

**СЎЗ САНЪАТИ
ХАЛҚАРО ЖУРНАЛИ
4 ЖИЛД, 1 СОН**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ
ИСКУССТВО СЛОВА
ТОМ 4, НОМЕР 1**

**INTERNATIONAL JOURNAL
OF WORD ART
VOLUME 4, ISSUE 1**



Бош муҳаррир:

Холбеков Муҳаммаджон

ф.ф.д., профессор (Ўзбекистон)

Бош муҳаррир ўринбосари:

Тўхтасинов Илҳом

ф.ф.д., доцент (Ўзбекистон)

Тахрир хайъати:

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 Рахыммамед

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)

Юсупов Ойбек

масъул котиб, доцент (Ўзбекистон)

Главный редактор:

Холбеков Муҳаммаджон

д.ф.н., профессор (Ўзбекистон)

Заместитель главного редактора:

Тухтасинов Илҳом

к.ф.н., доцент (Ўзбекистон)

Редакционная коллегия:

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)

Джураев Маматкул

д.ф.н., профессор (Ўзбекистон)

Куренов Рахыммамед

к.ф.н. (Туркменистон)

Кристофер Джеймс Форт

Университет Мичигана (США)

Умархаджаев Мухтар

д.ф.н., профессор (Ўзбекистон)

Мирзаев Ибодулло

д.ф.н., профессор (Ўзбекистон)

Балтабоев 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)

Юсупов Ойбек

отв. секретарь, доцент (Ўзбекистон)

Editor in Chief:

Kholbekov Muhammadjan

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

Deputy Chief Editor

Tuhtasinov Ilhom

Ph.D. Ass. Prof. (Uzbekistan)

Editorial Board:

Bakhtiyor Nazarov

academician. (Uzbekistan)

Yakub Umarogli

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

Almaz Ulvi Binnatova

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

Bakieva Gulandom

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

Minnulin Kim

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

Mahmudov Nizomiddin

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

Kerimov Ismail

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

Juraev Mamatkul

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

Kurenov Rakhimmamed

Ph.D. Ass. Prof. (Turkmenistan)

Christopher James Fort

University of Michigan (USA)

Umarkhodjaev Mukhtar

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

Mirzaev Ibodulla

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

Boltaboev Hamidulla

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

Dustmuhammedov Khurshid

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

Lixodzievsky A.S.

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

Siddiqova Iroda

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

Shiukashvili Tamar

Doc. of philol. sci. (Georgia)

Yusupov Oybek

Ass. prof. (Uzbekistan) - Senior Secretary

МУНДАРИЖА \ СОДЕРЖАНИЕ \ CONTENT

Адабиётшунослик

- 1. Rahmonqulova Zarnigor**
DISTOPIK EKZISTENSIALIZMNI INGLIZ YOZUVCHISI UILYAM
GOLDINGNING “LORD OF THE FLIES” ROMANIDA AKS ETISHI.....6
 - 2. Усманова Севара**
ЧЎЛПОН ВА С.РИЧАРДСОН РОМАНЛАРИДА БАДИИЯТ МУАММОЛАРИ.....14
 - 3. Yusupova Dildora**
SHOIR SO‘Z AYTGANI KELAR DUNYOGA...(Shoira Halima
Xudoyberdiyevaning hayot va ijod yo‘liga chizgilar).....21
 - 4. Қўчқорова Садоқат**
ФОЛЬКЛОРДА ЭРТАКНИНГ ЖАНР СИФАТИДАГИ ХУСУСИЯТИ.....28
 - 5. Мажитова Сабоҳат, Ярашов Шариф**
АМИР ХУСРАВ ДЕҲЛАВИЙ ВА АБДУРАҲМОН ЖОМИЙ ҒАЗАЛЛАРИДА
«ДИЛ» (ЮРАК, ҚАЛБ) ТИМСОЛИНИНГ ЛЕКСИК-СЕМАНТИК МАЙДОНИ.....35
 - 6. Джалилова Зарнигор**
ИНГЛИЗ ШЕЪРИЯТИДА ИНСОН СИФАТЛАРИНИНГ ГУЛЛАР
ТИМСОЛИДА ИФОДАЛАНИШИ.....42
- ## Лингводидактика
- 7. Khamzaev Sobir, Gilyazetdinov Eldar**
THE PRINCIPLES OF SPEAKING SKILLS’ FORMATION AMONG
STUDENTS OF LINGUISTIC UNIVERSITIES.....48
- ## Тилшунослик
- 8. Boygucheva Lola**
THE ROLE OF DISCOURSE ANALYSIS IN MODERN LINGUISTIC THEORY.....54
 - 9. Кобилова Азиза**
ШИФОБАХШ ЎСИМЛИКЛАР ЎРНИДА ИШЛАТИЛАДИГАН
ПЕРИФРАЗЛАР ТАҲЛИЛИ (ИНГЛИЗ ВА ЎЗБЕК ТИЛЛАРИ МИСОЛИДА).....59
 - 10. Mamanazarova Gulhayo, Paluanova Halifa**
ISSUES ON OIL AND GAS TERMINOLOGY IN LINGUISTICS.....65
 - 11. Melikova Elnora, Abdullayeva Nargiza**
PRIMARY AND SECONDARY PREDICATION.....71
 - 12. Saidova Farida**
O‘ZBEK TILI LEKSIK SINONIMLARINING LEKSIKOGRAFIK TALQINI.....77
 - 13. Turniyazov Behzod**
FAKTIVLIK ASOSIDAGI KAUZATSIYA.....83


14. Гафуров Бахтиёр АНАЛИЗ ЛИНГВОСТАТИСТИЧЕСКОЙ ХАРАКТЕРИСТИКИ АКЦЕНТНЫХ ФОНОВАРИАНТОВ ИМЕН СУЩЕСТВИТЕЛЬНЫХ РУССКОГО ЯЗЫКА.....	91
15. Ибаев Анвар КОМПАРАТИВ СИНТАКТИК ҚУРИЛМАЛАР ТИЛШУНОСЛИКНИНГ ЎРГАНИЛИШ ОБЪЕКТИ СИФАТИДА.....	97
16. Исмаилов Тургун ПРЕЗИДЕНТ ДИСКУРСИДА HOMELAND КОНЦЕПТИНИНГ МИЛЛИЙ ҚАДРИЯТЛАР ОРҚАЛИ ТАСАВВУР ЭТИЛИШИ (FREEDOM, LIBERTY ҚАДРИЯТИ МИСОЛИДА).....	103
17. Камолиддинова Вазира ОКСЮМОРОН ВА УНИНГ УСЛУБИЙ - СЕМАНТИК ХУСУСИЯТЛАРИНИНГ ҚЎЛЛАНИЛИШИ.....	111
18. Нигматова Лолахон ЎЗБЕК ТИЛИ МАХСУС КОРПУСЛАРИ – ТИЛ ВА МАДАНИЯТ МУШТАРАКЛИГИНИ ЎРГАНИШ ВОСИТАСИ.....	116
19. Сабирова Ёъзоа ЎЗБЕК ТИЛИНИНГ ИЗОҲЛИ ЛУҒАТЛАРИДА ЎЗЛАШМА ЮРИДИК ТЕРМИНЛАР ТАЛҚИНИ.....	121
20. Сулаймонова Нигора ЗАМАХШАРИЙ ГЎЗАЛ НУТҚ ВА САБР ОДОБИ ҲАҚИДА.....	130
21. Тўева Зулфия, Юлдашева Дилором БАЁН ВА УНИНГ ТУРЛАРИ ҲАҚИДА.....	137
22. Xolova Muyassar BOYSUN SHEVASI ONLAYN MA'LUMOTLAR BAZASI (QIDIRISH VA SARALASH IMKONIYATLARI ASOSIDA).....	146
23. Abdullayeva Oqila Holmo'minovna O'ZBEK TILI KORPUSINI YARATISH BOSQICHLARI VA MUAMMOLI JHATLARI....	153
24. Abdunazarova Zarina Islomovna THE FORMATION OF SPEECH COMPETENCE.....	159
25. Nilufar Abduraxmonova O'ZBEK TILI KORPUSINI YARATISHDA LINGVISTIK ANNOTATSIYALASH TAMOIYILLARI.....	164
26. Shamuradova Naima O'ZBEK VA INGLIZ XALQ MAQOLLARINING QIYOSIY O'RGANILISHI HAQIDA (ASOSIY QISMLARINI QO'LLASH).....	171
27. Яхёева Гулнора Бахтиёровна ОСОБЕННОСТИ РАЗВИТИЯ ТВОРЧЕСКОЙ АКТИВНОСТИ СТУДЕНТОВ В ПРОЦЕССЕ ОБУЧЕНИЯ ИНОСТРАННОМУ ЯЗЫКУ.....	175

28. Irgasheva Umida INGLIZ TILI DARSLARIDA TALABALAR KASBIY NUTQ KOMPETENTSIYASINI TAKOMILLASHTIRISHDA STEAM TEXNOLOGIYALAR AHAMIYATI.....	179
29. Jumaniyozova Nuriya Axmedovna YURIDIK DISKURS LINGVISTIK TADQIQOT OBYEKTI SIFATIDA.....	184
30. Khayitova Feruza Abdikhalikovna THE PRACTICAL ANALYSIS OF TOPONYMS IN GREAT BRITAIN BASED ON THE STORY OF A. CONAN DOYLE "THE HOUND OF THE BASKERVILLES".....	188
31. Z. Masodiqova, Sh. Xatamqulova O'ZBEK VA NEMIS MAQOLLARINI QIYOSHLASH.....	193
32. Teshaboyev Dilmurod Rahmadjonovich, Axmadjonova Mohinur Bahodirovna SON SO'Z TURKUMINING BOSHLANG'ICH SINF DARSLIKLARIDA IFODALANGAN SHAKLLARINI AKT ORQALI TUSHUNTIRISH.....	193
33. Maxmurova Mavjuda Halimovna METAFORALAR NUTQ SHAKLI SIFATIDA.....	204
34. Nazarov Rustam Irkinovich THE IMPORTANCE OF PAIR WORK AND GROUP WORK ACTIVITIES IN TEACHING ENGLISH IN TECHNICAL UNIVERSITIES.....	210
35. Saidbekova Raykhona Rustambekovna, Ruzmatova Dilnoza Ramatjonova FAMILY PERSPECTIVES IN THE NOVEL "AND THE MOUNTAIN ENCHOED" BY KHALED HUSSEINI.....	217
36. Алижон Раҳматович Қосимов, Хумора Муҳаммадиса қизи Иминова ТИЛНИНГ БИОПСИХОСОЦИАЛ ОНТОЛОГИК МАҚОМИ МУАММОЛАРИ.....	222

СЎЗ САЊАТИ ХАЛҚАРО ЖУРНАЛИ МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА INTERNATIONAL JOURNAL OF WORD ART

Nilufar Abduraxmonova,
fil.f.d.(PhD), Alisher Navoiy nomidagi Toshkent davlat
o'zbek tili va adabiyoti universiteti dotsenti
anilufar@navoiy-uni.uz

O'ZBEK TILI KORPUSINI YARATISHDA LINGVISTIK ANNOTATSIYALASH TAMOYILLARI

 <http://dx.doi.org/10.26739/2181-9297-2021-1-25>

ANNOTATSIYA

Maqolada o'zbek tili elektron korpusini yaratishda matnlarni qayta ishlash uchun lingvistik annotatsiyalashning asosiy talablari va tamoyillari haqida ma'lumot berilgan. Lingvistik annotatsiyalash tilning turli sathlari: morfologik, sintaktik, semantik jihatdan farqlanadi. Matnlar korpusining lingvistik bazasiga o'zbek tilidagi turli uslubga doir yozma materiallar kiritildi. Lingvistik jihatdan annotatsiyalash va matnlarga korpusga tayyorlashda avtomatik, yarim avtomat va qo'l mehnati usulidan foydalanildi. Korpusning elektron resurslari yaratilib, statistik natijalari o'rganildi.

Ishimizda o'zbek tili elektron korpusining kontenti, ya'ni matnlar hajmi va proporsional jihatdan turlari ixtiyoriy ravishda metama'lumotlari bilan birgalikda ularni to'ldirish, qayta ishlash hamda o'zgartirish imkoniyatiga ega.

Kalit so'zlar: o'zbek tili korpusi, lingvistik annotatsiya, o'zbek tili, kompyuter lingvistikasi, semantik annotatsiya, morfologik annotatsiya, sintaktik annotatsiya

Nilufar Abdurakhmonova,
PhD., Associate Professor of Alisher Navo'i Tashkent State
University of the Uzbek language and literature
anilufar@navoiy-uni.uz

PRINCIPLES OF LINGUISTIC ANNOTATION IN THE CREATION OF THE UZBEK LANGUAGE CORPUS

ANNOTATION

The article is devoted to on the basic requirements and principles of linguistic annotation for text processing in the creation of the electronic corpus of the Uzbek language. Linguistic annotation differs at different levels of language: morphologically, syntactically, semantically. The linguistic base of the corpus of texts includes written materials on various styles in the Uzbek language. Linguistic annotations and the preparation of texts for the corpus used automatic, semi-automatic and manual ways. Electron resources of the corpus were created and statistical results were investigated.

In our work showed the content of the electronic corpus of the Uzbek language could be modified filled, and processed as the size and proportional types of texts, along with metadata.

Key words: corpus of the Uzbek language, linguistic annotation, the Uzbek language, computational linguistics, semantic annotation, morphological annotation, syntactic annotation.

Нилуфар Абдурахмонова,

PhD, доцент Ташкентский Государственный Университет
узбекского языка и литературы им. Алишера Навои

ПРИНЦИПЫ ЯЗЫКОВОЙ АННОТАЦИИ ПРИ СОЗДАНИИ УЗБЕКСКОГО ЯЗЫКОВОГО КОРПУСА

АННОТАЦИЯ

Статья посвящена основным требованиям и принципам лингвистической аннотации для обработки текста при создании электронного корпуса узбекского языка. Лингвистическая аннотация различается на разных уровнях языка: морфологически, синтаксически, семантически. Лингвистическая база корпуса текстов включает письменные материалы по различным стилям узбекского языка. Лингвистические аннотации и подготовка текстов к корпусу используются автоматическими, полуавтоматическими и ручными способами. Созданы электронные ресурсы корпуса и исследованы статистические результаты.

В нашей работе показано, что содержание электронного корпуса на узбекском языке может быть модифицировано, заполнено и обработано как по размеру, так и по пропорциональным типам текстов вместе с метаданными.

Ключевые слова: корпус узбекского языка, лингвистическая аннотация, узбекский язык, компьютерная лингвистика, семантическая аннотация, морфологическая аннотация, синтаксическая аннотация.

Kirish

Korpusshunoslik (korpus lingvistikasi) turkiy tillar bo'yicha Aksan, Deniz Zeyrek, Kemal Oflazer, Umut Özge Bular; uyg'ur tili bo'yicha Yusup Aibaidulla, Kim-Teng Lua; boshqird tili bo'yicha Buskunbaeva L.A., Sirazitdinov Z.; hakas tili bo'yicha Sheimovich, tatar tili bo'yicha J.Suleymanov, A.Gatiatullin, O.Nevzorova, R.Gilmullin, B.Hakimov; qrimtatar tili bo'yicha L.Kubedinova hamda tuva tili bo'yicha Salchak kabi olimlar tomonidan izlanishlar olib borilgan. [1, 95] da qayd etilishicha, yuqorida sanab o'tilgan tillar orasida turk va uyg'ur tilining korpusi barcha til sathlarining annotatsiyalangani va talablarga to'liqroq javob berishi bilan ham farqlanadi. Tatar tilining "Tugen Tel" korpusi tatar tilining turli sathlari bo'yicha konseptual va funksional modellarning majmuasida sifatida qarash mumkin. Konseptual va funksional modellarning sinflari o'z navbatida muayyan til sathining struktural va funksional tavsifidan, shuningdek, tabiiy tilni qayta ishlash uchun texnologiyasi va axborot tizimlarini rivojlantirish uchun zarur bo'lgan umumiy ma'lumotlardan ham iborat. O'zbek tili korpusi bo'yicha muayyan darajada tadqiqotlar amalga oshirilgan.

manba	Hujjat soni	Gaplar soni	So'zlar	Token	Lemma
Badiiy adabiyot	8	35416	292015	74108	11221
guncha.uz	11	266	1945	1160	524
hidoyat.uz	20	778	11019	5382	1757
kun.uz	18	281	5047	2588	1071
lex.uz	33	1450	422241	8032	1602
marifat.uz	20	462	8592	4235	1544
minsport.uz	7	124	2013	1304	637
xs.uz_xalq so'zi	80	1536	23602	8911	2478
yoshlikjurnali.uz	38	8138	92304	34374	6922

O'zbek tili korpusining kontenti

Tadqiqotimizda o‘zbek tili korpusini yaratishda lingvistik resurslar sifatida web sahifalari, lug‘atlar (bir tilli va ikki tilli), o‘zbek tilining morfologik bazasi, badiiy adabiyotlar olingan (1-jadval). Avvalgi ishlarimizda uning morfologik va sintaktik tahliliga doir erishilgan natijalar e‘lon qilingan [2, 3, 4].

Matnlarni kompyuterda qayta ishlash jarayoni avtomat, yarim avtomat va qo‘l mehnati yordamida amalga oshirilgan. Korpus strukturasi shakllantirish uchun muvozanatlashgan korpus tamoyillariga asoslanildi. Ya‘ni olingan matnlarning janri, yaratilgan davri, mavzusi va uslubi proporsional tarzda teng taqsimlandi. Lingvistik resurs sifatida olingan manbalarning original holatini saqlagan holda o‘zbekcha matnlarning ham kiril, ham lotin yozuvidagi shakllari bazaga kiritildi. Matnni kompyuterda qayta ishlash uchun unga tegishli kontekstga doir ma‘lumot muayyan darajada annotatsiyalanishi kerak. Ishning lingvistik annotatsiyalash (razmetka) bosqichi sifatli natijaga erishishda muhim omil hisoblanadi. Razmetka quyidagi jihatlariga ko‘ra tasniflanadi [5, 28]:

- Ektralingvistik (metarazmetka) – matn (muallifi, nashr yili, nashriyot nomi, yili, janri, mavzusi) va uning muallifi haqida ma‘lumot;
- Matn tuzilishi – sarlavha, satrboshi, gap, so‘z shakli;
- Lingvistik sathiga ko‘ra:
 1. Morfologik (POS-tagging) razmetka;
 2. Sintaktik razmetka – bunda so‘zlar o‘rtasidagi sintaktik aloqalar va gap turlariga ko‘ra ma‘lumotlar keltiriladi;
 3. Semantik razmetka – matnda ifodalangan tushunchalarning predmet va predmet bo‘lmagan nomlari, faoliyat turlari hamda semantik munosabatlari tavsiflanadi.
- Anaforik – matnning biror elementi boshqa matnda anglashilgan mazmun bilan aloqadorlikda aniqlanadi va lingvistik jihatdan tahlil qilinadi;
- Prosodik – urg‘u, ritm, mantiqiy urg‘u va boshqlariga ko‘ra matn lingvistik jihatdan ifodalanadi.

Dunyo tajribasida lingvistik annotatsiyaning turlari xilma-xil. Masalan, Leech qoidasiga asoslangan annotatsiyalash [6] tamoyillariga e‘tibor berilsa, quyidagilarga guvoh bo‘lamiz:

1. Asos matnga qidirish uchun korpusdan olingan annotatsiya o‘sha matndan aynan olingan bo‘lish kerak. Buning uchun morfologik teglash usulidan foydalaniladi. Morfologik teglash turli tamoyillarga asoslanadi: “Mushtariy_NP1kitob_NN2o‘qiydi_VVZ” => Mushtariy kitob o‘qiydi. Ushbu gapda har bir so‘z so‘z turkumiga ko‘ra tavsiflanadi: NP (nounphrase) – ot shakli, NN (noun) – 2-ot shakl, VV (Verb) – fe‘l shakli.
2. Shuningdek, matn qolipiga mos bo‘lgan boshqa matndan annotatsiyalar olinishi mumkin.
3. Annotatsiya strukturasi oxirgi foydalanuvchiga yo‘naltirilgan bo‘lishi mumkin. Aksariyat korpuslarda annotatsiyaning to‘liq shakli keltirilgan yo‘riqnoma bo‘ladi. Bu foydalanuvchiga annotatsiya haqida batafsil ma‘lumot berishga xizmat qiladi.
4. Annotatsiya kim tomonidan va qanday yaratilganligi ko‘rsatiladi. Korpus annotatsiyasi xoh inson yoki avtomatik tarzda amalga oshirilsa ham inson tomonidan tuzatilishi kerak.
5. Annotatsiya strukturasi umum tomonidan qabul qilingan me‘yorlarga asoslanadi.
6. Hech bir annotatsiya o‘zidan oldingi annotatsiyalarni standart, deb olishi shart emas. Standart me‘yorlar tajriba davomida yuzaga chiqadi.

Korpusning lingvistik annotatsiyalash bosqichlari

Korpus annotatsiyasi GML (Generalized Markup Language), SGML (->HTML-Standard Generalized Markup Language, standart umumiydagi kodli til), XML (Extensible Markup Language-kengaytirilgan razmetkalar tili), RDF (Resource Description Framework) kabi kodlovchi tillar orqali amalga oshiriladi.

Korpus matnining formati ishlash yo‘nalishiga yordam beradi. Masalan, SGML/XML format korpus lingvistikasiga doir tadqiqotlar uchun, CoNLLU formatidagi matnlar NLP bo‘yicha tadqiqotlarga xizmat qiladi.

To‘liqroq tasavvurga ega bo‘lish uchun XML tilining orqali manbadan keltirilgan namunasini ilova qilamiz [7]:

```
<anthology>
  <poem>
    <heading>The SICK ROSE</heading>
    <stanza>
      <line>O Rose thou art sick.</line>
      <line>The invisible worm,</line>
      <line>That flies in the night</line>
      <line>In the howling storm:</line>
    </stanza>
    <stanza>
      <line>Has found out thy bed</line>
      <line>Of crimson joy:</line>
      <line>And his dark secret love</line>
      <line>Does thy life destroy.</line>
    </stanza>
  </poem>
  <!-- more poems go here -->
</anthology>
```

Shuningdek, korpus annotatsiyasi quyidagi ma’lumotlardan iborat bo‘lishi mumkin:

- metadata (metama’lumot), sarlavha, paragraf
- tokenlar
- lemmalar
- so‘z turkumi tegi (part of speech-POS)
- morfologik belgilar
- sintaktik tahlil

Ayrim manbalarda annotatsiyaning ikki xil shakli sifatida berilib, quyidagicha guruhlanadi: so‘z turkumlarini teglash (POS-parts of speech tagging) va razmetka (razmetka). Birinchi shaklda korpusdagi yoki segmetlardagi har bir leksik birlik grammatik belgilari (ot, sifat, fe‘l) teg bilan belgilanadi. Birinchi shaklda avtomatik ravishda har bir segment birlik grammatik jihatdan belgilanib olinadi, so‘ng mutaxassis tomonidan qayta tahrirlanadi. Chunki tabiiy tilning istisnoli vaziyatlari ayrim me‘yoriy qoidalarga mos tushavermasligi mumkin va ular tahlilda, shubhasiz, hisobga olinish kerak.

Korpusni teglash matnni lingvistik analiz qilish uchun foydalaniladi. Matnni razmetka qilish turli jihatlarini qamrab olishi mumkin. Masalan, yozma matnlar razmetkasida matnning struktur birliklari, sarlavhasi hisobga olinsa, og‘zaki matnlar uchun tekshirish obyekt sifatida olingan so‘zlovchilarning yoshi, jinsi, kasbi kabi tomonlarini ham ifodalashi mumkin. Kiritilgan ma’lumotlarni mashina o‘qiy oladigan formatda berish uchun matnni kodlash yo‘riqnomasi (Text Encoding Initiative (T.E.I.)) orqali standart tilda beriladi. Agar shu kungacha yaratilgan korpuslarga e’tibor berilsa, Brown korpusi so‘z turkumlari jihatidan teglanganligini, Penn Treebanki gaplarning sintaktik analiz qilish bo‘yicha annotatsiyalanganligi, SemCor korpusida esa so‘zlarning ma’nolariga ko‘ra semantik jihatdan farqlangan annotatsiyalanganini, Bangla, Hindi, Marathi, Telugu tillarini o‘z ichiga olgan hind korpusida so‘z turkumlariga ko‘ra lingvistik annotatsiyalanganini kuzatish mumkin.

Endilikda korpus lingvistikasi tilshunoslikning ajralmas qismiga aylandi. Negaki kompyuter lingvistikasi, sotsiolingvistika, pedagogika, tarjimashunoslik, diskurs analiz kabi sohalarda korpuslardan unumli foydalanilib, ijobiy natijalarga erishib kelinmoqda.

O‘zbek tilining elektron korpusini yaratishda dunyo miqyosidagi korpuslar taksonomiyasidan foydalaniladi, shu bois ishimizda ularning ilmiy tavsifiy tomonlariga e’tiborimizni qaratmoqchimiz.

Makkeneri tadqiqot uchun korpusning quyidagi uch ma'lumotdan iborat bo'lishini ta'kidlaydi [8, 29]: metama'lumot, matnga oid razmetka va lingvistik annotatsiya. Metama'lumot matn haqida ma'lumot beradi. Agar korpus yozma materiallardan iborat bo'lsa, u holda uning metama'lumoti manbaning muallifi, qachon nashr qilingani va qaysi tilda yozilganligi haqida ma'lumot beradi. Masalan, Amerika milliy korpusi (<http://www.anc.org/data/masc/corpus/>) (Open American National Corpus (OANC)) 1990-yildan buyon amalda qo'llanilib kelinmoqda. Uning hajmi 15 mln. so'zni tashkil qiladi. Korpus quyidagi birliklardan tashkil topgan: DOCUMENT NAME asosiy ma'lumot fayli uchun berilgan nom va barcha annotatsiya qilingan bog'langan ma'lumotlar. WORD COUNT hujjatlardagi mavjud so'zlarning miqdori, TYPE matnning yozma yoki og'zaki shakli, GENRE matnni tasniflash uchun korpusning lug'atidan olingan ma'lumotlarning janri, SUB-CATEGORY mavjud ma'lumotlarning qo'shimcha tasnifi, ANNOTATIONS mantiqiy tuzilishi (matnning asosiy qismi, paragrafi, sarlavhasi)dan asosiy to'plamdan iborat hujjatlar, lists annotations for the document beyond the basic set, which includes logical structure (body, paragraph, headings, etc.); jumalarning chegaralari, token va so'z turkumlari, ot va fe'l so'z turkumini hamda nomlar (shaxs, hudud, sana va tashkilot)ni aniqlovchi annotatsiyalar mavjud [4, 57].

Semantik razmetkada boshqa razmetkalarda bo'lganidek, yagona standart shakl bo'lmasa ham, harf va raqam yoki faqat raqamdan iborat kodlardan foydalaniladi. Birinchi harf yoki raqam umumiy semantik ma'noni, keyingi belgi esa so'z ma'nosini yanada maxsuslashtiruvchi kichik semantik guruhni ifodalaydi. Semantik razmetka nafaqat so'z, balki ko'plab birikmalarni ham semantik guruhlariga birlashtiradi, bunday paytda turli birikuvdagi bir ma'noni bildiruvchi birikmalar bitta belgi bilan kodlanadi. Idiomatik ibora tarkibidagi so'zlar miqdorini bildiruvchi axborot ham razmetkadan joy oladi. Semantik razmetka korpusdagi so'z ma'nosining ixtisoslashuvi, omonimlik va sinonimlik, ma'noviy guruhga ajratish kabi muammoni hal qiladi. V.P. Zaxarov, S.Yu. Bogdanovlar rus tili milliy korpusini tuzishda semantik razmetkalashning o'z variantini taklif qiladi [9]. Bu korpusda har bir so'zga uch xil – so'z razryadi, leksik-semantik tavsif, derivatsion izoh beriladi.

Anaforik razmetkalash jarayonida matnga ishlov berishda ayrim holatlarda qiyinchilik tug'diruvchi olmosh so'z turkumi va matn birliklariga ishora qiluvchi nutq birliklariga e'tibor qariladi. Ishora so'zning matndagi ma'nosini ajratib olish uchun razmetkaning alohida ko'rinishiga ehtiyoj tug'iladi. Anaforik razmetka shu xildagi ma'lumotni qo'shish uchun zarur. Ma'nosi olmosh bilan ifodalanayotgan so'z alohida kodlanib, keyingi o'rinda shu so'zga ishora qilayotgan olmosh yoniga shu kod biriktiriladi. Natijada olmoshning matndagi ma'nosi aniqlanadi hamda kerakli tadqiqotlarda korpusga asoslanish imkoniyati paydo bo'ladi.

Prosodik razmetkalash ovoz transkripsiya qilingan korpusda urg'u va ohangni ifodalovchi izoh kiritiladi. Razmetkaning diskurs deb ham ataluvchi ushbu turi sharh, izoh, eslatma, takrorlardagi to'xtamlarni bildirish uchun ishlatiladi. Korpusni razmetkalash (annotatsiyalash) dasturlashtirilgan yo'llar bilan amalga oshiriladi. Bunda, avvalo, vaqtni tejash va mehnatni kamaytirish nazarda tutilsa, ikkinchidan, matnga avtomatik ishlov berish muammosiga yechim topiladi. Hozircha anaforik va prosodik razmetka qiyinligicha qolib ketmoqda hamda ularning aksariyati qo'l mehnati bilan amalga oshirilmoqda.

Morfologik va sintaktik razmetka esa tegger va parsing yordamida amalga oshirilsa ham, bu dasturlarning ham aksariyati avtomatik razmetkadan keyingi tuzatishni talab qiladi. O'zbek tili uchun aynan istisnoli vaziyatlar ko'p bo'lganligi sababli qayta tahrirlash zaruriyatini keltirib chiqaradi. Chunonchi, morfologik omonimiya (ko'proq flektiv tillarga xos) va sintaktik ko'p ma'nolilik holatida dastur natijasining bir nechta shakllarini taklif qiladi, mutaxassi bularning orasidan eng munosibini tanlaydi.

Yangi avlod korpuslari hajmining haddan ziyod kengayib borayotgani mutaxassislar oldiga razmetkaning to'liq avtomatlashtirilgan turiga o'tish, yangi tegger va parsinglar yaratish vazifasini qo'ymoqda. Avtomatik morfologik tahlil (tegger) yordamida har bir leksik birlikka (so'z turkumi, lemma, grammema guruhi) alohida grammatik kategoriyalari (shaxs-son, kelishik va boshq.) beriladi. Masalan, Braun korpusida so'zning chastotasini aniqlash oson. Faqat bu so'z shaklining

(korpus tilida token) chastotasi bo'lad. Leksemaning chastotasini aniqlash uchun esa har bir so'zga uning lemmasi biriktirilgan bo'lishi kerak. V.P.Zaxarov fikricha, lingvistik razmetkaning morfologik, sintaktik, semantik, anaforik, prosodik kabi turlari quyidagi tamoyillar asosida amalga oshiriladi [9]:

- 1) razmetka sxemasini tavsiflash (asoslash);
- 2) umumiy lingvistik tushunchalar tizimini aniqlash;
- 3) foydalanuvchi uchun ma'lum bo'lgan tahlil sxemasini shakllantirish ;
- 4) razmetka sxemasining nazariy an'anaviyligiga erishish;
- 5) xalqaro andozalarga amal qilish.

Bundan tashqari, ayrim adabiyotlarda so'z turkumlarini annotatsiyalashning boshqa bir qator turlarini keltiriladi [10, 23]: bordiyu muayyan so'z uning so'z turkumlari sinfi bilan ma'lumot keltirilishi zarur bo'lsa, ichkiqator (inline) annotatsiya yaratiladi: The inhabitants of Egypt were called mummies. Ushbu gap quyidagicha teglanadi: The/Determiner inhabitants/Noun of/Preposition Egypt/Name were/Verb called/Participle mummies/Noun /Punct. Ya'ni gapda ishtirok etgan har bir so'z turkumi nominal yoki verbal guruhdan qaysi biriga tegishli ekanligiga ko'ra ajratiladi va / belgisi qo'yiladi. Olim qo'shimcha sifatida yana bir annotatsiyalash turi sifatida yonma-yon (stand-off) annotatsiyani keltirib o'tadi. w1:Det w2:Noun w3:Preposition w4:NE w5:Verb w6:Participle w7:Noun w8:Punct. Ushbu annotatsiyalash bosqichida so'zni qayta original matndan topib, u haqidagi ma'lumotni chiqarish uchun foydalaniladi.

Annotatsiya jarayonida muayyan xulosani chiqarish uchun annotatsiya modellaridan (annotation scheme) foydalaniladi. Ular annotatsiya qo'llanmalarida qayd etiladi. Masalan, ot son kategoriyasi bo'yicha annotatsiya qilinadi, degan xulosaga kelish uchun qo'llanmada keltirilgan tavsialardan foydalanish o'rinli. Chunki ayrim matn tahlili bilan bog'liq murakkab vaziyatlarda u yoki bu so'z haqida lingvistik ma'lumot olish uchun undagi berilgan ma'lumot juda muhim. Masalan, o'zbek tilida ayrim atoqli otlar turdosh otga (Nyuton) yoki turdosh otlar atoqli otga aylangan hollari ko'p kuzatiladi. Ularni bosh harf bilan kelish kelmasligi mavhum, chunki ular gapning boshida ham kelishi mumkin. Ushbu vaziyatda annotatsiyadan foydalaniladi.

Annotatsiyalashning keyingi talabalaridan biri bu matn haqidagi ma'lumotni qay tarzda ifodalashdir. Odatda matnni annotatsiyalash qo'l mehnati bilan, avtomatik va yarim avtomat tarzida amalga oshiriladi. Qo'l mehnati bilan amalga oshirishda inson idroki va uning shaxsiy tajribasiga tayaniladi. Avtomatik usulda esa kompyuter lingvistikasidagi metodlardan foydalangan holda natijaga erishiladi. Yarim avtomat yordamida annotatsiyalashda avtomatik dasturiy metod yordamida bajariladi, erishilgan natijalari inson tomonidan tekshirib chiqiladi.

Shuningdek, temporal (vaqt bilan bog'liq) annotatsiya turi ham mavjud [10, 24]. Ushbu annotatsiyada voqealar o'rtasidagi munosabatlar va tildagi vaqt va voqeani ifodalovchi ifodalar aniqlashga e'tibor beriladi. Bu kabi ma'lumotlarning berilishi voqealar izchilligini topishda yordam beradi. Temporal annotatsiya to'rt turdagi ma'lumotlarni beradi: voqealar, vaqtni bildiruvchi so'zlar, kalit (siganl) so'zlar hamda havolalar. Ingliz tilida so'zdan oldingi qism aniq vaqni ifodalamaydi, balki qolgan yoki davom etayotgan ikki hodisa o'rtasidagi vaqtinchalik munosabatni bildiradi. Vaqtni ifodalashda signal so'zlar havola berilganda qo'llaniladi, havolalar esa hodisalar orasidagi bog'lanish bo'lib hizmat qiladi. O'zbek tilida quyidagi misolni namuna sifatida keltirish mumkin: Covid 19ning dunyo bo'ylab tarqalishi davlat va jamiyatning turli sohalariga jiddiy ta'sir ko'rsatib, rivojlanish ko'rsatkichlarining pasayishiga olib keldi. [B1Covid 19]ning [B2dunyo bo'ylab tarqalishi] [B3davlat va jamiyatning turli sohalariga] [B4jiddiy ta'sir ko'rsatib], [B5rivojlanish ko'rsatkichlarining pasayishiga olib keldi]. Bunda [B1Covid 19]-> [B5rivojlanish ko'rsatkichlarining pasayishiga olib keldi] komponent biror voqeaning sodir bo'lishidan avvalgi holar va shu ikki tarkibning o'rtasidagi signallar [B2dunyo bo'ylab tarqalishi], [B4jiddiy ta'sir ko'rsatib]dan iborat.

Gaplarning ma'nolarini ifodalashning yana bir usuli bu semantik freymlardan foydalanishdir. Semantik freymlarda ensiklopedik bilimlar bilan bog'lanadi [10, 29]. Masalan, ot yoki fe'l so'z turkumi doirasida u yoki bu so'z ifodalagan ma'no munosabatlari (sinonimlar, uyadosh so'zlar, giponim va giperonimlar kabi) orqali uning semantik maydonlarini hosil qiladi. Shu kabi birliklar kontekstning ichidan semantik annotatsiya uchun model vazifasini bajaradi.

Semantik freymlar bilan annotatsiyalangan korpus FreymNet deb yuritiladi. Masalan, o'qimoq_harakat-holat, kitob, kitobxon, kutubxona_yosita kabi birliklar freymlar hisoblanib, o'qimoq, mutolaa qilmoq, ko'z yugurtirmoq, ko'zdan kechirib chiqmoq kabi fe'llar ma'no jihatidan farq qilsada, bitta semantik maydonni tashkil qiladi. Yani, ushbu fe'llar freymlar o'rtasidagi munosabatni yuzaga chiqaruvchi vositalari bo'lib, FreymNetning leksik birliklari hisoblanadi. Demak, freymlar semantikaga asoslanadi. Gapning sintaktik strukturalari darajasida lingvistik tavsifini berish diskurs annotatsiyada ifodalanadi. Ushbu holatda annotatsiya mantiqliy tamoyiliga asoslanadi. Ya'ni mantiqiylik deyilganda matn tarkibida so'zlovchilarning og'zaki va yozma nutqiy jarayonida ajratilgan yoki o'zaro bog'lanmagan so'z, fraza va gaplarning yig'indisi tushuniladi. Qisqa paragraflar turli darajadagi mantiqiylikni ko'rsatadi.

Xulosa

O'zbek tilining elektron korpusini yaratishda uning strukturasi va matnlardan qay tarzda foydalanishlik juda muhim masalalardan biridir. Negaki uni nafaqat foydalanuvchilar uchun, balki kompyuter lingvistikasi, ya'ni til texnologiyasi uchun ham tayyorlash asosiy maqsadlarimizdan biridir. Yuqorida qayd etilgan ilmiy tasniflar va qoidalar o'zbek tili korpusining lingvistik annotatsiyalash jarayoni uchun muhim manba bo'lib xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Dzhavdet Suleymanov et al. National Corpus of the Tatar L Grammatical Annotation and Implementation / Procedia - Social and Behavioral Sciences 95 (2013) 68 – 74 5th International Conference on Corpus Linguistics (CILC2013)
2. Abdurakhmonova N. A two-level morphological analysis of the Uzbek Corpus МАТЕРИАЛЫ IV МЕЖДУНАРОДНОГО НАУЧНОГО КОНГРЕССА ИНОСТРАННАЯ ФИЛОЛОГИЯ. СОЦИАЛЬНАЯ И НАЦИОНАЛЬНАЯ ВАРИАТИВНОСТЬ ЯЗЫКА И ЛИТЕРАТУРЫ// 2019/4/1 Симферополь, -P. 425-431
3. Abdurakhmonova N. Syntactic parsing based on Uzbek corpus “Language technology for all” International conference by UNESCO France, Paris 4-6 December
4. Абдурахмонова Н. Замоновий корпусларнинг компьютер моделлари Ўзбекистонда хорижий тиллар. —2020. — № 1(30). — Б. 50–58. [https://doi.org/ 10.36078/](https://doi.org/10.36078/)
5. К. К. Боярский Введение в компьютерную лингвистику. Санкт-Петербург 2013, - С. 28.
6. www.lancaster.ac.uk/fss/course/ling/corpus/corpus2/2MAXIMS>HTM
7. <https://tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>
8. Charlez M. English corpus linguistics: An introduction. Cambridge, 2004. 12., 29
9. Zaxarov V.P., Bogdanova S.Yu. Korpusnaya lingvistika. - Irkutsk: IGLU, 2011.
10. Sandra Kübler and Heike Zinsmeister CORPUS LINGUISTICS AND LINGUISTICALLY ANNOTATED CORPORA Bloomsbury Academic, New York, 2015, P. 23
11. <https://framenet.icsi.berkeley.edu/fndrupal/WhatIsFrameNet>

СЎЗ САНЪАТИ
ХАЛҚАРО ЖУРНАЛИ
4 ЖИЛД, 1 СОН

МЕЖДУНАРОДНЫЙ ЖУРНАЛ
ИСКУССТВО СЛОВА
ТОМ 4, НОМЕР 1

INTERNATIONAL JOURNAL
OF WORD ART
VOLUME 4, ISSUE 1