

UZBEKISTAN
LANGUAGE & CULTURE
O'ZBEKISTON

ISSN 2181-922X

TIL VA MADANIYAT

**KOMPYUTER
LINGVISTIKASI**

2025 Vol. 2 (6)

www.compling.tsuull.uz

O'ZBEKISTON

TIL VA MADANIYAT

KOMPYUTER LINGVISTIKASI

2025 Vol. 2 (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), , Vladimir Benko (Slovakiya), Ayrat Gatiatullin (Tataristan), Rinat Gilmullin (Tataristan), 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), Umid Yodgorov (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 tarjimai, 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: compling.tsuull.uz

Alisher Navoi'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), Vladimir Benko (Slovakia), Ayrat Gatiatullin (Tataristan), Rinat Gilmullin (Tataristan), Suyun Karimov (Uzbekistan), Abduvali Karshiyev (Uzbekistan), Mukhammadjon Musayev (Uzbekistan), Kamoliddin Shukurov (Uzbekistan), Utkir Hamdamov (Uzbekistan), Talat Zuparov (Uzbekistan), Bahadir Muminov (Uzbekistan), Fakhriddin 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), Umid Yodgorov (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: compling.tsuull.uz

MUNDARIJA**Zilola Xusainova, Rashidova Nazira**

Semantik role labeling (srl)ning zamonaviy yondashuvlari.....6

Zilola Xusainova, Norbekova Bahora

Ko'p ma'noli so'zlarning ma'nosini kontekst asosida aniqlash (WSD)ning nazariy konsepsiyasi.....18

Ruhillo Alayev, Mullaboyeva Xurmatoy

Video lavhalarga o'zbekcha avtomatik subtitrlar yaratish mavzusining o'rganilganlik darajasi.....30

Kenjayeva Marjona

Matn davri tushunchasi va uning lingvistik ahamiyati.....43

Aynura Axmedova

WordNetning psixolingvistik asosi.....54

Olimjanova Feruza

Sentiment tahlili va emotsiya aniqlashning nazariy asoslari.....66

Nozima Umarova

Lex.uz platformasi va uning huquqiy diskursga oid savol-javob tizimlari uchun ma'lumotlar bazasi sifatidagi roli.....77

CONTENT

Zilola Xusainova, Rashidova Nazira
Modern approaches to semantic role labeling (srl).....16

Zilola Xusainova, Norbekova Bahora
The theoretical concept of context-based word sense disambiguation (WSD).....28

Ruhillo Alayev, Mullaboyeva Khurmatoy
The study level of automatic uzbek subtitle generation for video materials.....41

Kenjayeva Marjona
The concept of text and its linguistic significance.....53

Aynura Akhmedova
The psycholinguistic foundations of WordNet.....64

Olimjanova Feruza
The theoretical foundations of sentiment analysis and emotion detection.....75

Nozima Umarova
The Lex.uz platform and its role as a database for question-and-answer systems related to legal discourse.....86

KO'P MA'NOLI SO'ZLARNING MA'NOSINI KONTEKST ASOSIDA ANIQLASH (WSD)NING NAZARIY KONSEPSIYASI

Zilola Xusainova¹,
Norbekova Bahora²

Annotatsiya. Ushbu maqolada tabiiy tilni qayta ishlash (Natural Language Processing, NLP) sohasining muhim muammolaridan biri hisoblangan ko'p ma'noli so'zlarning ma'nosini kontekst asosida aniqlash (Word Sense Disambiguation (WSD)) masalasining nazariy konsepsiyasi yoritiladi. Shuningdek, polisemiya hodisasining lingvistik tabiati, kontekst tushunchasining talqini, WSD jarayonining shakllanishi va rivojlanish bosqichlari ilmiy-nazariy jihatdan tahlil qilinadi. WSD sun'iy intellekt tizimlarida semantik noaniqliklarni bartaraf etish, matnni tahlil qilish, mashina tarjimasini, axborot izlash, savol-javob tizimlari, sintaktik hamda semantik annotatsiya jarayonlarida muhim ahamiyat kasb etadi. Maqolada lug'aviy model, korpusga asoslangan metodlar, zamonaviy neyron yondashuvlar haqida fikr yuritiladi.

Kalit so'zlar: *Ko'p manolilik, polisemiya, ma'no aniqlash, WSD, NLP*

Kirish

Til murakkab va ko'p qatlamli tizim bo'lib, undagi birliklar o'zaro semantik, sintaktik va pragmatik munosabatlar orqali bog'langan. Leksik birliklarning eng muhim xususiyatlaridan biri ko'p ma'nolilik hodisasi bo'lib, u tabiiy tillarning deyarli barchasida uchraydi. Ko'p ma'noli so'zlar nutq jarayonida turli kontekstlarda turlicha ma'nolarni ifodalashi mumkin. Mazkur holat inson uchun ko'pincha tabiiy va sezilmas bo'lsa-da, avtomatik til qayta ishlash tizimlari uchun murakkab muammo hisoblanadi. Tabiiy tilni qayta ishlash (Natural Language Processing – NLP) sohasida ko'p ma'noli so'zlar (polysemous words)ning ma'nosini aniqlash masalasi eng murakkab va dolzarb muammolardan biri hisoblanadi. Ko'p ma'noli so'zlar – bu bir xil shaklda bo'lib, turli kontekstlarda turli ma'nolarga

¹ Xusainova Zilola Yuldashevna – filologiya fanlari falsafa doktori (PhD), dotsent v.b. Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti.

E-pochta: xusainovazilola@navoiy-uni.uz

ORCID: 0000-0003-4357-7515

² Norbekova Bahora – ToshDO'TAU, Kompyuter lingvistikasi mutaxassisligi I bosqich magistranti.

ega bo'lgan so'zlar. Ushbu masalani hal qilish uchun ishlab chiqilgan Word Sense Disambiguation (WSD) – so'zning kontekst asosida to'g'ri ma'nosini aniqlash jarayoni bo'lib, u kompyuter tilshunosligining asosiy vazifalaridan biri sanaladi. WSDning nazariy konsepsiyasi inson miyasining tilni idrok etish mexanizmlariga asoslanadi, chunki insonlar kontekst orqali so'z ma'nolarini avtomatik ravishda tanlaydilar.

Tabiiy tillardagi so'zlarning ko'p ma'nolilik xususiyati uning semantik tizimini boyitadi, ammo kompyuter tizimlari uchun bu murakkab masalani yuzaga keltiradi. So'zning qaysi ma'noda qo'llanilganini matn asosida anglash inson uchun oddiy hol bo'lsa-da, lekin kompyuter tarmoqlari uchun bu qiyinchilik tug'diradi. Bir so'zning bir nechta ma'noga egaligi uni kontekstdan kelib chiqib to'g'ri talqin qilishni talab etadi. Masalan, *“quloq”* so'zini bir nechta ma'noga egaligi quyida keltiriladi:

- odam hamda umurtqali hayvonlarning eshitish hamda muvozanat a'zosi va uning tashqi qismi;
- ba'zi narsalarning ushlash, ilib qo'yish yoki mahkamlash uchun xizmat qiladigan qismi;
- bosh kiyimlarning quloqlarni va iyakni yopib, issiq tutib turish uchun tushirib qo'yiladigan uzun qismi;
- ba'zi mexanizm va musiqa asboblarning burab ishga tushirish yoki sozlash uchun xizmat qiladigan qismi;
- avtomobil, traktor va shu kabilar mashinalarni boshqarish uchun xizmat qiladigan qurilmasi, rul;
- o'simlikning barg yoki nihol yozadigan yeri; dastlabki niholcha, bargcha kabi oltita ma'noga egadir.

Ushbu so'zning berilgan ma'nolardan aynan qaysi birini ifodalayotganini konteks asosida aniqlash NLPning eng dolzarb masalalaridan biri bo'lib, uzoq yillardan buyon tadqiqotchilar diqqat markazida bo'lib kelmoqda. Tabiiy tilni qayta ishlashda ushbu muammo **Word Sense Disambiguation (WSD)** deb ataladi. So'z ma'nosini aniqlashtirish bu jumla yoki boshqa kontekst bo'lagida so'zning qaysi ma'nosi nazarda tutilganini aniqlash jarayonidir.

Asosiy qism

Word Sense Disambiguation (WSD) – noaniq, ya'ni ko'p ma'noli so'zning muayyan nutqiy vaziyatda aynan qaysi ma'noda qo'llanilganini avtomatik tarzda aniqlashga qaratilgan jarayon hisoblanadi. Inson tilni idrok etish jarayonida so'z ma'nosini kontekstga tayanib tez va tabiiy ravishda anglay oladi. Biroq

kompyuter tizimlari uchun bunday vazifa murakkab hisoblanib, katta hajmdagi lingvistik bilimlar, statistik ma'lumotlar va hisoblash resurslarini talab etadi [Jurafsky, Martin, 2020, 488; Manning, Schütze, 1999, 235].

WSD muammosi 1940-yillardan boshlab ilmiy tadqiqot obyekt sifatida shakllana boshlagan. Mazkur muammo dastlab mashina tarjimai bilan bog'liq izlanishlar doirasida o'rganila boshlangan. Xususan, amerikalik tadqiqotchi Varren Weaver 1949-yilda e'lon qilgan mashina tarjimai haqidagi mashhur memorandumida so'z ma'nosini kontekst asosida aniqlash zaruratini alohida ta'kidlab o'tgan [Weaver, 1949, 4]. Ushbu g'oya keyinchalik WSD masalasining mustaqil ilmiy muammo sifatida rivojlanishiga turtki bo'ldi.

Biroq WSD masalasining murakkabligi 1960-yillarda Bar-Hillel tomonidan baholab, bunday vazifani *“elektron kompyuter”* yordamida to'liq hal etish imkonsiz deb ta'kidlagan. Ushbu qarash uzoq vaqt davomida WSD bo'yicha *pessimist yondashuv*ning shakllanishiga sabab bo'ldi.

1970-yillarga kelib WSD sun'iy intellekt sohasida ishlab chiqilgan semantik talqin tizimlarining tarkibiy qismi sifatida qarala boshlandi. Bu davrda Wilks tomonidan taklif etilgan *“preference semantics”* yondashuvi muhim nazariy asoslardan biri bo'ldi. Mazkur yondashuvda so'z ma'nosi **semantik cheklovlar** va **kontekstual mosliklar** orqali aniqlangan. Shunga qaramay, o'sha davrdagi WSD tizimlari asosan qoida asosida va qo'lda kodlangan bo'lgani sababli, ularning moslashuvchanligi hamda qamrovi cheklangan edi.

1980-yillarda WSD tadqiqotlarida muhim burilish – **OALD (Oxford Advanced Learner's Dictionary of Current English)** kabi yirik elektron leksik resurslarning paydo bo'lishi qo'lda kodlash jarayonlarini qisman avtomatlashtirish imkonini berdi [OALD, 1980, XV-XVIII]. Ushbu davrda WSD jarayonida lug'atga asoslangan yondashuvdan keng foydalanila boshlandi, biroq aniqlashtirish mexanizmlari asosan qoidalarga va leksik manbalarga tayangan holda amalga oshirilgan.

1990-yillarda hisoblash lingvistikasi va mashinali o'qitish sohasidagi yutuqlar WSD muammosini yangi bosqichga olib chiqdi [Ng, Lee, 1996, 44]. Aynan shu davrda WSD nazorat ostida mashinali o'qitish metodlarini sinovdan o'tkazish uchun asosiy vazifalardan biriga aylandi. *Annotatsiyalangan korpuslar* asosida so'z ma'nolarini avtomatik o'rganish imkoniyati paydo bo'ldi va bu WSD aniqligini sezilarli darajada oshirdi.

2000-yillarga kelib nazoratli WSD tizimlari aniqlik ko'rsatkichlari bo'yicha ma'lum bir barqarorlikka erishdi.

Shuningdek, tadqiqotlar yo'nalishi yanada kengaydi: qo'polroq ma'no darajalari, domenlarga moslashuv, yarim-nazoratli va nazoratsiz korpusga asoslangan usullar, turli metodlarni kombinatsiyalash hamda grafik asosidagi yondashuvlar faol rivojlana boshladi. Shunga qaramay, nazorat ostidagi usullar eng yuqori natijalarga erishishda yetakchi bo'ldi.

2010-yillardan boshlab NLP sohasida *representation learning*, ya'ni so'z va kontekstlarni vektor ko'rinishida ifodalash hamda chuqur o'rganish (deep learning) texnologiyalarining jadal rivojlanishi WSD tadqiqotlari qayta jonlana boshladi. Neyron tarmoqlarga asoslangan modellar kontekstni chuqurroq va ko'p qatlamli tarzda modellashtirish imkonini berdi.

So'nggi yillarda olib borilgan tadqiqotlar WSD masalasining dolzarbligini yanada kuchaytirdi. Jumladan, Moro, Agirre va ularning hamkorlari (2014) qoidalarga asoslangan metodlarning samaradorligini ko'rsatib bergan bo'lsa, Panchenko va boshqalar (2017) nazoratsiz yondashuvlar imkoniyatlarini yoritib berdilar. Hadivinoto, Kumar, Huang va boshqa tadqiqotchilar (2019) nazorat ostidagi usullar yordamida yuqori aniqlikka erishish masalalarini yoritishgan.

Skarlini va boshqa tadqiqotchilar (2020) tomonidan klassik nazorat ostidagi usullar bilan samarali natijalarga erishish mumkin bo'lgan metodlarni ishlab chiqdilar. Keyingi tadqiqotlarda Sanaa Kaddoura va Reem Nassar (2024) arab tili uchun *EnhancedBERT* modelining WSD vazifalarida yuqori samaradorlik ko'rsatishini qayd etdilar. Deshan Sumanathilaka (2025) esa GPT-4 Turbo modeli WSD vazifalarini bajarishda ochiq kodli va tijorat modellari orasida yetakchi o'rinni egallashini ilmiy dalillar asosida ko'rsatib berdilar.

O'zbek tili uchun WSD tadqiqotlarining istiqbollari

O'zbek tili uchun Word Sense Disambiguation (WSD) muammosini tadqiq etish hozirgi kunda dolzarb va istiqbolli ilmiy yo'nalishlardan biri hisoblanadi. O'zbek tilida ko'p ma'nolilik hodisasi keng tarqalgan bo'lib, bu holat avtomatik matn tahlili jarayonlarida sezilarli semantik noaniqliklarni keltirib chiqaradi. Shu sababli WSD konsepsiyasini o'zbek tili materiallari asosida ishlab chiqish tabiiy tilni qayta ishlash tizimlarining sifatini oshirishga xizmat qiladi:

– o'zbek tiliga mos leksik va semantik resurslarni yaratish muhim ahamiyat kasb etadi. *Izohli lug'atlar, chastotali lug'atlar, semantik tarmoqlar va annotatsiyalangan korpuslar* WSD tizimlari uchun asosiy bilim manbai bo'lib xizmat qiladi. Jumladan, o'zbek

tilining milliy korpusini ma'no teglari bilan boyitish nazorat ostida va yarim-nazoratli WSD usullarini joriy etish imkonini beradi;

– zamonaviy chuqur o'rganish modellarini o'zbek tili uchun moslashtirish istiqbolli yo'nalishlardan biri hisoblanadi. Multilingual BERT, XLM-R, shuningdek, katta til modellari (LLM) asosida o'zbek tiliga mos WSD modellarini ishlab chiqish kontekstni chuqurroq tahlil qilish imkonini beradi. Bu esa mashina tarjimai, avtomatik referatlash, savol-javob tizimlari va sentiment tahlil kabi amaliy vazifalarda yuqori natijalarga erishishga xizmat qiladi.

Shuningdek, o'zbek tilida WSD tadqiqotlari leksik semantika va pragmatika o'rtasidagi bog'liqlikni chuqurroq o'rganish imkonini yaratadi. Kontekstning lingvistik va ekstralingvistik omillarini hisobga olgan holda ma'no aniqlash modellari ishlab chiqilishi o'zbek tilshunosligi uchun yangi ilmiy yondashuvlarni ochadi.

Metodlar

So'z ma'nosini kontekst asosida aniqlash (WSD) masalasi tabiiy tilni qayta ishlash sohasida dolzarb muammolardan biri bo'lib, uni hal etishda turli nazariy yondashuvlar va hisoblash metodlari qo'llaniladi. So'zning aynan qaysi ma'noda qo'llanilganini aniqlash uchun *kontekstual, leksik va statistik axborotlardan* foydalanish WSD muammosini samarali yechishga xizmat qiladi. WSD tadqiqotlarida keng qo'llaniladigan asosiy yondashuvlarni tahlil qilamiz:

Qoidalarga asoslangan (lug'aviy-semantik) yondashuv.

Qoidalarga asoslangan yondashuvlar WSD sohasidagi dastlabki va klassik metodlar hisoblanib, bunda bir matnda birga uchraydigan so'zlar o'zaro semantik jihatdan bog'liq bo'ladi hamda bu bog'liqlik ularning **lug'aviy-semantik tavsiflari (glosslar)**da aks etadi [Lesk, 1986, 22]. Lesk algoritmi mazkur yondashuvning mashhur algoritmi hisoblanib [Banerjee, Pedersen, 2002, 7], u kontekstdagi so'zlar glosslari orasidagi eng katta leksik moslikni aniqlash orqali so'z ma'nosini belgilaydi. Lesk algoritmi klassik lug'atga asoslangan usuldir.

Lug'aviy-semantik tavsiflari (glosslar)ga asoslangan yondashuvlardan tashqari, WordNet kabi leksik bilim bazalaridan foydalanib, so'z ma'nolari o'rtasidagi semantik o'xshashlikni hisoblashga asoslangan usullar ham mavjud. Shuningdek, semantik tarmoqlar asosida ishlovchi grafik yondashuvlar, xususan, *spreading activation* modellarining WSD vazifasida samaradorligi ko'rsatilgan [Agirre, Soroa, 2009, 37; Moro, Navigli, 2014, 4].

So'nggi yillarda Wikipedia kabi ochiq bilim manbalaridagi semantik munosabatlarni WordNetga avtomatik integratsiya

qilish lug'atga asoslangan tizimlarning aniqligini sezilarli darajada oshirgan. Masalan, "quloq" so'zi ko'p ma'noli bo'lib, "Bolaning qulog'i og'riganda darhol shifokorga borish kerak" gapida quloq so'zi buyumlardagi tutqich ma'nosida emas, balki inson tana a'zosi ma'nosida qo'llanilganini kontekst asosida aniqlash mumkin.

Yarim-nazoratli yondashuv (Semi-supervised methods). Tegl原因an ma'lumotlarning cheklanganligi sababli WSD tadqiqotlarida yarim-nazoratli o'rganish yondashuvlari keng qo'llaniladi. Ushbu metodlar teglangan va teglanmagan korpuslardan birgalikda foydalanish imkonini beradi. Yarowsky algoritmi ushbu yondashuvning dastlabki va mashhur namunalaridan biri bo'lib, u "har bir kollokatsiyada bitta ma'no" hamda "har bir nutq oqimida bitta ma'no" kabi tabiiy til qonuniyatlariga asoslanadi[Yarowsky, 1995, 192].

Bootstrapping yondashuvida har bir so'z uchun kichik hajmdagi boshlang'ich ma'lumotlar (qo'lda teglangan misollar yoki qoidalar to'plami) tanlanadi[Diab, 2002, 3]. Masalan, "oyoq" so'zi "og'rimoq" fe'li bilan birga kelganda ko'pincha inson yoki hayvon tana a'zosi ma'nosida ishlatiladi. Ushbu boshlang'ich ma'lumotlar asosida dastlabki klassifikator o'qitilib, keyinchalik u teglanmagan korpusdan ishonchli natijalarni tanlab, yangi o'quv ma'lumotlarini hosil qiladi[Panchenko, Morozova, 2017, 14]. Jarayon butun korpus tugaguncha yoki maksimal iteratsiya chegarasiga yetguncha takrorlanadi.

Shuningdek, katta hajmdagi teglanmagan matnlardan hamkorlik axborotlarini olish orqali teglangan korpusni boyitishga qaratilgan yondashuvlar ham mavjud. Ikki tilli moslangan korpuslarga asoslangan yondashuvlar ham yarim-nazoratli tizimlarga kiradi, chunki bir tildagi noaniq so'z ikkinchi tilda turli so'zlar orqali tarjima qilinib, ma'no farqlarini aniqlash imkonini beradi.

Nazorat ostidagi yondashuv (Supervised methods). Nazorat ostidagi yondashuvda kontekstual axborotning o'zi so'z ma'nosini aniqlashga asoslanadi. Ushbu metodlarda so'z ma'nolari oldindan belgilangan va teglangan korpuslar asosida modellar o'qitiladi. WSD vazifasida *Support Vector Machines (SVM)* algoritmlari yuqori samaradorlik ko'rsatgan[Tjong Kim Sang, de Meulder, 2003, 463].

Nazorat ostidagi yondashuvlarning asosiy afzalligi yuqori aniqlik bo'lsa-da, katta hajmdagi qo'lda teglangan korpuslarni yaratish katta mehnat va vaqt talab qiladi, bu esa ushbu yondashuvlarning amaliy qo'llanilishini cheklaydi.

Nazoratsiz yondashuv (Unsupervised methods).

Nazoratsiz WSD yondashuvlari tadqiqotchilar uchun eng murakkab, ammo istiqbolli yo'nalishlardan biri hisoblanadi. Ushbu yondashuvlarning asosiy g'oyasi shundan iboratki, o'xshash ma'nolar o'xshash kontekstlarda namoyon bo'ladi. Shu sababli, so'z ma'nolari kontekstual o'xshashlikka asoslangan klasterlash orqali avtomatik tarzda hosil qilinadi. Bu jarayon *word sense induction* yoki *word sense discrimination* deb ataladi [Schütze, 1998, 104].

Nazoratsiz yondashuv natijalari ko'pincha nazorat ostidagi modellarga nisbatan pastroq bo'lsa-da, ularni baholash murakkab, chunki hosil qilingan ma'nolarni lug'atdagi ma'nolar bilan moslashtirish talab etiladi. Agar bunday moslashtirish shart bo'lmasa, klasterlash sifatini *entropiya* va *tozalik* kabi metrikalar yordamida baholash mumkin.

Zamonaviy NLP tizimlarida so'z vektorlari (word embeddings) orqali ifodalash WSD samaradorligini oshirishga xizmat qiladi. An'anaviy embeddinglar so'zning barcha ma'nolarini bitta vektorda jamlasa-da, ma'no klasterlari uchun markaziy vektorlarni hisoblash orqali WSD natijalarini yaxshilash mumkin.

WordNet, **ConceptNet** va **BabelNet** kabi leksik bilim bazalarini embeddinglar bilan integratsiya qiluvchi tizimlar – **AutoExtend** va **Most Suitable Sense Annotation (MSSA)** bunga misol bo'la oladi [Rothe, Schütze, 2015, 182]. **AutoExtend** modelida so'z vektorlari grafik tuzilma asosida so'z va uning ma'nolari bo'yicha ajratiladi, **MSSA** [Pevlevina et al., 2016, 4] esa WordNet glosslaridagi so'zlar embeddinglarining o'rtacha qiymati asosida kontekstga eng mos ma'noni tanlaydi.

WSD muammosini yechishda bir necha metodlardan foydalangan holda so'zning kontekstdagi ma'nosi samarali usulda aniqlash mumkin. Quyidagi jadvalda ayrim WSD metodlarining umumiy tavsifi keltiriladi:

1-jadval. WSD metodlari

Yondashuv turi	Tavsif	Foydalaniladigan resurslar	Afzalliklari	Cheklovlari
Qoidalariga asoslangan yondashuvlar	Kontekstdagi so'zlar o'zaro semantik bog'liq bo'lib, bu bog'liqlik glosslar (lug'aviy-semantik tavsiflar) orqali aniqlanadi	Izohli lug'atlar, WordNet glosslari	Teglangan korpus talab qilmaydi, tushunarli va interpretatsiyalanuvchi	Keng qamrovli qoidalar yaratish qiyin, domenlarga sezgir

Lesk algoritmi	Kontekstdagi so'zlar glosslari orasidagi leksik moslikni maksimal darajada aniqlash	Izohli lug'atlar, WordNet	Oddiy va klassik metod, nazariy jihatdan asosli	Qisqa glosslar sabab aniqlik past bo'lishi mumkin
Leksik qoida bazalariga asoslangan usullar	So'z ma'nolari o'rtasidagi semantik masofani yoki o'xshashlikni hisoblash	WordNet, BabelNet	Til qoidalariga tayangan, resurslar boy	Resurslarga qaramlik yuqori
Grafik yondashuvlar	Semantik tarmoqda semantic guruhlar orqali ma'no aniqlash	Semantik tarmoqlar, WordNet	Murakkab kontekstlarda samarali	Hisoblash murakkabligi yuqori
Nazoratli mashinali o'qitish	Teglangan (annotatsiyalangan) ma'lumotlar asosida modelni o'rgatish	Annotatsiyalangan korpuslar	Yuqori aniqlik	Katta hajmdagi teglangan ma'lumot talab etadi
Nazoratsiz va yarim nazoratli usullar	Statistik va klasterlash usullari orqali yashirin ma'nolarni aniqlash	Korpuslar	Resursga kam talab	Natijalar barqaror emas
Chuqur o'rganish va LLM asosidagi yondashuvlar	Kontekstual vektorlar va transformer modellar orqali ma'no aniqlash	BERT, GPT, LLMlar	Eng yuqori natijalar, kontekstni chuqur tushunadi	Hisoblash resurslari talabi yuqori

Xulosa

Ushbu maqolada ko'p ma'noli so'zlarning kontekst asosida aniqlanishi (Word Sense Disambiguation, WSD) masalasining nazariy konsepsiyasi va metodologik yondashuvlari tahlil qilindi. WSD tabiiy tilni qayta ishlash sohasidagi uzoq yillik muammo bo'lib, uning yechimiga qator yondashuvlar ishlab chiqilgan. Jumladan, qoidalarga asoslangan, nazorat ostidagi, yarim-nazoratli, nazoratsiz va chuqur o'rganish asosidagi yondashuvlar. Har bir yondashuv o'ziga xos afzallik va cheklovlarga ega bo'lib, *kontekstual xususiyatlar, lug'aviy-resurslar* va *statistik metodlarga* tayanadi.

Qoidalarga asoslangan yondashuvlar, xususan Lesk algoritmi, so'zlarning glosslaridagi semantik bog'liqlikni aniqlash orqali WSDni amalga oshiradi. Yarim-nazoratli va nazoratsiz usullar kichik hajmdagi teglangan ma'lumotlar yoki korpuslardagi yashirin ma'nolar asosida so'z ma'nosini aniqlash imkonini beradi. Nazorat ostidagi va chuqur o'rganish metodlari esa katta hajmdagi ma'lumotlardan foydalangan holda yuqori aniqlikni ta'minlaydi. So'zlarni kontekstual vektorlar yordamida ifodalash esa zamonaviy NLP tizimlarida WSD samaradorligini sezilarli darajada oshiradi.

Shuningdek, maqolada WSD tadqiqotlaridagi tarixiy rivojlanish, 1940-yillardan boshlab bugungi kunga qadar erishilgan yutuqlar va ilg'or tadqiqot natijalari yoritildi. Xususan, **WordNet**, **BabelNet** va **ConceptNet** kabi leksik resurslarning WSD tizimlarida qo'llanilishi, shuningdek, nazoratsiz va yarim-nazoratli yondashuvlarning kontekstual tushunishga qo'shayotgan hissasi tahlil qilindi.

Umuman olganda, WSD muammosi hali ham NLP sohasida dolzarb bo'lib qolmoqda. Turli metodlar va ularning kombinatsiyalari so'z ma'nosini kontekstga mos aniqlashda samarali vosita bo'lib xizmat qiladi. Kelajakda, ayniqsa o'zbek tili uchun WSD tizimlarini rivojlantirish, chuqur o'rganish va representation learning metodlarini qo'llash orqali lingvistik resurslarni kengaytirish ilmiy hamda amaliy jihatdan muhim istiqbol hisoblanadi.

Foydalanilgan adabiyotlar

- Jurafsky D., Martin J.H. Speech and Language Processing. Pearson, 2020. – P. 487-495.
- Manning C., Schütze H. Foundations of Statistical Natural Language Processing. MIT Press, 1999. – P. 229-242.
- Weaver W. Translation. Memo, 1949. – P. 3-5.
- Wilks Y. Preference Semantics. Cambridge University Press, 1975. – P. 41-45.
- Oxford Advanced Learner's Dictionary (OALD). Oxford University Press, 1980.
- Ng H.T., Lee H. Integrating Machine Learning Methods for Word Sense Disambiguation. ACL, 1996. – P. 40-45.
- Lesk M. Automatic Sense Disambiguation Using Machine Readable Dictionaries. ACL, 1986. – P. 20-24.
- Banerjee S., Pedersen T. Extended Lesk Algorithm for Word Sense Disambiguation. CICLing, 2002. – P. 5-9.
- Agirre E., Soroa A. Personalizing PageRank for Word Sense Disambiguation. EACL, 2009. – P. 35-40.
- Moro A., Navigli R. Integrating Knowledge-Based and Corpus-Based WSD Methods. ACL, 2014. – P. 3-7.
- Yarowsky D. Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. ACL, 1995. – P. 189-193.
- Diab M. Semi-Supervised WSD Using Parallel Corpora. ACL, 2002. – P. 2-6.
- Panchenko A., Morozova O. Bilingual Corpora for Semi-Supervised WSD. LREC, 2017. – P. 12-17.

- SVM for WSD. Tjong Kim Sang E.F., de Meulder F., 2003. – P. 459–462.
- Schütze H. Automatic Word Sense Discrimination. Computational Linguistics, 1998. – P. 102–106.
- Rothe S., Schütze H. AutoExtend: Extending Word Embeddings to Embeddings for Synsets and Lexemes. ACL, 2015. – P. 179–183.
- Pelevina M., et al. MSSA: Most Suitable Sense Annotation for WSD. LREC, 2016. – P. 2–6.

THE THEORETICAL CONCEPT OF CONTEXT-BASED WORD SENSE DISAMBIGUATION (WSD)

**Zilola Xusainova³,
Norbekova Bahora⁴**

Abstract. This article highlights the theoretical concept of Word Sense Disambiguation (WSD), a significant problem in the field of Natural Language Processing (NLP), which involves determining the meaning of polysemous words based on context. Furthermore, the article provides a scholarly and theoretical analysis of the linguistic nature of polysemy, the interpretation of the concept of context, and the formation and development stages of the WSD process. WSD plays a crucial role in artificial intelligence systems for resolving semantic ambiguities, text analysis, machine translation, information retrieval, question-answering systems, and both syntactic and semantic annotation processes. The article also discusses lexicon-based models, corpus-based methods, and modern neural approaches.

Keywords: *Polysemy, word sense, disambiguation, WSD, NLP.*

References

- Jurafsky D., Martin J.H. Speech and Language Processing. Pearson, 2020. – P. 487-495.
- Manning C., Schütze H. Foundations of Statistical Natural Language Processing. MIT Press, 1999. – P. 229-242.
- Weaver W. Translation. Memo, 1949. – P. 3-5.
- Wilks Y. Preference Semantics. Cambridge University Press, 1975. – P. 41-45.
- Oxford Advanced Learner's Dictionary (OALD). Oxford University Press, 1980.
- Ng H.T., Lee H. Integrating Machine Learning Methods for Word Sense Disambiguation. ACL, 1996. – P. 40-45.
- Lesk M. Automatic Sense Disambiguation Using Machine Readable Dictionaries. ACL, 1986. – P. 20-24.
- Banerjee S., Pedersen T. Extended Lesk Algorithm for Word Sense Disambiguation. CICLing, 2002. – P. 5-9.

³ *Khusainova Zilola Yuldashevna* – Doctor of Philosophy (PhD) in Philology, Associate Professor; Alisher Navoi Tashkent State University of the Uzbek Language and Literature.

E-pochta: xusainovazilola@navoiy-uni.uz

ORCID: 0000-0003-4357-7515

⁴ *Norbekova Bahora* – Tashkent State University of Uzbek Language and Literature (TSUULL), Master's student (Year 1) in Computer Linguistics.

- Agirre E., Soroa A. Personalizing PageRank for Word Sense Disambiguation. EACL, 2009. – P. 35–40.
- Moro A., Navigli R. Integrating Knowledge-Based and Corpus-Based WSD Methods. ACL, 2014. – P. 3–7.
- Yarowsky D. Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. ACL, 1995. – P. 189–193.
- Diab M. Semi-Supervised WSD Using Parallel Corpora. ACL, 2002. – P. 2–6.
- Panchenko A., Morozova O. Bilingual Corpora for Semi-Supervised WSD. LREC, 2017. – P. 12–17.
- SVM for WSD. Tjong Kim Sang E.F., de Meulder F., 2003. – P. 459–462.
- Schütze H. Automatic Word Sense Discrimination. Computational Linguistics, 1998. – P. 102–106.
- Rothe S., Schütze H. AutoExtend: Extending Word Embeddings to Embeddings for Synsets and Lexemes. ACL, 2015. – P. 179–183.
- Pelevina M., et al. MSSA: Most Suitable Sense Annotation for WSD. LREC, 2016. – P. 2–6.

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 01.11.2023-yilda ruxsat etildi.
Bichimi 70x100 1/16, Ofset bosma. "Cambria" garniturası.
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.