olib kelish hodisasini otlashish deb nomlaymiz. Eganning modellari son soʻz turkumi doirasida quyidagichadir:

3-jadval

S=NUM+b.k	<i>Oltovlon</i> ola boʻlsa, ogʻzidagini oldirar. <i>Ikkov</i> hamisha inoq yurar edi, qachonki oralariga Said tushmaguncha.
S=NUM+b.k, e.g	Qiziq, <i>biri</i> – qochoq, <i>biri</i> –qahramon.
S=NUM+b.k, k	Katta farzandim maktabdan hamisha <i>beshlar</i> olib kelardi.
S=NUM+b.k, k, e.g	Bu yil <i>toʻrtlarim</i> ancha koʻpayib ketdi.

Miqdor songa - ov, -ala, -ovlon qoʻshimchalarini qoʻshish yoʻli bilan hosil boʻlgan *beshov, ikkala, toʻrtovlon* kabi jamlovchi sonlar otlarga yaqin turgani uchun otlashgan va otlashmagan holatda ham ega vazifasini bajarishi mumkin.

Eganning sifat bilan ifodalanishi (JJ). Ega sifat bilan ifodalanganda ham otga xos qoʻshimchalar bilan keladi va u bogʻlanib

kelgan ot so'z turkumi tushib qolganligi kontekstdan anglashib turadi. Sifat bilan ifodalanganda eganing modellarini ko'rib chiqsak:

4-jadval

S= JJ+b.k	<i>Aqlli o'zini ayblar, aqlsiz do'stini ayblar.</i>
S=JJ+b.k, k	Sinfda <i>faollar</i> ko'payganidan ancha hursand bo'ldim.
S=JJ+b.k, e.g	<i>Go'zalim</i> -bevafo gulistonim.
S=JJ+b.k, k. e.g	Sinfning eng <i>yaxshilari</i> hozir shu yerda.

Eganing ravish bilan ifodalanishi (RR). Bosh bo'lakning ravish bilan ifodalanish holatlari kam bo'lsa-da, ba'zida uchrab turadi. Misol uchun ushbu modellarda ravish so'z turkumi gapda ega vazifasida keladi:

5-jadval

S=RR+b.k, e.g	<i>Ko'pi</i> ketib, <i>ozi</i> qoldi. <i>Bugunim</i> ham o'tdi, <i>ertam</i> ham o'tadi.
S=RR+b.k, k	<i>Ko'plar</i> bu kunlarga yetishni orzu qiladi.
S=RR+b.k, k, e.g	<i>Ozlaring</i> bu tarafga qarab yuringlar, <i>ko'plaring</i> esa bu tarafga.

Eganing fe'l bilan ifodalanishi (VB). Fe'l so'z turkumi gapda ega vazifasida ham kelishi mumkin. Uning vazifa shakllaridan harakat nomi hamda sifatdosh shakllari otlarga xos egalik, kelishik, ko'plik qo'shimchalarini olib keladi va gapda eganing so'roqlariga javob bo'ladi. Misol uchun:

6-jadval

S=VB+b.k	<i>Eshitgan</i> ko'rganni yengar. <i>Ishlamoq</i> - ishtaha ochar.
S=VB+b.k, k	<i>O'qiganlar</i> ko'paysa, xalq ma'rifatli bo'ladi. Bu <i>yugurib-yelishlar</i> u uchun baribir foyda bermaydi.
S=VB+b.k, e.g	Imtihonda <i>tayyorlanganim</i> tushdi. Bu <i>borishim</i> eng oxirgisi bo'ladi.
S=VB+b.k, k, e.g	Menga <i>qilganlaring</i> uchun rozi bo'l, do'stim. <i>Nolishlaring</i> o'z boshingga yetadi.

Sifatdosh va harakat nomi ham fe'l shakli hisoblanadi. Ularning farqi ega sifatdosh bilan ifodalanganda u bog'lanib kelgan ot tushib qolgani gap mazmunidan sezilib turadi. Harakat nomi fe'lning otga xoslangan shakli bo'lganligi uchun bu hodisa kuzatilmaydi.

So'z turkumlaridan alohida olingan so'zlar ham borki, ular ham gap tarkibida kelib ega bajargan vazifani bajarishi mumkin.

Bularning birinchisi taqlid soʻzlar boʻlib, ular ham otlashib, yaʼni otga xos qoʻshimchalarni olib keladi.

Eganning taqlid soʻz bilan ifodalanishi (IM). Taqlid soʻzlar ega vazifasida kelganida u bogʻlanib kelgan ot tushib qoladi va undagi qoʻshimchalar oʻz-oʻzidan taqlid soʻzga qoʻshiladi.

7-jadval

S=IM+b.k	Hovlida ertalabdan buyon <i>“taq-taq”</i> tinmaydi.
S=IM+b.k, k	Qulogʻimga qayerlardandir <i>shivir-shivirlar</i> eshitilyapti.
S=IM+b.k, e.g	<i>Taqir-tuquring</i> bilan hammani bezovta qilyapsan.
S=IM+b.k, k, e.g	Suvning <i>shildir-shildirlari</i> odamga xush yoqadi.

Eganning undov soʻz bilan ifodalanishi (UH). Alohida olingan soʻz turkumlaridan undov soʻzlar ham gapda ega boʻlib kela oladi:

8-jadval

S=UH+b.k	<i>“Dod-voy”</i> borgan sari kuchaydi.
S=UH+b.k, e.g	<i>“Oh”im</i> osmonga yetdi.
S=UH+b.k, k	<i>“Oh-uf”lar</i> endi foyda bermaydi.
S=UH+b.k, k, e.g	Uning <i>“hoy-hoy”lari</i> oʻgʻliga eshitilmadi.

Eganning modal soʻz bilan ifodalanishi (MD). Eganning modal soʻz bilan ifodalanishi koʻp ham uchramaydigan hodisa boʻlsa-da, tilimizda mavjud. Aynan *“bor-yoʻq”* modal soʻzlari orqali bu hodisa amalga oshadi:

9-jadval

S=MD+b.k	<i>Bor</i> soʻylaydi, <i>yoʻq</i> oʻylaydi.
S=MD+b.k, e.g	<i>Borim</i> shu farzandlarimniki-da.
S=MD+b.k, k	<i>Borlar</i> bugungi mavzuni oʻzlashtirdi.
S=MD+b.k, k, e.g	<i>Yoʻqlaring</i> darsni qayta topshirish uchun kelasizlar.

Soʻz turkumlar doirasida eganing ifodalanishi shu modellar asosida tahlil qilinadi. Demak, ega gapda yordamchi soʻz turkumlaridan tashqari barcha soʻz turkumlari bilan ifodalanish xususiyatiga ega ekan. Endi eganing ibora bilan ifodalanishini koʻrib chiqsak:

Eganning ibora bilan ifodalanishi (IB).

S=IB+b.k	<i>Gapning po'stkallasi shuda.</i>
S=IB+b.k, k	Chumchuq yuraklar ketishi mumkin, albatta.
S=IB+b.k, e.g	"O'n gulidan bir guli ochilmagan"ing hozir kelib ketdi.
S=IB+b.k, k, e.g	"Yuzi qoralar"im kelishga ham uyalishadi bu yerga.

Ko'rinib turganidek, iboraning ham ba'zi turlari uchraydiki, gapda bosh bo'lak bo'lib kela olish uchun egaga xos bosh kelishik, ko'plik yoki egalik qo'shimchalarining biri bilan ifodalanadi. Iboraning o'zi, asosan, ko'chma ma'noda ishlatilgani bois yuqoridagi misollarda ham ushbu ma'no yetakchilik qiladi.

2.1. Eganing dependency parsing modellari.

Dependency parsing – bu tabiiy tilni qayta ishlashda gapning grammatik tuzilishini tahlil qilish hamda so'zlar orasidagi bog'lanishlarni aniqlash uchun foydalaniladigan usullar tizimi. Eganing modellarini tuzishda uning ahamiyati haqida to'xtalsak. Eganing modellarini yaratishdagi ikkinchi usul sifatida aynan dependency parsingni afzalligi shundaki, u so'zlar orasidagi sintaktik bog'lanishlarni aniqlab, eganing gapdagi boshqa bo'laklar bilan aloqasini aniq belgilab beradi. Bu, ayniqsa, eganing birikmali holatda ifodalangan vaziyatlarida qo'l keladi.

O'zbek tilida eganing dependency parsing modellarini yaratish uchun, avvalo, o'zbek tilidagi matnlarni to'plash, sintaktik bog'lanishlardan (masalan, nsubj, csubj kabi teglar) foydalanib, egani tahlil qilish va tahlil asosidagi modellarni mashinaga o'qitish talab qilinadi. Tayyorlanadigan ma'lumotlar quyidagilarni o'z ichiga oladi:

1. Tokenlar – bunda gapdagi har bir so'z alohida token shaklida ajraladi.
2. Teglar – gapdagi har bir so'z sintaktik tahlil qilinadi va natijalari yoziladi (nsubj, obj va h.k).
3. Bosh so'z – gapdagi so'zlar yo egaga, yoki kesimga bog'lanishiga ko'ra aniqlanadi.

Natijada dependency parsing eganing bir nechta turlarini bemalol aniqlay oladi:

Bir so'z turkumi bilan ifodalangan egalar: "**Anvar (subj)** bugun keladi".

Birikmali egalar: "Uning **yaxshi o'qigani (subj)** hammaga ma'lum".

Yashirin ega: "**(Sen-subj)** Keldingmi?"

Passiv ega: “**Kitob (subj)** Ali tomonidan o‘qildi”.

Aniqlovchi bilan kelgan ega: “Yaxshi o‘qiyotgan **Said (subj)** bugun darsda ko‘rinmadi”.

Ega (subject) gapning markaziy tarkibiy elementlaridan biri bo‘lib, uning aniqlanishi sintaktik analizatorning to‘g‘ri ishlashini ta‘minlaydi. Dependency parsing yordamida eganing modellarini yaratishda bir nechta muhim qadamlar bor, ularning birinchisi, eganing roli va vazifasini dastavval tushunish. Ega gapda kesim orqali anglashilgan harakat va holatning bajaruvchisini bildiradi. Eganing roli dependency parsingda nsubj (nominal subject) shaklida bo‘ladi. Har bir gapda egaga tegishli so‘zlar o‘rtasidagi sintaktik bog‘lanishlarni ushbu model aniqlaydi. Misol: “Ali kitob o‘qiyapti”.

Dependency parsing:

o‘qiyapti (root)
├— Ali (nsubj)
└— kitob (obj)

Ega — Ali, kesim — o‘qiyapti, obyekt — kitob.

Bundan tashqari, to‘plangan ma‘lumotlar bazasi yordamida dependency parsing vositalaridan foydalanish ham keyingi muhim qadamlardan biridir. Bu vositalar so‘zlar o‘rtasidagi bog‘lanishlarni aniqlashga yordam beradi. Ularga quyidagilarni kiritishimiz mumkin: SpaCy –keng tarqalgan, o‘rganish uchun yengil va yuqori samarali model; StanfordNLP (Stanza) –ochiq manba modeli, keng tarqalgan va o‘zbek tilida ham ishlatilishi mumkin; Transformers (BERT, RoBERTa) – katta til modellaridan foydalanish, yuqori aniqlikdagi tahlil qilish imkonini beradi. Bu vositalar yordamida o‘zbek tilidagi gaplar bo‘yicha dependency parsingni amalga oshirish hamda modellar yaratish mumkin. Undan so‘ng o‘qitilgan til modellarini sinash va baholash ham zarur. Sinovda modelning egani to‘g‘ri aniqlashini tekshirish kerak. Modelning aniqlik (accuracy), recall va F1-score kabi ko‘rsatkichlari orqali baholash mumkin. Shuningdek, modelni optimallashtirish va tuzatish zarur bo‘ladi. Modelni yaratishdan keyin uning natijalarini vizualizatsiya qilish yordamida sintaktik daraxt diagrammalarini chizish mumkin. Bu tahlilni yaxshilash va foydalanuvchilarga tushunarli qilish uchun yordam beradi.

Dependency parsingda eganing modellarini yaratish bir qancha afzalliklarga ega: egani aniq aniqlash va tahlil qilish, eganing gapdagi boshqa komponentlar bilan aloqalarini ko‘ra olish, eganing bir nechta turlarini tushunish va aniqlay olish hamda murakkab grammatik tuzilmalarda egani tahlil qila olish shular jumlasidandir.

Xulosa

Gap bo'laklarining modellarini tuzishda sintaktik tahlil jarayoni muhim hisoblanadi. Hozirgi kunda yaratilgan va yaratilayotgan sintaktik analizatorlarning ham asosi sintaktik tahlilga asoslangan katta miqdordagi ma'lumotlar bazasidir. Shuning uchun ham eganing yuqoridagi kabi modellarini tuzishda besh yuzta gap tahlilidan foydalandik hamda shu asosda eganing modellarini yaratdik. Bundan tashqari analizatorning to'g'ri ishlashi uchun tilning sintaktik va grammatik tuzilmalarini o'rganish ham zarur (ega, kesim, to'ldiruvchi, aniqlovchi, hol va ularning bog'lanish qoidalari, masalan, o'zbek tilida, ega + to'ldiruvchi + kesim). Amaliy qadam esa tilning sintaktik xususiyatlarini o'rgatish uchun katta hajmli, sifatli va belgilangan ma'lumotlar to'plash orqali amalga oshadi. Bu ma'lumotlar to'plami (corpus) annotatsiyalangan bo'lishi, ya'ni har bir gapdagi so'zlar va ularning sintaktik bog'lanishlari ko'rsatilgan bo'lishi zarur. Yana qo'lda tayyorlangan, sintaktik bog'lanishlar belgilangan matnlar to'plamidan, mavjud dependency parsing vositalaridan (masalan, Universal Dependencies) foydalanish ham mumkin. Umumiy olganda ushbu bosqichlar orqali sintaktik analizator yaratish imkoniyati mavjud:

1. Tilning grammatik qoidalarini aniqlash.
2. Annotatsiyalangan ma'lumot to'plamini tayyorlash.
3. Dependency parsing vositalaridan foydalanish.
4. Mashina o'qitish modeli yordamida analizatorni yaratish talab qilinadi.

Eganing yuqoridagi modellarini yaratishda ham o'zbek tili nazariy grammatikasi asos vazifasini bajargan. Qoidalarga asoslangan yondashuvga ko'ra eganing qaysi so'z turkumlari orqali ifodalana olishi, gapda qanday grammatik shakllarda kelishi hamda o'zidan avvalgi va keyingi so'zlar bilan bog'lanishi uning modellarini tuzishda muhim o'ringa ega.

Mazkur maqolada tabiiy tilni qayta ishlash jarayonida eganing modellarini yaratish masalasi keng yoritilgan. Avvalo, eganing qoidalarga asoslangan modellarini shakllantirishda so'z turkumlari (ot, fe'l, sifat va boshqalar) va ularning gapdagi sintaktik vazifalarini tahlil qilish asosiy yondashuv sifatida belgilandi. Bu jarayonda eganing turli ko'rinishlari (yashirin ega, murakkab ega va h.k) izohlangan va qoidalarga asoslangan nazariy asoslar ishlab chiqilgan. Maqolaning ikkinchi qismida dependency parsing texnologiyasining eganing modellarini yaratishdagi ahamiyati tahlil qilindi. Ayniqsa, dependency parsing yordamida gapdagi so'zlar

o'rtasidagi bog'lanishlarni aniqlash va eganing turli shakllarini samarali identifikatsiya qilish imkoniyati ko'rsatib o'tildi. Ushbu yondashuv o'zbek tili uchun sintaktik tahlil vositalarini ishlab chiqishning muhim bosqichlaridan biri ekanligi ta'kidlandi. Yakuniy natijalarga ko'ra, o'zbek tilida eganing qoidalarga asoslangan va dependency parsing texnologiyasiga asoslangan modellarini ishlab chiqish sintaktik analizatorni yaratish uchun ilmiy asos bo'lib xizmat qiladi. Ushbu tadqiqot tabiiy tilni qayta ishlash, tilshunoslik va kompyuter lingvistikasidagi yangi ishlanmalar uchun amaliy ahamiyatga ega.

Foydalanilgan adabiyotlar

- Abzalova M. Hozirgi o'zbek tili sintaksisi. – Toshkent: Sharq nashriyoti, 2005.
- Bolshakov I. Computational Linguistics. First Edition: 2004.
- Lisa S. Computational modeling for language acquisition: A tutorial with syntactic islands.
- Nurmonov A. Sistem tilshunoslik va uning asoslari. – Toshkent: O'zbek Milliy Ensiklopediyasi NMIU, 2010.
- Po'latov A. Kompyuter lingvistikasi. –Toshkent, 2009.
- Radford, A. An Introduction to English Sentence Structure. Cambridge University Press, 2010
- Sayfullayeva R. Hozirgi o'zbek adabiy tili. – Toshkent: O'qituvchi, 2007.
- G'ulomov A va Asqarova M. Hozirgi o'zbek tili: Gap bo'laklari va ularning funksiyalari. – Toshkent: Fan nashriyoti, 1985.

MODELS OF SUBJECT IN NATURAL LANGUAGE PROCESSING

Gulira'no Nuriddinova¹

Abstract. This article focuses on modeling the subject, one of the main components of a sentence, based on parts of speech (noun, verb, adjective, numeral, adverb, pronoun, phrase, onomatopoeia, modal words, and interjections). The expression of the subject within each part of speech category is explained with examples, and models intended for machine learning are developed accordingly. The second part of the article discusses the importance and advantages of dependency parsing models in natural language processing, as well as how these models can be used to create subject models. Developing such models based on parts of speech and dependency parsing is aimed as one of the initial steps in creating a syntactic analyzer for the Uzbek language. A dataset of 500 sentences was selected for the research, and subject models were created based on their analysis. This article serves as a scientific basis for new research in linguistics and computational linguistics.

Keywords: *subject, language modeling, NLP, syntactic analysis, applied linguistics, dependency parsing, token.*

References

- Abzalova M. Syntax of Modern Uzbek Language. – Tashkent: Sharq Publishing House, 2005.
- Bolshakov I. Computational Linguistics. First Edition: 2004.
- Lisa S. Computational Modeling for Language Acquisition: A Tutorial with Syntactic Islands.
- Nurmonov A. System Linguistics and Its Foundations. – Tashkent: Uzbek National Encyclopedia NMIU, 2010.
- Polatov A. Computational Linguistics. – Tashkent, 2009.
- Radford, A. An Introduction to English Sentence Structure. Cambridge University Press, 2010.
- Sayfullayeva R. Modern Uzbek Literary Language. – Tashkent: Teacher, 2007.
- Gulomov A. and Asqarova M. Modern Uzbek Language: Sentence Components and Their Functions. – Tashkent: Science Publishing House, 1985.

¹Nuriddinova Gulira'no Baxtiyor qizi - Master of degree. Alisher Navo'i Tashkent State University of Uzbek Language and Literature.

E-pochta: guliranonuriddinova@gmail.com

Jurnal 2017-yil 26-oktyabrda O'zbekiston Respublikasi Matbuot va axborot agentligi tomonidan 0936-raqam bilan ro'yxatdan o'tgan.

Jurnal O'zbekiston Respublikasi Oliy Attestatsiya Komissiyasi tomonidan filologiya fanlari bo'yicha falsafa doktori (PhD) va fan doktori (DSc) dissertatsiyalari asosiy ilmiy natijalari chop etilishi lozim bo'lgan ro'yxatga kiritilgan (30.10.2021. № 308/6).

Tahririyatga kelgan maqolalar mualliflarga qaytarilmaydi.

Manzil: Toshkent shahri, Yakkasaroy tumani, Yusuf Xos
Hojib ko'chasi 103-uy.
Telefonlar: +99871 281-45-11, +99871 281-41-93.
Website: compling.tsuull.uz
E-mail: kompling@navoiy-uni.uz

Bosishga **.**.*-yilda ruxsat etildi.
Bichimi 70x100 1/16, Ofset bosma. "Cambria" garniturasida.
Shartli b.t. 7,51. Nashr b.t. 7,62.

"O'zbekiston: til va madaniyat" jurnali tahririyatida tayyorlandi va sahifalandi.
"YASHNOBOD NASHR" bosmaxonasida chop etildi.
Adadi 300 nusxa. Buyurtma №2.
Bosmaxona manzili: Toshkent shahar Yashnobod tumani,
58-a harbiy shaharcha.