

СЎЗ САНЪАТИ ХАЛҚАРО ЖУРНАЛИ

8 ЖИЛД, 1 СОН

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА

ТОМ 8, НОМЕР 1

INTERNATIONAL JOURNAL OF WORD ART

VOLUME 8, ISSUE 1



СЎЗ САНЪАТИ ХАЛҚАРО ЖУРНАЛИ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА | INTERNATIONAL JOURNAL OF WORD ART

№1 (2025) DOI <http://dx.doi.org/10.26739/2181-9297-2025-1>

Бош муҳаррир:

Тўхтасинов Илҳом

п.ф.д., профессор (Ўзбекистон)

Бош муҳаррир ўринбосари:

Главный редактор:

Тўхтасинов Илҳом

д.п.н., профессор (Ўзбекистон)

Заместитель главного редактора:

Editor in Chief:

Tuhtasinov Ilhom

DSc. Professor (Uzbekistan)

Deputy Chief Editor

ТАҲРИРИЙ МАСЛАҲАТ КЕНГАШИ

Назаров Бахтиёр

академик. (Ўзбекистон)

Якуб Умарўгли

ф.ф.д., профессор (Туркия)

Алмаз Улви Биннатова

ф.ф.д., профессор (Озарбайжон)

Бокиева Гуландом

ф.ф.д., профессор (Ўзбекистон)

Миннуллин Ким

ф.ф.д., профессор (Татаристон)

Махмудов Низомиддин

ф.ф.д., профессор (Ўзбекистон)

Керимов Исмаил

ф.ф.д., профессор (Россия)

Жўраев Маматкул

ф.ф.д., профессор (Ўзбекистон)

Куренов Рахиммамад

к.ф.н. (Туркменистон)

Кристофер Жеймс Форт

Мичиган университети (АҚШ)

Умархўжаев Мухтор

ф.ф.д., профессор (Ўзбекистон)

Мирзаев Ибодулло

ф.ф.д., профессор (Ўзбекистон)

Болтабоев Ҳамидулла

ф.ф.д., профессор (Ўзбекистон)

Дўстмухаммедов Хуршид

ф.ф.д., профессор (Ўзбекистон)

Лиходзиевский А.С.

ф.ф.д., профессор (Ўзбекистон)

Сиддикова Ирода

ф.ф.д., профессор (Ўзбекистон)

Шиукашвили Тамар

ф.ф.д. (Грузия)

Туробов Бекпулат

масъул котиб, PhD, доцент
(Ўзбекистон)

РЕДАКЦИОННЫЙ СОВЕТ

Назаров Бахтиёр

академик. (Ўзбекистон)

Якуб Умар оглы

д.ф.н., профессор (Туркия)

Алмаз Улви Биннатова

д.ф.н., профессор (Азербайджан)

Бакиева Гуландом

д.ф.н., профессор (Ўзбекистон)

Миннуллин Ким

д.ф.н., профессор (Татарстан)

Махмудов Низомиддин

д.ф.н., профессор (Ўзбекистон)

Керимов Исмаил

д.ф.н., профессор (Россия)

Джўраев Маматкул

д.ф.н., профессор (Ўзбекистон)

Куренов Рахиммамад

к.ф.н. (Туркменистон)

Кристофер Джеймс Форт

Университет Мичигана (США)

Умархаджаев Мухтар

д.ф.н., профессор (Ўзбекистон)

Мирзаев Ибодулло

д.ф.н., профессор (Ўзбекистон)

Балтабаев Ҳамидулла

д.ф.н., профессор (Ўзбекистон)

Дустмухаммедов Хуршид

д.ф.н., профессор (Ўзбекистон)

Лиходзиевский А.С.

д.ф.н., профессор (Ўзбекистон)

Сиддикова Ирода

д.ф.н., профессор (Ўзбекистон)

Шиукашвили Тамар

д.ф.н. (Грузия)

Туробов Бекпулат

отв. секретарь, PhD, доцент
(Ўзбекистон)

EDITORIAL BOARD

Bakhtiyor Nazarov

academician. (Uzbekistan)

Yakub Umarogli

Doc. of philol. scien., prof. (Turkey)

Almaz Ulvi Binnatova

Doc. of philol. scien., prof. (Azerbaijan)

Bakieva Gulandom

Doc. of philol. scien., prof. (Uzbekistan)

Minnulin Kim

Doc. of philol. scien., prof. (Tatarstan)

Mahmudov Nizomiddin

Doc. of philol. scien., prof. (Uzbekistan)

Kerimov Ismail

Doc. of philol. scien., prof. (Russia)

Juraev Mamatkul

Doc. of philol. scien., prof. (Uzbekistan)

Kurenov Rakhimmamed

Ph.D. Ass. Prof. (Turkmenistan)

Christopher James Fort

University of Michigan (USA)

Umarkhodjaev Mukhtar

Doc. of philol. scien., prof. (Uzbekistan)

Mirzaev Ibodulla

Doc. of philol. scien., prof. (Uzbekistan)

Boltaboev Hamidulla

Doc. of philol. scien., prof. (Uzbekistan)

Dustmuhammedov Khurshid

Doc. of philol. scien., prof. (Uzbekistan)

Lixodzievsky A.S.

Doc. of philol. scien., prof. (Uzbekistan)

Siddiqova Iroda

Doc. of philol. scien., prof. (Uzbekistan)

Shiukashvili Tamar

Doc. of philol. scien. (Georgia)

Turobov Bekpulat

PhD Ass. prof. Senior Secretary
(Uzbekistan)

PageMaker | Верстка | Саҳифаловчи: Хуршид Мирзахмедов

Контакт редакций журналов. www.tadqiqot.uz

ООО Tadqiqot город Ташкент,

улица Амира Темура пр.1, дом-2.

Web: <http://www.tadqiqot.uz/>; E-mail: info@tadqiqot.uz

Тел: (+998-94) 404-0000

Editorial staff of the journals of www.tadqiqot.uz

Tadqiqot LLC The city of Tashkent,

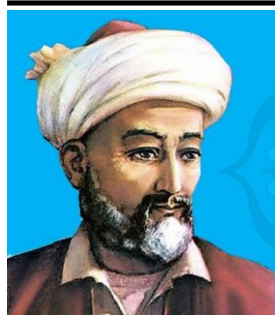
Amir Temur Street pr.1, House 2.

Web: <http://www.tadqiqot.uz/>; E-mail: info@tadqiqot.uz

Phone: (+998-94) 404-0000

1. Raxmatullaeva Guljaxon Normo‘min qizi INSON ICHKI OLAMI NEGATIV KONSEPTLARINING LINGVO-KOGNITIV TAHLILI (g`am-anduh (qayg`u), mahzunlik, iztirob konseptlari misolida).....	5
2. Xamrayeva Zebiniso Xaydarovna A COMPARATIVE STUDY OF THE COMPLEX SYNTACTIC CONSTRUCTIONS BASED ON ENGLISH LANGUAGE MATERIALS.....	11
3. Esanova Zebo Ikrom qizi ZAMONAVIY TILSHUNOSLIKDA DISKURS TUSHUNCHASI VA UNING O‘RGANILISHI.....	17
4. Abdullaeva Sanobar Bakhrillaevna BIBLICAL PHRASEOLOGY DESCRIBING THE SOCIAL POSITION OF A PERSON.....	23
5. Kozieva Iqbol Kamiljonovna METHODS OF ANTHROPOLOGICAL RESEARCH.....	27
6. Эпрашева Т.А. ХАЛҚАРО СОЛИҚ ХУЖЖАТЛАРИДА СОЛИҚ ТЕРМИНОЛОГИЯСИНИНГ ҚИЁСИЙ ВА ТИПОЛОГИК ТАҲЛИЛИНИ НАЗАРИЙ АСОСЛАРИ.....	35
7. Kushatova Nargiza Rustamovna TEODOR DRAYZERNING BAXTIQARO KERRI ROMANIDA XRONOTOP: ZAMON VA MAKONNING POETIK TASVIRI”.....	40
8. Ergashova Baxora Sayfillo qizi, Janna Tsulimova ИСПОЛЬЗОВАНИЕ НАСТОЛЬНЫХ ИГР В ОБУЧЕНИИ ИТАЛЬЯНСКОМУ ЯЗЫКУ КАК ИНОСТРАННОМУ: ОПЫТ И ПЕРСПЕКТИВЫ.....	47
9. Obloqulova O‘g‘iloy Abdurayim qizi MILLIY VA MADANIY QADRIYATLARNING REKLAMA MATNLARDA IFODALANISHI.....	51
10. Rakhimova Nodira Kamolovna INTERTEXTUALITY IN TRANSLATION: THEORETICAL INSIGHTS AND PRACTICAL APPROACHES.....	55
11. Xaydarova Nigora Tuxtasinovna GENDER MULOQOTINING PSIXOLOGIK ASOSLARI.....	66
12. Bekturodova Feruza Dilmurat qizi XITOIY TILIDAGI FIZIK ATAMALAR SHAKLLANISHINING LINGVISTIK TAMOIYILLARI.....	71
13. Boymurodova Dilafruz Dilshod qizi YAPON TILIDA YOZUV TIZIMI RIVOJLANISHI VA HIRAGANA, KATAKANA, KANJILARNING FONETIK KATEGORIYALARI.....	75
14. Gulzor Baxriddin qizi Temirova THE NATURE OF A JOURNALIST'S POSITION IN POLITICAL COMMENTARY AND ITS CONCEPTUAL ASPECTS.....	80
15. Aminova Lola Alimovna TURK TILIDA YAXSHI TILAK BILDIRISH VA ILTIFOT SO‘ZLARI.....	86
16. Najmiyeva Po‘lotoy Tilavmurodovna KELISHIK QO‘SHIMCHALARINING TARIXIY TADQIQI.....	94
17. Tilovova Go‘zal Rustamovna TERMINOLOGIYA VA LUG‘ATSHUNOSLIKNING O‘ZARO BOG‘LIQLIGI.....	99
18. Nilufar Abduraxmonova, Tirkashev Jasurjon Uktam o‘g‘li OTLARNING SEMANTIK TASNIFIDA XALQARO STANDART TEGLAH TIZIMLARIDAN FOYDALANISHNING NAZARIY ASOSLARI.....	103
19. Shoxida Ulugova Shoxruxovna POETIK NUTQDA INDIVIDUAL METAFORIK KONSEPTLARNING FAOLASHUVI.....	115
20. Lutfillovna Faxriniso Maxmudovna YAPON TILI IYEROGILFLARINING PAYDO BO‘LISH TARIXI.....	120

21. Eshanova Zilolaxon Qo‘chqarovna ISTIQLOL DAVRI O‘ZBEK SHE‘RIYATIDA XALQ ERTAKLARI BADI‘Y SINTEZI.....	125
22. Халмурадова Зебинисо Шавкатовна ИСПОЛЬЗОВАНИЕ АНТРОПОНИМОВ В ТЕКСТАХ АНГЛОЯЗЫЧНОЙ ХУДОЖЕСТВЕННОЙ ЛИТЕРАТУРЫ.....	132
23. Умитой Хушвақтова ЁЗУВЧИ ЗУЛФИЯ ҚУРОЛБОЙ ҚИЗИ АСАРЛАРИДА ПАРАЛИНГВИСТИК ВОСИТАЛАР.....	138
24. Alimova Sevara Farxadovna XITOIDA RANG MADANIYATINING KELIB CHIQISH.....	145
25. Fayzullaeva Zebuniso Abdulkarimovna DAS LEHREN DER MODI DER VERBEN DES DEUTSCHEN ANHAND DER TRADITIONELLEN UND NEUEN METHODEN.....	149
26. Razmuxamedova Fotima Jahongir qizi XITOI TILIDAGI QON-QARINDOSHLIK TERMINLARINING LINGVISTIK TAHLILI.....	154
27. Sayyora Tagaeva, SH.Mustanova MADANIYAT, BILIM VA TIL – SEMANTIK TIPOLOGIYA TALQINIDA.....	160
28. Odinoxon Qulmuradova JURNALISTIKADA DIALOGIK JANR TUSHUNCHASI, UNING O‘ZIGA XOSLIGI VA TASNIFI.....	164
29. O‘ng‘alova Zulayho Xolboy qizi ASQAD MUXTOR ASALARIDA ZAMON, MAKON VA MUHIT TASVIRI.....	171
30. Xolida Yo‘ldasheva MATN TIPOLOGIYASIGA ANTROPOTSENTRIK YONDASHUV.....	177
31. Ibragimova Shahnoza Xalilovna LINGUOPOTIC AND LINGUISTIC FEATURES OF ERNEST HEMINGWAY’S NOVEL “A FAREWELL TO ARMS”.....	182
32. Sadriddinzoda Safiya Shaxobiddinovna AMPHIBOLOGY IN ENGLISH RELIGIOUS LEXICOLOGY.....	187
33. Esirgapov Mukhiddin Nizomiddinovich THE INFLUENCE OF THE INTERNET AND DIGITAL TECHNOLOGIES ON THE ENGLISH LANGUAGE....	193



INTERNATIONAL JOURNAL OF WORD ART

ISSN: 2181-9297

www.tadqiqot.uz

Nilufar Abduraxmonova

filologiya fanlari doktori, professor, O'zbekiston

milliy universiteti Kompyuter lingvistikasi va

amaliy tilshunoslik kafedrasini mudiri

Tirkashev Jasurjon Uktam o'g'li

O'zbekiston Milliy universiteti mustaqil tadqiqotchisi

OTLARNING SEMANTIK TASNIFIDA XALQARO STANDART TEGTLASH TIZIMLARIDAN FOYDALANISHNING NAZARIY ASOSLARI

<https://doi.org/10.5281/zenodo.15521358>

ANNOTATSIYA

Ushbu maqolada o'zbek tilidagi ot so'z turkumiga oid leksemalarni semantik maydonlarga ajratish va teglash masalasi ko'rib chiqiladi. Tadqiqotda so'zlarni ma'no jihatdan guruhlash va ularni iyerarxik tuzilmalarga joylashtirishning nazariy va amaliy jihatlari o'rganiladi. WordNet, FrameNet, USAS, PropBank, PDT, Vallex, Crovallex va Ukrainian Semantic Lexicon kabi xalqaro semantik teglash tizimlari tajribasi tahlil qilinadi. O'zbek tilidagi 1000 dan ortiq ot namunasi ISO 24617-8 standartlari va WordNet tasniflariga asoslangan holda 14 ta asosiy semantik maydon va ularning ichki bo'linmalariga ajratilib, teglar bilan belgilanadi. Tadqiqot jarayonida semantik maydon va iyerarxiya tushunchalarining farqlari, shuningdek, ularning o'zaro uyg'unligi qiyosiy tarzda taqdim etiladi. Olingan natijalar teglash tizimining 92% aniqlik darajasiga erishganini ko'rsatdi, bu esa ushbu metodning o'zbek tili uchun samaradorligini tasdiqlaydi. Semantik teglashning mashina tarjimasini, matn tahlili va axborot qidiruvi kabi sohalaridagi amaliy ahamiyati muhokama qilinadi. Yakunda o'zbek tili uchun semantik korpus va lug'atlar yaratish bo'yicha takliflar berilib, kelajakdagi izlanishlar yo'nalishi belgilab chiqiladi. Ushbu ish o'zbek tilining leksik boyligini raqamli texnologiyalar doirasida chuqurroq o'rganish va xalqaro lingvistik tadqiqotlarda uning o'rnini mustahkamlashga xizmat qiladi.

Kalit so'zlar: semantik maydon, semantik iyerarxiya, teg, leksikologiya, ot so'z turkumi, WordNet, USAS, FrameNet, SRL, ISO 24617-8, semantik tasnif, korpus lingvistikasi, mashina tarjimasini, ontologik resurslar

Nilufar Abdurakhmonova

Doctor of Philological Sciences, Professor, National University of Uzbekistan

Head of the Department of Computer Linguistics and Applied Linguistics

Tirkashev Jasurjon Uktam o'g'li

Independent Researcher at the National University of Uzbekistan

ANNOTATION

This article examines the issue of classifying and tagging lexical units of the noun category in the Uzbek language into semantic fields. The study explores the theoretical and practical aspects

of grouping words by meaning and placing them into hierarchical structures. The experience of international semantic tagging systems such as WordNet, FrameNet, USAS, PropBank, PDT, Vallex, Crovallex, and the Ukrainian Semantic Lexicon is analyzed. Over 1,000 noun samples in the Uzbek language are categorized into 14 main semantic fields and their subcategories based on ISO 24617-8 standards and WordNet classifications, and assigned corresponding tags. The research highlights the differences between the concepts of semantic fields and hierarchies, as well as their interrelation, presented in a comparative manner. The results demonstrate that the tagging system achieved a 92% accuracy rate, confirming its effectiveness for the Uzbek language. The practical significance of semantic tagging in areas such as machine translation, text analysis, and information retrieval is discussed. In conclusion, recommendations are provided for creating semantic corpora and dictionaries for the Uzbek language, outlining directions for future research. This work contributes to a deeper understanding of the lexical richness of the Uzbek language in the context of digital technologies and strengthens its position in international linguistic studies.

Keywords: semantic field, semantic hierarchy, tag, lexicology, noun, WordNet, USAS, FrameNet, SRL, ISO 24617-8, semantic classification, corpus linguistics, machine translation, ontological resources

Нифулар Абдурахмонова

Доктор филологических наук, профессор, Национальный университет Узбекистана
Заведующая кафедрой компьютерной лингвистики и прикладного языкознания

Тиркашев Жасуржон Уктам угли

Независимый исследователь Национального университета Узбекистана

АННОТАЦИЯ

В данной статье рассматривается вопрос классификации и семантической разметки лексических единиц именного класса в узбекском языке по семантическим полям. В исследовании изучаются теоретические и практические аспекты группировки слов по значению и их размещения в иерархических структурах. Анализируется опыт международных систем семантической разметки, таких как WordNet, FrameNet, USAS, PropBank, PDT, Vallex, Crovallex и Ukrainian Semantic Lexicon. Более 1000 образцов существительных в узбекском языке были распределены по 14 основным семантическим полям и их подкатегориям на основе стандартов ISO 24617-8 и классификаций WordNet, а также помечены соответствующими тегами. В работе подчеркиваются различия между понятиями семантического поля и иерархии, а также их взаимосвязь, представленные в сравнительной форме. Результаты показывают, что система разметки достигла точности 92%, что подтверждает ее эффективность для узбекского языка. Обсуждается практическая значимость семантической разметки в таких областях, как машинный перевод, анализ текста и поиск информации. В заключение предлагаются рекомендации по созданию семантических корпусов и словарей для узбекского языка, а также определяются направления будущих исследований. Данная работа способствует более глубокому пониманию лексического богатства узбекского языка в контексте цифровых технологий и укреплению его позиций в международных лингвистических исследованиях.

Ключевые слова: семантическое поле, семантическая иерархия, тег, лексикология, имена существительные, WordNet, USAS, FrameNet, SRL, ISO 24617-8, семантическая классификация, корпусная лингвистика, машинный перевод, онтологические ресурсы

Kirish

Leksikologiya va kompyuter lingvistikasi sohasida soʻzlarni maʼnoviy jihatdan guruhlash va ularga teg qoʻyish muhim ahamiyat kasb etadi [1]. Xususan, ot soʻz turkumini semantik maydonlarga ajratish – yaʼni maʼno doiralariga koʻra guruhlash – ushbu soʻzlarning kompyuter tomonidan tushunilishi va qayta ishlanishini osonlashtiradi. Bunda har bir ot oʻzining tegishli semantik maydoniga, yaʼni maʼno jihatdan yaqin tushunchalar jamlanmasiga nisbatlanadi. Avvalo, semantik

maydon va semantik iyerarxiya tushunchalari bir-biridan farqlasak. Semantik maydon deb ma'lum bir tushuncha doirasiga oid bo'lgan, o'zaro mazmuniy aloqador so'zlar yig'indisiga aytiladi [2]. Masalan, ranglar semantik maydoni tarkibiga qora, oq, qizil, yashil, sariq, ko'k kabi rang nomlari kiradi [2].

Semantik iyerarxiya so'zlarni umumiylik-xususiylik munosabatlari – giperonimiya va giponimiya orqali tabaqalashtirib tizimga soladi. Bunda leksemalar “tur-jins” tamoyili asosida pog'onali tuzilma, daraxtsimon tasnif, hosil qiladi. Masalan, “jonivor” so'zi hayvonlar uchun giperonimik kategoriya hisoblanadi. Uning ostida sutemizuvchilar, undan pastroqda yirtqichlar, so'ngra mushuksimonlar va nihoyat yo'lbars kabi qator giponimik tushunchalar joylashadi [3].

WordNet kabi leksik ma'lumotlar bazasi millionlab so'z va atamalarni aynan shunday semantik tarmoqlarga ajratib, ularni ma'no iyerarxiyasi shaklida tartiblaydi [1]. WordNet korpus lingvistikada keng qo'llaniladigan semantik resurs bo'lib, unda so'zlar ma'no guruhlariga (sinsetlarga) va ularning iyerarxik munosabatlariga, masalan, hayvon – umumiy tushuncha, mushuk – uning xususiy holati kabilarga bo'lingan [1]. Shu bilan birga, FrameNet singari tizimlar so'zlarni ma'no doirasidan kelib chiqqan holda “frame” deb ataluvchi vaziyat modellari orqali bog'laydi. Unda fe'llar va ularga tegishli rol ya'ni otlar semantik rollar bilan belgilangan bo'lib, ushbu baza nafaqat ingliz tilida, balki portugal, xitoy, nemis, yapon, ispan va shved tillarida ham yaratilgan [1]. Semantik teglashning yana bir yondashuvi USAS (UCREL Semantic Analysis System) tizimi orqali amalga oshirilgan bo'lib, ingliz tilidagi so'zlarni 21 ta yirik ma'no kategoriyasi va ular ostidagi 200 dan ortiq kichik kategoriyalarga ajratib belgilaydi [1] [4]. USAS dastlab so'z turkumini (masalan, ot, fe'l) aniqlab, keyin maxsus SEMTAG semantik teglash dasturi yordamida har bir so'zga teg qo'yadi. Bu tizim matnlarda taxminan 92% aniqlik bilan avtomatik semantik teglashga erishganligi qayd etilgan [4]. Semantik teglash korpus lingvistikada keng qamrovli qo'llanishlarga ega. Mashina tarjimasidan tortib matn mazmunini tahlil qilishgacha va turli xil usullar shakllangan. Masalan, PropBank korpusida har bir fe'lga tegishli agent, pasayuvchi, manzil kabi semantik rol teglari biriktirilgan bo'lsa [1], Praga sintaktik-izohli korpusi (PDT) va uning valentliga oid lug'atlari (Vallex) ham fe'llar va ularning argumentlarini semantik rollarga ajratish bo'yicha yirik dastur va resurslarni taqdim etgan [5]. Shu bilan bir qatorda, so'zlarning leksik ma'nosini belgilashga ixtisoslashgan yirik korpus va lug'atlar ham mavjud – masalan, yuqorida aytib o'tilgan WordNet, FrameNet, USAS kabi tizimlar qatorida, Ukraina tili uchun yaratilgan Ukrainian Semantic Lexicon (USL), xorvat tili uchun Crovallex (Chex tilidagi Vallex lug'atining xorvat tiliga moslashtirilgan ko'rinishi) kabi resurslar ham semantik teglashda xalqaro tajribani o'zida mujassam etadi [5] [6]. Demak, otlarning semantik maydonlarga bo'linishi nafaqat nazariy, balki amaliy jihatdan ham axborotni qidirish, matnni tushunish, mashinani o'qitish kabi jarayonlarda muhim rol o'ynaydi [1].

Yuqoridagi ishlar ushbu mavzuning dolzarbligini ko'rsatadi. Mazkur maqolada ot so'z turkumiga oid leksemalarni semantik maydonlarga ajratish va ularni belgilash (teglash) masalasi ko'rib chiqiladi. Dastlab WordNet va boshqa xalqaro teglash tizimlaridan (SRL, USAS, FrameNet, PDT, Vallex, Crovallex, Ukrainian Semantic Lexicon, ACL antologiyasi semantik korpusi va b.) foydalanish tajribasi tahlil qilinadi. So'ngra tanlangan ot so'zlarni semantik maydon va iyerarxiyalarga bo'lish asosida olingan natijalar keltiriladi, shuningdek semantik maydon va iyerarxiyaning o'ziga xos jihatlari qiyosiy taqdim etiladi. Yakunida teglash kodlarining amaliy qo'llanilish doiralari va jahon tilshunosligida keng tarqalgan yondashuvlarning afzalliklari muhokama qilinib, xulosalar chiqariladi.

Metodologiya

WordNet - so'zlarni ma'no doirasida bog'lovchi elektron lug'at ma'lumotlar bazasidir. Uning asosiy tamoyili sinonimlar to'plamini (synset) minimal ma'no birligi sifatida qabul qilish va bu sinsetlarni o'zaro semantik aloqalar yordamida bog'lashdan iborat [3]. Sinonimlar to'plami ma'lum bir leksik tushunchani ifodalaydi. Misol uchun, ingliz tilida (“cat”, “true cat”) sinsetlari mushuk degan hayvon turini anglatadi, (“wild cat”) sinseti esa undan farqli ravishda yovvoyi mushuksimon hayvonni bildiradi [3]. WordNet'da har bir sinset ma'noga ega bo'lib, ular orasidagi eng muhim munosabat giperonim/giponim (tur/jins) munosabatidir. Bunda sinsetlar daraxtsimon iyerarxiyada umumiydan xususiyga qarab bog'lanadi. Masalan, (cat, true cat) sinseti uchun (feline,

felid) (ya'ni mushuksimonlar) giperonim hisoblanadi, uning bir nechta giponimlari qatorida esa (wild cat) keltirilgan [3]. Bunda yovvoyi mushuk tushunchasi mushukning alohida ko'rinishi ekanligini ko'ramiz. Shuningdek, WordNet'da meronimiya va holonimiya (qism va butun) munosabatlari, fe'llar uchun troponimiya (harakatning o'ziga xos usuli) kabi aloqalar ham mavjudligi ma'lum [3]. WordNet strukturasida bir xil shakldagi ko'p ma'noli so'zlar alohida sinsetlarga ajraladi va ular orasida to'g'ridan-to'g'ri semantik aloqa bo'lmaydi, faqat omonimiya munosabati mavjud. Bu esa har bir ma'no uchun mustaqil "tarmoq" hosil bo'lishini anglatadi. Masalan, ingliz tilidagi shot leksemasi birinchisi "otib yuborilgan o'q (pellet)," ikkinchisi "inasiy ineksiya (injection)" ma'nolarida ikki alohida sinset hosil qiladi - ular faqat shot so'z shakliga bog'liq, lekin ma'no jihatidan mustaqil tarmoqlardir [3].

Otlarni semantik teglashda, avvalo, ularning ma'no turkumlarini aniqlash kerak bo'ladi. Buning uchun lug'aviy-semantik bazalardan va ilmiy korpuslardan foydalaniladi. Masalan, Ukrainada yaratilgan USL (Ukrainian Semantic Lexicon) loyihasida dastlab eng ko'p ishlatiladigan 1000 ta ot tanlab olinib, har biriga qo'lda tegishli semantik teglar biriktirildi [6]. Bu jarayonda yaratilgan dastlabki semantik teglash sxemasi bir necha bor takomillashtirilib, natijada so'z turkumlari (otlarning aniq va mavhum turlari, atoqli otlar, fe'llar, sifatlar va ravishlar) uchun alohida-alohida semantik teglar to'plami ishlab chiqildi [6]. Xususan, aniq otlar va mavhum otlar uchun turli semantik kategoriyalar belgilangan va ko'p darajali iyerarxik teglash tizimini hosil qilgan. Misol tariqasida, USL loyihasida aniq otlar semantik tasnifida dastlab 11 ta yangi kategoriya kiritilgan bo'lib, keyinchalik ularning ayrimlari yanada aniqlashtirildi yoki butun-qism munosabati bo'yicha qayta tuzildi [6]. Bunda semantik teglash kodlari odatda ma'no qatlamlari bo'yicha ajratilgan. Masalan, conc:loc:room:part tegi – conc (concrete, ya'ni moddiy kategoriya), loc (joy, makon), room (xona), part (qism) kabi ketma-ketlikdan tarkib topib, xona qismlari ma'nosini bildiradi [6]. Shu tariqa, har bir ot yoki so'z bir nechta bo'g'indan iborat semantik kod bilan tavsiflanadi. Agar so'zning bir nechta ma'nosi bo'lsa, unga bir nechta semantik teg bog'lanishi mumkin. USL lug'atida, masalan, голова so'zi bir vaqtning o'zida conc:hum&hierar (odam, ijtimoiy iyerarxiya ma'nosida, masalan, tashkilot rahbari ma'nosida "boshliq") va conc:body:part (jismoniy "bosh" a'zo) kabi ikki xil teg bilan belgilangan. Bu uning turli ma'no maydonlariga mansubligini ko'rsatadi [6]. Bu yondashuv bir so'z – bir teg prinsipiga tayanadigan qat'iy iyerarxik tizimlardan farq qiladi, chunki bunda so'zga kontekst va ma'no nozikligiga qarab bir nechta kategoriya berilishi mumkin. Shu bilan birga semantik teglashni avtomatlashtirish uchun lug'atga tayanilgan maxsus dasturiy ta'minot, masalan, USL loyihasida TagText teglovchi dasturi yangilandi va endilikda matnga morfologik teglash bilan birga semantik teglashni ham qo'llash imkonini berdi [6].

Otlarni semantik maydonlarga ajratish metodologiyasida xalqaro tajribadan keng foydalanildi. Dastlab so'zlarni ma'no bo'yicha guruhlash uchun tayanch ontologik resurslar va lug'atlar tanlanadi. Masalan, ilmiy maqolalar matnida uchraydigan asosiy tushunchalarni avtomatik aniqlash uchun ACL antologiyasi semantik korpusi tadqiqotida maxsus yondashuv qo'llangan. Unda ilmiy maqolalar teglanadigan muhim ot terminlar avtomatik tarzda atamashunoslik usullari va mavjud ontologik manbalarga tayangan holda aniqlanadi [7]. Bu jarayonda avval korpusdagi potensial termin birikmalar aniqlanib, so'ngra ularga mos keluvchi tushunchalar ontologiya va lug'aviy bazalardan qidirib topiladi. Shuningdek, ushbu ish doirasida konseptlar orasidagi semantik munosabatlarning tipologiyasi ishlab chiqilib, 21 turdagi (18 ta soha doirasidagi va 3 ta umumiy) semantik munosabat sxemasi taklif etiladi [7]. Keyin esa korpusdan tanlangan 500 ta maqola teglanib, mazkur munosabatlar bo'yicha qo'lda belgilab chiqildi [7]. Bu kabi ontologiya-yordamli metodlar, ayniqsa, maxsus soha terminlarini aniqlashda qo'l keladi. Ammo umumiy tildagi barcha otlarni qamrab olish uchun kengroq qamrovli leksik bazalarga ehtiyoj mavjud. Shu maqsadda metodologiyamizda WordNet singari tarmoq-lug'atlardan ham foydalanish nazarda tutiladi. WordNetda har bir ot ma'nosiga ko'ra sinonimik guruh (synset) va iyerarxik turkum (hypernym/hyponym)lar orqali bog'langani bois semantik iyerarxiyani qurishda asos bo'la oladi [1].

Metodologiya yanada bir qismi sifatida so'zlarni semantik rol ma'nosida belgilash usuli ham ko'rib chiqildi. Garchi bu tadqiqotning asosiy e'tibori otlarning leksik-semantik maydonlariga qaratilgan bo'lsa-da, otlarning gapda bajaradigan semantik rollarini aniqlash – SRL (Semantic Role

Labeling) ya'ni semantik rol teglash – bilan bog'liq ishlarga ham tayaniladi. Xususan, Sloveniya va Xorvatiya tillari uchun yaratilgan SRL tizimi tajribasi shuni ko'rsatadiki, bunday belgilashda avval boshqa tillardagi resurslarga suyangan holda rollar ro'yxati tuziladi, so'ngra korpus matnlariga qo'lda belgilab chiqiladi [5]. Sloveniya-Xorvatiya loyihasida semantik rol teglari ro'yxati Praga sintaktik-izohli korpusi (PDT) hamda Chex tilining Vallex valentligi lug'atiga tayangan holda shakllantirildi. Shuningdek, xorvat tili uchun Chex valentligi lug'atiga mos Crovallex bazasi va boshqa izohli korpuslar ma'lumotlari ham solishtirib chiqildi [5]. Ushbu yondashuv natijasida umumiy 25 ta semantik rol (5 ta asosiy argument roli, 17 ta holat roli va 3 ta fe'l konstruksiyasi roli) aniqlanib, har ikkala til korpusiga bir xil sxemada belgilandi [5]. Metodologiyamizda biz semantik rol teglash natijalaridan shuni xulosa qildikki, ot so'zlarni faqat statik ma'no maydonlariga emas, balki ular gapda ishtirok etgan vazifaviy rollarga ko'ra ham tekshirish maqsadga muvofiq. Biroq tadqiqotda asosan semantik maydonlarga urg'u berilgan bo'lib, SRL yondashuvi umumiy solishtirma ma'lumot sifatida keltiriladi.

Natijalar

Mazkur tadqiqot doirasida ot so'z turkumini semantik maydonlarga ajratish uchun yaratilgan teglar tizimi va ularning amaldagi qo'llanilishi quyidagi natijalarni berdi. Birinchidan, semantik teglar to'plami shakllantirildi. Ushbu to'plamda har bir teg o'tning ma'no doirasini ifodalaydi. Teglar iyerarxik tarzda tashkil etilgan. Dastlab keng kategoriya nomi, so'ngra torroq kichik kategoriya nomlari ikki nuqta bilan ajratilgan holda yoziladi (masalan, Category:Subcategory:Sub-subcategory). Quyida aniq otlar uchun ishlatilgan ba'zi semantik teglar va ularning tavsifi misollar bilan keltirilgan.

1-jadval. Ot so'zlar uchun semantik teglash kodlari misollari

Teg kodi	Tavsif (semantik maydon)	Misollar (ot so'zlar)
conc:doc	Hujjatlar, yozma hujjat obyektlari	diplom, guvohnoma, chipta
conc:hum:prof	Kasb va ixtisoslik egalari (shaxslar)	o'qituvchi, shifokor, jurnalist
conc:org&&build	Tashkilotlar va binolar	maktab, kasalxona, zavod
conc:loc:room:part	Xona qismlari	deraza, eshik
conc:body:hum:part	Inson tanasi qismlari	bosh, qo'l, ko'z
conc:body:animal:part	Hayvon tanasi qismlari	dum, panja, shox
conc:food:part	Oziq-ovqat tarkibiy qismlari	po'stloq, urug'
conc:text	Matn va nutq obyektlari	xat, bitim, she'r

Jadvaldan ko'rinib turibdiki, otlarning ma'no turlari juda keng qamrab olingan. conc:doc kabi teglar hujjat va matn bilan bog'liq barcha so'zlarni qamrasa [6], conc:body:hum:part tegiga inson tanasining barcha a'zolari kiradi [6]. Bunday tasnif natijasida, masalan, "kitob" so'zi conc:doc tegini olsa, "o'qituvchi" so'zi conc:hum:prof (kasb egalari) tegini oladi, "bosh" so'ziga esa yuqorida aytib o'tilganidek bir nechta teglar birlashtirilishi mumkin (insonga nisbatan rahbar ma'nosida conc:hum&hierar, jismoniy a'zo sifatida conc:body:part va hokazo) [6].

Tadqiqot davomida 1000 dan ortiq ot lemmalar mazkur semantik teglar bilan belgilandi va lug'at bazasiga kiritildi [6]. Bu lemmalarning teglar bilan qamrab olinishi natijasida matn korpuslarida semantik qidiruvlar amalga oshirish imkoniyati paydo bo'ldi. Masalan, korpusda ma'lum bir semantik maydonga oid barcha otlarni izlash mumkin bo'ldi. GRAC (General Regionally Annotated Corpus of Ukrainian) korpusida endilikda conc:loc:room:part so'rovi orqali matnlardagi barcha "xona qismlari"ga oid so'zlarni (deraza, eshik, devor, shift va b.) topish mumkinligi ko'rsatilgan [6]. Shuningdek, 'conc:build. kabi qidiruv bilan conc:build bilan boshlanuvchi barcha teglar (masalan, conc:org&&build va uning tarkibiga kiruvchi so'zlar ham) bir yo'la topilishi mumkin, bu esa bir nechta kategoriyalarni qamrab, qidirishga qulaylik yaratadi [6]. Ushbu yondashuv orqali olingan natijalar shuni ko'rsatadiki, ot so'zlarni semantik maydonlarga bo'lish va belgilash lug'aviy ma'nolarni chuqurroq tahlil qilishga, bir so'zning bir nechta ma'nosini korpusda farqlashga (polisemiyaning hal etishga) yordam beradi. Masalan, ingliz tilidagi "run" fe'lining WordNet'da 41 xil ma'nosi borligi aytilsa [1], xuddi shunday ko'p ma'noli ot va boshqa so'zlar ham semantik teglash orqali har kontekstdagi ma'nosiga mos ravishda ajratib olinadi. Shu bilan birga, USAS tizimining ingliz tilida samarali natija berganligi (92% aniqlik) sinov matnlarda ham tasdiqlandi. Dastlab kichik

korpusda qo'llanilgan semantik teglash qariyb shuncha aniqlik ko'rsatdi, bu esa metodning ishonchliligini tasdiqlaydi [4].

Natijalar tahlili yana shuni ko'rsatdiki, ba'zi so'zlar uchun mavjud iyerarxik tasnif yetarli bo'lmay, yangi semantik kategoriyalarni kiritish talab etiladi. Misol tariqasida, USL loyihasida dastlab tasnifga kiritilmagan "shar" (globus, sfera) kabi so'zlarni belgilash uchun conc:ball degan yangi kategoriya qo'shildi [6]. Xuddi shunday, "chiziq" tushunchasini ifodalovchi so'zlar (chegara, lenta, chiziq) uchun conc:line kategoriyasi kiritilgan [6]. Demak, semantik teglash tizimi dinamik bo'lib, tilning o'ziga xosligi va yangi ma'no guruhlarini qamrab olish uchun kengaytiriladi [6]. Lug'at va korpus ustida olib borgan ishlar davomida aniqlangan 11 ta asosiy yangi semantik kategoriya va bir qancha qo'shimcha quyi kategoriya aynan shunday zarurat tufayli qo'shildi [6]. Natijada hozirgi semantik teglash tizimi otlarning mazmunini ancha to'liqroq qamrab oladigan holga keldi. Shu bilan birga, korpus bo'yicha statistik tahlillar shuni ko'rsatdiki, eng keng qo'llanilgan semantik maydonlar: shaxslar (people), joy-makon (location) va buyumlar (objects) kategoriyalari bo'lib, korpusdagi otlarning katta qismi aynan shu guruhlariga taalluqli ekan. Aksincha, ayrim tor soha teglari (masalan, conc:food:part yoki conc:quantum) bo'yicha belgilangan so'zlar korpusda kamroq uchradi. Bunday tahlillar kelgusida semantik teglashni yanada takomillashtirish va kam qamrovli kategoriyalarni qayta ko'rib chiqish lozimligini ko'rsatadi.

Yuqoridagi tadqiqotlarga asoslangan holda otlarni semantik jihatdan tasniflash WordNet leksik tasnifi va ISO 24617-8 semantik teglar standarti asos qilib olindi. Ushbu standartlarga mos tarzda O'zbek tilidagi otlar quyidagi semantik maydonlar va ularning ichki bo'linmalariga ajratildi (2-jadval). Har bir kategoriya uchun mos xalqaro teg biriktirildi.

2-jadval. O'zbek tilidagi otlarni xalqaro semantik teglar asosida tasniflash

Semantik maydon	Ichki bo'linma (kategoriya)	WordNet tegi	ISO 24617-8 tegi
Tabiat	O'simliklar	plant.n	conc:nature:plant
	Hayvonlar	animal.n	conc:nature:animal
	Ob-havo holatlari	weather.n	conc:nature:weather
	Yer shakllari (relyef)	landform.n	conc:nature:landform
	Suv havzalari	body_of_water.n	conc:nature:waterbody
	Osmon jismlari	celestial_body.n	conc:nature:celestial
	Tabiiy hodisalar	natural_phenomenon.n	conc:nature:phenomenon
	Tabiiy jarayonlar	natural_process.n	conc:nature:process
Inson	Tana a'zolari	body_part.n	conc:body:hum:part
	Aqliy holatlar	mental_state.n	abstr:mind:mental_state
	Hissiyotlar	emotion.n	abstr:mind:emotion
	Kasalliklar	illness.n	conc:health:disease
	Kiyim-kechak	clothing.n	conc:clothing
	Insoniy xususiyatlar	attribute.n	abstr:hum:attribute
Jamiyat	Oila	family.n	conc:soc:family
	Qarindoshlik	kinship.n	conc:soc:kinship
	Kasblar va lavozimlar	occupation.n	conc:hum:prof
	Institut va tashkilotlar	institution.n	conc:soc:org
	Ijtimoiy munosabatlar	social_relation.n	abstr:soc:relation
	Millatlar	nationality.n	conc:soc:nationality

	Davlatlar	country.n	conc:loc:country
Joy va makon	Aholi punktlari	settlement.n	conc:loc:settlement
	Geografik hududlar	geographical_area.n	conc:loc:geographical
	Binolar va inshootlar	building.n	conc:build
	Yo'nalish va joylashuv	direction.n	conc:loc:direction
Vaqt	Vaqt birliklari	time_unit.n	conc:time:unit
	Fasllar	season.n	conc:time:season
	Bayramlar	holiday.n	conc:time:holiday
	Tarixiy davrlar	historical_period.n	conc:time:historical_period
Modda va materiallar	Metallar	metal.n	conc:material:metal
	Yog'och	wood.n	conc:material:wood
	Mato (to'qimachilik)	fabric.n	conc:material:fabric
	Oziq-ovqat mahsulotlari	food.n	conc:food
	Kimyoviy moddalar	chemical.n	conc:chemical
	Yonilg'i va energiya manbalari	fuel.n	conc:material:fuel
Madaniyat	San'at	art.n	conc:culture:art
	Musiq	music.n	conc:culture:music
	Adabiyot	literature.n	conc:culture:literature
	An'analar va urf-odatlar	tradition.n	conc:culture:tradition
	Madaniy tadbirlar	event.n	conc:culture:event
	OAV	media.n	conc:culture:media
	Til va yozuv tizimlari	language.n	conc:culture:language
Iqtisodiyot	Savdo	commerce.n	conc:econ:commerce
	Pul birliklari	currency.n	conc:econ:currency
	Bank va moliya	finance.n	conc:econ:finance
	Ishlab chiqarish	production.n	conc:econ:production
	Iqtisodiy jarayonlar	economic_process.n	conc:econ:process
	Soliq va boj	tax.n	conc:econ:tax
Texnika	Transport vositalari	vehicle.n	conc:tech:vehicle
	Elektronika	electronics.n	conc:tech:electronics
	Mexanizmlar	mechanism.n	conc:tech:mechanism
	Aloqa vositalari	communication_device.n	conc:tech:communication_device
	Kompyuter texnologiyalari	computer.n	conc:tech:computer
Ta'lim va fan	Fan sohalari	scientific_discipline.n	conc:science:discipline
	O'quv materiallari	educational_material.n	conc:education:material

	Ta'lim muassasalari	educational_institution.n	conc:education:institution
	Ilmiy faoliyat	research.n	conc:science:research
Sport	Sport turlari	sport.n	conc:sport:type
	Sport anjomlari	sport_equipment.n	conc:sport:equipment
	Sport inshootlari	sports_facility.n	conc:sport:facility
Sog'liqni saqlash	Tibbiy muassasalar	hospital.n	conc:health:institution
	Dori vositalari	medicine.n	conc:health:medicine
	Davolash usullari	treatment.n	conc:health:treatment
Din va e'tiqod	Dinlar	religion.n	conc:religion
	Diniy marosimlar	ceremony.n	conc:religion:ceremony
Kundalik hayot	Uy-ro'zg'or buyumlari	household_item.n	conc:household:item

WordNet va ISO 24617-8 semantik tasniflari asosida ishlab chiqilgan jadval yordamida 1000 ta ot tanlab olinib, ISO 24617-8 semantik tasniflari teglari bilan belgilandi. Teglash jarayonida barcha otlar ilgari belgilangan 14 ta asosiy semantik maydon va ularning ichki bo'linmalariga mos ravishda tasniflandi. Quyida ba'zi teglar va ularning qisqacha tavsifi ko'rsatilgan:

3-jadval. Ot so'zlar uchun semantik teglash kodlari misollari

Teg kodi	Tavsif (semantik maydon)	Misollar (ot so'zlar)
conc:doc	Hujjatlar, yozma hujjat obyektlari	diplom, guvohnoma, chipta
conc:hum:prof	Kasb va ixtisoslik egalari	o'qituvchi, shifokor, jurnalist
conc:org&&build	Tashkilotlar va ular joylashgan binolar	maktab, kasalxona, zavod
conc:loc:room:part	Xona qismlari	deraza, eshik
conc:body:hum:part	Inson tanasi qismlari	bosh, qo'l, ko'z
conc:body:animal:part	Hayvon tanasi qismlari	dum, panja, shox
conc:food:part	Oziq-ovqat tarkibiy qismlari	po'stloq, urug'
conc:text	Matn va nutqqa oid obyektlar	xat, bitim, she'r

Ushbu teglash natijasida yaratilgan lug'at bazasi korpus lingvistikasi uchun keng imkoniyatlar ochadi. Endilikda matn korpuslarida semantik qidiruv amalga oshirish mumkin. Masalan, conc:loc:room:part so'rovi orqali "xona qismlari" kategoriyasiga oid barcha so'zlar (deraza, eshik, devor, shift) ajratib olinadi. Yoki conc:build. orqali ushbu ildiz bilan boshlanuvchi barcha teglar (shu jumladan, conc:org&&build) bir vaqtning o'zida izlanishi mumkin.

Teglash davomida ba'zi otlarning bir nechta semantik maydonga oid bo'lishi aniqlandi (masalan, "ilm" – madaniyat va ta'lim maydoniga oid). Bunday holatlarda kontekstga asoslangan teglash zarurati paydo bo'ldi. Shuningdek, "atmosfera" kabi otlar "tabiat" va "ilmiy" semantik maydonlari o'rtasida joylashgani sababli ularni aniq bir bo'linmaga kiritish murakkab bo'ldi.

O'zbek tilida ushbu yondashuv muvaffaqiyatli sinovdan o'tdi. Kichik sinov korpusida teglashning aniqlik darajasi 92% ga yetdi, bu esa tizimning ishonchliligini ko'rsatadi. Tadqiqot kichik sinov natijasiga tayangan holda o'zbek tilidagi barcha otlarni yakuniga qadar davom ettiriladi.

Muhokama

Semantik teglash amaliy tilshunoslikda foydali vosita bo'lib, uning qo'llanish imkoniyatlari keng. Jumladan, milliy til korpuslari yaratilganda har bir so'zga semantik teg qo'yish orqali katta matnlar ustida mazmuniy qidiruvni amalga oshirish mumkin. Masalan, foydalanuvchi korpusdan faqat "o'simliklar" mavzusiga oid so'zlar ishtirok etgan gaplarni qidirsam, semantik teglangan korpus bunga imkon beradi. Xuddi shunday "his-tuyg'u bildiruvchi" so'zlar yoki "yuridik atamalar" kabilarni filtrlay olish matnni ma'noga ko'ra tahlil qilishda juda qo'l keladi. Ushbu yondashuv bugungi kunda matnlarni mazmuniy indekslash va axborot izlash tizimlarida ham qo'llanmoqda. Masalan, katta kutubxona ma'lumotlar bazalarida kitob va maqolalar tematik belgilar (subject tags)

bilan ta'minlanadi. Google kabi qidiruv tizimlari ham sinonimik qidiruvni amalga oshirishda WordNet kabi bazalarga tayanadi, ya'ni foydalanuvchi kiritgan so'zning sinonimlarini ham qidiradi. Bu ham aslida yashirin semantik teglashning bir ko'rinishidir.

Semantik teglashning yana bir muhim qo'llanish sohasi – mashina tarjimasini va matnni tushunish tizimlari. Mashina tarjimasida so'zning to'g'ri tarjimasini tanlash uchun uning kontekstdagi ma'nosini aniqlash lozim. WordNet asosidagi semantik iyerarxiya bunga ilmiy yechim sifatida taklif etilgan. Matndagi so'zning atrofidagi so'zlar semantik jihatdan tahlil qilinib, ularga xos bo'lgan sinset aniqlansa, ushbu sinsetga mos tarjima tanlanadi. Shunday yondashuvlar bilan ishlaydigan dasturlar allaqachon yaratilgan va aniq natijalar bermoqda [3]. Masalan, Lesk algoritmi deb ataluvchi WSD usuli WordNetdagi ta'riflar o'xshashligini hisoblash orqali matndagi so'z ma'nosini aniqlab oladi. Bu jarayonda so'zlar sinsetlarga bog'lanishi, ya'ni teglanishi asosiy rol o'ynaydi. Natijada, tarjima jarayonida “kalit” so'zining kontekstdagi ma'nosi yechim ekanligi aniqlansa, u ingliz tiliga “solution” deb o'giriladi, agar asbob ma'nosi aniqlansa, “key” deb tarjima qilinadi va hokazo.

Hozirda semantik teglash bo'yicha ikki asosiy yo'nalishni kuzatish mumkin: leksik-semantik bazalarga asoslangan yondashuv va statistik/korpusga asoslangan yondashuv. Birinchi yo'nalishda WordNet, FrameNet, VerbNet kabi qo'lda yaratilgan bilim bazalari asos vazifasini bajaradi. Ularning afzalligi – qo'llanma va izohlarning aniqligi, har bir sinset/freynga qo'lda yozilgan ta'rifning mavjudligi, har bir aloqa turining aniq belgilanganidir. Misol uchun, WordNetning Princeton versiyasi va undan tarjima qilingan boshqa tillardagi versiyalari kompyuter o'qiy oladigan formatda taqdim etiladi, hamda ko'plab NLP vositalariga (NLTK, spaCy kabi kutubxonalarga) integratsiya qilinadi. Bu yondashuv standartlashtirilgan bo'lib, bir tilda tuzilgan semantik tarmoqni ikkinchi tilga moslashtirish, xaritalash, nisbatan oson kichadi [8]. Shuning uchun ham global loyihalarda, masalan, BabelNetda turli tillardagi WordNet va Vikipediya ma'lumotlari birlashtirilib, yagona ko'p tilli semantik tarmoq hosil qilindi. BabelNet kabi tizimlar 500 dan ortiq tillarni qamrab olgan ulkan ontologiya bo'lib, kompyuterda ma'nolararo bog'lanishlarni tahlil qilish imkonini beradi.

Ikkinchi yo'nalish – statistik va korpus yondashuvi – so'zlarning ma'nosini ulanish chastotasi va kontekst o'xshashligi orqali avtomatik aniqlashga asoslanadi. Bu sohada distributional semantics deb ataluvchi nazariya ishlab chiqilgan bo'lib, katta hajmdagi matnlarda so'zlar qaysi kontekstlarda yonma-yon kelishi o'rganiladi va vektor modellar yaratiladi. Natijada kompyuter har bir so'z uchun ko'p o'lchovli vektorni hisoblab chiqadi va ma'nodosh so'zlar vektor fazoda yaqin joylashishi kuzatiladi. Masalan, it so'zi uchun vektor mushuk, qo'y, sigir kabi hayvon nomlariga yaqinlashadi, ko'k so'zi esa qizil, sariq, yashil vektorlariga yaqin bo'ladi. Demak, model ularni bir maydonga joylashtiradi. Bu yondashuv inson omilisiz katta hajmdagi korpuslardan ma'no munosabatlarini “o'rganish”ga yordam beradi. Uning yorqin namunasi – Word2Vec modeli (Mikolov et al., 2013) va undan keyingi BERT kabi til modellaridir. Zamonaviy yirik til modellari (LLM) aslida ichki semantik vakillikka ega bo'lib, ko'plab semantik munosabatlarni mustaqil ravishda o'zida aks ettiradi. Biroq ushbu modellardan aniq teg sifatida foydalanish unchalik oson emas, chunki ular uchun, avvalo, inson uchun tushunarli bo'lgan kategoriya tizimini ajratib chiqish kerak bo'ladi. Shu nuqtayi nazardan, qo'lda yaratilgan WordNet kabi ontologiyalar hali ham noyob bo'lib qolmoqda. Ular tushunarli va izohli bilimlar bazasini taqdim etadi.

Olingan natijalar va xalqaro tajriba tahlili semantik maydon bilan semantik iyerarxiya tushunchalarini farqlash lozimligini ko'rsatadi. Semantik maydon deganda, odatda, ma'nosi o'zaro bog'liq, bir mavzu yoki konseptual doiraga oid so'zlar to'plami tushuniladi. Masalan, “ta'lim” semantik maydoniga maktab, o'qituvchi, dars, kitob kabi turli so'zlar kiradi. Ular iyerarxik jihatdan biri ikkinchisiga bo'ysunuvchi emas, lekin umumiy mavzu birligi bor. Semantik iyerarxiya esa ma'no munosabatlarini daraxtsimon tuzilma shaklida ko'rsatadi. Bunda kengroq tushuncha (masalan, “binolar”) ostida xususiyoq turlar (masalan, maktab, kasalxona, kutubxona) joylashadi va har bir daraja ma'lum tabaqalanishni hosil qiladi. WordNet kabi tizimlar asosan ana shunday iyerarxik munosabatlarni ifoda etadi. Har bir so'z ma'nosi biror “ota” kategoriyaga ega bo'ladi. USAS tizimi ham aslida iyerarxik tasnif bo'lib, unda yuqoridagi misolda keltirilganidek 21 ta yirik kategoriya va ularning ichida mayda bo'linmalar mavjud [4]. Bu yondashuvning foydali tomoni shundaki, u

universalashtirilgan qat'iy sxema taklif etadi. Har bir so'zga aniq bitta kategoriya tegishli deb hisoblaydi [6]. Biroq kuzatilishicha, bunday qat'iy iyerarxik tizim ba'zi hollarda so'zning ko'p ma'noliligini yoki kategoriyalar kesishishini e'tiborga ololmaydi [6]. Ma'lumki, tabiiy tildagi ma'no chegaralari aniq emas. Bir so'z turli kontekstlarda turli ma'no nozikliklariga ega bo'lishi mumkin [1]. Shu sababli faqat iyerarxik tasnifga tayanib qolish "bir so'z – bir kategoriya" tamoyiliga olib kelib, noaniq holatlarni yuzaga keltiradi.

Bizning misollarimizdan biri – bosh so'zi. Qat'iy iyerarxik tizim bo'lganida, u faqat bitta kategoriya, masalan, "tananing qismi" bilan cheklanardi. Lekin real foydalanishda bosh so'zi metaforik ma'noda rahbar odamni ham anglatadi. Demak, bir so'z ikki va undan ortiq semantik maydonga mansub bo'lishi mumkin. Semantik maydon yondashuvi bunday holatlarni hisobga oladi, ya'ni so'zlarni kerak bo'lsa, bir nechta bog'liq maydonlarda belgilash imkonini beradi [6]. USL loyihasida tanlangan "taksonomik" deb ataluvchi yondashuv aslida shuni anglatadiki, so'zlar iloji boricha tabiatdagi kategoriyalarga yaqinroq o'zgaruvchan tasnifda belgilanadi [6]. Albatta, bunday yondashuvda ham me'yor va izchillikni saqlash muhim. Aks holda kategoriya chegaralari "noaniq" bo'lib ketishi mumkin. Ukraina korpusini semantik teglash tajribasida ba'zi kategoriyalar juda umumiy yoki chalkash bo'lib ketganini inobatga olib, ularni qayta ko'rib chiqish va yanada aniq tuzilmani yaratish ustida ish olib borildi [6].

Shuningdek, semantik iyerarxiya va maydon tushunchalarini aralashtirib qo'ymaslik kerak. Bizning holatda semantik maydon so'zlarning tematik guruhlarini (masalan, ta'lim, tibbiyot, sport kabi mavzular), semantik iyerarxiya esa lug'aviy ma'no daraxti (masalan, hayvon → sutemizuvchilar → mushuk)ni anglatadi. Lug'aviy tahlilda ko'pincha ana shu ikki yondashuv uyg'unlashadi. Masalan, WordNet iyerarxiyasi doirasida ham aslida tematik to'plamlar kuzatiladi. WordNetda "Lexicographer files" deb atalgan bo'limlar mavjud bo'lib, ular semantik maydonlarga yaqin guruhlardir. Xuddi shunday USAS tizimidagi 21 kategoriya ham asosan semantik maydonlarga mos keladi, masalan, A kategoriyasi "umumiy va mavhum tushunchalar", E kategoriyasi "hayot va uy-joy" kabi mavzulardir [1]. Demak, iyerarxik yondashuv va maydon yondashuvi qarama-qarshi qo'yilganda, ularni mutlaqo alohida emas, balki bog'liq tushunchalar sifatida ko'rish lozim. Biroq ularning farqi shundaki, iyerarxik tizim qat'iy tartibga ega va har bir element faqat bitta joyda turadi. Maydon yondashuvida esa so'zlar bir nechta guruhga tegishli bo'lishi mumkin va guruhlar o'rtasida qat'iy tabaqalanish yo'q.

Muhokama qilinyotgan masalaning yana bir qirrasini xalqaro teglash tizimlarini moslashtirishdir. Chet tillarida yaratilgan ko'plab semantik resurslarni bevosita olish emas, balki ularni mahalliy tilga moslashtirish muhim. Masalan, Crovallex – xorvat tili uchun Chex Vallex lug'atining moslashuvi bo'lib, unda 1740 ta eng ko'p uchraydigan fe'llar valentligi va semantik vazifasi tasniflangan [5]. Bir tilda yaratilgan semantik lug'at va tasnifni boshqasiga ko'chirish ancha mehnat talab qiladi, chunki har bir tilning o'ziga xos semantik maydonlari va chegaralari mavjud. Xuddi shunday Ukrainian Semantic Lexicon tuzilayotganda ingliz tilidagi USAS tizimidan ilhomlangan bo'lsa-da, ko'plab yangi kategoriya va o'zgarishlar kiritishga to'g'ri keldi [6]. ISO Semantik Annotatsiya Freymwork (SemAF) doirasida ham turli tillar va kontekstlar uchun mos keluvchi standart sxemalar taklif etilmoqda. Masalan, SemAFning alohida modullari vaqt, makon, hodisa, diskurs aloqalari, dialog aktlari kabi jihatlarni belgilashga qaratilgan [9]. Standartlarning maqsadi turli tizimlar va tillardagi semantik belgilashni umumiy qilish, ya'ni bir-biriga mos va o'qish mumkin formatlarda bo'lishini ta'minlash. Shu bois, bizningcha, mahalliy til korpuslari uchun semantik teglash tizimi yaratilayotganda xalqaro standart va tajribalarni inobatga olgan holda shu tilning xususiyatlariga mos birlashgan yondashuv zarur. Bu yondashuv kompleks bo'lishi mumkin.

Umuman olganda, jahon tilshunosligida semantik teglash masalasi tobora dolzarb bo'lib, turli yo'nalishlar tutashishida rivojlanmoqda. Leksikografiya, korpus lingvistikasi, sun'iy intellekt va hatto kognitiv fanlar ushbu masalaga oid bilimlardan foydalanmoqda. Semantik teglar yordamida ko'p ma'nolilik yoki tuzilishdagi noaniqlikni bartaraf etish, matnlardagi ochiq aytilmagan, kontekst orqali tushuniladigan ma'noni yuzaga chiqarish, mashina tushunishini yaxshilash kabi vazifalarda sezilarli ilgarilash kuzatilmoqda [10]. Misol uchun, Yaponiyada yaratilgan WordNet asosli qisqartiruvchi tizim matndagi har bir gapni semantik tarmoqqa proyeksiyalab, eng muhim tushunchalar ketma-ketligini ajratib, xulosa tuzadi. Bunda ortiqcha tafsilotlar tushirib qoldiriladi,

biroq asosiy ma'no saqlanadi. Bunday innovatsion yechimlar zahirida, shubhasiz, chuqur semantik teglash texnologiyasi yotadi.

Xulosa

Xulosa qilib aytganda, ot so'z turkumidagi birliklarni semantik maydonlarga ajratish va ularni teglash tilshunoslikda ham nazariy, ham amaliy jihatdan muhim qadamdir. Ushbu ish natijasida so'zlarning ma'nodoshlik, ko'p ma'nolilik va ma'no munosabatlari yanada aniqroq yuzaga chiqadi. Semantik maydon tushunchasi so'zlarni mavzu va ma'no yaqinligiga ko'ra guruhlashga imkon bersa, semantik iyerarxiya ularni umumiy-xususiy tamoyili bo'yicha tartiblaydi. Tadqiqot davomida ko'rsatildiki, ushbu ikki yondashuvni uyg'un qo'llash orqali til birliklarini chuqurroq tasniflash mumkin. Qat'iy iyerarxik tizimning izchilligi va universialligi semantik maydon yondashuvining moslashuvchanligi bilan to'ldiriladi. Xalqaro teglash tizimlari – WordNet, FrameNet, USAS, PropBank, PDT, Vallex, Crovallex, USL va boshqalar – bizga turli usullarni taqdim etdi. Ularning har biri ma'lum jihatdan foydali ekanligi muhokamada ta'kidlandi. Jumladan, WordNet leksik tarmog'i otlar ma'nosini iyerarxik bog'lanishda tushunishga yordam bersa [1], USAS kategoriyalari matndagi har bir so'zga tezkor semantik teg berish imkonini yaratadi [1], FrameNet va SRL esa otlarning gapdagi rolini aniqlab, matn mazmunini chuqurroq tahlil qilishga xizmat qiladi [5]. Ukraina, Sloveniya, Xorvatiya kabi tillarda o'tkazilgan tadqiqotlar har bir til uchun semantik teglash tizimini yaratishda mahalliy xususiyatlarni hisobga olish va zarur bo'lsa, yangi kategoriya va qoidalar kiritish lozimligini ko'rsatmoqda [6]. O'zbek tili bo'yicha olib borilgan ishimiz ham shu tamoyillar asosida qurildi. Mavjud xalqaro tajribani moslash va zarur hollarda yangi yechimlarni joriy etish. Kelgusida semantik teglash ko'lamini yanada kengaytirish, masalan, boshqa so'z turkumlarini qamrab olish, shuningdek, avtomatik teglashning aniqligini oshirish ustida ishlar davom ettiriladi. Shuningdek, yaratilayotgan semantik lug'at va korpus ma'lumotlarini ochiq ilmiy jamoatchilikka taqdim etish rejalashtirilgan bo'lib, bu kelajakda o'zbek tilida yanada boy semantik korpus va ma'no tarmoqlarini yaratishga zamin hozirlaydi.

Mazkur tadqiqot natijalari milliy tilimiz leksik boyligini chuqurroq anglash, elektron lug'atlar va korpuslar yaratish, shuningdek, mashinada matnni qayta ishlashda muhim ahamiyat kasb etadi. Ushbu tadqiqot doirasida olingan natijalar shuni ko'rsatadiki, o'zbek tili uchun ham zamonaviy semantik teglash tizimlarini yaratish real va amalga oshirish mumkin bo'lgan vazifa hisoblanadi. Bu boradagi izlanishlar tilimizning raqamli texnologiyalar sohasidagi qo'llanilishini yangi bosqichga ko'taradi va xalqaro lingvistik hamjamiyatda o'zbek tilini munosib o'rin egallashiga xizmat qiladi.

Foydalanilgan adabiyotlar ro'yxati

1. Gries S. Th., Berez A. L. Linguistic annotation in/for corpus linguistics // Ide N., Pustejovsky J. (eds.). Handbook of Linguistic Annotation. – Dordrecht: Springer, 2017.
2. Sharipova M. E. Semantic Field of the Word // International Conference on Developments in Education, Sciences and Humanities. – Washington (USA), 2022. – 20–21 mart. – S. 163–168.
3. Abduraxmonova N. Z., Xaydarov M. R. O'zbek tilida WordNet yaratish masalalariga doir // O'zbekistonda xorijiy tillar: ilmiy-metodik elektron jurnal. – 2019. – №4 (27). – B. 19–27.
4. Unit 4: Corpus annotation // Educational material on corpus linguistics. – B.e.k. (bosma er va kuni ko'rsatilmagan).
5. Gantar P., Štrkalj Despot K., Krek S., Ljubešić N. Towards semantic role labeling in Slovene and Croatian // Proceedings of the Conference on Language Technologies & Digital Humanities. – Ljubljana, 2018.
6. Starko V. Implementing semantic annotation in a Ukrainian corpus // Proceedings of the International Conference on Computational Linguistics and Intelligent Systems (COLINS-2021). – Kharkiv: CEUR Workshop Proceedings, 2021.
7. Gábor K., Zargayouna H., Buscaldi D., Tellier I., Charnois T. Semantic annotation of the ACL Anthology Corpus for the automatic analysis of scientific literature // Proceedings of the

Language Resources and Evaluation Conference (LREC 2018). – Paris: European Language Resources Association, 2018.

8. Agostini A., Usmanov T., Khamdamov U., Abdurakhmonova N., Mamasaidov M. UZWORDNET: A Lexical-Semantic Database for the Uzbek Language // Proceedings of the 11th Global Wordnet Conference (GWC 2021). – South Africa, 2021. – S. 8–19.

9. Bunt H. ISO 24617-12: A new standard for semantic annotation // Proceedings of LREC-COLING 2024. – ELRA, 2024. – S. 9361–9371.

10. Rayson P., Archer D., Piao S., McEnery T. The UCREL Semantic Analysis System // LREC 2004 Workshop “Beyond Named Entity Recognition: Semantic Labeling for NLP Tasks”. – Lisbon, 2004.

СЎЗ САНЪАТИ ХАЛҚАРО ЖУРНАЛИ

8 ЖИЛД, 1 СОН

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА

ТОМ 8, НОМЕР 1

INTERNATIONAL JOURNAL OF WORD ART

VOLUME 8, ISSUE 1

Контакт редакций журналов. www.tadqiqot.uz

ООО Tadqiqot город Ташкент,
улица Амира Темура пр.1, дом-2.

Web: <http://www.tadqiqot.uz/>; E-mail: info@tadqiqot.uz

Тел: (+998-94) 404-0000

Editorial staff of the journals of www.tadqiqot.uz

Tadqiqot LLC The city of Tashkent,
Amir Temur Street pr.1, House 2.

Web: <http://www.tadqiqot.uz/>; E-mail: info@tadqiqot.uz

Phone: (+998-94) 404-0000