

СЎЗ САНЪАТИ ХАЛҚАРО ЖУРНАЛИ

8 ЖИЛД, 4 СОН

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА

ТОМ 8, НОМЕР 4

INTERNATIONAL JOURNAL OF WORD ART

VOLUME 8, ISSUE 4



СЎЗ САНЪАТИ ХАЛҚАРО ЖУРНАЛИ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА | INTERNATIONAL JOURNAL OF WORD ART

№4 (2025) DOI <http://dx.doi.org/10.26739/2181-9297-2025-4>

Бош муҳаррир:
Тўхтасинов Илҳом
п.ф.д., профессор (Ўзбекистон)

Бош муҳаррир ўринбосари:

Главный редактор:
Тухтасинов Илхом
д.п.н., профессор (Ўзбекистан)

Заместитель главного редактора:

Editor in Chief:
Tuhtasinov Ilhom
DSc. Professor (Uzbekistan)

Deputy Chief Editor

ТАҲРИРИЙ МАСЛАҲАТ КЕНГАШИ

Назаров Бахтиёр
академик. (Ўзбекистон)

Якуб Умарўғли
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 Рахиммамед
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)

Шнукашвили Тамар
ф.ф.д. (Грузия)

Туробов Бекпулат
масъул қотиб, PhD, доцент
(Ўзбекистон)

Мўминова Маъсума Ғолиб қизи
ф.ф.д., (PhD) вб. доцент

РЕДАКЦИОННЫЙ СОВЕТ

Назаров Бахтиёр
академик. (Ўзбекистан)

Якуб Умар оғли
д.ф.н., профессор (Туркия)

Алмаз Улви Биннатова
д.ф.н., профессор (Азербайджан)

Бакиева Гуландом
д.ф.н., профессор (Ўзбекистан)

Миннуллин Ким
д.ф.н., профессор (Татарстан)

Махмудов Низомиддин
д.ф.н., профессор (Ўзбекистан)

Керимов Исмаил
д.ф.н., профессор (Россия)

Джўраев Маматкул
д.ф.н., профессор (Ўзбекистан)

Куренов Рахиммамед
к.ф.н. (Туркменистан)

Кристофер Джеймс Форт
Университет Мичигана (США)

Умархаджаев Мухтар
д.ф.н., профессор (Ўзбекистан)

Мирзаев Ибодулло
д.ф.н., профессор (Ўзбекистан)

Балтабаев Ҳамидулла
д.ф.н., профессор (Ўзбекистан)

Дўстмухаммедов Хуршид
д.ф.н., профессор (Ўзбекистан)

Лиходзиевский А.С.
д.ф.н., профессор (Ўзбекистан)

Сиддикова Ирода
д.ф.н., профессор (Ўзбекистан)

Шнукашвили Тамар
д.ф.н. (Грузия)

Туробов Бекпулат
отв. секретарь, PhD, доцент
(Ўзбекистан)

Муминова Масума Ғолиб қизи
PhD и.о. доцент

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)

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

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

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

Dustmuhammadov 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)

Mo'minova Ma'suma
PhD Acting Docent
(Uzbekistan)

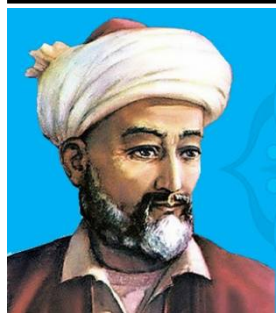
PageMaker | Верстка | Саҳифаловчи: Хуршид Мирзахмедов

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. Tajibayeva Latofat Farxodovna TABIAT VA INSON MUNOSABATLARINING ADABIYOTDA AKS ETISHI.....	5
2. Syuy Myao КРАТКОЕ РАССМОТРЕНИЕ ЯЗЫКОВОЙ И РИТОРИЧЕСКОЙ ЭСТЕТИКИ В «ЗОЛОТОЙ ЦЕПИ» ЧЖАН АЙЛИН.....	9
3. Najmiyeva Po‘lotoy Tilavmurodovna O‘ZBEK VA YAPON TILLARIDA KELISHIK QO‘SHIMCHALARI VA YAPON TILIDAGI KELISHIK PARTIKULALARINING MORFOLOGIK HAMDA SINTAKTIK MOHIYATI.....	14
4. Lutfilloeva Faxriniso Maxmudovna ЯПОН МИЛЛИЙ МАФКУРАСИДА ТИЛ ОМИЛИ ВА УНИНГ ИДЕОЛОГИК КОНЦЕПЦИЯСИ.....	20
5. Shaniyazov Jenisbay Ungarbaevich TURKIY VA INGLIZ TILLARIGA MUSHTARAK SO‘ZLAR HAQIDA.....	25
6. Ergasheva Tursunoy Abdukaюmovna ҚЎШИЛГАН ҚИЙМАТ СОЛИФИ ВА SALES TAX АТАМАЛАРИНИНГ ПРАГМАЛИНГВИСТИК ҚИЁСИЙ ТАХЛИЛИ.....	29
7. Xolikov Bahodir DEV O‘ZBEK MIFOLOGIYASI DOMINANT OBRAZALARIDAN BIRI SIFATIDA.....	34
8. Nilufar Abduraxmonova, Mahliyo Eshmamatova PARALLEL KORPUS ASOSIDA RAVISHDOSHLI BIRIKMALARNING SEMANTIKASI VA STRUKTUR MODELLARI TADQIQI.....	40
9. Hasanova Sevinch Hotam qizi THE MULTILEVEL STRUCTURE OF THE LANGUAGE SYSTEM AND ITS CONCEPTUAL INTERPRETATION IN TRANSLATION THEORY.....	52
10. Namazova Manzura Urakovna O‘ZBEK ADABIYOTIDA PSIXOLOGIZM (O‘ZBEK ROMANLARI ASOSIDA).....	58
11. Mavloнова Наргиза Алишеровна ПРИНЦИПЫ ОТБОРА ФРАЗЕОЛОГИЧЕСКИХ СООТВЕТСТВИЙ И ИДЕОГРАФИЧЕСКОЕ ОПИСАНИЕ АНГЛИЙСКИХ ФРАЗЕОЛОГИЗМОВ И ИХ СООТВЕТСТВИЙ В РУССКОМ И УЗБЕКСКОМ ЯЗЫКАХ.....	63
12. Ismailov Anvar Rustamovich CHET TILI O‘QITUVCHILARNING PEDAGOGIK KOMPETENSIYASINING MOHIYATI.....	70
13. Achilova Zuxra Qurbonovna “MALLA SAVDOGAR” DOSTONI VARIANTLARIDA AVAZXON BILAN BOG‘LIQ SAFAR MOTIVI VA VAQT KULTI.....	75
14. Abduraxmonova Nilufar Zaynobiddin qizi, Xoljurayeva Yulduz Sobir qizi O‘ZBEK TILIDAGI NORP OBYEKTLARINI BIOES ANNOTATSIYA SXEMASI ASOSIDA ANIQLASH.....	88
15. Kim Татьяна Сергеевна ДИСКУРСИВНЫЙ АНАЛИЗ НОВОСТНОГО ТЕКСТА КОРЕЙСКИХ ГАЗЕТ.....	96
16. Laziza Djhiyanbayeva РАНГ КОМПОНЕНТЛИ ТУРФУН БИРИКМАЛАР.....	102
17. Laziza Dzhiiyanbayeva SOCIAL CHARACTERISTICS OF COLOR SEMANTICS.....	108
18. Makhammatkulova Diyora Nodirovna, Makhammatkulova Iroda Nodirovna, Rahimov Nodir Makhammatkulovich DIGITAL FOOTPRINT AS A PHENOMENON OF MODERN EDUCATION.....	114

19. Boboyev Ulash Ne'matovich OMMAVIY AXBOROT VOSITALARI TILINING LEKSIK IMKONIYATLARI (GAZETA TILI MISOLIDA).....	120
20. Файзуллаева Мохинорабегим Бехзодбек кизи СРАВНИТЕЛЬНЫЙ АНАЛИЗ РЕЧЕВЫХ АКТОВ ИЗВИНЕНИЯ В КИТАЙСКОМ И УЗБЕКСКОМ ЯЗЫКАХ.....	124
21. Rashidova Muxlisa, Sobirova Madina Mahmudjon qizi UNIVERSAL DEPENDENCIES MODELI DOIRASIDA SINTAKTIK ANNOTATSIYALANGAN KORPUSLARNI TAHLIL QILISH.....	129
22. Ишниязова Шахноз Ашировна МИФОЛОГИЧЕСКОЕ МЫШЛЕНИЕ И ТОТЕМИЧЕСКИЕ МОТИВЫ В УЗБЕКСКОЙ ПРОЗЕ XX ВЕКА: К ПРОБЛЕМЕ ТИПОЛОГИЧЕСКОЙ БЛИЗОСТИ С ПРОИЗВЕДЕНИЯМИ ЧИНГИЗА АЙТМАТОВА.....	135
23. Тошбекова Дилором Исмоиловна СОПОСТАВИТЕЛЬНЫЙ АНАЛИЗ СТРУКТУРНОЙ ОРГАНИЗАЦИИ ИРОНИЧЕСКИХ ФРАЗЕОЛОГИЗМОВ В РУССКОМ И УЗБЕКСКОМ ЯЗЫКАХ.....	143
24. Умурова Гузал Хотамовна ПСИХОЛОГИЯ ХУДОЖЕСТВЕННОГО ТВОРЧЕСТВА - ЗНАЧИМОСТЬ КОМПЛЕКСНОГО ПОДХОДА К ИЗУЧЕНИЮ И ОРГАНИЗАЦИИ ТВОРЧЕСКОЙ ДЕЯТЕЛЬНОСТИ.....	149
25. Даминова Фазода Ирисовна ТЕОРЕТИКО-ИСТОРИЧЕСКИЕ И МИРОВОЗЗРЕНЧЕСКИЕ ОСНОВЫ ДРЕВНЕРУССКОЙ ЛИТЕРАТУРЫ.....	154
26. Усманова Салиха Юлдашевна ИЗ ИСТОРИИ ИЗУЧЕНИЯ ВВОДНО-МОДАЛЬНЫХ ЕДИНИЦ В РУССКОМ И УЗБЕКСКОМ ЯЗЫКОЗНАНИИ.....	161
27. Hakimov Muxamadali Rafikovich ITALYAN TILIDA "ERKAKLIK" GENDER TUSHUNCHASINING LEKSIK-SEMANTIK VA LINGVOMADANIY TAHLIL.....	170
28. Lola Ibragimova INGLIZ VA O'ZBEK FRAZELOGIYASI: STRUKTUR-SEMANTIK VA LINGVOMADANIY XUSUSIYATLAR (SOG'LIQ-SALOMATLIK MAQOLLARI MISOLIDA).....	176
29. Nazirova Munisa Mirzakamalovna MODAL BIRLIKLARNING LINGVISTIK VA PRAGMATIK ASOSLARI.....	182
30. Sharipov Raufjon Elmurod o'g'li AKADEMIK G'ANIJON ABDURAHMONOVNING "TARIXIY SINTAKSIS" ASARIDA QO'SHMA GAPLAR EVOLYUTSIYASI.....	186
31. N.Z.Abduraxmonova, Salomat Otamurodova MASHINA TARJIMASI VA LINGVISTIK JARAYONLAR.....	191
32. Davlatova Hulkaroy Uktamovna BOLALARDA TIL VA TAFAKKURNING SHAKLLANISHI: LINGVOKOGNITIV YONDASHUVDA NAMOYON BO'LISHI.....	198



ISSN: 2181-9297

www.tadqiqot.uz

INTERNATIONAL JOURNAL OF WORD ART


Rashidova Muxlisa

O'zbekiston Milliy universiteti, Jurnalistika va o'zbek filologiyasi fakulteti, Kompyuter lingvistikasi yo'nalishi magistranti
Email: tolliboyevamuxlisa@gmail.com | ORCID: 0009-0003-5281-1905

Sobirova Madina Mahmudjon qizi

O'zbekiston Milliy universiteti, Jurnalistika va o'zbek filologiyasi fakulteti, Kompyuter lingvistikasi yo'nalishi magistranti
Email: madinasobirova063@gmail.com | ORCID: 0009-0004-0471-3769

UNIVERSAL DEPENDENCIES MODEL DOIRASIDA SINTAKTIK ANNOTATSIYALANGAN KORPUSLARNI TAHLIL QILISH

 <https://doi.org/10.5281/zenodo.19322614>

ANNOTATSIYA

Ushbu maqola zamonaviy tilshunoslikning jadal rivojlanayotgan sohalaridan biri — korpus lingvistikasida sintaktik annotatsiyalangan korpuslarni shakllantirish va ulardan foydalanishning metodologik asoslarini tadqiq etishga bag'ishlangan. Tadqiqot doirasida korpus tushunchasining mazmuni, annotatsiya turlari hamda sintaktik tahlilning nazariy modellari, xususan, bog'liqlik grammatikasi (dependency grammar) tamoyillari chuqur tahlil qilingan. Maqolada annotatsiyalangan korpuslar bilan ishlashning bosqichma-bosqich metodologik tizimi ishlab chiqilgan bo'lib, unda ilmiy va rasmiy uslubdagi matnlarning sintaktik xususiyatlari Universal Dependencies (UD) standartlari asosida o'rganilgan. Amaliy qismda gaplardagi so'zlarning o'zaro sintaktik munosabatlari maxsus teglar (tags) yordamida belgilangan va ushbu jarayonning lingvistik tahlildagi samaradorligi asoslab berilgan.

Tayanch so'zlar: korpus lingvistika, sintaktik annotatsiya, treebank, analitik tahlil, metodologiya, Universal Dependencies.

Rashidova Mukhlisa

Master's Student in Computational Linguistics, Faculty of Journalism and Uzbek Philology, National University of Uzbekistan
Email: tolliboyevamuxlisa@gmail.com | ORCID: 0009-0003-5281-1905

Sobirova Madina Mahmudjon kizi

Master's Student in Computational Linguistics, Faculty of Journalism and Uzbek Philology, National University of Uzbekistan
Email: madinasobirova063@gmail.com | ORCID: 0009-0004-0471-3769

ANALYSIS OF SYNTACTICALLY ANNOTATED CORPORA WITHIN THE FRAMEWORK OF THE UNIVERSAL DEPENDENCIES MODEL

ABSTRACT

This article is devoted to the study of the methodological foundations for creating and utilizing syntactically annotated corpora within corpus linguistics, one of the rapidly developing branches of modern linguistics. The research provides an in-depth theoretical analysis of the concept of "corpus," types of annotation, and dependency models of syntactic analysis. A systematic methodological framework for working with annotated corpora is developed, focusing on the syntactic characteristics of scientific and official texts based on Universal Dependencies (UD) standards. In the practical part, the syntactic relations between words in sentences are identified and labeled using specific tags, demonstrating the effectiveness of this process in linguistic analysis. The results of the research are of great importance for improving the syntactic level of Uzbek language corpora and developing machine translation systems.

Keywords: corpus linguistics, syntactic annotation, treebank, analytical analysis, methodology, Universal Dependencies.

Рашидова Мухлиса

Магистрант направления «Компьютерная лингвистика», факультет журналистики и узбекской филологии, Национальный университет Узбекистана Email: tolliboyevamuxlisa@gmail.com | ORCID: 0009-0003-5281-1905

Собирова Мадина Махмуджон кизи

Магистрант направления «Компьютерная лингвистика», факультет журналистики и узбекской филологии, Национальный университет Узбекистана Email: madinasobirova063@gmail.com | ORCID: 0009-0004-0471-3769

АНАЛИЗ СИНТАКСИЧЕСКИ АННОТИРОВАННЫХ КОРПУСОВ В РАМКАХ МОДЕЛИ UNIVERSAL DEPENDENCIES**АННОТАЦИЯ**

Данная статья посвящена исследованию методологических основ создания и использования синтаксически аннотированных корпусов в рамках корпусной лингвистики. В работе проводится теоретический анализ понятия «корпус», видов аннотирования, а также моделей синтаксического анализа, основанных на грамматике зависимостей (dependency grammar). Систематизированы методологические этапы работы с аннотированными корпусами, в ходе которых на основе стандартов Universal Dependencies (UD) были изучены синтаксические особенности текстов научного и официального стилей. В практической части исследования определены синтаксические отношения между словами в предложениях и зафиксированы с помощью специальных тегов, что подтверждает эффективность данного процесса для лингвистического анализа.

Ключевые слова: корпусная лингвистика, синтаксическая аннотация, treebank, аналитический анализ, методология, Universal Dependencies.

Kirish.

Jahon amaliy tilshunosligida korpus lingvistikasi va korpus texnologiyasi muammolarini o'rganish XX asrning ikkinchi yarmidan rivojlanish bosqichiga ko'tarildi. Ingliz tili korpusining paydo bo'lishi boshqa tillarda ham elektron korpuslarning yaratilishida muhim ro'l o'ynadi. Zamonaviy amaliy tilshunoslik va kompyuter lingvistikasining jadal rivojlanishi natijasida til birliklarini tizimli o'rganishda elektron korpuslar markaziy o'ringa chiqdi. Korpus — bu tilni empirik (real matnlar asosida) tadqiq etish uchun mo'ljallangan ma'lumotlar bazasi bo'lib, u shunchaki matnlar to'plami bo'lishdan tashqari, lingvistik annotatsiyalar bilan boyitilgan tahlil vositasi hamdir. Morfologik annotatsiya so'zning shakliy xususiyatlarini aniqlab bersa, tilning "skeleti" hisoblangan sintaktik annotatsiya gapdagi so'zlar o'rtasidagi grammatik munosabatlarni belgilaydi. Bu jarayon matnni tizimli tahlil vositasiga aylantirib, tilshunoslarga gap strukturasi, so'z tartibini va subordinatsiyani aniq o'rganish imkonini yaratadi. Sintaktik annotatsiyangan korpus bo'yicha qilingan ishlar ham birmuncha ilgariladi. Xususan, ingliz tilidagi sintaktik annotatsiyalarga

korpuslar, nemis sintaktik annotatsiyalangan korpuslar, rus sintaktik annotatsiyalangan korpuslari, PKU Chinese Dependency Treebank, UD Chinese GSD kabi.

Adabiyotlar tahlili. Zamonaviy tilshunoslik va tabiiy tilni qayta ishlash (NLP) sohalarida til ma'lumotlarini tizimli tahlil qilishning ahamiyati kundan-kunga ortib bormoqda. Sintaktik annotatsiyalangan korpuslar til modellarini yaratish, avtomatik tarjima tizimlari va matnlarni chuqur tahlil qilishda fundamental resurs bo'lib xizmat qiladi. Yuqori sifatli annotatsiyalangan ma'lumotlar bazasiga bo'lgan ehtiyoj, ayniqsa, sun'iy intellekt texnologiyalarining rivojlanishi bilan yanada dolzarblashdi.

XX asrning oxiridan boshlab dunyoning ko'plab tillari uchun sifatli sintaktik korpuslar yaratila boshlandi. Bu yo'nalishdagi eng mashhur loyiha ingliz tilidagi Penn Treebank bo'lib, u sintaktik tuzilmalarni tizimli belgilash orqali tahlil algoritmlari uchun poydevor vazifasini bajardi. Shuningdek, nemis tilidagi TIGER, fransuzcha French Treebank, xitoycha Chinese Treebank va rus tilidagi SynTagRus kabi loyihalar ham milliy tillarning sintaktik qonuniyatlarini raqamli formatga o'tkazishda muhim rol o'ynadi.

Bugungi kunda eng yirik global loyiha hisoblangan Universal Dependencies (UD) turli tillarning sintaktik annotatsiyasini yagona standartga keltirishga intilmoqda. UD — bu bog'liqlik grammatikasi (dependency grammar) modeliga asoslangan tizim bo'lib, barcha tillarda so'zlararo munosabatlarni bir xil asosda ifodalashni maqsad qilgan [1]. Loyiha doirasida annotatsiya jarayoni uchta asosiy darajadan iborat:

1. Lemma: So'zning lug'aviy asosi (agglutinativ tillarda, masalan, o'zbek tilida, bu odatda affikssiz o'zak shaklidir).
2. Universal so'z turkumlari (UPOS): O'zgaras 17 ta belgi (tag)dan iborat grammatik kategoriyalar ro'yxati.
3. Xususiyatlar to'plami (Features): Muayyan so'z shakliga xos lisoniy va morfologik ko'rsatkichlar.

UD sxemasida har bir gapning sintaktik tuzilishi "daraxt" shaklida ifodalanib, bitta markaziy so'z (predikat) atrofida iyerarxik bog'liqliklar shakllantiriladi. Bu yondashuv ko'p tilli NLP tizimlarini yaratishda tilshunoslik va kompyuter texnologiyalarini uzviy bog'lashga imkon beradi.

Sintaktik korpuslar nazariy jihatdan ikki modelga bo'linadi: konstituentiyaviy (constituency-based) va bog'liqlik (dependency-based) modellari [2]. Konstituentiyaviy model gap strukturasi guruhlar (frazalar) ko'rinishida yaxshi ifodalasa-da, morfologik jihatdan boy va erkin so'z tartibiga ega tillar (rus, xitoy va turkiy tillar) uchun bog'liqlik modeli ancha samaraliroq hisoblanadi. O'zbek tilida ham sintaktik annotatsiyalangan korpuslar yaratish ishlari hozirda rivojlanish bosqichida bo'lib, bu jarayonda Universal Dependencies modeliga asoslangan loyihalar markaziy o'rin tutadi. O'zbek UD korpusi (Uzbek-UT) so'zlar orasidagi grammatik bog'liqliklarni belgilash orqali tilning morfologik va sintaktik xususiyatlarini tizimli o'rganish imkonini beradi. Mazkur korpus tadqiqotchilarga gap tarkibidagi subordinatsiyani, so'z tartibini va bog'lovchi vositalarni empirik usulda tahlil qilish uchun qulay manba bo'lib xizmat qiladi. Shuni ta'kidlash lozimki, korpus yaratish jarayoni o'ta mehnat va vaqt talab qiladigan murakkab vazifadir. Malakali mutaxassislar tomonidan amalga oshiriladigan annotatsiya jarayoni nafaqat ilmiy tadqiqotlar sifatini oshiradi, balki avtomatik tarjima va nutqni aniqlash kabi texnologik yechimlar uchun muhim lisoniy baza bo'lib xizmat qiladi.

Tadqiqot metodologiyasi. Mazkur tadqiqotda o'zbek tili matnlarini sintaktik annotatsiyalash jarayoni empirik metod hamda distributsion tahlil tamoyillariga tayanadi. Tadqiqot metodologiyasi xalqaro Universal Dependencies (UD) standartlari asosida ishlab chiqilgan bo'lib, quyidagi bosqichma-bosqich jarayonlarni o'z ichiga oladi:

1. **Gaplarni yig'ib olish** – tadqiqot obyekti sifatida internet nashrlari hamda rasmiy manbalardan olingan matn namunalaridan foydalanildi. Bunda asosiy e'tibor ilmiy va rasmiy uslubga xos bo'lgan sodda yoyiq gaplarni tanlab olishga qaratildi.
2. **Teglash** – har bir so'z morfologik va sintaktik xususiyatlariga ko'ra teglarga ajratiladi. Universal Dependencies standartlariga mos tarzda amalga oshirildi.

Tahlil va natijalar Sintaktik annotatsiyalangan korpuslar yaratish va ularni modellashtirish masalasi o‘tgan asrning ikkinchi yarmidan boshlab tilshunoslik va kompyuter fanlari kesishmasidagi asosiy muammolardan biri bo‘lib kelmoqda. Dastlabki nazariy asoslar N. Xomskiy tomonidan ishlab chiqilgan konstituentiyaviy model (Phrase Structure Grammar) va Z.Harrisning strukturaviy lingvistika metodlariga borib taqaladi. Biroq, amaliy korpus lingvistikasida haqiqiy burilish 1990-yillarda M. Markus va uning hamkasblari tomonidan yaratilgan Penn Treebank loyihasi bilan yuz berdi. Ushbu loyiha ingliz tili misolida sintaktik annotatsiyaning ilk keng ko‘lamli standartini belgilab berdi. Keyinchalik, tilshunoslikda sintaktik tahlilning ikki asosiy yo‘nalishi — konstituentiyaviy va bog‘liqlik (dependency) modellari o‘rtasidagi farqlar va ularning turli tillar uchun samaradorligi tadqiqotchilar diqqat markazida bo‘ldi. D. Jurafsky va J. Martin kabi olimlar o‘z tadqiqotlarida bog‘liqlik grammatikasi (Dependency Grammar) erkin so‘z tartibiga ega bo‘lgan va morfologik jihatdan boy tillar (masalan, turkiy, slavyan tillari) uchun ancha qulay ekanligini isbotladilar. Shu asosda, nemis tili uchun TIGER, rus tili uchun SynTagRus kabi nufuzli treebanklar shakllantirildi. So‘nggi o‘n yillikda ushbu sohadagi eng yirik global tashabbus — Universal Dependencies (UD) loyihasidir. J. Nivre va uning jamoasi tomonidan 2014-yilda asos solingan ushbu loyiha tillararo sintaktik annotatsiyani standartlashtirishni maqsad qilgan[4]. UD tizimi nafaqat G‘arb tillarini, balki Osiyo va Afrika tillarini ham yagona teglash tizimiga birlashtirdi. Tadqiqotchi C. Biemann ta’kidlaganidek, UD modeli NLP tizimlarining ko‘p tilli muhitda ishlash samaradorligini keskin oshirdi[5].

O‘zbek tilshunosligida korpus lingvistikasining nazariy va amaliy masalalari so‘nggi yillarda N. Abdurahmonovning tadqiqotlarida fundamental tarzda yoritib berildi. Uning ishlarida o‘zbek tili elektron korpusining lingvistik modellari, xususan, morfologik va sintaktik analizator algoritmlari ilmiy asoslab berilgan bo‘lib, u o‘zbek tili sintaktik tahlilining lisoniy modellarini ishlab chiqqan[3]. Shuningdek, B. Madaminov o‘zbek tili uchun treebank yaratishning metodologik muammolarini o‘rganib, ilk bor UD standartlarini o‘zbek tili sintaktik annotatsiyasiga tatbiq etish masalalarini ko‘tardi. T. Xoliyorovning tadqiqotlarida esa gaplarni avtomatik tahlil qilishning matematik modellari va bog‘liqlik grammatikasiga asoslangan algoritmlari tahlil etilgan. Yuqorida bayon etilgan nazariy yondashuvlar hamda sintaktik annotatsiyalash tamoyillariga tayangan holda, tadqiqotning amaliy bosqichida o‘zbek tilidagi real matnlardan iborat tanlanma shakllantirildi. Tahlil jarayonida gap tarkibidagi lisoniy birliklar (tokenlar) aniqlanib, ularning o‘zaro iyerarxik bog‘lanishlari va sintaktik munosabatlari xalqaro standartlar asosida teglar (tags) yordamida belgilandi. O‘zbek tili korpusining sintaktik qatlamini yaratishda jahon miqyosida tan olingan Universal Dependencies (UD) tizimining annotatsiya bazasi va metodologik ko‘rsatmalaridan foydalanildi. O‘zbek tili agglyutinativ qurilishga ega ekanligi, ya’ni so‘z shakllarining murakkab morfologik tarkibi va erkin so‘z tartibi kabi xususiyatlari sintaktik annotatsiya jarayonida yuksak aniqlik hamda chuqur lisoniy tajribani talab etadi.

Gaplar	Tahlil
Bu uning hayotidagi birinchi ma’muriy ish.	ega[olmosh]+aniqlovchi[olmosh+ning]+aniql ovchi[sifat]+aniqlovchi[son+inchi]+aniqlovch i[sifat]+kesim[ot]
Oksana Chusovitina — gimnaziya direktori.	ega[ot]+aniqlovchi[ot]+kesim[ot]
Bunday nomdor sport maktablari O‘zbekistonda atigi ikkita.	aniqlovchi[olmosh+sifat+ot]+ega[ot+lar+i]+h ol[ot+da]+ravish+kesim[son+ta]
Ikkinchisi bokschi Bahodir Jalolov sharafiga nomlangan.	ega[son+nchi+si]+aniqlovchi[ot]+aniqlovchi[ot]+to’ldiruvchi[ot+i+ga]+kesim[fe’l]

Oksana Chusovitina nomidagi ixtisoslashtirilgan sport gimnastikasi maktabi zalida uchrashdik.	aniqlovchi[ot+ot+i+dagi]+ aniqlovchi[sifatdosh]+aniqlovchi[ot]+ aniqlovchi[ot+ot+si]+aniqlovchi[ot+i]+ hol[ot+i+da] + kesim[fe'l]
Soʻzning asosiy qismlari tahlil qilinadi.	aniqlovchi[ot+ning]+aniqlovchi[sifat]+ega[ot +lar+i]+kesim[ot+fe'l]
Bahodir Jalolov – oʻzbekning yengilmas bokschisi.	ega[ot]+aniqlovchi[ot+ning]+aniqlovchi[sifat] +kesim[ot+si]

Tadqiqotning dastlabki bosqichida internet resurslari hamda rasmiy nashrlardan ilmiy va rasmiy uslubga mansub sodda, yoyiq gaplar saralab olindi. Tanlab olingan matnlar korpus tahlili talablariga muvofiq quyidagi bosqichlarda qayta ishlandi:

1. Morfologik tahlil va teglash: Har bir soʻzning leksik-grammatik xususiyatlari, soʻz turkumi va morfologik kategoriyalari aniqlanib, maxsus teglar bilan belgilandi. Morfologik sathning aniqligi sintaktik bogʻlanishlarni toʻgʻri shakllantirish uchun poydevor vazifasini oʻtadi.
2. Sintaktik annotatsiyalash: Morfologik tahlil asosida soʻzlar oʻrtasidagi iyerarxik bogʻliqliklar aniqlandi. Annotatsiya jarayonida yordamchi soʻzlar, xususan, koʻmakchi va yuklamalarning sintaktik vazifasini belgilashda maʼlum murakkabliklar kuzatildi. Biroq, bu muammolar Universal Dependencies (UD) standartlari asosida tizimlashtirildi.

Olib borilgan izlanishlar oʻzbek tilining boy sintaktik imkoniyatlarini raqamli formatda aks ettirishga xizmat qiladi. Shakllantirilgan annotatsion korpus tilning strukturaviy xususiyatlarini va sintaktik naqshlarini (syntactic patterns) chuqur oʻrganishda muhim resurs hisoblanadi.

Xulosa

Sintaktik annotatsiyalangan korpuslar zamonaviy tilshunoslik va tabiiy tilni qayta ishlash (NLP) sohalarida chuqur strukturaviy tahlillarni amalga oshirish uchun eng muhim resurs hisoblanadi. Bunday korpuslarni tahlil qilishda qoʻllaniladigan metodologik yondashuvlar — xususan, statistik tahlil, graf strukturalari va mashina oʻrganish modellari — ilmiy izlanishlarning aniqlik darajasini sifat jihatidan yangi bosqichga koʻtaradi. Bugungi kunda oʻzbek tili uchun sintaktik annotatsiyalangan korpuslar yaratish va ularni tizimli tahlil qilish strategik zaruriyatga aylangan. Zero, annotatsiyalangan maʼlumotlar bazasi yordamida tilning oʻziga xos soʻz birikmalari, sintaktik naqshlari va murakkab gap tuzilmalarini empirik usulda oʻrganish imkoniyati tugʻiladi. Bu esa nafaqat nazariy tilshunoslik, balki, avtomatik tarjima tizimlarining aniqligini oshirish, nutqni aniqlash va generatsiya qilish texnologiyalarini takomillashtirish, mashina oʻrganishi modellarini oʻzbek tili grammatikasiga moslashtirishda hal qiluvchi ahamiyatga ega.

Shu bilan birga, sintaktik korpuslarni shakllantirish jarayoni oʻziga xos qiyinchiliklardan xoli emas. Annotatsiya jarayonining yuqori sarf-xarajat va katta mehnat talab qilishi, shuningdek, turli annotatsiya standartlari oʻrtasidagi nomuvofiqliklar sohadagi asosiy toʻsiqlar sanaladi. Biroq, Universal Dependencies (UD) kabi standartlashtirilgan yondashuvlarning joriy etilishi turli tillar uchun yagona sintaktik model yaratish imkonini bermoqda. Bu kabi xalqaro standartlar oʻzbek tilining global NLP ekotizimiga integratsiyalashuvini taʼminlab, koʻp tilli tillararo tadqiqotlarni samarali amalga oshirishga keng yoʻl ochadi.

Foydalanilgan adabiyotlar.

1. Nivre J., et al. Universal Dependencies v1: A Multilingual Treebank Collection // Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016). – Portorož, Slovenia, 2016. – P. 1659–1666.
2. Marcus M. P., Santorini B., Marcinkiewicz M. A. Building a large annotated corpus of English: The Penn Treebank // Computational Linguistics. – 1993. – Vol. 19, No. 2. – P. 313–330.

3. Abdurahmonova N.Z. O‘zbek tili elektron korpusining kompyuter modellari: Monografiya. – Toshkent: Nodirabegim, 2021. – 202 b.
4. Nivre J., et al. Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection // Language Resources and Evaluation. – 2020. – Vol. 54. – P. 403–442.
5. Biemann C., et al. Corpus Linguistics and NLP: Methods and Applications. – Springer, 2013. – 240 p.
6. Jurafsky D., Martin J. H. Speech and Language Processing (3rd ed. draft). – Pearson, 2023. – 610 p.
7. Abdurahmonova N.Z. Korpus lingvistikasi. Darslik. – GlobeEdit, 2023. – 357 b.
8. Internet manbalari:Uzdaily.uz, Google scholar.
9. Abdurahmonova N. MorphUz: O‘zbek tili uchun morfologik analizator // International Journal of Computational Linguistics. – 2021. – № 3. – B. 67–74.
10. Abdurahmonova N., Ismailov A., Yulbarsov O. O‘zbek tili uchun sintaktik analizator algoritmini ishlab chiqish // Axborot texnologiyalari va lingvistika. – Toshkent, 2023. – B. 45–52.
11. Harris Z. Methods in Structural Linguistics. – Chicago: University of Chicago Press, 1951. – 384 p.
12. Chomsky N. Syntactic Structures. – The Hague: Mouton & Co., 1957. – 117 p.
13. <https://catalog.ldc.upenn.edu/LDC99T42>,
14. <https://catalog.ldc.upenn.edu/LDC97S62>,https://en.wikipedia.org/wiki/Brown_Corpus
15. <https://www.ims.uni-stuttgart.de/en/research/resources/corpora/tiger/>
16. <https://www.sfs.untuebingen.de/en/ascl/resources/corpora/tuba-dz.html>
17. https://universaldependencies.org/treebanks/ru_syntagrus/index.html
18. <https://ruscorpora.ru/new/>, <https://iling.spb.ru/>
19. <http://ccl.pku.edu.cn>
20. https://universaldependencies.org/treebanks/zh_gsd/

СЎЗ САНЪАТИ ХАЛҚАРО ЖУРНАЛИ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА
INTERNATIONAL JOURNAL OF WORD ART

Контакт редакций журналов. www.tadqiqot.uz
ООО Тадқиқот 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