

ISSN 2181-922X

# O‘ZBEKISTON

TIL VA MADANIYAT

## AMALIY FILOLOGIYA MASALALARI

2022 Vol. 2 (5)

[www.aphil.tsuull.uz](http://www.aphil.tsuull.uz)

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

**Bosh muharrir:** Saodat Muhamedova  
**Bosh muharrir o'rinbosari:** Botir Elov  
**Mas'ul kotib:** Xurshida Kadirova

### **Tahrir kengashi**

Aynur O'zjan (Turkiya), Baydemir Husayn (Turkiya), Alfiya Yusupova (Rossiya), Luiza Samsitova (Rossiya), Almaz Ulvi (Ozarbayjon), Abdulhay Sobirov, Muyassar Saparniyazova, Manzura Abjalova, Nargiza Musulmonova, Yekaterina Shirinova, Shoira Isayeva, Oqila Turaqulova, Ikrom Islomov, Munira Shodmonova, Oqila Abdullayeva, Dilrabo Elova.

### **Jurnal haqida ma'lumot**

“O'zbekiston: til va madaniyat. Amaliy filologiya masalalari” seriyasi - Oliy attestatsiya komissiyasi ilmiy nashrlar ro'yxatidagi “O'zbekiston: til va madaniyat” akademik jurnalining ilovasi hisoblanib, unda professor-o'qituvchilar, doktorantlar, stajor-tadqiqotchilar, mustaqil izlanuvchilar, magistrantlarning amaliy tilshunoslik, amaliy adabiyotshunoslik, kompyuter lingvistikasi, o'zbek tilini davlat tili va xorijiy til sifatida o'qitish, noshirlik ishi kabi sohalarga oid tadqiqotlari nashr qilinadi.

Jurnal ilovasi bir yilda ikki marta chop etiladi.

O'zbek, rus va ingliz tillaridagi, shuningdek, boshqa turkiy tillarda yozilgan maqolalar qabul qilinadi.

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

Mualliflar fikri tahririyat nuqtayi nazaridan farq qilishi mumkin.

“O'zbekiston: til va madaniyat. Amaliy filologiya masalalari” seriyasi 2022-yildan chiqa boshlagan.

Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti.

O'zbekiston, Toshkent, Yakkasaroy tumani, Yusuf Xos Hojib ko'chasi, 103.

**Email:** [aphil@tsuull.uz](mailto:aphil@tsuull.uz)

**Website:** <http://www.aphil.tsuull.uz>

## MUNDARIJA

### Amaliy filologiya

#### **Ikrom Islomov**

Geografik terminlarning ayrim semantik va grammatik xususiyatlari.....4

#### **Shoira Isayeva**

Methods of text analysis in literary reading lessons (example of the works of abdulla kakhhor).....11

### O'zbek tilini xorijiy til sifatida o'qitish

#### **Kadirova Xurshida**

O'zbek tilini xorijiy til sifatida o'qitishda o'qituvchi shaxsi va kreativ yondashuvi.....17

### Lingvokulturologiya

#### **Гульбахор Искандарова**

Роль языковой среды в формировании лингвокультурологических элементов в детской речи.....26

### Tabiiy tilni qayta ishlash (NLP)

#### **Botir Elov, Nizomaddin Xudayberganov, Zilola Xusainova**

Tabiiy tilni qayta ishlashda bag of words algoritmidan foydalanish.....35

#### **Botir Elov, Shahlo Hamroyeva, Oqila Abdullayeva, Mohiyaxon Uzoqova**

O'zbek tilida pos tegging masalasi: muammo va takliflar.....51

#### **Nizomaddin Xudayberganov, Shaxboz Hasanov**

Tabiiy tilni qayta ishlashda so'zlar orasidagi masofani aniqlash algoritmlaridan foydalanish.....69

### Korpus lingvistikasi

#### **Dilrabo Elova**

Fe'l grammatik shakllarining uslubiy xoslanishiga ko'ra lingvistik annotatsiyalash masalasi xususida.....84

#### **Botir Elov, Ma'rufjon Amirqulov**

O'zbek-ingliz tillarining teglangan parallel korpusini yaratish bosqichlari.....97

#### **Madina Samatboyeva**

Matnlarda NER obyektlarini aniqlashdagi muammolar.....110

## AMALIY FILOLOGIYA

**Geografik terminlarning ayrim semantik va grammatik xususiyatlari**Ikrom Islomov<sup>1</sup>**Abstrakt:**

Ushbu maqolada termin va unga asos bo'lgan umumleksik birlik orasidagi munosabat masalasiga doir ayrim fikr-mulohazalar bayon qilingan. Shuningdek, asos-hosila munosabatida bo'lgan termin va leksemaning semantik strukturasidagi umumiy sema, farqlovchi sema, asos sema, o'zbek geografik terminlari yasash sistemasi va morfemik tarkibi, morfologik tuzilishiga oid grammatik xususiyatlar ayrim misollar asosida tahlil qilingan. Termin va umumleksik birlikning o'zaro sistemaviy tadqiqi natijalari ularning leksikografik talqinlari mukammallashishiga xizmat qilishi ta'kidlangan.

**Kalit so'zlar:** *termin, terminosistema, geografik termin, leksema, relyef oti, semantik struktura, sema, asos sema, yasama termin, tub termin, yasovchi asos, yasovchi vosita.*

**Kirish**

Tilning leksik sistemasi tarkibida terminologik birliklar ham o'ziga xos mikrosistema sifatida alohida o'rin egallaydi. Terminlar leksik sistemaning taraqqiyotida muhim omil sanaladi. Tilning sohaviy qo'llanishida terminologik birliklarning semantik va grammatik xususiyatlari o'ziga xos. Terminologik birliklarning hosil bo'lishi, terminosistemaning shakllanishi va taraqqiyotida terminlar mansub tilning umumleksik birliklari muhim ahamiyat kasb etadi. Tildagi umumleksik birlik va u asos bo'lgan terminning o'zaro munosabati semantik aspektda tekshirilsa, ulardagi semantik umumiylik va farqli belgilar yaqqol namoyon bo'ladi. Shuningdek,

---

<sup>1</sup>*Islomov Ikrom Xushboqovich* – Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti Amaliy filologiya kafedrasi dotsenti, filologiya fanlari doktori.

**E-mail:** [islomovikrom@navoiy-uni.uz](mailto:islomovikrom@navoiy-uni.uz)

**ORCID:** 0000-0002-5011-1572

**Iqtibos uchun:** Islomov, I. 2022. "Geografik terminlarning ayrim semantik va grammatik xususiyatlari". *O'zbekiston: til va madaniyat. Amaliy filologiya*. 2 (5): 4-10.

umumiste'moldagi birlikning lisoniy qiymati va semantik strukturasi aniq belgilash, ularning leksikografik talqinini mukammallashtirish omili til birligini terminologik va umumleksik aspektda tadqiq qilish bilan bevosita bog'liq.

Borliq hodisalari ko'p qirrali mohiyatga ega bo'lganligi bois olamning lisoniy manzarasi va ilmiy terminologiya o'rtasidagi o'zaro munosabat dialektik qarama-qarshiliklar birligi qonuniyati asosida namoyon bo'ladi. Leksemalarga terminologik, onomastik, dialektal va badiiy qo'llanishlar xos bo'lib, bu qo'llanishlarning har birida mazkur atov birliklar o'ziga xos belgilari bilan farqlilik kasb etadi.

Termin bilan umumleksik birlik orasidagi munosabat masalasi tilshunoslikda turli daraja va aspektlarda o'rganilib kelinmoqda. Bu boradagi tadqiqotlarda termin va leksemaning umumiy hamda farqli xususiyatlari yuzasidan turlicha munosabat va xulosalar bildiriladi. "Terminlarning qo'llanish, tarqalish ko'lami muayyan terminologik sistema bilan cheklangan bo'lib, ular insonning faoliyati doirasidagi aniq uslubda harakat qiladi, voqelanadi. Terminologik leksika o'z tarkibiga tor mutaxassislik doirasidan ommaviy muloqot doirasiga ko'chgan noprofessional nutqiy kontekstda keng qo'llanadigan so'z va so'z birikmalarini qamrab oladi" [Dadaboyev, 2020, 6]. Bundan ma'lum bo'ladiki, terminlar va ularning xususiyatlarini tekshirish va baholashda, albatta, termin bilan unga asos bo'lgan umumleksik birlik orasidagi munosabatning sistemaviy o'rganilishi ham masalaning ilmiy yechimida alohida o'rin egallaydi.

### **Asosiy qism**

Terminologik birlik va unga asos bo'lgan umumleksik birlik orasidagi munosabatni o'zbek tilining ayrim relyef otlari va o'zbek geografik terminlari misolida kuzatishga harakat qilamiz. O'zbek tilida yer yuzasi ko'rinishlari, yer relyefi bilan bog'liq tushunchalarni ifodalovchi leksemalarning semantik strukturasi yer relyefining turli shakllari, ko'rinishlariga xos belgilar asosida shakllangan. Demak, yer relyefi bilan bog'liq tushunchalarni ifodalovchi leksik birliklarning semantik strukturasi "sathdan baland", "sathdan past", "sathga teng" kabi semalar mavjud. Ana shu semalar asosida bitta semantik guruhni tashkil qilgan leksemalar mikrosistemi o'zbek geografik terminlar sistemasining shakllanish manbai hamdir.

O'zbek tili leksikasida yer sathining baland, past, tekis ko'rinishlarini ifodalovchi leksemalar sistemasiga yuzdan ortiq birliklar mansub. Ular forma jihatidan turli-tuman bo'lsa-da,

semantik jihatdan uchta guruhni tashkil qiladi: *balandliklar, tekisliklar, pastliklar*. Yer yuzasining shakl va ko'rinishlari bilan bog'liq leksemalar sememasida, albatta, "Yer sathidan baland", "Yer sathidan past", "Yer sathiga teng" semalardan biri mavjud, ana shu sema asosida qayd qilingan mikrosistemalardan biriga mansub bo'ladi.

Yer sathining turli ko'rinish va shakllarini ifodalovchi leksemalar o'z-o'zidan relyef otlari maqomini ham egallaydi. Relyef otlari sifatida umumleksik tizimda barqarorlashgan mazkur atov birliklar keyingi bosqichlarda soha terminlari sistemasining shakllanishida ham muhim manba bo'lib xizmat qiladi. Bu holatda ham yer relyefi tushunchasiga aloqador sema termin semantik strukturasi asosini tashkil qiladi. Leksemaning umumiste'mol ma'nosiga nisbatan terminologik ma'no aniq va ixtisoslashgan xususiyati bilan o'ziga xoslikka ega bo'ladi. Demak, umumiste'mol ma'noning ixtisoslashuvi natijasida terminologik ma'no shakllanadi, bu ma'no sohaviy, ilmiy tushunchalar ifodachisi sifatida alohida talqini va tavsifiga ega bo'ladi.

O'zbek geografik terminlarining yasalishiga xos xususiyatlar ham o'ziga xoslik kasb etadi. Geografik leksemalar sistemasi tarkiban turli ko'rinishlarga ega bo'lib, ular dastlab tub yoki yasamaligi jihatidan tub geografik terminlar va yasama geografik terminlar guruhiga ajraladi.

Birinchi guruhga *vodiy, gaza, adir, tepa, qir, jar, dovon, tog', cho'qqi* kabilar mansub bo'lsa, *balandlik, beltog', ochgaza, past-tekislik, supatog', jarlik, yetimtog', yassi tepa* kabilar esa yasama geografik terminlar guruhini tashkil qiladi.

Tub geografik terminlar deb qayd qilinadigan *tog', tepa, adir, qir, dara, gaza, dovon* kabilar umumxalq tilida ham tub leksemalar mavqeyida bo'lgan. Ular ixtisoslashib, soha tushunchalarini ifodalovchi terminlar tizimiga o'tganda ham grammatik o'zgarishsiz tub geografik birliklar tarkibidan joy olgan, ya'ni o'zlarining tub leksema ekanligini soha birligi mavqeyiga o'tganda ham saqlab qolgan. Bunday aynanlikni ularning tub-yasamaliligi nuqtayi nazaridan aytish mumkin, xolos. Semantik strukturada esa, albatta, farq mavjud bo'ladi. Termin sifatida terminologik ma'no kasb etadi, umumleksik birlik esa umumiy, ijtimoiy ma'nosi bilan farqlanib turadi.

Yer sathi ko'rinishlari bilan bog'liq yasama geografik leksemalar, asosan, affiksatsiya va kompozitsiya usullari orqali yasalgan.

*Tekislik, jarlik, balandlik, tog'lik, do'nglik* kabilar qo'shimcha qo'shish usuli bilan yasalgan yasama terminlar bo'lsa; *beltog', yassi tepa, yassi jar, qoratog', yonbag'ir, ochgaza, do'ngqum, pasttekislik* kabilar kompozitsiya usuli orqali hosil qilingan yasama terminlar hisoblanadi. Darhaqiqat, "Yasama so'zning ma'nosi va boshqa xususiyatlari uning tarkibiy qismlari ma'nosi, shu qismlarga xos xususiyatlar asosida yuzaga keladi". [Hojiyev, 2010]

Yer sathi ko'rinishlarini ifodalovchi atov birliklari asos bo'lgan yasama geografik terminlarning vujudga kelishida [*ot+lik=asosdan anglashilgan qism ma'nosiga aniq ega joy, ko'rinish*] hamda [*sifat+lik=asosdan anglashilgan belgiga ega joy, ko'rinish*] derivatsion qoliplar muhim o'rin tutadi. Masalan, *jarlik, tog'lik, tepalik, sayhonlik, adirlik* kabilar [*ot+lik*] qolipi hosilalari bo'lsa, *pastlik, balandlik, tekislik, chuqurlik* kabi terminlar ikkinchi qolip asosida yuzaga kelgan. Ko'rinadiki, [*ot+lik*] ko'rinishidagi qolipning [*ot*] qismini yer sathi ko'rinishlarini ifodalovchi atov birliklar egallaydi. Ya'ni affiksatsiya usuli bilan hosil qilingan yasama geografik terminlarda yasovchi asos umumleksik birlikdan bo'lib, yasovchi qo'shimcha mavqeida, asosan, *-lik* ot yasovchi qo'shimchasi ishtirok etadi. *-lik* qo'shimchasidan yasama sodda geografik atamalar tuzishda o'rin-joy ko'rinishini ifodalash uchun keng foydalaniladi. Ma'lumki, *-lik* affiksi, asosan, ot, sifat va ravish turkumiga oid so'zlardan ot yasovchi ko'p ma'noli qo'shimchadir. [Nurmonov, Mahmudov 1987]

*-lik* affiksi yasovchi asosdan anglashilgan qismga egalik ma'nosidagi ot yasaydi: *balandlik, pastlik, jarlik, do'nglik, qiyalik, yalanglik* kabi. Bunday leksemalarning ayrimlari *-lik* affiksisiz qo'llanilishi ham mumkin. Masalan, *jarlik, tepalik, do'nglik* kabi leksemalariga xos ma'noda *jar, tepa, do'ng* leksemalarning ishlatilishi kuzatiladi. Lekin *-lik* affiksi qo'llanganda, ko'rsatilgan ma'no aniq ifodalanadi. Qiyoslang: *tepa* – makonning yuqori tomoni (yuqorisi); *tepalik* – er sathining yuqoriga ko'tarilgan qismi. [Hojiyev, 1989]. Yuqoridagi *-lik* affiksi orqali yasalgan yasama so'zda asosdan anglashilgan joy, ko'rinish aniqlik kasb etmoqda. Bu aniqlik yasama so'zning yasalish strukturasi tashkil etuvchilaridan biri bo'lgan *-lik* komponenti ma'nosi yordamida yuzaga chiqadi.

Yasama geografik terminlar tarkibida yasovchi asos, asosan, umumiste'moldagi leksemalar hisoblanadi, ba'zan *baland, tekis, past* kabi sifat turkumidagi birliklar ham ot mavqeyida kelib, polifunksionallik belgisi orqali yer sathi ko'rinishini ifodalovchi geografik terminlar hosil qilishda qatnashadi. Masalan, *balandlik, tekislik, pastlik* каби.

Geografik terminlarning tuzilish jihatidan qo'shma hisoblanadigan tipdagi terminlar tarkibida yer sathi ko'rinishlarini ifodalovchi umumiste'mol leksemaning miqdori ancha. Xususan, *olatog'*, *supatog'*, *yonbag'ir*, *yonartog'*, *yassi jar*, *do'ngqum* kabi. Bu kabi geografik terminlarning yasamaligini ularning morfologik tarkibini kuzatish orqali quyidagicha izohlash mumkin.

Qo'shma yasama geografik terminlar tarkibi, asosan, ikki o'zakli bo'lib, alohida ma'nolariga ega bo'lgan birliklar birikuvidan hosil qilingan. Ikki birlikning shunchaki birikishidan yuzaga kelgandek ko'rinadigan bu kabi qo'shma leksemalarning yasalishida o'ziga xos so'z yasalish strukturasi belgilash mumkin. *Olatog'*, *supatog'*, *yetimtog'*, *yonar tog'*, *yassi tog'*, *beltog'* kabi qo'shma geografik leksemalar kompozitsiya usuli bilan hosil qilingan yasama terminlar hisoblanadi. Aynan *tog'* leksemasi kompozitsiya usuli bilan yasalgan bunday tipdagi qo'shma leksemalarning bir nechasi tarkibida ishtirok etmoqda. Lekin, shunga ko'ra, mazkur leksemalarning so'z yasalish strukturasi yasovchi asos va yasovchi shakl komponentlariga ajratish mushkullik tug'diradi. Bu borada akademik A.Hojiyev tomonidan ham munosabat bildirilgan. Bunday qo'shma leksemalarning tarkibidagi *tog'* komponenti so'z yasalish asosi yoki yasovchi shakl vazifasida bo'la olmaydi. Masalan, *olatog'*, *supatog'*, *yetimtog'*, *yonartog'* qo'shma leksemalari *tog'*ning turli belgilari asosida hosil qilingan. Shu tufayli ham mazkur qo'shma leksemalar ma'nosini komponentlari ma'nosi bilangina asoslab bo'lmaydi. Bunday qo'shma leksemalarning ma'nosi yana boshqa leksemalar bilan birga asoslanadi: *olatog'* - ko'p yillik qor va muzliklari bo'lgan *tog'*; *yetimtog'* - *tog'* tizmasidan ajralgan holda alohida joylashgan yolg'iz *tog'*; *yonartog'* - so'nmagan vulkani bor *tog'*.

### **Xulosa**

Zamonaviy tendensiya va yondashuvlar termin va unga asos bo'lgan umumleksik birlik munosabatini yangicha aspekt va metodlarda o'rganishni taqozo qiladi. Termin va leksema o'zaro semantik jihatdan umumiy semalarga ega bo'lsa-da, ularning semantik strukturasi jiddiy farqli belgilar mavjud, ularning umumleksik va terminologik mazmunini baholash, leksik tizimdagi o'rni va ahamiyatini ochib berish orqali leksik birlikning lisoniy qiymatini aniqlash imkoniyati qo'lga kiritiladi. O'zbek tilining umumiste'moldagi relyef otlari tizimi o'zbek tili geografik terminosistemasining birlamchi boyish manbayi hisoblanadi.



Terminlarning semantik strukturasi, yasalish tizimi, morfemik tarkibi terminga asos bo'luvchi umumleksik birlik asosida talqin qilinsa, ularning o'zaro umumiy va farqli xususiyatlari aniq xarakter kasb etadi, pirovard natijada terminologik va umumleksik birlikka oid leksikografik talqin hamda tavsifning mukammallashuvini ta'minlovchi asosiy omil bo'lib xizmat qiladi.

**Adabiyotlar:**

- Dadaboyev, H. 2020. *O'zbek terminologiyasi*. O'quv qo'llanma. Yoshlar nashriyot uyi, Toshkent: 136 b.
- Hojiyev, A. 2010. *O'zbek tili morfologiyasi, morfemikasi va so'z yasalishining nazariy masalalari*. Toshkent: Fan.
- Hojiyev, A. *O'zbek tili so'z yasalishi*. 1989. Oliy o'quv yurtlarining filologiya fakultetlari studentlari uchun qo'llanma. Toshkent: O'qituvchi.
- Nurmonov, A., Mahmudov, N. 1987. "*O'zbek tili so'z yasalishiga bir nazar*". *O'zbek tili va adabiyoti* 5. 22 - 24.

## **Some semantic and grammatic characteristics of geographical terms**

Ikrom Islomov<sup>1</sup>

**Abstract:**

This article discusses some considerations regarding the relationship between a term and the general lexical unit on which it is based. The general seme in the semantic structure of the term and lexeme in relation to the derivative stem, the distinctive seme, the main seme, the system and morphemic structure of Uzbek geographical terms, and the grammatical features of the morphological structure are also analyzed. based on some examples. It was noted that the results of a systematic study of the term and the general lexical unit will serve to improve their lexicographic interpretations.

**Key words:** *term, terminological system, geographical term, lexeme, relief name, semantic structure, seme, basic seme, artificial term, primitive term, constructive basis, constructive tool.*

---

<sup>1</sup>*Islomov Ikrom Xushboqovich* – Doctor of Philological sciences, Associate Professor of the Department of Applied Philology Tashkent State University of Uzbek Language and Literature named after Alisher Navoi.

**E-mail:** [islomovikrom@navoiy-uni.uz](mailto:islomovikrom@navoiy-uni.uz)

**ORCID:** 0000-0002-5011-1572

**For reference:** Islomov, I. 2022. "Some semantic and grammatic characteristics of geographical terms". *Uzbekistan: language and culture. Applied philology*. 2 (5): 4-10.

**References**

- Dadaboyev, H. 2020. *O'zbek terminologiyasi*. O'quv qo'llanma. Yoshlar nashriyot uyi, Toshkent: 136 b.
- Hojiyev, A. 2010. *O'zbek tili morfologiyasi, morfemikasi va so'z yasalishining nazariy masalalari*. Toshkent: Fan.
- Hojiyev, A. *O'zbek tili so'z yasalishi*. 1989. Oliy o'quv yurtlarining filologiya fakultetlari studentlari uchun qo'llanma. Toshkent: O'qituvchi.
- Nurmonov, A., Mahmudov, N. 1987. "*O'zbek tili so'z yasalishiga bir nazar*". *O'zbek tili va adabiyoti* 5. 22 - 24.

## Methods of text analysis in literary reading lessons (example of the works of Abdulla Kakhhor)

Shoira Isayeva<sup>1</sup>

*“You have fire in your soul, spread your fire,  
burn wet and dry together”  
(Abdulla Kakhhor)*

### Abstract:

The works of literature are analyzed from a variety of perspectives in the methodology for teaching literature. It is imperative that such approaches are carried out in accordance with the purpose of the lesson. A number of methods are available for analyzing the development of events in the works, analyzing their content and form, studying the writer's ideas and thoughts through their expression in the work, and drawing a conclusion. This article analyzes the ideological and artistic aspects of Abdulla Kakhhor's stories. In addition, attempts have been made to portray the tragedy of the period in the writer's style and narratives. Abdulla Kakhhor is one of the writers who vividly reflected the tragedy of time, the suffering of people, social problems in all his works and left an indelible mark on our literature with his works of genius. In the classes, the ideological and artistic content of the writer's works was determined through plotting and analysis of images.

**Key words:** *poetics, tragedy, aspect, sociality, style of writer, horizon.*

### Introduction

Abdulla Kakhhor is a real writer, a talented storyteller who has a place in life and art, and who won the hearts of people with his unique talent and sharp pen. He is one of the most influential figures in Uzbek literature with his inexhaustible works, unique artistic style and ability to express many meanings in a few words. In his works, the author was able to reveal the truth of life, sometimes

---

<sup>1</sup>*Isayeva Shoira Boltaboyevna* – candidate of philological sciences, assistant professor of the department of Uzbek language education in foreign language groups, Tashkent State University of Uzbek Language and Literature named after Alisher Navoi.

**E-mail:** shoiraisayeva@navoi-uni.uz

**ORCID:** 0000-0003-1535-2108

**For reference:** Isayeva Sh. B.2022. “Methods of text analysis in literary reading lessons (example of the works of Abdulla Kakhhor)”. *Uzbekistan: language and culture. Applied philology.* 2 (5): 11-16.

with irony, sometimes with laughter, and it was not for nothing that he was recognized as the “King of Fairy Tales”. In recent years, a lot of creative work has been done to reveal the creative path of Abdulla Kakhhor and the essence of his works. In particular, Naim Karimov in his work “Landscapes of the 20th century literature” gave information about the life and work of A.Kakhhor, which we do not know. Literary critic Rahmon Kochkor prepared the author’s program “Astonishment”, dedicated to the life of the writer. Also, young artists such as Umarali Normatov, Ibrahim Hakkul, Markhabo Kochkarova studied the life and work of Abdulla Kakhhor. The reason for the interest in the unparalleled creativity and life of the author is his courage. What courage can you say?!

We all know that the author’s work was carried out under the pressure of an authoritarian regime that opposed any news. Abdulla Kakhhor was not afraid to reveal the plight of the oppressed people in his works at a time when the so-called Soviet poets of that time were singers of truth and truth, and at the same time persecuted writers. He sought to bring people out of spiritual poverty and depravity and warn them of the dark days of colonialism. As the author himself put it: “Better a poor horse than no horse.” This bitter truth is the truth of the past. The writer managed to convey this to his contemporaries and to us, generations, not simply, but through his satirical and humorous works. The condition of our people, the oppression to which our country is subjected, corruption and deception of the government, whose intentions boil down only to robbing the people, are clearly demonstrated in their masterpieces. In a word, Abdulla Kakhhor is a translator of the national language.

### **Main part**

We want to prove this by analyzing several of the author’s stories throughout the article. First of all, let’s talk about the story of “The Thief”. In the story, we become witnesses of the adversity of the time in which the writer lived and worked, the fact that the spiritually poor were a puppet in the hands of the colonialists. The robbery of the house of Cain, the protagonist of the story, depicts the actions of dishonest, corrupt officials. This is evidenced by the epigraph to the story “Death of a Horse, Dog Feast”. According to the plot, the cow of Cain’s grandfather was kidnapped. Here the author quotes a proverb: “Let the house of the farmer burn down, and the ox will not be lost. “Only this proverb can tell us about the financial situation of people of that time. Reading this story, we see that Cain’s

grandfather was robbed again, and Cain was the first to rob him by making fifty promises to his grandfather. When fifty did not work, Cain ran to his grandfather. Through Amir and his speech, our writer described the officials of that time who were interested only in their own interests and were indifferent to the plight of the people. Let's take a look at the next sentence in history:

He will return? Would someone volunteer to come up with a beautiful template? Why are you crying? Do not Cry! Cain looked at the ground and spoke.

Already in this proposal, we see that our ordinary people have always been a plaything in the hands of corrupt officials. Throughout history, Cain is also robbed by his grandfather's Tilmoh and Prestav. Obviously, the thief is not the one who stole the bull, but the real thief of the tyrannical government. This is how the then government deceived the people. It is no coincidence that at the end of the story, the bull's father-in-law gave him two oxen. Cain tells his grandfather that there is a "small" condition in return, and that it will become known in the fall. This one sentence is a sign that Cain is still being robbed. In the stories of Abdulla Kakhhor, the past is directly condemned. Speaking about the tragic days of the past, one cannot but dwell on the author's story "Horror". In this story, the author reflects the atmosphere of that time, the spiritual backwardness of people, the tragedy of the time, that is, horror. We see the unprecedented skill of Abdulla Kakhhor from the very beginning of the story. In the epigraph to the story, we see proof of this: "Women who do not know the day when women saw him in the past do not believe what they say." Indeed, in the past, women had no rights and were looked down upon. The fact that Unsin's father, the protagonist of the story, was married to the same man as the eighth wife was a nightmare in itself. The disgusting atmosphere of that time made the poor do it. Our people suffered a lot during the Soviet era, and it was financial difficulties that forced them to do what they did not want to do. We can see this in the story of how young Unsin was given to Olimbek Dodho, an old and humble man, in exchange for his father's debts. For a young girl, life in the dojo house was hell. Because in this story, if the dodho blames one wife, he hits them all the same. Tired of living in such an environment, the depressed never tired of going to the cemetery at night to return home and return for tea. Because for this young girl, the horror of the cemetery was nothing compared to the horror of the dodho house. In that moment, he was able to do anything to escape the terror of the dodho house and live a free,

free life. At the end of the story, Unsin performs this terrible task, but his heart cannot bear this horror. This is a worthy act, and this should be the end of it. But the spirit goes free. As for the analysis of A. Kakhhor's story "The Patient", the story reflects the invisible shortcomings of a sick woman in the whole society. Thus, through the events of the story, the author reveals that the original patient is society. It's as if the writer was thinking about our society through storytelling. The story skillfully describes the suffering of our people from poverty, the fact that they could not even find a loan to treat a sick woman, and the idea of a hospital with "money depicting a white padsha".

In an interview with *Mushtum* magazine, Abdulla Kakhhor said: He has destroyed the art, the most important educational tool". It makes every writer feel proud and proud. Because it's not for nothing, it's up to each writer to understand. Therefore, the writer cannot explain his feelings in any way. He tries to explain the sincere words of his heart to young readers and all his fans through his works. This requires great courage, contentment and patience from the writer. Abdulla Kakhhor was such a brave, contented and, of course, patient man.

In today's process of globalization and integration, there is no need to explain how important it is for each nation to know its past, the heritage of its ancestors, and the history of the formation of the nation. At the same time, the works of art that play an important role in shaping the worldview of today's man, in particular, the works that reflect the way of life, thoughts and aspirations of our people in a particular period, are especially invaluable. By reading and studying such works, we can better understand the spiritual values and changes in the spiritual worldview of our people. From this point of view, if we talk about the works of Abdulla Kakhhor, a writer who created his own great school of literature in the literature of the twentieth century, we will be able to put forward a very pure truth. Abdulla Kakhhor's stories, first and foremost, amaze with their sincerity and persuasive power. Everything in the author's image is a life event, an event that happened, a part of real life, an episode; most of the stories are based on real life, they are taken from the events that the writer saw and heard in his life, from the lives of acquaintances. But they are not exactly a copy of life. "If writing was about copying from life, there would be no easier job in the world," he said. Copying from life is like copying from a book. Copies will remain. You can't expect originality from such things. Originality

comes from experiencing the realities of life, feeling them, absorbing what you are thinking, and expressing your desires”.

Abdulla Kakhhor was also an effective translator. He skillfully translated the centuries of Pushkin, Tolstoy, Gogol and Chekhov into Uzbek. In particular, Chekhov translated his works with special interest and experience. Of course, the writer's services and hard work paid off. In 1966, he was awarded the Hamza State Prize, in 1967 the People's Writer of Uzbekistan, and in 2000 the Order of Merit. Abdulla Kakhhor is named after several streets, schools and collective farms in Tashkent and Kokand, as well as houses of culture and the Republican Satire Theater. In 1987, the Abdulla Kakhhor House-Museum was opened in Tashkent. His works have been translated into Russian, Kazakh, Ukrainian, Belarusian, Kyrgyz, Karakalpak and Tajik languages. Some of his stories have been published in English, German, French, Czech, Polish, Bulgarian, Romanian, Arabic, Hindi and Vietnamese.

### **Conclusion**

In conclusion, we can say that Abdulla Kakhhor is one of the writer who vividly reflected the tragedy of time, the suffering of people, social problems in all his works and left an indelible mark on our literature with his works of genius. In his works, writer openly described the situation of people. With his works, he strove to bring people oppressed by Soviet oppression out of the spiritual quagmire. We can see this from the stories analyzed above. In the following articles, we will focus on the work of Abdulla Kakhhor and many other writers.

### **References:**

- Karimov N. 1999. *Landscapes of 20th century literature*. Tashkent: Academic publication, p 580.
- Sharafiddinov O. 1988. *Abdulla Kahor*. Tashkent: p 326.
- Koshjanov M. 1984. *Abdulla Kahor skills*. Tashkent: Publishing House of Literature and Art named after Gafur Gulom, p 322.
- Abdulla Kahor. 1987–1989. *Works (5 volumes)*. Tashkent: Publishing House of Literature and Art named after Gafur Gulom, p 250.
- Mamiraliyev Q. *Some reviews on the mutation of genres in uzbek poetry*. International journal for innovative engineering and management research. ISSN 2456 – 5083. Vol 10 Issue03, Feb 2021. [www.ijiemr.org](http://www.ijiemr.org)

# Adabiy o'qish mashg'ulotlarida matnni tahlil qilish usullari (Abdulla Qahhor asarlari misolida)

Shoira Isayeva<sup>1</sup>

## Abstrakt:

Adabiyot o'qitish metodikasida asarlar turli nuqtayi nazardan tahlilga tortiladi. Bunday yondashuvlar darsning maqsadiga ko'ra amalga oshiriladi. Adabiy o'qish mashg'ulotlarida asarlardagi voqealar rivojini o'rganuvchi, asar mazmuni va shakliga ko'ra tahlil qiluvchi, asardagi ifoda vositalari orqali yozuvchining g'oyasini, fikrlarini o'rganuvchi va tahliliy xulosalar beruvchi bir necha turli usullar mavjud. Ushbu maqolada Abdulla Qahhor qissalarining g'oyaviy-badiiy jihatlari tahlil qilinadi. Yozuvchi uslubi va hikoyalarida davr fojiasini tasvirlashga harakat qilingan. Abdulla Qahhor o'zining barcha asarlarida zamon fojiasi, odamlar dardi, ijtimoiy muammolarni yorqin aks ettirgan, daho asarlari bilan adabiyotimizda o'chmas iz qoldirgan adiblardan biridir. Mashg'ulotlarda syujet qurilishi va obrazlar tahlili orqali yozuvchi asarlarining g'oyaviy-badiiy mazmunining aniqlanishi tahlil qilindi.

**Kalit so'zlar:** *poetika, tragediya, aspekt, ijtimoiylik, yozuvchi uslubi, xarakter chegarasi.*

## Adabiyotlar:

- Karimov N. 1999. *XX asr adabiyoti manzaralari*. Toshkent: Akademnashr, 580 b.
- Sharafiddinov O. 1988. *Abdulla Qahhor*. Toshkent: 326 b.
- Qo'shjonov M. 1984. *Abdulla Qahhor mahorati*. Toshkent, G'afur G'ulom nomidagi Adabiyot va san'at nashriyoti, 322 b.
- Abdulla Qahhor. 1987–1989. *Asarlar* (5 tomlik). Toshkent: G'afur G'ulom nomidagi Adabiyot va san'at nashriyoti, 250 b.
- Mamiraliyev Q. *Some reviews on the mutation of genres in uzbek poetry*. International journal for innovative engineering and management research. ISSN 2456 – 5083. Vol 10 Issue03, Feb 2021. [www.ijiemr.org](http://www.ijiemr.org)

---

<sup>1</sup>*Isayeva Shoira Boltaboyevna* – filologiya fanlari nomzodi, Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti O'zga tilli guruhlarda o'zbek tili ta'limi kafedrasi dotsenti.

**E-pochta:** [shoiraisayeva@navoi-uni.uz](mailto:shoiraisayeva@navoi-uni.uz)

**ORCID:** 0000-0003-1535-2108

**Iqtibos uchun:** Isayeva Sh. B.2022. "Adabiy o'qish mashg'ulotlarida matnni tahlil qilish usullari (Abdulla Qahhor asarlari misolida)". *O'zbekiston: til va madaniyat. Amaliy filologiya*. 2 (5): 11-16.



**O‘ZBEK TILINI XORIJIY TIL SIFATIDA O‘QITISH****O‘zbek tilini xorijiy til sifatida o‘qitishda o‘qituvchi shaxsi va kreativ yondashuvi**Xurshida Kadirova<sup>1</sup>**Abstrakt:**

Til ta’limida o‘qituvchining shaxs sifatidagi roli eng muhim omil sanaladi. Maqolada o‘qituvchi shaxsining so‘nib qolish sabablari va undan ozod bo‘lish usullari bayon qilingan. O‘qituvchining ruhiy holati dars samaradorligiga qanchalik ta’sir qilsa, uning obro‘si va professional ixlosini shunchalik tushurishi tahlilga tortilgan. Shuningdek, baxtli o‘qituvchilarning belgilari sanab o‘tilgan. Unga hayotiy misollar keltirilgan. O‘qituvchining aktyor bo‘lishi kerakligi haqidagi tushunchaga fikr bildirib, samimiyligni saqlash muhimligi ta’kidlangan.

**Kalit so‘zlar:** *O‘qituvchining toliqishi, aktyor, jamoaviy ish, baxtli o‘qituvchilar, ta’lim falsafasi.*

**Kirish**

Bugungi kunda o‘zbek tiliga bo‘lgan qiziqish ortib borayotganligini hisobga olgan holda uni boshqa millat vakiliga tilimizning go‘zal taraflarini ko‘rsatish, til ta’limni qulaylashtirish siyosat darajasiga ko‘tarilar ekan, buning ustida ishlashni davr taqozo qilmoqda. Bunda o‘qituvchining roli, til o‘rganuvchi yoki o‘zga tilli talabaning roli haqidagi muayyan nazariy va amaliy bilimlar muhim ahamiyat kasb etadi. So‘ngi 20 yil ichida juda katta muammo yuzaga kelmoqda. Yaxshi o‘qituvchilar, buyuk o‘qituvchilar o‘z kasbini tark etayotganligining guvohi bo‘lyapmiz. Gap shundaki, vijdonan mehnat qilgan o‘qituvchilarni bu kasb juda toliqtiradi. Ammo olingan bilimlardan foydalanish muhimligini va talabalarning tilni mashq qilishi zarurligini anglagan o‘qituvchilar bu jarayonga jalb qilinishi, rag‘batlantirilishi kerak.

<sup>1</sup>Kadirova Xurshida Batirovna – filologiya fanlari nomzodi, Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti O‘zga tilli guruhlarda o‘zbek tili ta’limi kafedrasida dotsenti.

**E-mail:** qodirova@navoiy-uni.uz

**ORCID:** 0000-0001-5194-7155

**Iqtibos uchun:** Kadirova, X. 2022. “O‘zbek tilini xorijiy til sifatida o‘qitishda o‘qituvchi shaxsi va kreativ yondashuvi”. *O‘zbekiston: til va madaniyat. Amaliy filologiya.* 2(5): 17-25.

O'qituvchi aktyor, murabbiy, psixolog, gid, yo'l boshlovchi va hokazo. Lekin o'qituvchi bu haqda o'ylashdan allaqachon charchagan! Kasbiy charchash degani, o'qituvchilarning yonayotgan sham kabi, bor nurini berishlari, hech narsa qolmaguncha bor kuchlarini berishlari demakdir. Oxirgi 3 yildan 5 yoki 10 yilgacha ishlagan va o'qituvchilikni tashlab ketgan ko'plab o'qituvchilarni ko'rishimiz mumkin. Ular avtomobil savdosi, qurilish... o'qituvchilikdan boshqa hamma narsa bilan shug'ullanishga rozi. Nima uchun bu sodir bo'lyapti?

Albatta, ta'limda har doim maoshdan norozilik, qadrsizlik va hokazo mavjud. Lekin asosiy sabablardan biri kasbning o'zi ekanligini taxmin qilish mumkin. Tilni o'rgatish shunchalik qiyin va, shu bilan birga shunchalik qiziqarliki, undan voz kechish ham mushkul.

### Asosiy qism

Talabalarga o'zbek tilida muloqot qilishda yordam berishga intilar ekan, o'qituvchi talaffuz bo'yicha, o'qish, yozish, grammatika, nutq va hokazolar bo'yicha mutaxassis bo'lishi kerakligini anglaydi. Uchuvchilar, hamshiralalar, boshlang'ich sinf o'quvchilari, o'rta maktab o'quvchilari, kollej talabalari, mutaxassislar va ishbilarmonlarga tilni o'rgatishi mumkin. Va har yili yangi darsliklar, yangi standartlar paydo bo'ladi, jadval yangilanadi, ta'lim jarayoniga qarashlar va yondashuvlar o'zgaradi. O'qituvchilik sohasi doimo rivojlanib bormoqda, o'qituvchi o'zini karuselda o'tirgandek his qiladi va u tezroq va tezroq aylanadi. Qanday qilib professional toliqishdan qochish mumkin? Quyida mazkur masalani biroz yengillatadigan ba'zi omillarga to'xtalib o'tamiz.

**O'qitish – bu jamoaviy ish.** “O'qituvchining toliqishi” ko'plab topshiriqlar, mas'uliyat va doimiy tazyiqlardan kelib chiqqan tushuncha deb atash mumkin. O'qituvchini charchatadigan, limondek siqib suvini chiqaradigan kasbni tanlashga hech kim majbur qilmaganligini va endilikda mashaqqatni faqat o'zi uchun qulay qilishga intilish, kasbidan rohatlanish usullarini yaratish kerakligini ta'kidlash zarur. Kuch olish va ilhomlanishning ko'plab usullari mavjud.

Aksariyat yosh o'qituvchilarning shikoyatlariga duch kelamiz. Aslini olganda o'qituvchilik – bu **jamoaviy ish**. Barcha o'qituvchilar haddan tashqari ish bilan yuklangan. Biroq o'nlab yillar davomida ishlagan tajribali o'qituvchilar ham oz emas, ularning ko'plab tayyor dars rejalari va ajoyib g'oyalari bor. Tushkunlik, sabrsizlik muvaffaqiyatsizlikka olib kelmasdan avval shu tajribali

o'qituvchilardan yordam so'rashlari kerak. Zero, bunday o'qituvchilar savollarga mamnuniyat bilan javob berishadi. Chunki, ular savollarga javob berishni yaxshi ko'radilar. Masalan, hafta kunlarini talabalarga qanday o'qitish haqida kimning fikri bor? Biror o'qituvchi shu mavzuda qo'shiq berishi mumkin. Boshqa bir o'qituvchi haftaning ma'lum bir kunini ifodalovchi rangli qog'ozdan foydalanishini o'rtoqlashadi.

Yana misol: oddiy o'tgan zamonnini qanday qilib samaraliroq o'rgatish kerakligini so'rash mumkin. Tabiiyki, gapirishni, o'rgatishni, savollarga javob berishni xush ko'ruvchi bunday shaxslar yordam qo'lini cho'zadi. Biror o'qituvchi metodika bilan o'rtoqlashadi, boshqasi darslikdagi kerakli sahifani ko'rsatadi, uchinchisi o'zining kitob javonidan kerakli papkani topadi va auditoriyada ishlash uchun topshiriqlarni to'kib, ichidan tanlab beradi. O'qituvchi hamkasblari yordami bilan ko'p vaqtni tejashi mumkin. Qolaversa, har bir o'qituvchining fikrini inobatga olgan holda muammoni turli tomonlardan ko'rish imkonini paydo bo'ladi, bu bebaho tajriba. Ko'p o'tmay, kimningdir dialog topshiriqlarini bajarishda yordam berishini va grammatika uchun yana biror hamkasbiga murojaat qilishni bilib olish mumkin, ammo o'qitish yo'nalishi bundan ham ko'proq. Bunda til o'qituvchilari dunyodagi eng ochiq odamlar ekanligini bilib olish mumkin. Fikr almashishni, muhim savollarni berishni, murakkab muammolarga yechim izlashni o'qituvchi yaxshi ko'radi. Tajribali mutaxassislar bilan ijtimoiy tarmoqni qurish o'qituvchining o'ziga ham tajribali va ilg'or o'qituvchi bo'lishiga imkon beradi. Va bir xil maxsus bilimga ega bo'lgani uchun emas, balki u yoki bu savolni kimga berishni aniq bilib olish mumkin.

Bu Genri Fordning hikoyasini eslatadi, u aqlli bo'lish uchun hamma narsani bilish shart emas, balki u yoki bu ma'lumotni qayerdan olishni bilishdir, deb hisoblaydi. Uning so'zlariga ko'ra, pulda bir qator tugmalar bor va o'ng tugmani bosish orqali har qanday savolga javob olishi mumkin. Savollarga javob topish uchun turli xil umumiy ma'lumotlar bilan vaqt sarflash kerak emas, agar kerakli bilim va ma'lumotlarni taqdim etadigan mutaxassislar bo'lsa, boshqa narsalarga chalg'ish kerak emas.

Muvaffaqiyatli o'qituvchi har qanday savolga javob topa oladi, chunki u kimdan nimani so'rashni va javobni qayerdan qidirishni biladi. O'qituvchi sifatidagi faoliyati shu tarzda osonlashadi. Muhimi, yordam beruvchilar tajribaliroq bo'lishi, istalgan vaqtda o'z fikrlari bilan bo'lishishga tayyor odamlar bo'lishi kerak.

**Baxtli o'qituvchi.** Ikki tadqiqotchi Ellen Skinner va Ketlin Edj [Ellen Skinner] o'z-o'zini anglash bo'yicha kitob yozdilar. Va ular odamlarga toliqishdan qochishga yordam beradigan ikkita omil mavjudligini aniqladilar. Biz ularning birinchisi, ya'ni jamoalar tashkil etish yoki ijtimoiy qo'llab-quvvatlovchi yoki professional hamjamiyat deb ataladigan narsa haqida qayd etdik. Endi ikkinchisiga to'xtalamiz.

Aytishlaricha, toliqmaslik uchun o'qituvchilar o'z-o'zini nazorat qilish tuyg'usiga ega bo'lishi kerak. Nima uchun ba'zi o'qituvchilar o'zlarini nazorat qilinayotgandek his qilishadi, boshqalari esa yo'q?

Shubhasiz, o'qituvchida ba'zi narsalarni nazorat qilish mumkin bo'lgan va ba'zi narsalarni nazorat qilish mumkin bo'lmagan narsalar bor. O'qituvchi istalgan makon va zamonda dars berishi muqarrar. Nazorat tuyg'usini saqlab qolish uchun ko'pincha qanchalik nazoratga ega ekanligi yoki qayerda ekanligini emas, balki mavjud narsalar bilan qanday dars o'tish haqida o'ylash muhim.

Bunda o'qituvchilardan ijodkorlik talab etiladi. Baxtli o'qituvchilarning quyidagi jihatlari bo'ladi:

**Birinchiidan**, ular imkonsiz narsalarga emas, balki qila oladigan narsalarga e'tibor berishadi. O'qituvchilar ko'p vaqtni berilgan erkinlik miqdoridan bahramand bo'lish uchun o'tkazishi kerak. Hatto materiallar, kitoblar, testlar yoki maosh haqida o'ylamasa ham bo'ladi. O'qituvchi o'zi uchun yarata oladigan ko'p erkinliklar bor. O'zi yoqtirgan o'rgatish uslubini yaratishi, istalgan tajribani o'tkazishi, o'qitish imkoniyatlaridan o'zi rohatlanishi kerak. O'qituvchi, o'ziga yuklangan turli xil nazorat vositalariga qaramay, o'z auditoriyasida o'zi yetakchidir. Unda ajoyib imkoniyat borligi va boshqalarga ta'sir o'tkazish uchun erkin ekanligini his qilishi kerak. O'zlari qila olmaydigan narsalarga e'tibor qaratishga moyil bo'lgan o'qituvchilar odatda eng baxtsizdir. Va o'zlari qila oladigan narsalarga e'tibor qaratadigan o'qituvchilar ko'pincha eng ijodiy va eng qiziqarli bo'ladi.

**Ikkinchiidan**, baxtli o'qituvchilar o'zgarishlarni tez qabul qiluvchi bo'ladi. Aslini olganda, ba'zan o'qituvchi istagan yoki qilishingiz zarur bo'lgan ko'p narsalarni amalga oshirish uchun resurslar yetarli bo'lmaydi va bu ayanchli. Ko'pgina mamlakatlarda ba'zi o'qituvchilar darsliklar va yangi texnologiyalar kabi yetarli resurslarga ega emasligidan aziyat chekishadi. Ayrim chekka hududlarda maktablarning tomi bo'lmay, har safar yog'ingarchilik bo'lgani tufayli maktab yopilishi va belgilangan dastur o'zlashtirilmay

qolishi tufayli o'qituvchi maqdasiga erishmagani, o'z imkoniyatlarini ko'rsata olmagan uchun boshqa kasbni, balki o'zi xohlamagan ishni tanlashiga olib keladi. Ayrimlar esa shunchaki proyektor ishlamayotgani yoki internetga kirish imkoni yo'qligidan kuyinishadi. Bu esa dars samarasining tushib ketishiga sabab bo'ladi. Kafedradagi o'qituvchilarning darsdan oldin printer yoki kserokopiyaning buzilib qolgandagi stressini tasavvur qilib bo'lmaydi. Bularning barchasi o'qituvchini shashtini qaytarishi va o'rgatish qobiliyatiga ta'sir qilishi mumkin. Ammo, agar ayni damda shasht qaytgan bo'lsa, doimo shunday bo'ladi degani emas...

Har doim o'zgarishlarga yuzma-yuz kela olish katta kuch talab etadi. O'z-o'zini nazorat qilishni susaytiruvchi narsalar haqida o'ylashning o'rniga, o'zining muhitini iloji boricha xotirjam qilishga intilish kerak.

Bunga o'xshagan misollarni ta'lim sohasida ko'plab keltirish mumkin. Ayniqsa o'zbek o'qituvchilaridek jon fido insonlar borki, o'z maoshlaridan bo'lsa ham o'z muhitni yaratadi. Atrof-muhit ta'siriga e'tibor qaratmasdan, uni yaxshilashga harakat qilish orqali o'qituvchi optimist va ko'proq o'z-o'zini nazorat qilish tuyg'usiga ega bo'ladi.

***Uchinidan***, baxtli o'qituvchilar muvozanatni topadilar. Yana bir muammo borki, ba'zan til o'qituvchilarini, doimo o'zlarini o'z kasbiga butunlay baxshida qiluvchilar sifatida ko'rish mumkin. Ular uy vazifalarini tekshirish uchun bir necha kun vaqt sarflashadi. Har bir talaba uchun alohida izoh yozadigan va kechgacha dars rejalarini ko'rib chiqadigan o'qituvchilarni bilamiz. O'qituvchilar o'zlarini butunlay kasbga bag'ishlashsa juda yaxshi. Ammo, buning salbiy tomoni bor. Ayniqsa, har bir ish varag'ini takomillashtirish yoki ulkan afishadagidek kalit so'zini bo'yash uchun son-sanoqsiz soat sarflaydiganlar uchun.

O'qituvchi, avvalo, odam. Inson sifatida o'z qadrini, o'z sog'lig'i va jonini o'ylashga uning odam ekanligini eslatib turish uchun kimdir bo'lishi shart emas. Bolalari bilan muzqaymoq yeyish, turmush o'rtog'i bilan kinoga borish, do'stiga qo'ng'iroq qilish, sevimli mashg'uloti bilan shug'ullanishga haqli bo'lgan odam u. Dunyodagi eng yaxshi o'qituvchilar o'z qalblarini talabalarga berishadi, lekin shu bilan birga o'z hayotlarini ham yaxshi o'tkazadilar va o'zlarini qiziqarli odamlarga aylantiradilar. Sevimli mashg'ulotga ega bo'lish, qandaydir mashg'ulotdan zavqlanish o'qituvchini quvvatlantiradi va doimiy shikoyatlardan xalos etadi.

**To'rtinchidan**, baxtli o'qituvchilar o'qitishning moddiy manfaatini qadrlashadi. O'qituvchilik kasbi ko'pincha katta maoshni kafolatlamaydi. Ammo bu uning mukofotsiz qolganini anglatmaydi. Ba'zan mukofotlar, bonuslar bo'ladi. Talabalarining qozongan muvaffaqiyatlarga e'tibor berilsa, ularning hayotiga ijobiy zavq beradi, ilhomlantiradi. Oxir-oqibat, pul tufayli o'qituvchi bo'lishdan bosh tortganlarga havas qilmay qoladi. Va bu eng muhim narsa. Avvalo, tilning otash ekanligini va har bir o'quvchini olov yoqishga harakat qilayotganligini unutmaslik kerak.

**O'qituvchi artist emas.** O'qituvchi haqidagi metaforalardan biri o'qituvchini aktyor bilan solishtirish bo'lib kelgan. Ammo bu metafora bilan bog'liq muammo shundaki, ko'p jihatdan o'qituvchilarni aktyorlarga qiyoslab bo'lmaydi. O'qituvchilar imo-ishoralarni qo'llashi, butun tanasini, yuzini ishlatishi kerak. Ular o'z o'rganuvchilariga qanday ta'sir qilish haqida o'ylashlari kerak, xuddi aktyor o'z tomoshabinlariga ta'sir qilishi kerak bo'lganidek. Hatto, tuni bilan uy vazifasini tekshirgandan keyin ham o'zlarini yaxshi his qilayotgandek ko'rsatishlari kerak.

Ammo gap shundaki, agar o'qituvchi talabalarga o'zining haqiqiy siyratini ko'rsatmas ekan, o'zi haqida to'g'ri taassurot qoldirishi muammo bo'ladi, bu esa toliqishga to'g'ridan-to'g'ri yo'ldir. Ba'zi o'qituvchilar o'zini juda vazmin tuta bilishadi, ayrimlari jo'shqin va samimiydir. O'qituvchi kim bo'lishingizdan qat'iy nazar, o'z ichki dunyosining eng yaxshisini ko'rsata olishi kerak.

Sham singari o'chib qolmasligi uchun o'qituvchi o'z-o'zini to'g'ri anglashi, o'zini to'g'ri tushunishi kerak. Nima uchun ushbu kasbni tanlaganini, o'zini unda qanday his qilayotganini tushunish juda muhimdir. Yuqorida aytib o'tganimizdek, o'qituvchi o'zining maqsadiga bo'lgan ishonchini ko'rgandagina odamlar unga murojaat qilishadi. Shundagina talabalar, boshqa hamkasblar unga ergashadilar, unga o'xshashga intiladi, havas qiladi. Mabodo asosiy maqsad yo'lida harakat qilinsa, ma'muriy xodim bilan ham kelishish mumkin.

## **Xulosa**

Qanday qilib toliqishdan qutulish mumkin? Rol o'ynash kerak emas, samimiy bo'lish kerak. Buni yangi o'qitish usullarini o'rganishga nisbatan passiv munosabat bilan aralashtirish kerak emas. Albatta, o'qituvchi ikkilanishi, o'z qulaylik hududidan chiqib ketishi va natijada yangi narsalarni o'rganishi kerak. U o'zi ichki mohiyatini topishda o'qituvchi o'zining ta'lim falsafasini tushunishi

va o'zining professional sohasida eng yaxshisi bo'lish uchun qo'ldan kelganini qilishi kerak. Zero, o'qituvchining toliqishiga qarshi ko'rilgan eng samarali usul – bu iloji boricha yorqinroq yonishdir.

Ayni o'rinda o'qituvchi mehnatining niashaqqatli ekanligini ta'kidlab o'tgan Alisher Navoiyning quyidagi fikrini eslashni lozim topdik: “Uning ishi odam qo'ldan kelmas, odam emas, balki dev ham qila bilmas. Bir kuchli kishi bir yosh bolani saqlashga o'zlik qilardi, u esa bir to'da bolaga ilm va adab o'rgatadi, ko'rkim bunga nima yetsin...” [Alisher Navoiy, 1966.]

### **Adabiyotlar:**

Alisher Navoiy. 1966. *Mahjub ul-qulub*. Asarlar. O'n besh tomlik. 13 tom. Toshkent, Badiiy adabiyot nashriyoti, 192-193-betlar.

Kadirova, X.B. 2022. *O'zbek tilini xorijiy til sifatida o'qitishda ta'lim sifatini belgilovchi omillar* // International scientific journal of Biruni. №2. URL: <https://cyberleninka.ru/article/n/o-zbek-tilini-xorijiy-til-sifatida-o-qitishda-ta-lim-sifatini-belgilovchi-omillar>

Parenting, Motivation, and the Development of Children's Coping // Ellen Skinner and Kathleen Edge / [https://books.google.co.uz/books?id=Pa56zLQFealC&pg=PA77&lpg=PA77&dq=Skinner,+Kethlin+Edge&source=bl&ots=9Yumasdqz9&sig=ACfU3U3Nb1UacANac0\\_jllp9Gj90Po3mhw&hl=ru&sa=X#v=onepage&q=Skinner%2C%20Kethlin%20Edge&f=false](https://books.google.co.uz/books?id=Pa56zLQFealC&pg=PA77&lpg=PA77&dq=Skinner,+Kethlin+Edge&source=bl&ots=9Yumasdqz9&sig=ACfU3U3Nb1UacANac0_jllp9Gj90Po3mhw&hl=ru&sa=X#v=onepage&q=Skinner%2C%20Kethlin%20Edge&f=false)

В поисках структуры совладания: обзор и критика систем категорий для классификации способов совладания // Эллиен Скиннер, Кэтлин Эдж, Джеффри Альтман, Хейли Шервуд / DOI: 10.1037/0033-2909.129.2.216;

Размышления о преодолении трудностей и развитии на протяжении всей жизни // Эллиен Скиннер, Кэтлин Эдж / <https://doi.org/10.1080/016502598384414>

Эллиен, Скиннер, Кэтлин Эдж, Джеффри Альтман, Хейли Шервуд // В поисках структуры совладания: обзор и критика систем категорий для классификации способов совладания / DOI: 10.1037/0033-2909.129.2.216;

Эллиен, Скиннер, Кэтлин Эдж // Размышления о преодолении трудностей и развитии на протяжении всей жизни / <https://doi.org/10.1080/016502598384414>

# Personality of the teacher and creative approach in teaching uzbek as a foreign language

Khurshida Kadirova<sup>1</sup>

## Abstract:

The role of the teacher as a person is considered to be the most important factor in language education. The article describes the reasons for the disappearance of the teacher's personality and ways to get rid of it. It is analyzed how much the teacher's mental state affects the effectiveness of the lesson, how much it reduces his reputation and professional sincerity. The signs of happy teachers are also listed. He gives real life examples. Commenting on the idea that the teacher should be an actor, the importance of maintaining sincerity was emphasized.

**Key words:** *teacher burnout, actor, teamwork, happy teachers, pedagogical philosophy.*

## References:

- Alisher Navoi. 1966. Mahbub ul-Qulub. Works. Fifteen roofs. 13 roofs. — Tashkent, Literary Publishing House, pp. 192-193.
- Kadirova, Kh.B. 2022. Factors determining the quality of education in teaching Uzbek as a foreign language // International scientific journal of Biruni. #2. URL: <https://cyberleninka.ru/article/n/o-zbek-tilini-khorijiy-til-sifatida-o-kyutuda-lim-sifatini-begilovchich-millar>
- Parenting, Motivation, and the Development of Children's Coping // Ellen Skinner and Kathleen Edge / [https://books.google.co.uz/books?id=Pa56zLQFeaIC&pg=PA77&lpg=PA77&dq=Skinner,+Kathlin+Edge&source=bl&ots=9YymasdQz9&sig=ACfU3U3Hb1UacANac0\\_jllp9Gj90Po3mhw&hl=ru&sa=X#v=onepage&q=Skinner%2C%20Kathlin%20Edge&f=false](https://books.google.co.uz/books?id=Pa56zLQFeaIC&pg=PA77&lpg=PA77&dq=Skinner,+Kathlin+Edge&source=bl&ots=9YymasdQz9&sig=ACfU3U3Hb1UacANac0_jllp9Gj90Po3mhw&hl=ru&sa=X#v=onepage&q=Skinner%2C%20Kathlin%20Edge&f=false)
- V research structure of governance: overview and cri tique of the system category for classification of ways of governance // Ellen Skinner, Kathleen Edge, Jeffrey Altman, Haley Sherwood / DOI: 10.1037/0033-2909.129.2.216;

<sup>1</sup> Kadirova Khurshida Batirovna – Candidate of Philological Sciences, Alisher Navoi Tashkent State University of Uzbek Language and Literature, Associate Professor of the Department of Uzbek Language Education in Foreign Language Groups.

**E-mail:** [kadirova@navoiy-uni.uz](mailto:kadirova@navoiy-uni.uz)

**ORCID:** 0000-0001-5194-7155

**For reference:** Kadirova, Kh. 2022. "Personality of the teacher and creative approach in teaching uzbek as a foreign language". *Uzbekistan: language and culture. Applied philology*. 2(5): 17-25.



Размышления о преодолении трудностей и развитии на протяжении всей жизни // Ellen Skinner, Kathleen Edge / <https://doi.org/10.1080/016502598384414>

Ellen, Skinner, Kathleen Edge, Jeffrey Altman, Haley Sherwood // V poiskakh struktury sovladania: obzor i kritika sistem kategoriy dlya klassifikatsii sposobov sovladania / DOI: 10.1037/0033-2909.129.2.216;

Ellen, Skinner, Kathleen Edge // Reflections on coping and development throughout life / <https://doi.org/10.1080/016502598384414>

## LINGVOKULTUROLOGIYA

## Роль языковой среды в формировании лингвокультурологических элементов в детской речи

Гульбахор Искандарова<sup>1</sup>

### Аннотация

Статья посвящена влиянию национально-культурного фактора в процессе онтогенеза. В ней анализируются этапы формирования и социализация личности. Наблюдения онтогенетического процесса узбекских и китайских детей показывают, что культурно-общественная среда имеет важную роль в становлении мыслительных реакций ребенка. Проведенный анализ выявляет наиболее основные элементы в речи детей, которые несут в себе национально-культурное значение для носителей языка. Эти особенности выделяются в рамках межкультурной коммуникации.

**Ключевые слова:** социализация, национально-культурное значение, коммуникативный потенциал, речевое поведение, речевая ситуация.

### Введение

Речевое поведение человека проявляется в двух основных внешних факторах — его действиях и речи, общении с другими людьми. Посредством речевого тона, произношения, выбора слов, обогащения вербальной речи невербальными средствами, т.е. движением головы, рук и тела, передается определенная информация о психике, настроении, желании общаться, культурном уровне и общительности говорящего.

В мировой науке XXI века наблюдается возникновение нового антропоцентрического подхода к фактору человеческой речи. Это — перевод исследовательского внимания от объекта субъекту, т.е. человек привлекается к анализу внутри языка, а

---

<sup>1</sup>Искандарова Гульбахор Турабаевна – Старший преподаватель кафедры Преподавания узбекского языка в иноязычных группах Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои

**Э-почта:** iskandarovagulbahor@navoiy-uni.uz

**ORCID:** 0000-0002-4189-2725

**Для цитаты:** Искандарова, Г. Т. 2022. “Роль языковой среды в формировании лингвокультурологических элементов в детской речи”. *O'zbekiston: til va madaniyat. Amaliy filologiya*. 2(5): 26-34.

язык — внутри человека. Выражаясь словами И. А. Бодуэна де Куртенэ, «язык существует только в головах, только в сердцах, только в психике индивидов и личностей, составляющих определенную лингвистическую общность» [Маслова, 2001; С. 6]. В рамках данной парадигмы в лингвистике сформировались такие направления, как когнитивное языкознание и лингвокультурология. По мнению В. А. Масловой, лингвокультурология изучает язык в качестве культурного феномена. Это, по сути, своеобразное видение и восприятие мира сквозь призму языка, при этом язык является зеркалом, отображающим отдельную «национальную» ментальность. Как пишет В. Н. Телия, «в языке следует опираться на культурный фактор, в человеке — на лингвистический» [Телия, 1996; С. 271].

### **Основная часть**

С усвоением речи человек превращается в члена конкретного общества, т.е. социализируется. Социализация личности — это приобретение человеком в биологическом понимании качеств, необходимых для социальной жизни в обществе [Выготский, 1987; С. 56]. Деятельность человека в обществе исходит из возможности удовлетворения своих нужд на базе законов и норм, принятых в определенном лингвообществе. В связи с этим ребенок приобретает родной язык в конкретной лингвокультуре, формируется как личность в социально-культурном аспекте.

С рождением ребенок оказывается в определенной культурной среде. Таким образом, социально-культурные меры усваиваются им в первую очередь посредством именно этих культурных признаков. Только в процессе развития он становится членом культурного социума. А этот процесс обусловлен как минимум двумя системами отношений: вступление ребенка в контакт с людьми и постепенное изучение окружающей его объективной реальности. Эти процессы происходят одновременно, во взаимосвязи между собой. Обе системы взаимоотношений передаются ребенку посредством основного средства общения — языка. Посему среди всех многочисленных способов приобретения общественно-культурных знаний язык считается самым главным. Такое особенное положение и значение языка объясняется тем, что в нем сохраняется вся концентрация культурного наследия и

общественного опыта прошлых поколений. Вместе с языком ребенок впитывает и этот опыт, превращая его в неотъемлемую часть своей личности. Но неправильно было бы утверждать, что ребенок, изучивший язык и знаки языковой структуры, сам по себе усвоит культурные признаки. Особенность приобщения к культуре социума посредством языка в том, что это реализуется через непосредственное и опосредованное общение [Гвоздев, 1961, С. 39]. Культурная среда, в которой воспитывается ребенок, характеризуется в первую очередь коммуникативными свойствами. Ребенок, сам не понимая того, усваивает комплекс свойств контактной деятельности и становится членом социума.

Общение — один из способов взаимоотношений между людьми, который является чисто человеческим образцом осуществления совместной деятельности. Ребенок, являющийся социальным существом либо объектом социализации, оказывается в ситуации, при которой он может удовлетворить свои потребности только посредством общения, в том числе, речевого общения [Леонтьев, 1974, С. 156]. Сила языка и усвоения речевой деятельности проявляется именно в этом.

Приступление ребенка к материально-практической деятельности (общение с предметами), функциональный подход к предметам из окружающей его среды составляет вторую из вышеназванных двух систем. Эти две системы можно рассматривать в качестве диалектической единицы: в определенной степени они различаются между собой, однако в общественной практике не могут существовать друг без друга. Например, для ребенка, воспитываемого в узбекской семье, такие национальные блюда, как лепешка, плов; такие национальные атрибуты, как тубетейка, “бешик” — колыбель, “тумар” — амулет имеют такое же особое значение, как для китайского ребенка — блюда типа риса и соевого молока; красные фонари, золотые рыбки и изображения дракона, палочки для еды, которые отображают его национальные особенности и посему дороги ему. Представители различных национальностей, живущие на определенной территории, лелеют ценности, традиции и обычаи этих мест так же, как и свои. К примеру, представитель русского или казахского этноса, проживающий в Узбекистане, обычно с любовью употребляет такие блюда, как сумалек или нарын. Уйгуры или

представители других меньшинств Китая в дни праздников одеваются в красную одежду, поскольку этот цвет у китайцев считается символом удачи и счастья. Ребенок, становясь частью общества, впитывает в свою речь также влияние национальных традиций.

Ребенок за краткое время переходит от простых выражений до сложных правильно в смысловом и грамматическом аспектах предложений и достигает мерной речи. Дети усваивают грамматику своего родного языка задолго до школьного возраста без специального обучения, полностью соблюдая морфологические и синтаксические закономерности в своей речи.

Потребность в общении и речи возникает у ребенка не сама по себе. Первая причина этому — система биогенетической информации передается в его мозг через отца и матери. В состав подобной системы входит также информация об общении и контактах. Вторая причина — ребенок старается выразить свое отношение к направленной ему речи. Ребенок развивается в зависимости от культурного и духовного уровня окружающей его среды, и это непосредственно отражается в его речевой деятельности.

При помощи народных пословиц, приговорок, фразеологических оборотов и прочих языковых единиц, отражающих в себе национально-культурные свойства языка, передается информация о языке и культуре, связанной с ним. В частности, в узбекской речи проявляется узбекская культура общения. Эти особенности хорошо видны при сопоставлении с языком и культурой других народов. Например, у узбеков и некоторых тюркских народов много пословиц, связанных с хлебом, среди них — «Хлеб — дорог, крошка хлеба — дороже». Узбекский народ традиционно бережно относится к хлебу. В свою очередь, китайцы с особым почитанием относятся к «манто» — небольшой лепешке из рисовой или пшеничной муки.

Ребенок с рождением оказывается под влиянием окружающего мира — семьи и бытовой среды. Психическое состояние и настроения, социальная ситуация в семье непременно отразятся на развитии подрастающего ребенка.

Наличие в семье бабушки и дедушки не может не отразиться в речи ребенка. Как известно, дедушки и бабушки обычно дарят детям еще больше любви и тепла, нежели

родители. Они много общаются с детьми, употребляют в их отношении много ласкающих слов и выражений. Если сравнить речь детей, выросших в подобной среде, с детьми из семей без бабушек и дедушек, выясняется следующее: в первом случае дети используют множество эмоционально-экспрессивно окрашенных слов: *“асалим”* - «медовый», *“ойижоним”* - «мамочка», *“адажонгинам”* - «папочка», *“опоқойижон”* «дорогая бабушка» и т.д. В речи же детей — выходцев из семей, живущих врозь с бабушками и дедушками, такие слова используются сравнительно реже. Дети, выросшие в руках у дедушек и бабушек, становятся их любимцами и нередко становятся баловнями, что отражается в их психике и языковом поведении. Они используют слова, которые «ловят» в речи дедушек и бабушек, даже если не понимают их значение. Например, *“барака топгур”, “кўнглимининг кўчаси”, “айланиб кетай”, “вой қоқиндиқ”* — выражения, характерные для речи взрослых. В речи ребенка, получившего религиозное воспитание в семье, формируются ассоциации, связанные с данной тематикой, и также отражаются в его словоупотреблении. Такие дети ведут себя очень свободно и чувствуют самоуверенность, когда вступают в вербальный контакт со своими ровесниками. Они выражают свое отношение к действиям ровесников такими выражениями, как *“Ундай қилма, гуноҳ бўлади”* - «Не делай так, грех» или *“Сўкма, сўксанг, дўзахга тушасан”* - «Не матерись, если будешь материться, попадешь в ад».

Согласно наблюдениям, дети, воспитывающиеся в похожей семейной среде, быстрее учатся говорить, быстро ловят речь, направленную на себя, и начинают понимать ее. Также в речи таких детей очень рано наблюдаются речевые особенности взрослых людей. Например, в речи трехлетнего Тимурбека Искандарова активно используются приветствие на взрослый манер (подача правой руки), а также свойственные взрослым ругание, мат, гнев, а когда он недоволен, вместо того, чтобы плакать, как свои ровесники, выражает свое недовольство словами «Я обиделся на тебя»; кроме того, когда кто-нибудь чихает, он тут же реагирует словами «Будьте здоровы», а после еды открывает руки для после трапезного благопожелания («Аминь»). Это можно наблюдать и в ходе игры: он обращается к куклам словами *“аллақўзим”* «баю-бай, мой козлик», называет их «своими детьми» и пр.

В проявлении языкового потенциала ребенка важное

место занимает то, каким по счету ребенком в семействе он является. Если он — старший в семье и количество людей, общающихся с ним, еще не многочисленно, то развитие его речи может значительно запаздать. Если же у него есть старшие сестры или братья, они много общаются с младшим: вместе играют, питаются, а это, естественно, способствует быстрому развитию речевых способностей. Если ребенок — старший в семье, он «учит» младших сестер и братьев разговаривать и этим совершенствует свои речевые способности. Например, Абдулазиз, 3 года и 9 месяцев от роду, обращается к младшему брату Абдуджалилу, придавая особое значение ласкательным формам: *“Жалил, «ойи» эмас «ойижон» десанг, ойижоним олдингга келадилар”* - «Джалил, если ты будешь говорить не «мама», а «мамочка», то мамочка придет к тебе».

В развитии детской речи немаловажную роль играет посещение детского сада. Согласно наблюдениям, у воспитанника детского садика повышенный потенциал к вступлению в контакт, он без стеснения общается и с воспитателем, и с незнакомцами. Среда детского садика непосредственно влияет и на речевое поведение ребенка: он много общается со своими ровесниками в детсаде, учится у них многим отрицательным и положительным действиям, хорошим и плохим словам и оборотам. А ребенок, воспитывающийся дома, стесняется общаться с людьми, встречающимися на улице, и стоит молча перед ними. Будет трудно стимулировать их к общению. Например, девочка по имени Турсуной, 2 года и 11 месяцев от роду, когда на улице соседские женщины говорят ей: «Какая милая девочка, как твое имя?», стеснительно прячется за спину мамы. Если мать призывает отвечать ее словами «Говори, скажи, что тебя зовут Турсуной», она приходит на грань плача.

Лексический багаж детей, воспитывающихся дома, обычно несколько более беден, нежели таковое у воспитанников садика. Как известно, основой обогащения словарного запаса является совершенное усвоение правил языка. Ребенок учится в детсаде играм, новым сказкам и стишкам, активно общается с воспитателями и ровесниками. В результате этого у ребенка возникают ассоциации типа «в детсаде — хорошо» либо «в детсаде — плохо». Детсад для ребенка — фактор, второй по значимости после семейной среды, он играет важную роль в умственном и физическом развитии детей.

В семейном общении, кроме региональных диалектальных особенностей, также отражается своеобразный быт. В городе Ташкенте, наряду с ташкентским диалектом, а также в Бухарской области, наряду с бухарским диалектом, употребляются лексические единицы, диминутивы и формы обращения, свойственные только для этой семьи. Например, ребенок из хорезмской семьи, проживающей в городе Ташкенте, дома говорит на хорезмском диалекте, а на улице, в детском садике в его речь постепенно проникают элементы ташкентского диалекта. В 2–3 года ребенок начинает осознавать данное противоречие и спрашивает у матери: *“Нега «ҳа» эмас «ҳова» дедингиз?”* - «Почему вы вместо *хова* («да», элемент хорезмского диалекта) сказали *ха* («да»)?» или *“Нима учун «сабзи»ни «гашир» деяпсиз?”* - «Почему вы морковь называете *гяшир* (элемент хорезмского диалекта)?”. Это получает отражение в речевой деятельности ребенка. Литературный язык для ребенка — понятие абстрактное, родной язык для него — единое средство общения.

В речевом развитии ребенка имеет значение даже профессия отца и матери. Посему речевое поведение ребенка, у которого родители имеют высшее образование, заметно отличается от такового у ребенка, не имеющего образованных родителей. Дети первой группы по сравнению со вторыми имеют связную речь, легко и непринужденно устанавливают контакт с ровесниками, общаются без стеснения.

У ребенка рано просыпается интерес к профессии или сфере, в которой заняты его родители, он быстро усваивает термины и понятия, свойственные этой сфере и начинает употреблять их в своей речи. Если в семье мать является домохозяйкой, в таком случае речь ребенка приобретает особые характеристики.

### **Вывод**

Говоря обобщенно на основе вышеприведенных мыслей, можем отметить, что развитие детской речи неразрывно связано с окружающей его средой. Ребенок каждым своим движением пытается подражать старшим и тем самым понять жизнь, при этом многие свои действия он выполняет произвольно. Это значительно облегчает общение с ребенком. Это также свидетельствует о постепенном приспособлении ребенка к окружающей его действительности и вовлечении его в



процесс социализации. В этот период следует максимально внимательно обращаться с детьми, дарить им ласку, нежность и любовь, и тогда его речь будет связной, а психика и поведение — красивыми и благородными.

### **Литература**

- Выготский, Л.С. 1982. Собрание сочинений в 6ти томах. Т 2. Москва.  
Гвоздев, А.Н. 1961. Вопросы изучения детской речи. Москва.  
Леонтьев, А.Н. 1974. Психология общения. Тарту.  
Маслова, В.А. 2001. Лингвокультурология. Москва.  
Ольшанский, В.Б. 1970. Социализация // Философская энциклопедия. Т.5. Москва.  
Пиаже, Ж. 1932. Речь и мышление ребенка. Москва: Учпедгиз.  
Телия, В.Н. 1996. Русская фразеология. Семантический, прагматический и лингвокультурологический аспекты. Москва: Школа “Языки русской культуры”.

## **The importance of language environment in the development of linguo-cultural features of child’s speech**

Gulbahor Iskandarova<sup>1</sup>

### **Abstract**

This article discusses the importance of language as the basis of a particular national culture and mentality, the role of intercultural cooperation as a means of communication, the emergence of national spirituality and national thinking through the national language, statistics on world languages, there is a brief discussion of the Uzbek language and its suggestions and recommendations for enhancing its value in the world of languages.

It has also been revealed through scientific considerations and analysis that language is a magnificent phenomenon that reflects the psyche and culture of a people, that it is impossible to study culture without its participation and assistance, and that the concepts of language and culture are inseparable.

**Keywords:** *language, mentality, intercultural cooperation,*

---

<sup>1</sup>*Iskandarova Gulbahor Turabaevna* – Department of Uzbek Language Education in Foreign Language Groups of Tashkent State University of Uzbek Language and Literature named after Alisher Navoi

**E-mail:** iskandarovagulbahor@navoiy-uni.uz

**ORCID:** 0000-0002-4189-2725

**For reference:** Iskandarova G. T. 2022. “The importance of language environment in the development of linguo-cultural features of child’s speech” *Uzbekistan: language and culture. Applied Philology* 5(2): 26-34.

*competence, Internet language, communication, means of communication, social phenomenon*

### **References**

- Vigotskiy, L.C. 1982. *Sobranie sochineniy v 6ti tomakh. T 2.* Moskva.
- Gvozdev, A.N. 1961. *Voprosi izucheniya detskoy rechi.* Moskva.
- Leontyev, A.N. 1974. *Psixologiya obsheniya.* Tartu.
- Maslova, V.A. 2001. *Lingvokulturologiya.* Moskva.
- Olshanskiy, V.B. 1970. *Cotsiolizatsiya // Filosofskaya ensiklopediya. T.5.* Moskva.
- Piaje, J. 1932. *Rech i mishleniye rebyonka.* Moskva: Uchpedgiz.
- Teliya, V.N. 1996. *Russkaya frazeologiya. Semanticheskiy, pragmaticheskiy i lingvokulturologicheskiy aspekti.* Moskva: Skola "Yaziki russkoy kulturi".

## TABIIY TILNI QAYTA ISHLASH (NLP)

## Tabiiy tilni qayta ishlashda bag of words algoritmida foydalanish

Botir Elov<sup>1</sup>Nizomaddin Xudayberganov<sup>2</sup>Zilola Xusainova<sup>3</sup>**Annotatsiya:**

Soʻzlar sumkasi modeli - mashinali oʻrganish algoritmlari tomonidan qayta ishlash lozim boʻlgan matnning raqamli koʻrinishi. Bag Of Words (BoW) modellashtirish algoritmidan foydalanib, matnni raqamli matritsalariga aylantirish va qayta ishlash mumkin. Soʻzlar sumkasi (BoW) - soʻzning hujjatdagi statistikasini hisoblaydigan algoritm. BoW algoritmidan hujjatlarni oʻzaro solishtirish, qidiruv tizimlarida maʼlumotlarni izlash, hujjatlarni tasniflash va tematik modellashtirish kabi NLP ilovalarida foydalaniladi. Ushbu maqolada oʻzbek tilidagi matnlarni BoW algoritmi vositasida raqamli shaklga oʻtkazish usullari tahlil qilinadi.

**Kalit soʻzlar:** *BoW, Bag of words, soʻzlar jamlanmasi, soʻz vektori, token, BoW algoritmi, TF-IDF usuli.*

**Kirish**

Soʻzlar jamlanmasi (“sumkasi”) (**Bag of words, BoW**) - bu tabiiy tilni qayta ishlashda matnni modellashtirish (raqamlashtirish) usuli. Ushbu maqolada BoW algoritmining (usuli) qoʻllanilish

---

<sup>1</sup>*Elov Botir Boltayevich* – texnika fanlari boʻyicha falsafa doktori (PhD), dotsent. Alisher Navoiy nomidagi Toshkent davlat oʻzbek tili va adabiyoti universiteti Kompyuter lingvistikasi va raqamli texnologiyalar kafedrasini mudiri.

**E-pochta:** elov@navoiy-uni.uz

**ORCID:** 0000-0001-5032-6648

<sup>2</sup>*Xudayberganov Nizomaddin Uktamboy oʻgʻli* – Alisher Navoiy nomidagi Toshkent davlat oʻzbek tili va adabiyoti universiteti Kompyuter lingvistikasi va raqamli texnologiyalar kafedrasini oʻqituvchisi.

**E-pochta:** nizomaddin@navoiy-uni.uz

**ORCID:** 0000-0002-6213-3015

<sup>3</sup>*Xusainova Zilola Yuldashevna* – Alisher Navoiy nomidagi Toshkent davlat oʻzbek tili va adabiyoti universiteti Kompyuter lingvistikasi va raqamli texnologiyalar kafedrasini oʻqituvchisi.

**E-pochta:** xusainovazilola@navoiy-uni.uz

**ORCID:** 0000-0003-4357-7515

Iqtibos uchun: Elov B., Xudayberganov N., Husainova Z. 2022. “Tabiiy tilni qayta ishlashda bag of words algoritmidan foydalanish”. *Oʻzbekiston: til va madaniyat. Amaliy filologiya*. 2 (5): 35-50.

sohasi, konsepsiyasi haqida fikr-mulohaza yuritiladi va o'zbek tilidagi matnlar uchun Python tilidagi tatbig'i keltiriladi. Axborot tizimlaridagi katta hajmdagi matnli ma'lumotlarni NLP vositalari yordamida qayta ishlash orqali muhim qarorlar qabul qilinadi. Ushbu katta hajmdagi ma'lumotlarni *tushunish* va muayyan *qarorlarni qabul qilish* uchun ularni **raqamli shaklga olib kelish** kerak [Rudkowsky, Haselmayer, Wastian, Jenny, Emrich, Sedlmair, 2018, 141; Zhang, Jin, Zhou, 2010, 44]. Tabiiy tilni qayta ishlash vositalari usbu vazifani amalga oshirishga yordam beradi.

**BoW algoritmi** yordamida matnli ma'lumotlardan zarur xususiyatlar ajratib olinadi. Ushbu yondashuv hujjatlardan xususiyatlarni olishning *oddiy va moslashuvchan* usulidir. BoW - bu hujjatdagi so'zlarning paydo bo'lishini tavsiflovchi matnning raqamli ko'rinishi. BoW usulida faqat matndagi so'zlar soni (statistikasi) aniqlanadi. Biroq grammatik tafsilotlar va so'z tartibi e'tiborsiz qoldiriladi. Usulning so'zlar "sumkasi" deb atalishi hujjatdagi so'zlarning *tartibi yoki tuzilishi* haqidagi har qanday ma'lumotning tashlab yuborilishiga asoslangan [Yan, Li, Gu, Yang, 2020, 82642].

Matn bilan bog'liq eng katta muammolardan biri shundaki, uning tartibsiz va strukturlanmaganligida. Mashinali o'rganish algoritmlari esa strukturlangan va aniq belgilangan qat'iy uzunlikdagi ma'lumotlarni qayta ishlashga asoslangan. BoW algoritmidan foydalanib, o'zgaruvchan uzunlikdagi matnlarni fiksirlangan uzunlikdagi matn (**vektor**)ga aylantiriladi [Rudkowsky, Haselmayer, Wastian, Jenny, Emrich, Sedlmair, 2018,143; Yan D., Li K., Gu S., Yang L, 2020, 82645; Qiu D., Jiang H., Chen S., 2020, 6]. Mashinali o'rganish modellarida raqamlashtirilmagan matnli ma'lumotlar qayta ishlanmaydi. Shu sababli, matnlarni raqamli ko'rinishga olib kelish lozim. BoW algoritmidan foydalanib, matnga ekvivalent raqamlar vektori hosil qilinadi. BoW usuli orqali matnni vektorga aylantirish jarayonini o'zbek tilidagi gaplar misolida ko'rib chiqamiz:

**1-gap:** "Adirlar ham bahorda lola bilan go'zal, chunki lola – bahorning erka guli"

**2-gap:** "Lola ham shifokorlik kasbini tanladi"

1-jadval. Gaplarni tokenlarga ajratish

1-gap	2-gap
Adirlar	Lola
Ham	ham
bahorda	shifokorlik
lola	kasbini
bilan	tanladi

go'zal	
,	
chunki	
lola	
-	
bahorning	
erka	
guli	

**1-qadam:** Yuqoridagi matndagi barcha so'zlarni ko'rib chiqib, umumiy so'zlardan iborat lug'atni shakllantiramiz.

- Adirlar
- ham
- bahorda
- lola
- bilan
- go'zal
- ,
- chunki
- -
- bahorning
- erka
- guli
- Lola
- shifokorlik
- kasbini
- tanladi

Hosil qilingan lug'atda "**Lola**" va "**lola**" so'zlari bir xil emas. Chunki ular gaplarda turli shaklda kelgan va shuning uchun takrorlanmaydi. Shuningdek, so'zlar ro'yxatidan vergul: "," ham o'rin olgan. Berilgan ikkita gapdan hosil qilingan lug'atda **14 ta** token mavjud bo'lib, har bir tokenni baholash uchun vektorda bitta pozitsiyaga ega bo'lgan 14 uzunlikdagi hujjat vektoridan foydalanish mumkin. Matnga mos vektorni hosil qilishda har bir so'zning mavjudligini qayd etishda **1** yoki **0** qiymatlaridan (chastota) foydalaniladi. 1-gapga mos vektor quyidagicha hosil qilinadi:

2-jadval. 1-gapga mos so'zlar chastotasi

So'z	Chastota
Adirlar	1
ham	1
bahorda	1

lola	1
bilan	1
go'zal	1
,	1
chunki	1
-	1
bahorning	1
erka	1
guli	1
Lola	0
shifokorlik	0
kasbini	0
tanladi	0

1-gapdagi chastotalardan foydalanib quyidagi vektorni hosil qilamiz:

**[1,1,1,1,1,1,1,1,1,1,0,0,0,0]**

2-gapga mos vektor quyidagicha hosil qilinadi:

*3-jadval. 2-gapga mos so'zlar chastotasi*

So'z	Chastota
Adirlar	0
ham	1
bahorda	0
bola	0
bilan	0
go'zal	0
,	0
chunki	0
-	0
bahorning	0
erka	0
guli	0
Lola	1
shifokor	1
kasbini	1
tanladi	1

Xuddi shunday, yuqoridagi chastotalardan 2-gapga mos vektorni hosil qilamiz:

**[0,1,0,0,0,0,0,0,0,0,1,1,1,1]**

Berilgan ikkita so'zga mos vektorni quyidagicha aniqlaymiz:

	dir lar	ham	Bahor da	lola	bilan	go'zal	,	chun ki	-	bahor ning	erka	guli	Lola	shifo korlik	kasbini	tan ladi
1-gap	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
2-gap	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1

Yuqoridagi misolda BoW usulining eng yaxshi jihatlari o'z aksini topmagan. Chunki, **“Lola”** va **“lola”** so'zlari bir xil ma'noga ega bo'lsa-da, ikki marta qayd etilgan. Shuningdek, hech qanday ma'lumotni bildirmaydigan vergul “;”, “-” belgisi ham lug'at tarkibiga kiritilgan va tahlil jarayonida ishtirok etgan. BoW usuliga ba'zi o'zgarishlarni amalga oshirish orqali samaradorlikni oshirishni ko'rib chiqamiz.

### **Boshlang'ich ishlov berish**

**1-gap:** *“Adirlar ham bahorda lola bilan go'zal, chunki lola – bahorning erka guli”*

**2-gap:** *“Lola ham shifokorlik kasbini tanladi”*

**1-qadam:** *Berilgan gaplarni kichik harf (registr)ga aylantirish.*

**2-qadam:** *Matndan maxsus belgilar va nomuhim so'zlar (stopwords)ni olib tashlash. Stopwords – “albatta”, “ba'zi”, “bilan”, “ham”, “chunki” kabi matnda muhim ahamiyat kasb etmagan so'zlar. Yuqoridagi amallar bajarilganidan so'ng, berilgan gaplarda quyidagi o'zgartirishlar amalga oshiriladi:*

**1-gap:** *“adirlar bahorda lola go'zal lola bahorning erka guli”*

**2-gap:** *“lola shifokorlik kasbini tanladi”*

Yuqoridagi gaplar ma'noga ega bo'lmasa-da, gap mazmuniga ta'sir qiluvchi maksimal ma'lumotlarga ega so'zlardan iborat.

**3-qadam:** *Berilgan gaplardagi barcha so'zlardan iborat lug'atni shakllantirish.*

- adirlar
- bahorda
- lola
- go'zal
- bahorning
- erka
- guli
- shifokorlik
- kasbini
- tanladi

Hosil qilingan lug'atda endilikda **10** ta so'z mavjudligi sababli, har bir so'zni baholash uchun vektorda bitta o'ringa ega bo'lgan 10 ta qat'iy uzunlikdagi hujjat vektoridan foydalanish mumkin. 1-gapga

mos vektor quyidagicha hosil qilinadi:

3-jadval. 1-gapga mos so'zlar chastotasi

So'z	Chastota
Adirlar	1
bahorda	1
lola	2
go'zal	1
bahorning	1
erka	1
guli	1
shifokorlik	0
kasbini	0
tanladi	0

1-gapdagi chastotalardan foydalanib quyidagi vektorni hosil qilamiz:

**[1,1,2,1,1,1,1,0,0,0]**

2-gapga mos vektor quyidagicha hosil qilinadi:

4-jadval. 2-gapga mos so'zlar chastotasi

So'z	Chastota
adirlar	0
bahorda	0
lola	1
go'zal	0
bahorning	0
erka	0
guli	0
shifokorlik	1
kasbini	1
tanladi	1

2-gapdagi chastotalardan foydalanib quyidagi vektorni hosil qilamiz:

**[0,0,1,0,0,0,0,1,1,1]**

Berilgan ikkita so'zga mos vektorni quyidagicha aniqlaymiz:

	Adirlar	bahorda	lola	go'zal	bahorning	erka	guli	shifokorlik	kasbini	tanladi
1-gap	1	1	2	1	1	1	1	0	0	0
2-gap	0	0	1	0	0	0	0	1	1	1



Mashinali o'rganishda ishlatiladigan ma'lumotlar to'plamlari juda katta hajmda bo'lib, bir necha ming yoki hatto millionlab so'zlardan iborat lug'atni o'z ichiga olishi mumkin. Shunday qilib, so'zlar jamlanmasidan foydalanishdan oldin matnni boshlang'ich qayta ishlash lozim. BoW usuli samaradorligini oshirishi mumkin bo'lgan turli xil qayta ishlash bosqichlari mavjud bo'lib, ularning ba'zilari ushbu maqolada batafsil keltiriladi.

Yuqoridagi misollarda BoW vektorini hosil qilish uchun lug'atdagi barcha so'zlardan foydalandik. Bu BoW modelini amalga tatbiq qilishda bir qator murakkabliklarni yuzaga keltiradi. Amalda vektorni yaratish uchun lug'atdan faqat bir nechta so'zlar (eng ko'p uchragan so'zlar) ishlatiladi.

### **BoW usulini Pythonda amalga oshirish**

BoW algoritmini Python tilida quyidagicha tatbiq qilamiz:

```
def vectorize(tokens):
    vector=[]
    for w in filtered_vocab:
        vector.append(tokens.count(w))
    return vector
```

```
def unique(sequence):
    seen = set()
    return [x for x in sequence if not (x in seen or seen.
```

```
add(x))]
```

```
# Nomuhim so'zlar ro'yxatini yaratish
stopwords=["ham","bilan","chunki",...]
```

```
# Maxsus belgilar ro'yxati
special_char=["",":",";","'","\"","?","-"]
```

```
# Korpusdagi gaplarni aniqlash
string1="Adirlar ham bahorda lola bilan go'zal, chunki lola -
bahorning erka guli"
string2="Lola ham shifokor bo'lib ishlaydi"
```

```
# so'zni kichik harfga aylantirish
string1=string1.lower()
string2=string2.lower()
```

```
# gaplarni tokenlarga ajratish
```

```
tokens1=string1.split()
tokens2=string2.split()
print(tokens1)
print(tokens2)
```

```
# lug'at ro'yxatini tuzish
vocab=unique(tokens1+tokens2)
print(vocab)
```

```
# lug'atlar ro'yxatini filtrlash
filtered_vocab=[]
for w in vocab:
    if w not in stopwords and w not in special_char:
        filtered_vocab.append(w)
print(filtered_vocab)
```

```
# gaplarni vektorlarga aylantirish
vector1=vectorize(tokens1)
print(vector1)
vector2=vectorize(tokens2)
print(vector2)
```

---

```
['adirlar', 'ham', 'bahorda', 'lola', 'bilan', "go'zal", ',', 'chunki', 'lola',
'-', 'bahorning', 'erka', 'guli']
['lola', 'ham', 'shifokor', "bo'lib", 'ishlaydi']
['adirlar', 'ham', 'bahorda', 'lola', 'bilan', "go'zal", ',', 'chunki', '-',
'bahorning', 'erka', 'guli', 'shifokor', "bo'lib", 'ishlaydi']
['adirlar', 'bahorda', 'lola', "go'zal", 'bahorning', 'erka', 'guli',
'shifokor', "bo'lib", 'ishlaydi']
[1, 1, 2, 1, 1, 1, 1, 0, 0, 0]
[0, 0, 1, 0, 0, 0, 0, 1, 1, 1]
```

### **Sklearn yordamida BoWni ishlab chiqish.**

Yuqoridagi BoW modelini Python yordamida osongina amalga oshirish uchun **Sklearn** kutubxonasidagi **Count Vectorizer** () funksiyasidan foydalanish mumkin.

```
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer,
TfidfVectorizer
```

```
sentence_1="Adirlar ham bahorda lola bilan gozal , chunki
```

```
lola - bahorning erka guli"
sentence_2="Lola ham shifokor"

# ngram_range=(2,2) bigramlardan foydalanish
CountVec = CountVectorizer(ngram_range=(1,1), stop_
words=["ham","bilan","chunki"])

# transformatsiya
Count_data = CountVec.fit_
transform([sentence_1,sentence_2])

# dataframeni shakllantirish
cv_dataframe=pd.DataFrame(Count_data.
toarray(),columns=CountVec.get_feature_names())
print(cv_dataframe)
```

---

	adirlar	bahorda	bahorning	erka	go'zal	guli	kasbini	lola	
	shifokorlik	tanladi							
1	1	1	1	1	1	1	1	0	
2	0	0							
	2	0	0	0	0	0	0	1	1
1	1								

### N-gramlar

1. N-gramlar nima va ulardan nima maqsadda foydalaniladi? Quyidagi misollarni ko'rib chiqamiz [Elov, 2022, 62; Yadav, Borgohain, 2015, 1771]:

**1-gap:** "Bu uylar narxi arzon, lekin ular shahar markaziga yaqin emas"

**2-gap:** "Bu mahsulot narxi arzon emas"

Ushbu misol uchun faqat 8 ta so'zdan iborat lug'atni shakllantiramiz:

- uylar
- narxi
- arzon
- shahar
- markaziga
- yaqin
- emas

- mahsulot

Shunday qilib, ushbu gaplar uchun mos vektorlar:

“Bu uylar narxi arzon, lekin ular shahar markaziga yaqin emas” = **[1,1,1,1,1,1,0]**

“Bu mahsulot narxi arzon emas” = **[0,1,1,0,0,1,1]**

Yuqorida hosil qilingan natijalarda qanday muammo mavjud?

2-gap salbiy, 1-gap esa ijobiy ma’noga ega. Yuqorida shakllantirilgan vektorlarda ushbu ma’lumot aks etmagan. Ushbu muammoni hal qilish uchun N-grammlarni aniqlab olish lozim.

**N-gramm** – N-tokenli soʻzlar qatoridir: 2 gramm (odatda **bigramm** deb ataladi) ikki soʻzdan iborat “*juda yaxshi*”, “*yaxshi emas*” yoki “*sizning vazifangiz*” va 3-gramm (koʻpincha **trigramm** deb ataladi) – “*bu mumkin emas*” yoki “*shahar markazidagi uylar*” kabi soʻzlar trigrammni anglatadi.

Masalan, oldingi misoldagi birinchi gapga mos bigrammalar (“*Bu mahsulot narxi arzon emas*”):

- “Bu mahsulot”
- “mahsulot narxi”
- “narxi arzon”
- “arzon emas”

Yuqoridagi misolda oddiy (bitta) soʻzlarni ishlatish oʻrniga, yuqorida koʻrsatilganidek, **bigrammlardan** (bag-of-bigrams) foydalanamiz. Yaratilgan model asosida 1-gap va 2-gaplar oʻzaro farqlanadi. Shunday qilib, bigrammlardan foydalanish tokenlarni yanada tushunarli koʻrinishga olib keladi. Xulosa sifatida bigrammlar jamlanmasi soʻzlardan koʻra kuchliroqdir va koʻp hollarda ushbu usuldan foydalanishni tavsiya qilamiz.

### **Tf-Idf usuli**

Yuqorida qoʻllaniladigan baholash usulida har bir soʻzning statistikasi aniqlanib, vektordagi soʻz soni boʻyicha ifodalandi. Hosil qilingan vektordan hujjatlar haqida ma’lumot olishda foydalanib boʻlmaydi. Agar biror soʻz hujjatda koʻp marta uchrasa, bu soʻzning hujjat mazmunida muhimligini yoki aktualligini anglatmaydi [Yadav, Borgohain, 2015; Stecanella, 2019.]

Oʻzbek tilidagi juda koʻp matnlarda “bilan”, “va”, “balki” va shunga oʻxshash soʻzlar juda koʻp ishlatiladi. Shu sababli ushbu soʻzga mos ball qiymatini kamaytirish lozim. Bu yondashuv term frequency-inverse document frequency yoki qisqacha Tf-Idf deb nomlanadi. TF-IDF usuli yordamida soʻzning berilgan hujjatda qanchalik muhimligi aniqlanadi. Hujjatga mos Tf-Idf qiymat

qanday hisoblanadi? Hujjatdagi so'z uchun TF-IDF qiymat ikki xil ko'rsatkichni ko'paytirish orqali hisoblanadi:

**TF-IDF=TF\*IDF.** Hujjatdagi so'zning chastotasi (**TF**, term frequency)ni hisoblashning bir necha yo'li mavjud. Eng oddiyi hujjatda so'z paydo bo'lgan holatlarini aniqlash usuli. Shuningdek, chastotani hisoblashning boshqa usullari mavjud. Masalan, so'z namunalarining dastlabki sonini hujjat uzunligiga yoki hujjatdagi eng ko'p uchraydigan so'zning chastotasiga bo'lish orqali hisoblash mumkin. TF chastotani hisoblash formulasi [Zhu Z., Liang J., Li D., Yu H., Liu G., 2019, 26997.]:

$$\mathbf{TF}(i, j) = n(i, j) / \Sigma n(i, j)$$

Bu yerda,

- $n(i, j)$  – hujjatda  $n$ -chi so'z necha marta ushrashi;
- $\Sigma n(i, j)$  – hujjatdagi so'zlarning umumiy soni.

Hujjatlar to'plamidagi so'zning IDF qiymati hujjatlar to'plamida so'zning qanchalik ko'p marotaba yoki kamdan-kam uchrashini ko'rsatadi. Qiymat **0** ga qanchalik yaqin bo'lsa, so'z shunchalik ko'p uchraydi. Ushbu ko'rsatkichni hujjatlarning umumiy sonini aniqlab, ulardagi so'zni o'z ichiga olgan hujjatlar soniga bo'lish va logarifmni hisoblash orqali hisoblash mumkin.

Shunday qilib, agar so'z hujjatlarda juda ko'p marotaba ishlatilgan bo'lsa, IDF qiymat **0** ga, aks holda **1** ga yaqinlashadi. Ushbu ikki qiymatni ko'paytirish orqali hujjatdagi so'zning **TF-IDF** qiymatini aniqlash mumkin. Qiymat qanchalik katta bo'lsa, ushbu hujjatda bu so'z shunchalik dolzarb (muhim ahamiyatga ega) bo'ladi. Matematik jihatdan aytganda, TF-IDF qiymat quyidagicha hisoblanadi:

$$\mathbf{IDF} = 1 + \log(N/dN)$$

Bu yerda,

- $N$  – ma'lumotlar to'plamidagi hujjatlarning umumiy soni;
- $dN$  –  $n$ -o'rindagi so'z mavjud hujjatlarning umumiy soni.

Yuqoridagi formulaga qo'shilgan **1** qiymat **IDFni silliqlash jarayoni** deb nomlanadi. TF-IDF usulining Python tilidagi tatbig'ini ko'rib chiqamiz. Yuqoridagi **BoW (Tf-IDF)** modelini amalga oshirish uchun **Sklearn** kutubxonasidagi **TfidfVectorizer ()** funksiyasidan foydalanish mumkin.

```
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer,
TfidfVectorizer
```

```
sentence_1="Bu uylar narxi arzon, lekin ular shahar  
markaziga yaqin emas"
```

```
sentence_2="Bu mahsulot narxi arzon emas"
```

```
# silliq IDF holida  
print("Silliq IDF holida:")  
# tf-idf qiymatni hiosblash  
tf_idf_vec = TfidfVectorizer(use_idf=True,  
                             smooth_idf=False,  
                             ngram_range=(1,1),stop_  
words=["bu","lekin","ular"]) # to use only bigrams ngram_  
range=(2,2)
```

```
# transformatsiya  
tf_idf_data = tf_idf_vec.fit_  
transform([sentence_1,sentence_2])
```

```
# dataframeni shakllantirish  
tf_idf_dataframe=pd.DataFrame(tf_idf_data.  
toarray(),columns=tf_idf_vec.get_feature_names())  
print(tf_idf_dataframe)  
print("\n")
```

```
# silliq IDF  
tf_idf_vec_smooth = TfidfVectorizer(use_idf=True,  
                                    smooth_idf=True,  
                                    ngram_range=(1,1),stop_  
words=["bu","lekin","ular"])
```

```
tf_idf_data_smooth = tf_idf_vec_smooth.fit_  
transform([sentence_1,sentence_2])
```

```
print("Silliq IDF:")  
tf_idf_dataframe_smooth=pd.DataFrame(tf_idf_data_smooth.  
toarray(),columns=tf_idf_vec_smooth.get_feature_names())  
print(tf_idf_dataframe_smooth)
```

---

Silliq IDF holida:

arzon    emas    mahsulot    markaziga    narxi    shahar

uylar	yaqin					
1	0.262912	0.262912	0.00000	0.445149	0.262912	0.445149
2	0.412859	0.412859	0.69903	0.000000	0.412859	0.000000

Silliq IDF:

uylar	yaqin	arzon	emas	mahsulot	markaziga	narxi	shahar
1	0.302873	0.302873	0.000000	0.425677	0.302873	0.425677	0.425677
2	0.448321	0.448321	0.630099	0.000000	0.448321	0.000000	0.000000

### **BoW usulidagi cheklovlar**

BoW usulini amalga oshirish oson bo'lsa-da, ba'zi kamchiliklar mavjud:

– modelda so'zning gapdagi joylashuvi inobatga olinmaydi. Misol uchun, “mahsulot arzon” va “mahsulot arzonmi” so'z birikmalariga mos BoW modelida aynan bir xil vektor hosil qilinadi.

– BoW so'zning semantikasini inobatga olmaydi. Masalan, “kurash” va “dzyudo” so'zlari ko'pincha bir xil kontekstda ishlatiladi. Biroq, bu so'zlarga mos keladigan vektorlardagi so'zlar modelida butunlay boshqacha. Gaplarni modellashtirishda bu muammo yanada jiddiylashadi. Masalan: “Ishlatilgan mashinalarni sotib oling” va “Eski mashinalarni sotib oling” BoW modelida mutlaqo boshqa vektorlar bilan ifodalanadi.

– lug'at hajmi – BoW modeli oldida turgan katta muammo. Misol uchun, agar model hali ishlatilmagan yangi so'zga duch kelsa, Biblioklept (kitob o'g'irlagan degan ma'noni anglatadi) kabi unikal, ammo ma'lumot beruvchi so'zni aytamiz. BoW modelida bu so'z inobatga olinmaydi.

### **Xulosa**

BoW – hujjatlarni bir butun sifatida tasniflashning oddiy va qulay usuli hisoblanadi. BoW modeli orqali mashinali o'rganish algoritmlariga ma'lumotlarni raqamli ko'rinishda taqdim etish uchun matn modellashtiriladi. BoW modelning aniqligini oshirishning bir necha yo'llari mavjud:

– BoW modelida unigrammlarni (so'zlar) ishlatish o'rniga, bigramm yoki trigrammdan foydalanish model

samaradorligini oshirishga xizmat qiladi.

– TF-IDF modelidan matndagi muhim va nomuhim soʻzlarni aniqlashda foydalaniladi. Odatda TF-IDF modelidan mashinali oʻrganishga asoslangan NLP vazifalarda qoʻllaniladi.

Nisbatan oddiy model boʻlishiga qaramay, BoW koʻpincha matn tasnifi NLP vazifalari uchun ishlatiladi.

### **Adabiyotlar:**

- Rudkowsky, E., Haselmayer, M., Wastian, M., Jenny, M., Emrich, Š., & Sedlmair, M. More than Bags of Words: Sentiment Analysis with Word Embeddings. *Communication Methods and Measures*, 12(2–3), 2018. <https://doi.org/10.1080/19312458.2018.1455817>
- Zhang, Y., Jin, R., & Zhou, Z. H. Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1–4), 2010. <https://doi.org/10.1007/s13042-010-0001-0>
- Yan, D., Li, K., Gu, S., & Yang, L. Network-Based Bag-of-Words Model for Text Classification. *IEEE Access*, 8, 2020. <https://doi.org/10.1109/ACCESS.2020.2991074>
- Qiu, D., Jiang, H., & Chen, S. Fuzzy information retrieval based on continuous bag-of-words model. *Symmetry*, 12(2), 2020. <https://doi.org/10.3390/sym12020225>
- Elov, B.B., Hamraeva, Sh., Axmedova, X., Methods for creating a morphological analyze // 14th International Conference on Intellegent Human Computer Interaction. 19-23 October 2022, Tashkent.
- Elov, B.B. N-gramm til modellari vositasida oʻzbek tilida matn generatsiya qilish // Kompyuter lingvistikasi: muammolar, yechim va istiqbollar / Xalqaro ilmiy-amaliy konferensiya toʻplami. Elektron nashr / ebook. – Toshkent, 2022.
- Yadav, A. K., & Borgohain, S. K. (2015). Sentence generation from a bag of words using N-gram model. *Proceedings of 2014 IEEE International Conference on Advanced Communication, Control and Computing Technologies, ICACCCT 2014*. <https://doi.org/10.1109/ICACCCT.2014.7019414>
- Stecanella, B. What is TF IDF? MonkeyLearn. 2019.
- Zhu, Z., Liang, J., Li, D., Yu, H., & Liu, G. Hot Topic Detection Based on a Refined TF-IDF Algorithm. *IEEE Access*, 7. 2019. <https://doi.org/10.1109/ACCESS.2019.2893980>
- Pietro, M. di. Text Classification with NLP: Tf-Idf vs Word2Vec vs BERT. 2020. Medium.



# Using the bag of words algorithm in natural language processing

Botir Elov<sup>1</sup>

Nizomaddin Xudayberganov<sup>2</sup>

Zilola Xusainova<sup>3</sup>

## **Abstract:**

A bag-of-words model is a digital representation of text to be processed by machine learning algorithms. Using the Bag Of Words (BoW) modeling algorithm, text can be converted and processed into digital matrices. Bag of Words (BoW) is an algorithm that calculates the statistics of a word in a document. The BoW algorithm is used in NLP applications such as document comparison, information retrieval in search engines, document classification, and thematic modeling. This article presents the methods of converting Uzbek texts into digital form using the BoW algorithm.

**Keywords:** *BoW, Bag of words, set of words, word vector, token, BoW algorithm, TF-IDF method.*

## **References:**

- Rudkowsky, E., Haselmayer, M., Wastian, M., Jenny, M., Emrich, Š., & Sedlmair, M. (2018). More than Bags of Words: Sentiment Analysis with Word Embeddings. *Communication Methods and Measures*, 12(2–3). <https://doi.org/10.1080/19312458.2018.1455817>
- Zhang, Y., Jin, R., & Zhou, Z. H. (2010). Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1–4). <https://doi.org/10.1007/s13042-010-0001-0>

---

<sup>1</sup>*Elov Botir Boltayevich* – doctor of philosophy in technical sciences (PhD), associate professor. Head of the Department of Computer Linguistics and Digital Technologies of Tashkent State University of Uzbek Language and Literature named after Alisher Navoi.

**E-mail:** [elov@navoiy-uni.uz](mailto:elov@navoiy-uni.uz)

**ORCID:** 0000-0001-5032-6648

<sup>1</sup>*Xudayberganov Nizomaddin Uktambov o'g'li* – Teacher of the Department of Computer Linguistics and Digital Technologies of Tashkent State University of Uzbek Language and Literature named after Alisher Navoi.

**E-mail:** [nizomaddin@navoiy-uni.uz](mailto:nizomaddin@navoiy-uni.uz)

**ORCID:** 0000-0002-6213-3015

<sup>2</sup>*Xusainova Zilola Yuldashevna* -Teacher of the Department of Computer Linguistics and Digital Technologies of Tashkent State University of Uzbek Language and Literature named after Alisher Navoi.

**E-mail:** [xusainovazilola@navoiy-uni.uz](mailto:xusainovazilola@navoiy-uni.uz)

**ORCID:** 0000-0003-4357-7515

**For reference:** Elov B., Xudayberganov, N., Husainova, Z. 2022. "Using the bag of words algorithm in natural language processing". *Uzbekistan: language and Culture. Applied philology*. 2 (5): 35-50.

- Yan, D., Li, K., Gu, S., & Yang, L. (2020). Network-Based Bag-of-Words Model for Text Classification. *IEEE Access*, 8. <https://doi.org/10.1109/ACCESS.2020.2991074>
- Qiu, D., Jiang, H., & Chen, S. (2020). Fuzzy information retrieval based on continuous bag-of-words model. *Symmetry*, 12(2). <https://doi.org/10.3390/sym12020225>
- Elov, B.B., Hamraeva, Sh., Axmedova, X., Methods for creating a morphological analyze // 14th International Conference on Intellegent Human Computer Interaction. 19-23 October 2022, Tashkent.
- Elov, B.B. N-gramm til modellari vositasida o'zbek tilida matn generatsiya qilish // *Kompyuter lingvistikasi: muammolar, yechim va istiqbollar / Xalqaro ilmiy-amaliy konferensiya to'plami. Elektron nashr / ebook. – Toshkent: 2022.*
- Yadav, A. K., & Borgohain, S. K. (2015). Sentence generation from a bag of words using N-gram model. *Proceedings of 2014 IEEE International Conference on Advanced Communication, Control and Computing Technologies, ICACCCT 2014*. <https://doi.org/10.1109/ICACCCT.2014.7019414>
- Stecanella, B. (2019). What is TF IDF? *MonkeyLearn*.
- Zhu, Z., Liang, J., Li, D., Yu, H., & Liu, G. (2019). Hot Topic Detection Based on a Refined TF-IDF Algorithm. *IEEE Access*, 7. <https://doi.org/10.1109/ACCESS.2019.2893980>
- Pietro, M. di. (2020). Text Classification with NLP: Tf-Idf vs Word2Vec vs BERT. *Medium*.

## O'zbek tilida pos tegging masalasi: muammo va takliflar

Botir Elov<sup>1</sup>,  
Shahlo Hamroyeva<sup>2</sup>,  
Oqila Abdullayeva<sup>3</sup>,  
Mohiyaxon Uzoqova<sup>4</sup>

### Abstrakt:

Til korpusi qurilishida lingvistik ta'minot masalasi muhim va murakkab hisoblanadi. Korpuslarda matnlardagi nutq bo'laklariga mos identifikatorini belgilash jarayoni muammolidir, sababi tilni modellashtirish teglash qoidasi va tilda mavjud qonuniyat bilan bog'liq. Teglash, xususan, grammatik teglash yoki PoS tegging o'zbek korpus lingvistikasi uchun ham dolzarb masaladir. Chunki maxsus "kodlangan" belgilar tizimi o'zbek tili bilan bog'liq NLP masalalarini yechishda birlamchi kalit bo'lib xizmat qiladi. Maqolada teg va PoS tegging masalasining jahon tilshunosligidagi o'rganilishi tahlil qilindi, o'zbek tilshunosligidagi hozirgi teglash jarayoniga nazar tashlandi. Jahonda keng qo'llanib kelinayotgan tegsetlardan oziqlangan holda o'zbek tili qonuniyatlari asosida muqobil tegset taklif qilindi.

---

<sup>1</sup>*Elov Botir Boltayevich* – texnika fanlari bo'yicha falsafa doktori (PhD), dotsent. Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti Kompyuter lingvistikasi va raqamli texnologiyalar kafedrasini mudiri.

**E-pochta:** elov@navoiy-uni.uz

**ORCID:** 0000-0001-5032-6648

<sup>2</sup>*Hamroyeva Shahlo Mirdjanovna* – filologiya fanlari doktori (DSc), Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti Kompyuter lingvistikasi va raqamli texnologiyalar kafedrasini dotsenti v.b.

**E-pochta:** shaxlo.xamrayeva@navoiy-uni.uz

**ORCID:** 0000-0002-5429-4708

<sup>3</sup>*Abdullayeva Oqila Xolmo'minovna* – filologiya fanlari bo'yicha falsafa doktori (PhD). Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti Kompyuter lingvistikasi va raqamli texnologiyalar kafedrasini katta o'qituvchisi.

**E-pochta:** abdullayeva.oqila@navoiy-uni.uz

**ORCID:** 0000-0002-2524-4832

<sup>4</sup>*Uzoqova Mohiyaxon Tuyg'un qizi* – Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti Kompyuter lingvistikasi va raqamli texnologiyalar kafedrasini magistranti.

**E-pochta:** uzoqovamohiyaxon@navoiy-uni.uz

**ORCID:** 0000-0001-7102-0824

**Iqtibos uchun:** Elov B., Hamroyeva Sh., Abdullayeva O., Uzoqova M. 2022. "O'zbek tilida PoS tegging masalasi: muammo va takliflar". *O'zbekiston: til va madaniyat. Amaliy filologiya*. 2 (5): 51-68.

**Kalit soʻzlar:** *Teg, razmetka, annotatsiya, tegset, NLP, korpus, Braun korpusi, CLAWS, PoS tepping.*

### **Kirish**

Maʼlumki, istalgan turdagi korpusni tuzish mavjud datani teglashdan boshlanadi. Teglash jarayoni annotatsiya, razmetka kabi soʻzlar bilan turli adabiyotlarda yonma-yon, baʼzan sinonim sifatida ishlatilganligiga tez-tez guvoh boʻlish mumkin [Zaxarov, 2005, 6]. Tadqiqotimizda til korpuslarida nutq boʻlaklarini teglash masalasini muhokama qilishdan oldin, teg, razmetka va annotatsiya tushunchalari toʻgʻrisidagi fikrlarni kuzatamiz. Mazkur terminlarni sinonim sifatida baholash toʻgʻrimi? Biri oʻrnida ikkinchisi qoʻllanilganda bir mazmuni bera oladimi? Annotatsiya, teg va razmetkaning bir-biridan farqlovchi aniq chegarasi mavjudmi?

**Annotatsiya** – Korpus lingvistikasi doirasida qaralganda berilgan matnga bevosita aloqasi boʻlmagan, ammo uning qaysidir qismi haqidalingvistikyoki ekstralingvistikaxborot beruvchi umumiy maʼlumot. Annotatsiya oʻz ichiga metamaʼlumot va teglarni qamrab olishi mumkin. Ayrim manbalarda [Leech, Vilson, 1994] korpus annotatsiyasi deganda, til korpusida matnning elektron shakliga kodlash qoʻshilgan izohlovchi, lingvistik maʼlumotlarni qoʻshish amaliyoti ham tushuniladi. McEnery annotatsiyalashni uch usulda amalga oshirish mumkinligini yozadi: toʻla avtomatlashtirilgan, yarim avtomatlashtirilgan va qoʻl mehnati yordamida. Ammo hech bir usul mukammal ish bermasligi, mutlaqo xatosiz natija olib boʻlmasligini aytgan.

**Razmetka** – McEnery razmetka va meta maʼlumotni annotatsiyalash jarayonining bir qismi sifatida baholagan [McEnery, Hardie, 2012, 47]. Boshqa manbalar bilan tanishganimizda esa razmetkaga annotatsiyalash jarayonining oʻzi sifatida baho berilganligiga guvoh boʻldik [Myfilology.ru]. Ingliz manbalarida razmetka markup termini bilan yuritiladi [McEnery, Hardie, 2012, 47]. Jahon amaliyotida razmetkalashning standart prinsiplari ishlab chiqilgan. U SGML yoki Standard Generalized Markup Language [<https://www.techtarget.com/>] deb ataladi. Eʼtibor qaratish lozim boʻlgan jihat shundaki, SGMLda standart annotatsiya emas, balki annotatsiyalash jarayonini tashkil etish metodologiyasi aks etadi. SGML asosida ishlab chiqilgan va keng ommalashgan tillarga HTML yoki XMLni misol qilish mumkin.

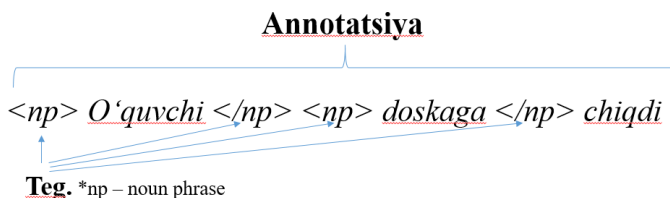
**Teg** – Kompyuter yordamida matn tahlilini amalga oshirish jarayonini tezlashtirish va osonlashtirishga xizmat qiluvchi shartli belgi yoki maxsus kod. Teglar bir necha turlarga boʻlinadi [<https://>

ucrel.lancs.ac.uk/]: semantik teg, sintaktik teg va grammatik teg. Grammatik teg, shuningdek, PoS (Part of speech) tegging nomi bilan ham mashhur.

Annotatsiya (razmetka) va tegning farqini quyidagi jadvalda aniqroq ko'rish mumkin [<https://knowledge.autodesk.com/>]:

### 1-jadval. Annotatsiya va teg terminining ta'riflanishi

Annotatsiya	Teg
Muayyan komponent yoki segment haqida jadval yoki grafik ko'rinishidagi ma'lumot	Muayyan komponent yoki segment uchun unikal identifikator
Komponent yoki segment uchun bir nechta annotatsiya bo'lishi mumkin	Komponent yoki segment uchun yagona nom beriladi va u teg deb ataladi
Bir qancha ta'riflar to'plamini matn shaklida o'zida jamlaydi va unda teglar ham aks etadi.	Gap emas, uning bo'laklari haqida ma'lumot beruvchi muayyan formatdagi data hisoblanadi



### 1-rasm. SGMLda annotatsiyalash va teglash jarayoni.

1-rasmda SGMLda ishlab chiqilgan standart asosida gap annotatsiyalangan. Ko'rish mumkinki, gapdagi har bir bo'lak o'zi tegishli bo'lgan turkum asosida teglanib, boshlanish va tugash jarayoni qirrali qavs </> yordamida aks ettirilmoqda. Bu jarayonni umumiy qilib annotatsiyalashning bir ko'rinishi sifatida baholash mumkin.

Demak, annotatsiya va razmetka kengroq tushuncha bo'lib, teg uning bo'lagi sifatida xizmat qilishi mumkin. Teg – unikal kod, ammo havola qilayotgan komponentasining qisqartmasi sifatida baholanmasligi lozim.

### Asosiy qism

Izlanishlarimiz davomida [Mitkov, 2003, 228] shu narsa ayonlashdiki, teglashga bo'lgan ehtiyoj, ya'ni nutq bo'laklarini unikal kodlashtirish kompyuterni avtomatik tahlil qilishga yordam beruvchi manbalar, jumladan, korpus tuzish ishlaridan keyin boshlangan. Shuning uchun ham teglash jarayoni birinchi marta jahon tajribasida

Braun korpusini tuzish bilan qayd etiladi. 1971-yilda 77 xil turdagi teglar TAGGIT dasturi uchun Braun korpusi mualliflari tomonidan ishlab chiqilgan [Mitkov, 2003, 228]. Bunday teglashning asosiy yutuq'i nafaqat asosiy grammatik kategoriyalar, balki ularning ichki guruhlarigacha teglanganida ko'ringan. Keyinroq, Lankaster universiteti tomonidan CLAWS teggerini yaratish uchun tegsetlar (teglar to'plami) ham ishlab chiqila boshlangan [<https://ucrel.lancs.ac.uk>]. Tegger – bu berilgan matnni yoki segmentni avtomatik teglashga xizmat qiluvchi dastur. Shunday qilib, 1980-yillarda tuzish boshlangan CLAWS1 tegseti [<https://ucrel.lancs.ac.uk>] jami 137 ta tegni o'zida qamrab olgan. Teglarining aksariyati Braun korpusida foydalanilgan teglardan oziqlangan, ammo ulardan farqli bo'lgan. CLAWS tegsetlarini ishlab chiqishda bevosita ishtirok etgan Robert Garside bu haqida shunday yozgan [<https://ucrel.lancs.ac.uk>]: *“CLAWS1da qo'llanilgan teglar to'plami Braun korpusini teglashda foydalanilgan teglardan oziqlangan. Biz teglangan LOB korpusini Braun korpusi bilan taqqosladik, ammo Braun tegsetidagi atoqli otlar va olmoshlar teglariga o'zgartirish kiritgan holda sintaktik birliklar (so'zlar va tinish belgilari) uchun jami 137 ta ehtimoliy teglarni ishlab chiqdik”*. CLAWS tegsetlari yillar davomida takomillashtirilib, bugungi kunda 8 taga yetdi [<https://ucrel.lancs.ac.uk>]. Bunda har bir tegset keyingisi uchun asos sifatida xizmat qilavergan.

Jahon miqyosida keng tarqalgan va mashhur bo'lgan tegsetlardan Pen Tree Bank Tegsetni [<https://www.sketchengine.eu>] ham e'tirof etishni lozim topdik. Mazkur tegset 1993-1994-yillarda Shtutgart universiteti tomonidan ishlab chiqilgan [<https://www.ims.uni-stuttgart.de>]. Keyinroq bu tegsetlardan Sketch Engine korpusida ham foydalanilgan [<https://www.sketchengine.eu>]. Hozirda Sketch Engine tarkibidagi tegsetlar versiyasi 3 taga yetgan, ammo ular Pen Tree Bank tegsetlarining aynan o'zi emas, biron takomillashgan versiyasi hisoblanadi. CLAWS va Pen Tree Bank tegsetlarini solishtirganimizda, aksariyat teglar bir xil shaklga ega ekanligiga guvoh bo'ldik (Masalan, *CC = Coordinating conjunction, JJ = adjective, MD = modal*). Demak, yaratilish yilini hisobga olgan holda Pen Tree Bank tegsetlari CLAWS tegsetlaridan oziqlangan, deyish mumkin.

Yana bir ommalashgan tegger va tegsetlar Python kutubxonasiga tegishli NLTK dasturida aks etgan. NLTK – Natural Language Toolkit tokenizatsiya, stemming, lemmatizatsiya va avtomatik teglay oluvchi universal hamda NLP olamidagi eng ommabop dastur bo'lib, unda taqdim etilgan tegsetlarni o'rganib chiqqanimizda Pen Tree Bank va CLAWS tegsetlaridan deyarli farq

qilmasligiga guvoh bo'ldik [<https://www.learnstek.org>]. Hatto BNC (British National Corpus) namunaviy tegsetlari [<https://ucrel.lancs.ac.uk>] ham CLAWS tegsetlari asosida tuzilgan.

Biroz farqli tegsetlarni 25 xil tegsetni taqqoslash asosida universal tegset ishlab chiqqan bir guruh olimlar tadqiqotida ko'rinadi [Petrov, Das, McDonald, <https://static.googleusercontent.com>]. Tadqiqotda bayon etilishicha, 22 xil tilga xos umumiy grammatik kategoriyalar o'rganib chiqilgan va universal teglar taklif qilingan. Universal tegsetdagi [<https://universaldependencies.org>] teglar yuqorida tahlilga olingan tegsetlardan mutlaqo yangi teglarni taklif qilishi bilan farq qiladi. Ammo faqat eng yirik grammatik guruhlar uchun teglar taklif etilgani bois muayyan til uchun to'liq bu tegsetlarni joriy qilish imkonsiz, degan fikrdamiz.

Shunday qilib, jahondagi eng ommalashgan tegsetlarni tadqiq va tahlil qilish natijasida shunday xulosaga keldikki, **har qanday til uchun joriy qilinishi mumkin bo'lgan standart teg mavjud emas.** Bu xulosani quyidagi sabablar bilan dalillasak:

1) Tegsetlarni ishlab chiqishda har bir maktab o'zining yondashuvlariga asoslangan. Davr o'tishi bilan biri ikkinchisidan oziqlangan va ilgari tegsetlar yangi teglar bilan takomillashtirib borilgan.

2) Dunyodagi hech bir til tuzilishi va tabiatiga ko'ra to'laqonli o'xshash emas, garchi o'xshashlik darajasiga qarab umumiy til oilalari va guruhlariga tasniflansa ham.

3) Annotatsiyalashning turlari kengaygani sari yangi tegsetlar ishlab chiqishga ehtiyoj tug'ilaveradi. Natijada tegsetlar statik holatdan dinamik holatga o'tib qoladi.

Shular sababli ham, bizning fikrimizcha, butun dunyoga tatbiq qilish mumkin bo'lgan mukammal va standart tegsetlar hanz mavjud emas va buni ishlab chiqish ham mantiqsizlik bo'ladi.

O'zbek tilshunosligida razmetka, annotatsiyalash masalasi turli tadqiqotlarda monografik planda o'rganilgan. Ba'zi tadqiqotlarda teglar to'plami taklif etilgan. Teg va teglash masalasini ba'zi ishlarda razmetkalash shaklida kuzatish mumkin [Hamroyeva, 2020, 83]. Sh.Hamroyeva tadqiqotida mukammal teg – keng imkoniyatli, universal korpus garovi ekanligini ta'kidlaydi. Razmetkalash jarayoni uchun bir nechta teg lozimligi, teglarni lingvistik modellashtirish maqsadga muvofiqligi va aynan mana shu modelda morfologik teg shartli qisqartma shaklini olishi ko'rsatilgan. Har bir nutq bo'lagi uchun morfologik va semantik teglarning o'zbek tilida lingvistik modellari taklif qilingan [Hamroyeva, 2020, 205].

O'zbek tilida avtomatik tahlil va tahrir uchun so'z shakllarini formallashtirish va modellashtirish yuzasidan e'lon qilingan ishlarda lingvistik modullar ishlab chiqilgan [Abjalova, 2019, 178]. A.Abjalova tadqiqotida har bir shaklga qo'yilgan belgini razmlar sifatida baholagan. Masalan, ko'plik affiksiga *k\_a*, egalik affiksiga *e\_a*, kelishik affiksiga *ke\_a* va hokazo ramzlarni kiritgan [Abjalova, 2019, 96]. M.Abjalova tadqiqotida grammatik kategoriyalarni formallashtirish uchun modullar taklif qilgan. Formal grammatika nuqtayi nazaridan kategorial shakllarni quyidagicha formallashtirgan: zamon kategoriyasini *z* deb belgilab, o'tgan zamon – *z1*, hozirgi zamon – *z2*, kelasi zamon – *z3* shaklida [Abjalova, 2019, 106]. Shuningdek, har bir kategorial shakl o'z ichida grammatik affiksilariga ko'ra belgilash davom ettirilgan. Oxirgi amalga oshirilgan ishlarda o'zbek tilining morfologik bazasini shakllantirish yuzasidan ham nazariyalar mavjud. O'Xoliyorov fe'l shakllarining morfologik lug'atini tuzib, unda fe'l shakllarining nisbat, zamon, mayl, shaxs-son, bo'lishli-bo'lishsizlik, vazifa shakllari aranjirovkasi uchun lingvistik modellar taklif qilgan [Xoliyorov, 2021, 22]. Masalan, Fe'l so'z turkumi uchun *F* belgisi, fe'l so'z turkumi bo'lishsizlik shakli uchun *bb2* yoki ravishdosh shakli uchun *v2* belgisi, ot so'z turkumi kichraytirish shakli uchun *kich*, birlik shakli uchun *s1* belgisi tavsiya etilgan. O.Abdullayeva tadqiqotida matnda nutq bo'laklarini teglash uchun 65 ta morfologik va 40 ta semantik teg taklif qilgan, foydalanuvchilar uchun qulay bo'lishi, korpus foydalanuvchilarining kutilayotgan keng xalqaro auditoriyasini hisobga olgan holda xalqaro standardlashtirilgan teglarni ham teg jadvalida berilgani ta'kidlangan [Abdullayeva, 2022, 93]. O'zbek tilshunosligida e'lon qilingan ko'plab ilmiy tadqiqot natijalarida teg va tegsetlarning bir-biriga yaqin va o'xshash jihatlari ko'zga tashlanadi. Albatta, oldingi ishlardan namuna olish, oziqlanish holatini kuzatish mumkin. Tadqiqot ishlarida grammatik shakl uchun qaysi tegning belgilanishi emas, balki korpus platformasida PoS teglash masalasining hal qilinganligi, lingvistik tahlillar amalga oshirilganda muhim natijalarning olinishi bilan belgilanishi ham aytib o'tilgan [Abdullayeva, 2022, 71].

O'zbek tilida olimlar tomonidan PoS uchun tegning belgilanishini quyidagi jadvalda ham ko'rish mumkin.

## 2-jadval. Fe'llarda nisbat kategoriyasi linivistik modelining teglanishi

	Aniq nisbat	O'zlik	Majhul	Orttirma	Birgalik
Sh.Hamroyeva	an.n.	o'zl.n.	maj.n.	ort.n.	bir.n.



O'Xoliyorov	n1	n2	n3	n4	n5
M.Abjalova	x1	x2	x3	x4	x5
O.Abdullayeva	Aniqn	O'zn	Majn	Ortn	Birn
Turklang	-	REFL	PASS	CAUS	RECP

Tegsetlarni shakliy tomondan ishlab chiqishda, guvoh bo'lamizki, shu vaqtga qadar uch usuldan foydalanilgan: 1) teglar faqat harflardan iborat; 2) teglar faqat sondan iborat; 3) teglar harf va sonning kombinatsiyasidan iborat. Ammo bunda teglarni abreviaturadan farqlash kerak [Abdullayeva, 2022, 70]. Teglar muayyan kategoriyaning qisqartmasi bo'lishi shart emas, eng muhimi inobatga olinishi lozim bo'lgan jihat ularning sodda, qisqa va eslab qolishga qulay bo'lishi lozim.

Biz yuqoridagi barcha tahlillar asosida chiqargan xulosalarimizga asoslangan holda esda saqlash uchun tushunarli bo'lgan, sodda va qisqa grammatik tegsetni taklif etishga qaror qildik. Ushbu tegsetni ishlab chiqishda, dastlab, o'zbek tilida mavjud bo'lgan barcha grammatik kategoriyalar o'rganib chiqildi va to'plandi. Dunyo miqyosida eng keng tarqalgan tegset CLAWS tegsetlari bo'lgani bois biz ham ulardan oziqlanishga qaror qildik. Teglarni nomlashda esa quyidagi prinsipga asoslanildi:

1. Xalqaro teg o'zbek tilidagi mavjud grammatik kategoriyaga to'liq mos tushsa, o'zgarishsiz qabul qilindi.

2. Xalqaro teg o'zbek tilidagi mavjud grammatik kategoriyaga qisman mos tushsa, uni to'ldirish uchun eslab qolishga qulay qilib inglizcha so'zlar qisqartmasiga tayangan holda yangi teg yasaldi.

I. Xalqaro teg o'zbek tilidagi mavjud grammatik kategoriyani o'zida umuman aks ettirmasa, tegsetda bir tillik prinsipini saqlash maqsad qilingani uchun inglizcha tushunchalarga tayangan holda mutlaqo yangi teg yasaldi.

	<b>Ot so'z turkumi</b>	N
<b>1.1.</b>	<b>Bevosita kuzatish mumkin bo'lgan/ bo'lmaganligiga ko'ra</b>	
1.1.1.	Mavhum ot	NAbs
1.1.2.	Aniq ot	NCnc
<b>1.2.</b>	<b>Atalish/atalmasligiga ko'ra</b>	
1.2.1	Atoqli ot	NP
1.2.2	Turdosh ot	NC (Noun Cognate)
<b>1.3</b>	<b>Atoqli otlarning leksik-semantik ma'no turlari</b>	
1.3.1	Shaxs nomi	NPTitl

1.3.2	Geografik nom	NPGeo
1.3.3	Tashkilot, muassasa, korxonasi nomi	NPOrg
1.3.4	Samoviy yoritgichlar nomi	NPCos
1.3.5	Tarixiy sana yoki bayram nomi	NPDt
1.3.6	Hayvon nomi	NPAnm
1.3.7	Mahsulot nomi	NPPrd
1.3.8	Ilohiy tushuncha nomi	NPDvn
1.3.9	Mansab nomi	NNA
1.3.10	Unvon nomi	NNB
1.3.11	Joy nomlari	NPTop
<b>1.4.</b>	<b>Turdosh otlarning leksik-semantik ma'no turlari</b>	
<b>1.4.1.</b>	<b>Narsa/buyum otlari</b>	NCPr
1.4.1.1	Hayvon va uning a'zolari otlari	NCAAn
1.4.1.2	Oziq-ovqat nomlari	NCPrd
1.4.1.3	O'simlik va uning qismlari otlari	NCPlt
1.4.1.4	Diniy-ilohiy otlar	NCDvn
1.4.1.5	Inson tana a'zolari nomlari	NBTitl
1.4.1.6	Fan nomlari	NCSub
<b>1.4.2.</b>	<b>Shaxs oti</b>	NCPrs
1.4.2.1	Qarindoshlikka ko'ra	NCRI
1.4.2.2	Yoshga ko'ra	NCAg
1.4.2.3	Kasbga ko'ra	NCJb
1.4.2.4	Jinsga ko'ra	NCSx
1.4.2.5	Joyga ko'ra	NCPlc
1.4.2.6	Ijtimoiy holatiga ko'ra	NCSoc
<b>1.5.</b>	<b>Makon/maskan otlari</b>	NCPlc
<b>1.6.</b>	<b>Zamon otlari</b>	NCTm
<b>1.7.</b>	<b>Faoliyat/jarayon otlari</b>	NCEv
<b>1.2.11.</b>	<b>Miqdor otlari (Hisob so'zlar)</b>	NNU
<b>1.3.</b>	<b>Predmetning o'zini yoki to'dasini bildirishiga ko'ra</b>	
1.3.1.	Yakka ot	NN1
1.3.2.	Jamlovchi ot	NN
	<b>Tuzilishiga ko'ra</b>	
1.3.4	Sodda ot	ND
1.3.5.	Qo'shma ot	NCmp
1.3.6.	Juft ot	NCpl
1.3.7.	Takroriy ot	NRep
<b>1.4.</b>	<b>Yasalishiga ko'ra</b>	
1.4.1.	Tub ot	ND0
1.4.2.	Yasama ot	ND1
<b>1.5.</b>	<b>Yasama bo'lsa</b>	

1.5.1.	Affiksatsiya	ND1Af
1.5.2.	Kompozitsiya	ND1Cmp
<b>1.6.</b>	<b>Kategoriyalari</b>	
<b>1.6.1</b>	<b>Kelishik kategoriyasi</b>	Cs (Case)
1.6.1.1	Bosh kelishik	Nom (Nominative)
1.6.1.2	Tushum kelishigi	Acc (Accusative)
1.6.1.3	Qaratqich kelishigi	Gen (Genetive)
1.6.1.4	Jo'nalish kelishigi	Dat (Dative)
1.6.1.5	O'rin-payt kelishigi	Loc (Locative)
1.6.1.6	Chiqish kelishigi	Abl (Ablative)
<b>1.6.2</b>	<b>Egalik kategoriyasi</b>	PossP
1.6.2.1	Birinci shaxs egalik birlik	PossP1s
1.6.2.2	Birinci shaxs egalik ko'plik	PossP1p
1.6.2.3	Ikkinchi shaxs egalik birlik	PossP2s
1.6.2.4	Ikkinchi shaxs egalik ko'plik	PossP2p
1.6.2.5	Uchinchi shaxs egalik birlik	PossP3s
1.6.2.6	Uchinchi shaxs egalik ko'plik	PossP3p
1.7.1	Ko'plik	p
1.7.2	Birlik	s
	<b>Sifat so'z turkumi</b>	JJ
2.1.1.	Oddiy daraja	JJOrd
2.1.2.	Qiyosiy daraja	JJR
2.1.3.	Orttirma daraja	JJT
2.1.4	Ozaytirma daraja	JJMin
<b>2.2</b>	<b>Tuzilishiga ko'ra</b>	
2.2.1.	Sodda sifat	JJD
2.2.2.	Qo'shma sifat	JJCmp
2.2.3.	Juft sifat	JJCpl
2.2.4.	Takror sifat	JJRep
<b>2.3</b>	<b>Yasalishiga ko'ra</b>	
2.3.1.	Tub sifat	JJD0
2.3.2.	Yasama sifat	JJD1
<b>2.4</b>	<b>Yasalgan bo'lsa</b>	
2.4.1.	Affiksatsiya	JJD1Af
2.4.2.	Kompozitsiya	JJD1Cmp
<b>2.5.</b>	<b>Semantik guruhleri</b>	
2.5.1	Xususiyat-holat bildiruvchi sifat	JJFt
2.5.3	Hid bildiruvchi sifat	JJSc
2.5.4	Rang-tus bildiruvchi sifat	JJCol
2.5.5	Maza-ta'm bildiruvchi sifat	JJTs
2.5.6	Shakl bildiruvchi sifat	JJM
2.5.7	O'lchov bildiruvchi sifat	JJS
2.5.8	O'rin bildiruvchi sifat	JJL

2.5.9	Payt bildiruvchi sifat	JJT
<b>III.</b>	<b>Fe'l so'z turkumi</b>	VB
3.1.	<b>Lug'aviy ma'noga ega/ega emasligiga ko'ra</b>	
<b>3.1.1</b>	Mustaqil fe'l	VBI
3.2.1	<b>Lug'aviy ma'noviy ifodasiga ko'ra</b>	
3.2.1.1	Harakat	VBAc (Action)
3.2.1.2	Nutq	VBSp (Speech)
3.2.1.3	Holat	VBCon (Condition)
3.2.1.4	Natijali faoliyat	VBRes (Result)
3.2.1.5	Tafakkur	VBTh (Thinking)
3.2.1.6	Munosabat	VBAtt (Attitude)
<b>3.2.2</b>	<b>Grammatik xususiyatiga ko'ra</b>	
3.2.2.1	O'timli	VBTran (Transitive)
3.2.2.2	O'timsiz	VBIntran (Intransitive)
<b>3.2.3</b>	<b>Tuzilishiga ko'ra</b>	
3.2.3.1	Sodda	VBD
3.2.3.2	Qo'shma	VBComp
3.2.3.3	Juft	VBcpl
3.2.3.4	Takror	VBRep
<b>3.2.4</b>	<b>Yasalish xususiyatiga ko'ra</b>	
3.2.4.1	Tub	VBD0
3.2.4.2	Yasama	VBD1
<b>3.2.5</b>	<b>Yasama bo'lsa</b>	
3.2.5.1	affiksatsiya	VB1Aff
3.2.5.2	kompozitsiya	VB1Cmp
<b>3.3</b>	<b>Nomustaqil fe'l</b>	VBD
<b>3.3.1</b>	<b>Vazifasiga ko'ra</b>	
3.3.1.1	Ko'makchi fe'l	VBH
3.3.1.2	Yetakchi fe'l	VBL
3.3.1.3	To'liqsiz fe'l	VBNotF
3.3.1.4	Ko'makchi fe'lli so'z qo'shilmasi	KFSQ
<b>3.4</b>	<b>Zamon qo'shimchalari</b>	T (Tense)
3.4.1	O'tgan zamon	TPast
3.4.2	Hozirgi zamon	TPres
3.4.3	Kelasi zamon	TFut
<b>3.5</b>	<b>Shaxs-son</b>	Prs
3.5.1	I shaxs birlik	Prs1s
3.5.2	I shaxs ko'plik	Prs1p
3.5.3	II shaxs birlik	Prs2s
3.5.4	II shaxs ko'plik	Prs2p
3.5.5	III shaxs birlik	Prs3s

3.5.6	III shaxs ko'plik	Prs3p
3.6.1	Fe'ning bo'lishli shakli	XY
3.6.2	Fe'ning bo'lishsiz shakli	XNOT
<b>IV.</b>	<b>Son so'z turkumi</b>	NUM
<b>4.1.</b>	<b>Tuzilishiga ko'ra</b>	
<b>4.1.1</b>	Sodda	NUMD
<b>4.1.2</b>	Murakkab	NUMCmp
<b>4.2.</b>	<b>Ma'noviy guruhlariga ko'ra</b>	
4.2.1	Dona son	NUMC (Count)
4.2.2	Jamlovchi	NUMCol (Collection)
4.2.3	Chama son	NUMG
4.2.4	Taqsim son	NUMDiv (Dividing)
4.2.5	Kasr son	MF
4.2.6	Sanoq son	NUM
4.2.7.	Tartib son	MD
<b>V.</b>	<b>Olmosh so'z turkumi</b>	P
<b>5.1</b>	<b>Ma'no turlari</b>	
5.1.1	Kishilik olmosh	PP
5.1.2	Ko'rsatish olmosh	PDem (Demosntrative)
5.1.3	O'zlik olmosh	PRef (Reflexive)
5.1.4	Gumon olmosh	PPred (Prediction)
5.1.5	So'roq olmosh	PQues (Question)
5.1.6	Belgilash olmoshi	PInd (Indefinitive)
5.1.7	Bo'lishsizlik olmoshi	PDis (Distributive)
<b>5.2</b>	<b>Tuzilishiga ko'ra turlari</b>	
5.2.1	Sodda olmosh	PD
5.2.2	Qo'shma olmosh	PCmp
5.2.3	Juft olmosh	PCpl
5.2.4	Takror olmosh	PREp
<b>5.3</b>	<b>Yasalishiga ko'ra</b>	
5.3.1	Tub olmosh	PD0
5.3.2	Yasama olmosh	PD1
<b>5.4</b>	<b>Yasalishiga ko'ra</b>	
5.4.1	Affiksatsiya	PD1Aff
5.4.2	Kompozitsiya	PD1Comp
<b>VI</b>	<b>Ravish so'z turkumi</b>	RR
<b>6.1</b>	<b>Lug'aviy ma'noviy guruhi</b>	
6.1.1	Payt ravishi	RT
6.1.2	O'rin ravishi	RL

6.1.3	Holat ravishi	RCon (Condition)
6.1.4	Miqdor-daraja ravishi	RMes (Measurment)
6.1.5	Maqsad ravishi	RPur (Purpose)
6.1.6	Sabab ravishi	RRs (Reason)
<b>6.2</b>	<b>Tuzilishiga ko'ra</b>	
6.2.1	Sodda ravish	RRD
6.2.2	Qo'shma ravish	RRComp
6.2.3	Juft ravish	RRCpl
6.2.4	Takror ravish	RRRep
<b>6.3</b>	<b>Yasalishiga ko'ra</b>	
6.3.1	Tub ravish	RRD0
6.3.2	Yasama ravish	RRD1
<b>VII</b>	<b>Bog'lovchi</b>	C
7.1	Sof bog'lovchi	C0
7.2	Vazifadosh bog'lovchi	C1
<b>7.3</b>	<b>Teng bog'lovchi</b>	CEq (Equeal)
7.3.1	Biriktiruvchi bog'lovchi	CC
7.3.2	Ayiruv bog'lovchi	CDis (Distributive)
7.3.3	Zidlov bog'lovchi	CCB
7.3.4	Inkor bog'lovchi	CNeg (Negative)
<b>7.4.</b>	<b>Ergashtiruvchi bog'lovchi</b>	CS
7.4.1	Aniqlovchi bog'lovchi	CSDet (Determine)
7.4.2	Sabab bog'lovchi	CSRs (Reason)
7.4.3	Shart bog'lovchi	CSIf (If)
7.4.4	Chog'ishtiruv bog'lovchi	CSCmp (Comparing)
7.4.5	Inkor bog'lovchi	CSNeg (Negative)
7.4.6	Maqsad bog'lovchi	CSPur (Purpose)
<b>VIII</b>	<b>Ko'makchi</b>	II
8.1	Sof ko'makchi	II0
8.2	Vazifadosh ko'makchi	II1
<b>IX</b>	<b>Yuklama</b>	Prt (Particle)
<b>9.1</b>	<b>Tuzilishiga ko'ra</b>	
9.1.1	So'z yuklama	PrtD
9.1.2	Qo'shimchasimon yuklama	PrtComp
<b>9.2</b>	<b>Ma'no turi</b>	
9.2.1	So'roq va taajjub yuklama	PrtQ
9.2.2	Kuchaytiruv-ta'kid yuklama	PrtEm (Emphasis)
9.2.4	Ayirish-chegaralash yuklama	PrtB
9.2.5	O'xshatish-qiyoslash yuklama	PrtCmp
9.2.6	Gumon yuklama	PrtPred

9.2.7	Inkor yuklama	PrtNeg
<b>9.3</b>	<b>Vazifasiga ko'ra</b>	
9.3.1	Sof yuklama	Prt0
9.3.2	Vazifadosh yuklama	Prt1
<b>X</b>	<b>Modal so'z</b>	MD
<b>10.1</b>	<b>Ma'nosiga ko'ra</b>	
10.1.1	Ishonch	MDTr (Trust)
10.1.2	Gumon	MDPred (Predicting)
10.1.3	Tasdiq ma'nosi	
10.1.4	Fikrni xulosalash	MDSum (Summary)
10.1.5	Fikrning tartibi	MDOOr (Order)
10.1.6	Fikrning dalillanishi	MDEvd (Evidence)
10.1.7	Achinish	MDPt (Pity)
10.1.8	Quvonch	MDJ (Joy)
10.1.9	Tasdiq so'zlar	MDCnf (Confirmation)
10.1.10	Inkor so'zlar	MDDnl (Denial)
10.1.11	Mavjudlik	MDEx (Existence)
10.1.12	Zaruriyat	MDNec (Necessety)
10.1.13	Ruxsat	MDPer (Permission)
10.1.14	Esga olish	MDRm (Rememebering)
<b>10.2</b>	<b>Tabiatiga ko'ra</b>	
10.2.1	Sof	MD0
10.2.2	Vazifadosh	MD1
<b>XI</b>	<b>Undov so'z</b>	UH
<b>11.1</b>	<b>His-hayajon undovlari</b>	UHEm (Emotion)
11.1.1	Quvonch	UHJ
11.1.2	Erkalash-erkalanish	UHSp (Spoiling)
11.1.3	Rohatlanish	UHEj (Enjoyment)
11.1.4	Hayron qolish	UHSrp (Surprize)
11.1.5	Norozilik	UHDiss (Dissatisfaction)
11.1.6	G'azab-nafrat	UHH (Hate)
11.1.7	Og'riq	UHPn (Pain)
11.1.8	Do'q-po'pisa	UHTh (Threat)
11.1.9	Afsuslanish	UHReg (Regret)
11.1.10	Cho'chitish, qo'rqitish	UHF (Fear)
11.1.11	Ta'kid-kuchayrtirish	UHEm (Emphasis)
<b>11.2</b>	<b>Buyruq-xitob undovlari</b>	UHImp (Imperative)
11.2.1	Shaxsga qaratilgan	UHPrs (Person)

11.2.2	Hayvon/parrandalarga qaratilgan	UHA <sub>n</sub> (Animal)
11.2.3	Nutqiy-odat	UHD <sub>ly</sub> (Daily)
<b>XII</b>	<b>Taqlid soʻzlar</b>	IM (Imitative)
<b>12.1</b>	Tovushga taqlid	IMV (Voice)
<b>12.2</b>	Holatga taqlid	IMCon

## **Xulosa**

Xulosa qilib aytganda, oʻzbek tilshunosligida teglovchi dasturlar uchun teglarning kelishib olingan va meʼyorga solingan varianti eʼlon qilingunga qadar tadqiqotlarda oʻxshash va farqli teg-modellarni kuzatish mumkin. Taklif qilinayotgan teglar leksik yozuvlardan farq qilishi tabiiy. Hozirda nafaqat grammatik teglar, semantik teglar, sintaktik teglar, shu bilan birga uslubiy va prosodik teglar ham ishlab chiqilishi kerak. Tildagi umumiy qonuniyatlarni oʻrgangan holda, til modelini elektronlashtirish ustida ishlanmoqda. Tillarda yaratilayotgan korpuslar koʻp mehnat va vaqt talab qiladi va aynan mana shu korpuslarning ahamiyati va qiymatini tilda ishlab chiqilgan tegsetlar oshiradi. Biz tadqiqotimizda taklif qilgan tegset koʻplab ilmiy izlanishlar natijasi sifatida vujudga keldi va keyingi tadqiqotlar uchun, korpuslar uchun yechim boʻlib xizmat qiladi.

## **Adabiyotlar:**

- Захаров В.П. Корпусная лингвистика. Санкт-Петербург, 2005. – С. 48.
- Leech G., Wilson A. Recommendations for the morphosyntactic annotation of corpora. // EAGLES Document EAG-TCWG-MAC/R, 1994. [www.ilc.cnr.it/EAGLES/browse.html](http://www.ilc.cnr.it/EAGLES/browse.html).
- McEnery T., Hardie A. Corpus linguistics: Method, theory and practice. Cambridge: Cambridge University Press, 2012.
- Корпусы: аннотированные и неаннотированные. Лингвистическая аннотация (разметка) и метаданные. — Текст: электронный // Myfilology.ru - информационный филологический ресурс: [сайт]. – URL: <https://myfilology.ru//177/korpusy-annotirovannye-i-neannotirovannye-lingvisticheskaya-annotacziya-razmetka-i-metadannye/> (дата обращения: 2.12.2022).
- [https://www.techtarget.com/whatis/definition/SGML-Standard-Generalized-Markup-Language#:~:text=SGML%20\(Standard%20Generalized%20Markup%20Language\)%20is%20a%20standard%20for%20how,It%20is%20metadata.](https://www.techtarget.com/whatis/definition/SGML-Standard-Generalized-Markup-Language#:~:text=SGML%20(Standard%20Generalized%20Markup%20Language)%20is%20a%20standard%20for%20how,It%20is%20metadata.)
- <https://ucrel.lancs.ac.uk/annotation.html#POS>
- <https://knowledge.autodesk.com/support/autocad-plant-3d/learn-explore/caas/CloudHelp/cloudhelp/2018/ENU/Plant3D-UserGuide/files/GUID-1E4D93B3-A6C0-4CF4-BA9A-23EC87522B54-htm.html>



- [https://e-library.namdu.uz/81%20%D0%A2%D0%B8%D0%BB%D1%88%D1%83%D0%BD%D0%BE%D1%81%D0%BB%D0%B8%D0%BA/%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%B0%20%D1%82%D0%B0%D1%80%D0%B6%D0%B8-D0%BC%D0%B0%D1%81%D0%B8/1mitkov\\_ruslan\\_editor\\_the\\_oxford\\_handbook\\_of\\_computational\\_li%20\(1\).pdf](https://e-library.namdu.uz/81%20%D0%A2%D0%B8%D0%BB%D1%88%D1%83%D0%BD%D0%BE%D1%81%D0%BB%D0%B8%D0%BA/%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%B0%20%D1%82%D0%B0%D1%80%D0%B6%D0%B8-D0%BC%D0%B0%D1%81%D0%B8/1mitkov_ruslan_editor_the_oxford_handbook_of_computational_li%20(1).pdf)
- <https://ucrel.lancs.ac.uk/annotation.html#POS>
- <https://ucrel.lancs.ac.uk/claws1tags.html>
- <https://ucrel.lancs.ac.uk/papers/ClawsWordTaggingSystemRG87.pdf>
- <https://ucrel.lancs.ac.uk/claws8tags.pdf>
- <https://www.sketchengine.eu/penn-treebank-tagset/>
- <https://www.ims.uni-stuttgart.de/en/research/projects/textkorpora-werkzeuge/>
- <https://www.sketchengine.eu/english-treetagger-pipeline-2/#toggle-id-1>
- <https://www.learn2tek.org/blog/categorizing-pos-tagging-nltk-python/>
- [https://ucrel.lancs.ac.uk/bnc2sampler/guide\\_c7.htm#m2num](https://ucrel.lancs.ac.uk/bnc2sampler/guide_c7.htm#m2num)
- <https://static.googleusercontent.com/media/research.google.com/ru//pubs/archive/37072.pdf>
- <https://universaldependencies.org/u/pos/>
- Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Монография. – Тошкент, 2020. – 229 б.
- Холийоров О'. О'zbek tili ta'limiy korpusini tuzishning lingvistik asoslari. Filol. fan. bo'yicha falsafa doktori PhD avtoref. – Termiz, 2021.
- Abdullayeva O. O'zbek tilining internet axborot matnlari korpusini shakllantirishning nazariy va amaliy asoslari. filol. fan. bo'yicha falsafa doktori (PhD) .... diss. Andijon, 2022.
- Abdullayeva O., Bozorqulova O'. Korpus lingvistikasida lingvistik tahlillar: muammo va yechimlar // O'zbek tili tadqiqi va ta'limi muammolari. Toshkent, 2021.27.10. 296-302-b.
- Abdullayeva O.X. O'zbek tili korpusini yaratish bosqichlari va muammoli jihatlari // So'z san'ati. – Toshkent, 2021. ISSN- 2181-9297, 4-jild, 1-son. – B. 153-158.

## Evaluation of part of speech tagging in uzbek language: problems and proposals

Botir Elov<sup>1</sup>  
Shahlo Hamroyeva<sup>2</sup>  
Oqila Abdullayeva<sup>3</sup>  
Mohiyaxon Uzoqova<sup>4</sup>

### Abstract:

Speaking of a language corpus, the issue of building a linguistic database becomes the subject of concern because of its complexity and importance at the same time. The process of assigning appropriate identifiers to speech fragments in corpus texts is problematic since language modeling is associated with the rules and patterns of tagging existing in the language. Tagging, especially grammatical tagging or PoS tagging, is also a topical issue for Uzbek corpus linguistics. Because a special “encoded” symbol system serves as the primary key in solving NLP problems related to the Uzbek language. The article analyzes the studies of tagging and PoS tagging in world linguistics and considers the current tagging process in Uzbek linguistics. Based on the rules of the Uzbek language, an alternative

---

<sup>1</sup>*Elov Botir Boltayevich* – doctor of philosophy in technical sciences (PhD), associate professor. Head of the Department of Computer Linguistics and Digital Technologies of Tashkent State University of Uzbek Language and Literature named after Alisher Navoi.

**E-mail:** elov@navoiy-uni.uz

**ORCID:** 0000-0001-5032-6648

<sup>2</sup>*Hamroyeva Shahlo Mirdjanovna* – doctor of philology(DSc), associate professor of the Department of computer linguistics and digital technologies of Tashkent State University of Uzbek language and Literature named after Alisher Navoi.

**E-mail:** shaxlo.xamrayeva@navoiy-uni.uz

**ORCID:** 0000-0002-5429-4708

<sup>3</sup>*Abdullayeva Oqila Xolmo‘minovna* – doctor of philosophy in philology(PhD), Senior teacher. Senior lecturer of the Department of Computer Linguistics and Digital Technologies of Tashkent State University of Uzbek Language and Literature named after Alisher Navoi.

**E-mail:** abdullayeva.oqila@navoiy-uni.uz

**ORCID:** 0000-0002-2524-4832

<sup>4</sup>*Uzoqova Mohiyaxon Tuyg‘un qizi* – Graduate student of the Department of Computer Linguistics and Digital Technologies of Tashkent State University of Uzbek Language and Literature named after Alisher Navoi.

**E-mail:** uzoqovamohiyaxon@navoiy-uni.uz

**ORCID:** 0000-0001-7102-0824

**For reference:** Elov B., Hamroyeva Sh., Abdullayeva O., Uzoqova M. 2022. “Evaluation of Part of speech tagging in uzbek language: problems and proposals”. *Uzbekistan: language and culture. Applied philology.* 2 (5): 51-68.

set of tags was proposed taking sets of tags widely used in the world into consideration.

**Keywords:** *Tag, markup, annotation, tagset, NLP, corpus, Brown's corpus, CLAWS, PoS tagging.*

**References:**

- Zaxarov V.P. Korpusnaya lingvistika. Sankt-Peterburg, 2005. – 48 s.
- Leech G., Wilson A. Recommendations for the morphosyntactic annotation of corpora. // EAGLES Document EAG-TCWG-MAC/R, 1994. [www.ilc.cnr.it/EAGLES/browse.html](http://www.ilc.cnr.it/EAGLES/browse.html).
- McEnery T., Hardie A. Corpus linguistics: Method, theory and practice. Cambridge: Cambridge University Press, 2012.
- Korpusi: annotirovannie i neannotirovannie. Lingvisticheskaya annotatsiya (razmetka) i metadannie. — Tekst: elektronniy // Myfilology.ru – informatsionniy filologicheskii resurs: – URL: <https://myfilology.ru//177/korpusy-annotirovannye-i-neannotirovannye-lingvisticheskaya-annotatsiya-razmetka-i-metadannie/> (data obrasheniya: 2.12.2022).
- [https://www.techtarget.com/whatis/definition/SGML-Standard-Generalized-Markup-Language#:~:text=SGML%20\(Standard%20Generalized%20Markup%20Language\)%20is%20a%20standard%20for%20how,It%20is%20metadata.](https://www.techtarget.com/whatis/definition/SGML-Standard-Generalized-Markup-Language#:~:text=SGML%20(Standard%20Generalized%20Markup%20Language)%20is%20a%20standard%20for%20how,It%20is%20metadata.)
- <https://ucrel.lancs.ac.uk/annotation.html#POS>
- <https://knowledge.autodesk.com/support/autocad-plant-3d/learn-explore/caas/CloudHelp/cloudhelp/2018/ENU/Plant3D-UserGuide/files/GUID-1E4D93B3-A6C0-4CF4-BA9A-23EC87522B54-htm.html>
- [https://e-library.namdu.uz/81%20%D0%A2%D0%B8%D0%BB%D1%88%D1%83%D0%BD%D0%BE%D1%81%D0%BB%D0%B8%D0%BA/%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%B0%20%D1%82%D0%B0%D1%80%D0%B6%D0%B8%D0%BC%D0%B0%D1%81%D0%B8/1mitkov\\_ruslan\\_editor\\_the\\_oxfordhandbook\\_of\\_computational\\_li%20\(1\).pdf](https://e-library.namdu.uz/81%20%D0%A2%D0%B8%D0%BB%D1%88%D1%83%D0%BD%D0%BE%D1%81%D0%BB%D0%B8%D0%BA/%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%B0%20%D1%82%D0%B0%D1%80%D0%B6%D0%B8%D0%BC%D0%B0%D1%81%D0%B8/1mitkov_ruslan_editor_the_oxfordhandbook_of_computational_li%20(1).pdf)
- <https://ucrel.lancs.ac.uk/annotation.html#POS>
- <https://ucrel.lancs.ac.uk/claws1tags.html>
- <https://ucrel.lancs.ac.uk/papers/ClawsWordTaggingSystemRG87.pdf>
- <https://ucrel.lancs.ac.uk/claws8tags.pdf>
- <https://www.sketchengine.eu/penn-treebank-tagset/>
- <https://www.ims.uni-stuttgart.de/en/research/projects/textkorpora-werkzeuge/>
- <https://www.sketchengine.eu/english-treetagger-pipeline-2/#toggle-id-1>
- <https://www.learntek.org/blog/categorizing-pos-tagging-nltk-python/>
- [https://ucrel.lancs.ac.uk/bnc2sampler/guide\\_c7.htm#m2num](https://ucrel.lancs.ac.uk/bnc2sampler/guide_c7.htm#m2num)
- <https://static.googleusercontent.com/media/research.google.com/ru//>

[pubs/archive/37072.pdf](https://pubs/archive/37072.pdf)

<https://universaldependencies.org/u/pos/>

Hamroyeva Sh. O'zbek tili mualliflik korpusini tuzishning lingvistik asoslari: Monografiya. – Toshkent, 2020. – 229 b.

Xoliyorov O'. O'zbek tili ta'limiy korpusini tuzishning lingvistik asoslari.

Filol. fan. bo'yicha falsafa doktori PhD avtoref. – Termiz, 2021.

Abdullayeva O. O'zbek tilining internet axborot matnlari korpusini shakllantirishning nazariy va amaliy asoslari. filol. fan. bo'yicha falsafa doktori (PhD) ... diss. Andijon, 2022.

Abdullayeva O., Bozorqulova O'. Korpus lingvistikasida lingvistik tahlillar: muammo va yechimlar // O'zbek tili tadqiqi va ta'limi muammolari. Toshkent, 2021.27.10. 296-302-b.

Abdullayeva O.X. O'zbek tili korpusini yaratish bosqichlari va muammoli jihatlari // So'z san'ati. – Toshkent, 2021. ISSN- 2181-9297, 4-jild, 1-son. – B. 153-158.

## Tabiiy tilni qayta ishlashda soʻzlar orasidagi masofani aniqlash algoritmlaridan foydalanish

Nizomaddin Xudayberganov<sup>1</sup>  
Shaxboz Hasanov<sup>2</sup>

### Abstrakt:

Soʻzlar orasidagi masofa – soʻzlarning turli sifatlariga koʻra tavsiflanishi mumkin. Ularning aynan tarkibiy qismiga koʻra oʻxshashligini aniqlash uchun tabiiy tilni qayta ishlash yoʻnalishida turli algoritmlar taklif qilinadi. Bu algoritmlar oʻzining ishlatilish oʻrinlari hamda unumdorligi bilan bir-biridan farq qiladi. Asosan sohaning imlo tekshiruvchi, nutqni aniqlash hamda plagiatni aniqlash kabi yoʻnalishlarida foydalanish uchun qoʻllaniladi. Ushbu maqolada birdan ortiq soʻzlarning oʻxshashlik masofasini aniqlashda foydalaniladigan algoritmlar va ularga xos xususiyatlar tahlil qilinadi.

**Kalit soʻzlar:** *Hamming masofasi, Levenshteyn masofasi, Kosinus oʻxshashligi, oʻxshashlik jadvali, kodlash nazariyasi.*

### Kirish

Tabiiy tilni qayta ishlash hamda mashinali oʻrganish yoʻnalishlarida masofaviy oʻlchovlar muhim rol oʻynaydi. Masofa oʻlchovi-bumuammoli sohadagi ikkita obyekt oʻrtasidagi nisbiy farqni umumlashtiruvchi qiymat hisoblanadi. Odatda, bu ikkita obyekt tavsiflovchi maʼlumotlar qatoridan foydalanib maxsus algoritmlar yordamida aniqlanuvchi qiymat sifatida qaraladi. Maʼlumotlarning turlariga qarab turli xil masofa oʻlchovlari tanlanishi va ishlatilishi kerak. Shunday qilib, turli xil mashhur masofa oʻlchovlarini va natijada olingan qiymatlarni hisoblashni bilish muhimdir. Soʻzlar orasidagi oʻxshashlik masofasi – ularning mazmunidan koʻra shaklan

---

<sup>1</sup>Xudayberganov Nizomaddin Uktambov oʻgʻli – Alisher Navoiy nomidagi Toshkent davlat oʻzbek tili va adabiyoti universiteti Kompyuter lingvistikasi va raqamli texnologiyalar kafedrasida oʻqituvchisi.

**E-pochta:** nizomaddin@navoiy-uni.uz

**ORCID:** 0000-0002-6213-3015

<sup>2</sup>Hasanov Shaxboz Gʻolibjon oʻgʻli – Toshkentdagi INHA universiteti talabasi.

**E-pochta:** shaxbozhsh@gmail.com

**ORCID:** 0000-0002-2368-2863

**Iqtibos uchun:** Xudayberganov N., Hasanov Sh. 2022. “Tabiiy tilni qayta ishlashda soʻzlar orasidagi masofani aniqlash algoritmlaridan foydalanish”. *Oʻzbekiston: til va madaniyat. Amaliy filologiya*. 2 (5): 69-83.

bir-biriga qanchalik yaqinligini ko'rsatuvchi qiymat. Maqolada quyidagi algoritmlar orqali ushbu qiymatni aniqlash hamda undan qanday foyda olish ko'rib chiqiladi.

1. Hamming masofasi.
2. Levenshteyin masofasi.
3. Kosinus o'xshashligi.

Mavjud nazariyalarni o'rganish davomida shu narsa aniqlandiki, o'zbek tilidagi so'zlar raqamli ko'rinishga olib kelinishi lozim. Ushbu qiymatlar uchun Python tilidan foydalanish maqsadga muvofiq.

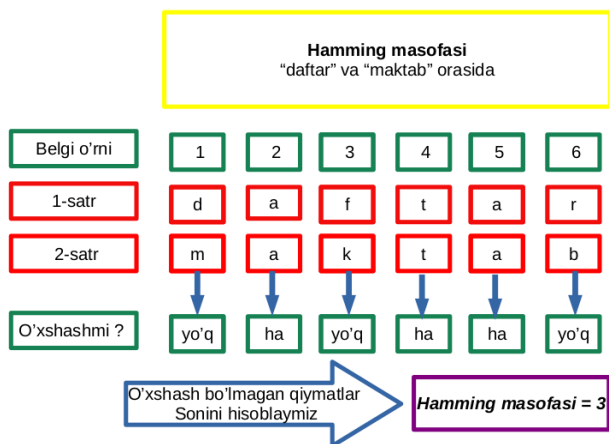
### **Asosiy qism**

Berilgan satrlarni taxminan moslashtirish, shuningdek, katta hajmdagi matn orasidan kerakli ma'lumotlarni qidirish kabilarni amalga oshirish mumkin, bu esa o'z navbatida imlo tekshiruv, maxsus belgilarni aniqlash uchun tuzatish tizimlari, nutqni aniqlash, spamni filtrlash, to'g'ri yozuvni tekshirish kabi turli xil ilovalarda qo'llaniladi. Bundan tashqari tabiiy tilni tushunishda aniqlikni oshirish, plagiatni aniqlash kabi yo'nalishlarda samarali qo'llash mumkin bo'ladi. Matnlardagi plagiat akademik hamjamiyatni tashvishga solayotgan masalalardan biri hisoblanadi. Endi eng keng tarqalgan matn plagiatni so'zlarni kiritish, o'chirish yoki almashtirishni o'z ichiga olgan turli xil kichik o'zgartirishlar kiritish orqali yuzaga keladi. Biroq, bunday oddiy o'zgarishlar plagiatni aniqlash jarayonida ortiqcha taqqoslashni talab qiladi. Ushbu maqola orqali jarayonning samaradorligini oshiruvchi bir qancha nazariyalar taqdim qilinadi.

**Hamming masofasi.** Hamming masofasi - tabiiy tilni qayta ishlash yo'nalishida teng uzunlikdagi ikki satr (so'z yoki so'zlar jamlanmasi) orasidagi mos belgilari turli xil bo'lgan o'rinlar sonidir. Boshqacha qilib aytganda, u bir qatorni boshqasiga o'zgartirish uchun zarur bo'lgan o'zgartirishlarning minimal sonini yoki bitta qatorni boshqasiga o'zgartirishi mumkin bo'lgan xatolarning minimal sonini o'lchaydi. Umumiyroq matnda Hamming masofasi ikki ketma-ketlik orasidagi tahrirlash masofasini o'lchash uchun bir nechta qator ko'rsatkichlaridan biridir. Hamming masofasi amerikalik matematik olim Richard Hamming sharafiga nomlangan. 1950-yilda Hamming kodlari, Xatolarni aniqlash va xatolarni tuzatish kodlari haqidagi fundamental maqolasida konsepsiyani kiritgan. Bitlarning Hamming masofasi tahlili axborot texnologiyalari, kodlash nazariyasi va kriptografiya kabi bir qancha fanlarda qo'llaniladi. [Bill, 2020, 206]

Algoritmni tahlil qilish jarayonida berilgan satrlarning mos

belgilarini o'zaro taqqoslash hamda Hamming masofasini aniqlovchi o'zgaruvchi olishimiz va uni doimo o'zgarishlar jarayonida oshirib borishimiz talab qilinadi. Misol qilib "daftar" hamda "maktab" so'zlardan foydalanilganda ular orasidagi Hamming masofasini hisoblash ko'rsatilgan.



### 1-rasm. Hamming masofasini aniqlashga oid misollar

Boshqacha qilib aytganda, "daftar" so'zidan "maktab" so'zini hosil qilishimiz uchun 3 ta belgini o'zgartirish yetarli bo'ladi. Satrlarda belgilar soni tengligi orqali natijaga erishishimiz uchun faqatgina belgilar almashuvini qo'llashimiz yetarli bo'ladi. Xuddi shu usulda quyidagi misollarda ham Hamming masofasi hisoblanadi.

1. "kun" va "tun" bir xil uzunlikdagi so'zlar aynan 1 ta turli belgiga ega demak, ular orasidagi Hamming masofasi 1 ga teng bo'ladi.

2. "uycha" va "uyat" so'zlarida o'zbek tilida ikkala so'z ham 4 ta harfdan iborat, lekin belgilar soni turli (mos ravishda 5 ta va 4 ta) bo'lganligi sababli Hamming masofasini aniqlashning iloji yo'q.

3. 00010011 va 01010111 ikkilik sanoq sistemasiga tegishli bo'lgan ushbu sonlar orasidagi 3-va 6-belgilar farq qilganligi sababli Hamming masofasi 2 ga teng.

O'zgarmas n uzunlikdagi satr uchun Hamming masofasi n uzunlikdagi so'zlar to'plamining ko'rsatkichidir (hamming fazosi deb ham ataladi). Chunki u musbat va simmetriya shartlarini bajaradi. Agarda ikki so'zning Hamming masofasi 0 ga teng bo'lsa, ular aynan bir xil bo'ladi. Bundan tashqari ushbu qiymatlar uchburchak tengsizligini ham qanoatlantiradi. Haqiqatan ham, agar uchta a, b va c so'zlari ko'riladigan bo'lsa, a satrning i-o'ringidagi qiymati

b satrning i-o'rindagi qiymati va c satrning xuddi shu qiymatlari solishtiriladigan bo'lsa nazariya to'g'riligini ko'rish mumkin bo'ladi. Demak, a va c orasidagi Hamming masofasi a va b hamda b va c orasidagi Hamming masofalarining yig'indisidan katta emas. Ikkita ikkilik sanoq sistemasidagi a va b qiymatlar uchun Hamming masofasi a XOR b dagi birlar soniga teng. Ushbu qiymat esa kodlash nazariyasida xatolikni aniqlash va xatolarni tuzatish yo'nalishida foydalaniladi [Derek, 2003, 255].

Tabiiy tilni qayta ishlash yo'nalishida qisqa uzunlikdagi so'zlar orasidagi masofalarni aniqlash hamda ularni to'g'irlash orqali foydalanish maqsadga muvofiq bo'ladi. Bundan tashqari kodlash nazariyasi sohasida ham ikkilik sanoq sistemasida berilgan xatoliklarni aniqlash va bartaraf etish uchun samarali yechim bo'la oladi. Katta hajmdagi matn yoki hujjatlar uchun foydalanish uchun yuqori samaradorlikka ega emas.

#### **Psevdokodi**

```
function hamming_masofasi(a,b)
    indicator ← 0
    len1 ← length(a)
    for n from 1 to len1 do
        if a[n] <> b[n] then
            indicator ← indicator + 1
    return indicator
end
```

#### **Algoritmining Python dasturlashga tatbiq qilinishi.**

Hamming masofasini hisoblovchi funksiya yaratiladi hamda unga 2 ta satr elementlari qabul qilinadi. Funksiya ishlash jarayonida berilgan qiymatlarning har bir belgisini o'zaro tekshiradi hamda yakuniy qiymatni qaytaradi.

```
def hamming_masofasi (satr1, satr2):
    masofa_hisoblovchi = 0
    for n in range(len(satr1)):
        if satr1[n]!=satr2[n]:
            masofa_hisoblovchi += 1
    return masofa_hisoblovchi
```

**Levenshteyin masofasi.** Levenshteyin masofasi (**Levenshteyin distance**) — tabiiy tilni qayta ishlash yo'nalishida Levenshteyin masofasi ikki satr (bir yoki bir nechta so'zlar) o'rtasidagi masofani o'lchash uchun qo'llanadigan ko'rsatkichidir. Boshqacha qilib aytganda, ikki so'z orasidagi Levenshteyin masofasi bir so'zni boshqasiga o'zgartirish uchun zarur bo'lgan tahrirlarning



(qo'shish, o'chirish yoki almashtirish) minimal soni hisoblanadi. Bu masofani 1965-yilda Vladimir Levenshteyin sharafiga nomlangan. Levenshteyin masofasini **tahrirlash masofasi** deb ham atash mumkin. Vladimir Iosifovich Levenshteyin - rus olimi bo'lib, u axborot texnologiyalari, xatolarni to'g'rilash nazariyalari va kombinator dizayni bo'yicha tadqiqotlar olib borgan. Fanga qo'shgan boshqa hissalar orasida u 1965-yilda ishlab chiqqan Levenshteyin masofasi va Levenshteyin algoritmi mashhur [Levenshteyin, 1966].

Demak algoritm a satrni b satrga aylantirish uchun zarur bo'lgan o'zgarishlarning minimal sonini aniqlashda foydalaniladi. Ushbu jarayonni amalga oshirishda 3 xildagi amallar qo'llaniladi: belgi qo'shish, o'chirish hamda almashtirish. Ushbu algoritmda satrlar uzunligi bir xil bo'lishi talab qilinmaydi. Ushbu nazariyaning samaradorligini oshirish maqsadida pastdan yuqoriga dinamik dasturlashning namunasi bo'lgan algoritm 1974-yilda Robert A. Vagner va Maykl J. Fisherning "Satrdan-satrga tuzatish muammosi" maqolasida muhokama qilingan.

Levenshteyin masofasidan tabiiy tilni qayta ishlash yo'nalishida so'zlarning o'zaro mosligini aniqlashda foydalanish mumkin. Bunda asosiy maqsad kam sonli farqlar kutilgan vaziyatlarda, berilgan matnlardagi satrlar o'xshashligini aniqlash bo'ladi. Bu esa masalan, imlo tekshiruv, optik belgilarni aniqlash uchun tuzatish tizimlari va tarjima xotirasi asosida tabiiy tilga tarjima qilishga yordam beradigan dasturiy ta'minotlar uchun asosiy qismlardan biri sifatida xizmat qila oladi. Optik belgilarni aniqlash yoki optik belgilarni o'quvchi dasturlar - skanerlangan hujjat, qo'lyozma fotosurati yoki subtitrdan bo'ladimi, qo'lda yozilgan yoki bosilgan matn tasvirlarini mashina tiliga elektron yoki mexanik aylantirish jarayonlaridan tashkil topadi. Bundan tashqari ushbu o'xshashlikdan plagiatni tekshirish kabi maqsadlarda foydalanish muvaffaqiyatli natija beradi. Bundan tashqari taxminiy satrlarni moslashtirish jarayonida ham keng qo'llaniladi. Taxminiy satrlarni moslashtirish (ko'pincha so'zlashuv tilida noaniq satrlarni qidirish deb ataladi) aniq bo'lmagan ko'rinishdagi namunalariga mos keladigan satrlarni topish usulidir. Taxminiy satrlarni moslashtirish muammosi odatda ikkita kichik muammoga bo'linadi: berilgan satr ichida taxminiy namuna mosliklarini topish va namunaga taxminan mos keladigan satrlarni topish.

Levenshteyin masofasi bir nechta oddiy yuqori va quyi chegaralarga ega. Bularga quyidagilar kiradi:

1. Berilgan ikki satrlar uzunliklari orasidagi farq.

2. Uzunligi kattaroq bo'lgan satrning uzunlik qiymatini aniqlovchi birlik.

3. Agar satrlar bir xil uzunlikka ega bo'lsa ular orasidagi farq 0 ga teng.

4. Agar satrlar bir xil o'lchamga ega bo'lsa, Hamming masofasi Levenshteyin masofasining yuqori chegarasi hisoblanadi.

5. Ikki satr orasidagi Levenshteyin masofasi ularning uchinchi satr bilan o'zaro Levenshteyin masofalarining yig'indisidan katta emas (uchburchak tengsizligi o'rinli).

Levenshteyin masofasini ikkita uzunroq satrlar orasida ham hisoblash mumkin, ammo uni hisoblash uchun sarflanadigan vaqt, bu ikki satrlar uzunligiga taxminan proporsional bo'lib, buni amaliy jarayonlarda foydalanish nisbatan kamroq unumdorlikka sabab bo'ladi. Shunday qilib, taqqoslash tezligini yaxshilash maqsadida odatda qisqa satrlardan foydalaniladi.

Quyidagi misollar orqali nazariy bilimlarni amaliy ko'rinishga olib kelish mumkin.

1. "olma" so'zidan "olmaxon" so'zini hosil qilish uchun so'zning davomiga "x", "o" va "n" harflarini qo'shish orqali natijaga erishiladi. Bunda umumiy 3 o'zgarish, ya'ni qo'shish amalga oshirildi. Demak Levenshteyin masofasi 3 ga teng bo'ladi.

2. Yuqoridagi misolni teskari holati qaraladigan bo'lsa, "olmaxon" so'zidan "olma" so'zini hosil qilish uchun 3 o'zgarishni, ya'ni so'z oxiridagi 3 harfni o'chirishni amalga oshirish talab qilinadi. Bu vaziyatda ham Levenshteyin masofasi 3 ga teng bo'ladi.

3. "men" so'zi xatolik tufayli "man" kabi yozilgan vaziyatda Levenshteyin masofasi nechta o'zgartirish orqali to'g'ri so'zga erishish mumkinligini ko'rsatadi. Bu vaziyatda faqat "a" harfini "e" harfiga o'zgartirish yetarli. Demak Levenshteyin masofasi 1ga teng.

Endi umumiy vaziyatda Levenshteyin masofasini yanada chuqurroq tushunish maqsadida quyidagi o'zbek tiliga oid bo'lgan "matematika" hamda "maktab" so'zlari ko'rib chiqiladi. Berilgan "matematika" so'zidan "maktab" so'zini hosil qilish uchun o'zaro mos tushuvchi harflarni belgilab olish hamda boshqa belgilar uchun o'zgarishlarni amalga oshirish lozim bo'ladi.

1. M, A harflaridan so'ng K harfini qo'shish (1 ta o'zgarish)

2. 1- T harfidan so'ng E, M harflarini o'chirish (2 ta o'zgarish)

3. 2- T harfini B harfiga almashtirish (1 ta o'zgarish)

4. 2- T harfidan so'ng I, K va A harflarini o'chirish (3 ta o'zgarish)

			BELGI QO'SHISH		BELGI O'CHIRISH	BELGI O'CHIRISH		BELGI ALMASHTIRISH	BELGI O'CHIRISH	BELGI O'CHIRISH	BELGI O'CHIRISH
a=	M	A		T	E	M	A	T	I	K	A
b=	M	A	K	T			A	B			

## 2-rasm. Levenshteyin masofasini aniqlashga oid misollar

O'zgarishlar soni: 1ta belgi qo'shish, 1ta belgi almashtirish va 5ta belgini o'chirish. Demak, Levenshteyin masofasi = 7ga teng bo'ladi.

**Levenshteyin jadvali** - bu 2 o'lchamli massiv bo'lib,  $lev[i][j]$  - Levenshteyin masofasining qiymatiga teng bo'ladi. Bunda:  $i = 'a'$  satr uzunligi va  $j = 'b'$  satr uzunligi [Grashchenko, 2022]

$$lev_{a,b}(i, j) = \begin{cases} \max(i, j) & \text{Agar } \min(i, j) = 0, \\ \min \begin{cases} lev_{a,b}(i-1, j) + 1 \\ lev_{a,b}(i, j-1) + 1 & \text{aks holda} \\ lev_{a,b}(i-1, j-1) + 1 & (a_i \neq b_j) \end{cases} \end{cases}$$

## 3-rasm. Levenshteyin jadvalini hosil qilish qoidasi

Berilgan formula orqali Levenshteyin jadvali hosil qilinadi.

	i→	M	A	T	E	M	A	T	I	K	A
j↓	0	1	2	3	4	5	6	7	8	9	10
M	1	0	1	2	3	4	5	6	7	8	9
A	2	1	0	1	2	3	4	5	6	7	8
K	3	2	1	1	2	3	4	5	6	6	7
T	4	3	2	1	2	3	4	4	5	6	7
A	5	4	3	2	2	3	3	4	5	6	6
B	6	5	4	3	3	3	4	4	5	6	7

## 4-rasm. Levenshteyin jadvali

### Pseudokodi

```
function LEVDIST(a,b)
  if a.size() == 0 then
    return b.size();
  end
  if b.size() == 0 then
    return a.size();
  end
  indicator ← a[0] 0 ≠ b[0] ? 1 : 0;
  return Minimum(
```

```

LEVDIST(a.substr(1),b)+1, //
o'chirish;
LEVDIST(b.substr(1),a)+1, //
qo'shish;
LEVDIST(a.substr(1),b.substr(1))+indicator); //
almashtirish;
end

```

**Python dasturlash tiliga tatbig'i.** Avvalambor, berilgan satrlar uzunligi orqali maxsus Levenshteyin jadvalimizni hosil qilamiz. Keyingi qadamlarni kerakli bo'lgan qiymatlarni hisoblash hamda ularni maxsus jadvalga saqlash orqali olib boriladi.

```

#Levenshteyn algoritmi
# 'a' va 'b' qiymatlarni aniqlash:
a = "matematika"
b = "maktab"

# lev o'zgaruchi massivini rows = len(a) + 1 va
columns = len(b) + 1 orqali kiritish
lev = [[0 for i in range(len(b) + 1)] for j in
range(len(a) + 1)]

# Birinchi qatordagi(row) sonlarni aniqlash:
for i in range(len(a) + 1):
lev[i][0] = i
# Birinchi ustundagi(column) sonlarni aniqlash:
for j in range(len(b) + 1):
lev[0][j] = j

for i in range(1, len(a) + 1):
    for j in range(1, len(b) + 1):
        if a[i - 1] == b[j - 1]:
            lev[i][j] = lev[i - 1][j - 1]
        else:
            insertion = 1 + lev[i][j - 1]
            deletion = 1 + lev[i - 1][j]
            replacement = 1 + lev[i - 1][j - 1]

```

# Eng foydali o'zgarishni

tanlash:

lev[i][j] = min(insertion,

deletion, replacement)

#Levenshtein Masofasini ekranga chiqar

print("Levenshtein Masofasi: ", lev[len(a)][len(b)])

**Kosinus o'xshashligi** – bu ikki obyekt bir-biriga qanchalik o'xshashligini aniqlash usuli. Bunda obyektlarning maxsus xususiyatlariga va ushbu xususiyatlarning qancha miqdoriga ega ekanligi qaraladi. Ikki obyekt o'xshash, agar ularning xususiyatlari va mos miqdorlari o'xshash bo'lsa. Kosinus o'xshashligining qiymatini hisoblashda chiziqli algebradan foydalaniladi. Demak, obyektlarning maxsus qiymatlari ularga mos son ko'rinishiga olib kelinadi hamda obyekt esa ushbu sonlar umumiydashgan maxsus vektor sifatida qaraladi va kosinus o'xshashligi vektorlar orasidagi burchakning kosinusi orqali aniqlanadi (xususiyatlar soniga qarab bu vektorlar n o'lchamli fazoda bo'lishi mumkin). Bundan kelib chiqadiki, kosinus o'xshashligi qiymati doimo [-1,1] oralig'iga tegishli bo'ladi. Kosinus o'xshashligi 0ga teng bo'lganda (burchak 90) hech qanday o'xshashlik topilmaydi, kosinus o'xshashligi 1ga yaqinlashgan sari vektorlar orasidagi burchak kichrayadi va o'xshashlik ortadi, aksincha burchak ortgan sari kosinus o'xshashligi qiymati kamayadi, natijada ular o'zaro farqli ekanligi aniqlanadi. Kosinus o'xshashligi murakkablik darajasi yuqori bo'lmaganligi va muammoga tatbiq qilinish jarayoni oson bo'lganligi bois turli sohalarda qo'llanilib keladi [<https://en.wikipedia.org>].

Matnlarning (hujjatning) o'xshashligini aniqlashda yuqoridagi berilgan usullarga nisbatan ancha samaraliroq usul hisoblanadi. Bundan tashqari matnlar o'xshashligini tekshirish plagiatni aniqlash hamda oldini olishda ham keng foydalaniladi. Bu esa mualliflik huquqini himoya qilish uchun ham samarali usul sifatida ko'rilishi mumkin. Maktab, litsey, kollej va universitet talabalarining vazifalar va imtihonlar jarayonida topshirgan ishlaridan ularning bu vazifalarga yondashuvini aniqlash mumkin bo'ladi. Kosinus o'xshashligi ma'lumotlarni qazib olish, axborot qidirish va matnni moslashtirish jarayonlarida ham qo'llaniladi. Vektorlar o'zgaruvchining xususiyatlariga tayinlangandan so'ng, o'lchov obyektlar orasidagi o'xshashlikni tushunish uchun qimmatli vositaga aylanadi. Ma'lumotni qazib olish - bu ma'lumotlarni tahlil qilish orqali biznes muammolarini hal qilishga yordam beradigan

namuna va munosabatlarni aniqlash uchun katta ma'lumotlar to'plamini saralash jarayonidir. Ma'lumotni qazib olish texnikasi va vositalari korxonalariga kelajakdagi tendensiyalarni bashorat qilish va ko'proq ma'lumotga ega biznes qarorlar qabul qilish imkonini beradi. Axborot qidirish - bu hujjatdagi ma'lumotlarni qidirish, hujjatlarning o'zini qidirish, shuningdek, ma'lumotlarni tavsiflovchi metama'lumotlarni, matnlar, tasvirlar yoki tovushlar ma'lumotlar bazalaridan qidirish haqidagi fandır.

**Xususiyatlari.** Kosinus o'xshashligining eng diqqatga sazovor xususiyati shundaki, u individual vektor o'lchamlarini mutlaq emas, balki nisbiy taqqoslashni aks ettiradi. Har qanday konstanta  $a$  va vektor  $V$  uchun  $V$  hamda  $aV$  vektorlari maksimal darajada o'xshashdir. Shunday qilib, o'lchov chastotasi mutlaq qiymatlardan muhimroq bo'lgan ma'lumotlar uchun mos keladi. Biroq, Jensen-Shannon, SED va ucburchak divergensiya kabi axborot nazariyasiga asoslangan so'nggi o'lchovlar hech bo'lmaganda ba'zi kontekstlarda yaxshilangan vaziyatda barqaror semantikaga ega ekanligi ko'rsatiladi.

Kosinus o'xshashligi Evklid masofasi bilan bog'liq ekanini ham ta'kidlab o'tish lozim. Vektorlar orasidagi Evklid masofasi, ular ichidagi kvadrat qiymatlarning birlik yig'indisining normallashtirilgan holati sifatida aniqlanadi. [Connor, 2016]

Kosinus o'xshashligi quyidagi maxsus matematik formula orqali aniqlanadi.

$$\text{cosine similarity} = S_C(A, B) := \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

### 5-rasm. Kosinus o'xshashligini aniqlovchi matematik formula

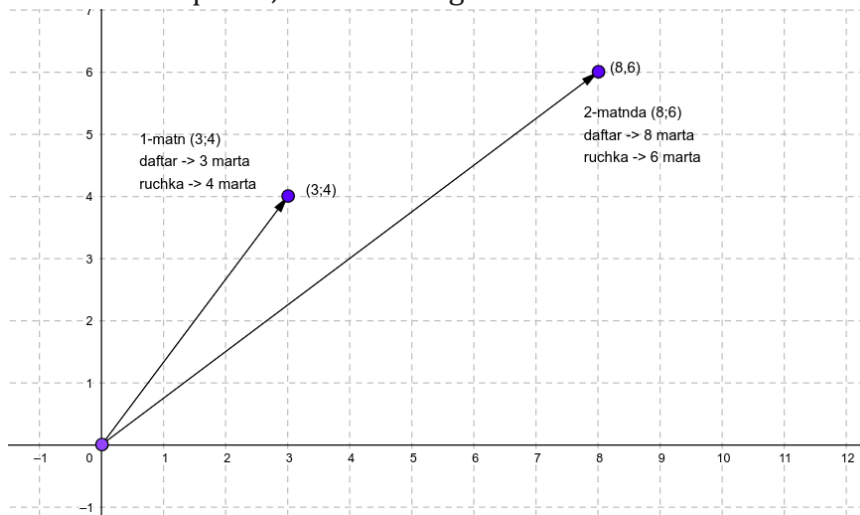
Quyida matnlar faqat 2 so'zdan tashkil topgan 2 xil matn va ularda qatnashgan so'zlarning soniga nisbatan 2 o'lchamli dekart koordinatalar sistemasida vektorlar sifatida tasvirlangan (berilgan misol tasavvur qilish va tasvirlash oson bo'lishi uchun tanlangan).

1- matn: {ruchka, daftar, daftar, ruchka, daftar, ruchka, ruchka}

2- matn: {daftar, ruchka, daftar, daftar, ruchka, daftar, daftar, ruchka, ruchka, ruchka, daftar, daftar, daftar} dan tashkil topgan bo'lsin.

So'zlarni matnda uchragan o'rinlariga nisbatan mos sonlar bilan almashtiramiz va maxsus vektorlarni hosil qilib olamiz.

1- matnda 3 ta daftar va 4 ta ruchka uchragani bois, daftar 3 va ruchka so'zi esa 4 soni bilan almashtiriladi hamda (3;4) nuqtasi hosil qilinadi. Xuddi shunday usulda 2-matndan mos ravishda (8;6) nuqtasi hosil qilinadi. So'ngra koordinatalar boshiga nisbatan vektorlar hosil qilinib, ular orasidagi burchak kosinusi hisoblanadi.



$$\text{Kosinus o'xshashligi} = \frac{3 \cdot 8 + 4 \cdot 6}{\sqrt{3^2 + 4^2} \cdot \sqrt{8^2 + 6^2}} = \frac{24}{25} = 0.96 \text{ ga teng.}$$

Bundan tashqari ko'proq so'zlardan tashkil topgan matnlar orqali so'zlarning takrorlanish darajalarini quyidagicha jadvalga mos ravishda joylashtirib chiqiladi.

Hosil bo'lgan so'zlardan tashkil topgan setlar hosil qilinadi hamda ular orqali umumiy kalit so'zlarga ega to'plam yasaladi. Qiymatlarni maxsus vektorlarga aylantirish orqali yuqorida berilgan formuladan foydalanib kosinus o'xshashligi hisoblab chiqiladi [<https://www.itl.nist.gov>].

Hujjat	jamo	murabbiy	jarima	futbol	basketbol	zarba	g'alaba
Hujjat 1	1	1	0	2	1	0	2
Hujjat 2	2	3	0	3	2	1	3
Hujjat 3	1	2	1	2	3	1	0
Hujjat 4	0	2	3	3	1	1	2

6-rasm. Kosinus o'xshashligi jadvali

**“Yumshoq” kosinus o'lchovi** - ikki vektor orasidagi berilgan obyektning xususiyatlar juftligi o'rtasidagi o'xshashlikni hisobga oladi. An'anaviy kosinus o'xshashligi vektor fazo modelining (VFM) xususiyatlarini mustaqil yoki butunlay boshqacha hisoblaydi, yumshoq kosinus o'lchovi esa VFMdagi xususiyatlarining

o'xshashligini hisobga olishni taklif qiladi, bu an'anaviy kosinus o'xshashligiga nisbatan qo'shimcha ma'lumotlar talab qiladi [Connor, 2016].

Masalan, tabiiy tilni qayta ishlash (NLP) sohasida so'zlarning xususiyatlari o'rtasidagi o'xshashlik juda intuitivdir. Shuning uchun ulardan qo'shimcha ma'lumotlar talab qilinishi aniqlik darajasini oshirishga yordam beradi. Misol uchun, "misol" va "boshqotirma" so'zlari tarkiban turli xil so'zlar va shuning uchun VFMning turli nuqtalarida ko'rsatiladi. Biroq ular ma'no jihatidan bir-biriga bog'liqdir.

Yumshoq kosinusni hisoblash uchun s o'zgaruvchi tanlanadi va matritsasi xususiyatlari orasidagi o'xshashlikni ko'rsatish uchun ishlatiladi. N o'lchamli fazoda a va b vektorlari orasida yumshoq kosinus o'xshashligi quyidagicha hisoblanadi:

$$\text{soft\_cosine}_1(a, b) = \frac{\sum_{i,j}^N s_{ij} a_i b_j}{\sqrt{\sum_{i,j}^N s_{ij} a_i a_j} \sqrt{\sum_{i,j}^N s_{ij} b_i b_j}},$$

7-rasm. Formula

$s_{(i,j)} = o'xshashlik(xususiyat_i, xususiyat_j)$  agar xususiyatlar o'rtasida o'xshashlik bo'lmasa 0ga, aks holda 1ga teng. Berilgan tenglama an'anaviy kosinus o'xshashlik formulasiga ekvivalentdir [Sidorov, Velasquez, Stamatatos, 2013, 1].

#### Psevdokodi

```
function CosineDist(a,b):
X_list <- TOKENIZE(LOWERCASE(a));
Y_list <- TOKENIZE(LOWERCASE(b));
X_set <- REMOVE_STOPWORDS(X_list);
Y_set <- REMOVE_STOPWORDS(Y_list);
l1 = CREATE_VECTOR(X_set);
l2 = CREATE_VECTOR(Y_set);
return COSINE(l1,l2)
end
```

#### Kosinus o'xshashligining python dasturlash tiliga tatbig'i

```
l1 = []; l2 = []
# berilgan gaplar to'plamlari
X_set = {'men', 'har', 'kuni', 'maktabga', 'boraman'}
Y_set = {'maktabga', 'bugun', 'men', 'barvaqt', 'bordim'}
# ikkala satrning kalit so'zlarini o'z ichiga olgan
```



to'plam hosil qilamiz

```
rvector = X_set.union(Y_set)
```

```
for w in rvector:
```

```
    if w in X_set: l1.append(1) # maxsus vektor
```

yaratamiz

```
    else: l1.append(0)
```

```
    if w in Y_set: l2.append(1)
```

```
    else: l2.append(0)
```

```
    c = 0
```

```
# kosinus formulasi
```

```
for i in range(len(rvector)):
```

```
    c+= l1[i]*l2[i]
```

```
    cosine = c / float((sum(l1)*sum(l2))**.5)
```

```
print("kosinus o'xshashligi: ", cosine)
```

```
Natijada: kosinus o'xshashligi 0.4
```

## Xulosa

So'zlar orasidagi masofa qiymatini tavsiflovchi bir qancha qiymatlar mavjud. Ulardan Hamming, Levenshteyin masofalari va Kosinus o'xshashligini yuqorida ko'rib chiqqan holda shuni xulosa qilish mumkinki, har bir nazariyada kelib chiqadigan natija hamda ularning samaradorligi orqali o'zining ishlatilish o'rnini aniqlash mumkin. Aynan bir xil belgilar bilan boshlanuvchi so'zlarni tekshirgan vaziyatda, Hamming masofasi boshqalarga nisbatan tabiiy tilni qayta ishlash jarayonida soddaroq hamda samaraliroq bo'ladi. Biroq, satrlar hajmi ortishi hamda farq qiluvchi nuqtalar o'zgarishi bilan Levenshteyin masofasi samaradorlik darajasi yuqoriga ko'tariladi. Bundan tashqari, yuqorida berilgan nazariyalar orasida, katta hajmdagi matn yoki hujjatlar bilan ishlagan vaziyatda Kosinus o'xshashligi eng munosib tanlov bo'ladi. Demak, Hamming masofasini boshlang'ich asos deb qaraladigan bo'lsa, Levenshteyin masofasini uning so'zlarga nisbatan mukammalroq vaziyati, shu bilan birgalikda, Kosinus o'xshashligi esa so'zlar jamlanmasi bo'lgan matnlar yoki hujjatlar bilan ishlash uchun eng munosib nazariya sifatida qaraladi.

## Adabiyotlar:

Waggener Bill. Pulse Code Modulation Techniques. Springer. p. 206. ISBN. Retrieved 13 June, 2020.

Robinson, Derek J. S. (2003). An Introduction to Abstract Algebra. Walter de Gruyter. pp. 255–257. ISBN.

- Levenshteyn, Vladimir I. (February 1966). "Binary codes capable of correcting deletions, insertions, and reversals". *Soviet Physics Doklady*.
- Levenshteyn Distance Computation by Sergey Grashchenko November 16, 2022. <https://www.baeldung.com/cs/Levenshteyn-distance-computation>
- Cosine similarity [https://en.wikipedia.org/wiki/Cosine\\_similarity](https://en.wikipedia.org/wiki/Cosine_similarity)
- Connor, Richard (2016). *A Tale of Four Metrics. Similarity Search and Applications*. Tokyo: Springer.
- Cosine distance, cosine similarity, angular cosine distance, angular cosine similarity. <https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/cosdist.htm>.
- Understanding cosine similarity and its application. Richmond Alake Sep 15, 2020. Connor, Richard (2016). *A Tale of Four Metrics. Similarity Search and Applications*. Tokyo: Springer.
- Sidorov, Grigori; Velasquez, Francisco; Stamatatos, Efstathios; Gelbukh, Alexander; Chanona-Hernández, Liliana (2013). *Advances in Computational Intelligence. Lecture Notes in Computer Science*. Vol.7630. LNAI 7630. pp.1-11.

## The use of word distance algorithms in natural language processing

Nizomaddin Xudayberganov<sup>1</sup>  
Shaxboz Khasanov<sup>2</sup>

### Abstract:

Distance between words - can be characterized according to different qualities of words. Different algorithms are proposed in the direction of natural language processing to determine their similarity according to their component. These algorithms differ from each other in their places of use and performance. It is mainly used in areas such as spell check, speech recognition and plagiarism detection. This article analyzes the algorithms used to determine the similarity distance of more than one word and their characteristics.

---

<sup>1</sup>*Xudayberganov Nizomaddin Uktamboy o'g'li* – Teacher of the Department of Computer Linguistics and Digital Technologies of Tashkent State University of Uzbek Language and Literature named after Alisher Navoi.

**E-mail:** [nizomaddin@navoiy-uni.uz](mailto:nizomaddin@navoiy-uni.uz)

**ORCID:** 0000-0002-6213-3015

<sup>2</sup>*Hasanov Shaxboz G'olibjon o'g'li* – student, The INHA University in Tashkent.

**E-mail:** [shaxbozhsh@gmail.com](mailto:shaxbozhsh@gmail.com)

**ORCID:** 0000-0002-2368-2863

**For reference:** Xudayberganov N., Khasanov Sh. 2022. "The use of word distance algorithms in natural language processing". *Uzbekistan: language and culture. Applied philology*. 2 (5): 69-83.

**Key words:** *Hamming distance, Levenshteyin distance, Cosine similarity, similarity table, coding theory.*

**References:**

- Waggener Bill. Pulse Code Modulation Techniques. Springer. p. 206. ISBN. Retrieved 13 June, 2020.
- Robinson, Derek J. S. (2003). An Introduction to Abstract Algebra. Walter de Gruyter. pp. 255–257. ISBN.
- Levenshteyin, Vladimir I. (February 1966). "Binary codes capable of correcting deletions, insertions, and reversals". Soviet Physics Doklady.
- Levenshteyin Distance Computation by Sergey Grashchenko November 16, 2022. <https://www.baeldung.com/cs/Levenshteyin-distance-computation>
- Cosine similarity [https://en.wikipedia.org/wiki/Cosine\\_similarity](https://en.wikipedia.org/wiki/Cosine_similarity)
- Connor, Richard (2016). A Tale of Four Metrics. Similarity Search and Applications. Tokyo: Springer.
- Cosine distance, cosine similarity, angular cosine distance, angular cosine similarity. <https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/cosdist.htm>.
- Understanding cosine similarity and its application. Richmond Alake Sep 15, 2020. Connor, Richard (2016). A Tale of Four Metrics. Similarity Search and Applications. Tokyo: Springer.
- Sidorov, Grigori; Velasquez, Francisco; Stamatatos, Efstathios; Gelbukh, Alexander; Chanona-Hernández, Liliana (2013). Advances in Computational Intelligence. Lecture Notes in Computer Science. Vol.7630. LNAI 7630. pp.1–11.

## KORPUS LINGVISTIKASI

## Fe'l grammatik shakllarining uslubiy xoslanishiga ko'ra lingvistik annotatsiyalash masalasi xususida

Dilrabo Elova<sup>1</sup>

### Abstrakt:

Jahon tilshunosligida jadal rivojlanayotgan kompyuter lingvistikasi yo'nalishida tabiiy tilga ishlov berish, so'rov berish va javob olish sifatini yaxshilash, lingvistik tadqiqot, lingvodidaktika, leksikografiya korpuslardan foydalanish dolzarb masalaga aylandi. Tilshunoslikda, xususan, kompyuter lingvistikasi sohasida korpus birliklarini tanlash, ularni lingvistik, jumladan, morfologik, semantik, sintaktik va stilistik teglash, stilistik teglar tizimini ishlab chiqish, til korpusi uchun maxsus axborotlarni qamrab olgan uslubiy xoslanishni farqlovchi lingvistik ta'minot yaratish muammosi dolzarb bo'lib qolmoqda.

Fe'l so'z turkumi grammatik kategoriyalarga boy so'z turkumi bo'lib, ana shu shakllar lug'aviy birliklar singari turli uslubiy ma'nolarni anglatadi. Birgina fe'lning o'zida ish-harakat jarayoni va uning tavsifi, harakatning tez yoki sekin bajarilishi, takrorlanishi, bajarilish-bajarilmasligi, harakat bajaruvchisining soni, bajarilish payti kabi qator semantik grammatik ottenkalar ifodalanadi.

Ushbu maqolada fe'lning lug'aviy va sintaktik shakl yasovchi qo'shimchalarini uslubiy xoslanishiga ko'ra annotatsiyalash tamoyillari, ularning teglar tizimi, o'zbek tili milliy korpusida uslubiy xoslanganlik belgisi bo'yicha avtomatik ishlov berishning ahamiyati va uslubiy xoslanganlik belgisi bo'yicha qidiruv parametrlari haqida fikr yuritiladi.

**Kalit so'zlar:** *fe'l, lug'aviy shakl yasovchi, sintaktik shakl yasovchi, fe'lning mayl shakllari, fe'l zamon shakllari, fe'lning vazifa shakllari, fe'l nisbat shakllari lingvistik annotatsiya, nutq uslublari, uslubiy teglashtirish.*

### Kirish

Til korpuslarida adabiy til bilan birga noadabiy leksika ham o'rin oladi. Buning sababi shuki, zamonaviy korpus birliklari –

---

<sup>1</sup>Elova Dilrabo Qudratilloevna - Alisher Navoiy nomidagi Toshkent davlat

o'zbek tili va adabiyoti universiteti Amaliy filologiya kafedrasida katta o'qituvchisi, filologiya fanlari bo'yicha falsafa doktori (PhD).

**E-pochta:** elova@navoiy-uni.uz

**ORCID:** 0000-0002-2329-1811

**Iqtibos uchun:** Elova D. 2022. "Fe'l grammatik shakllarining uslubiy xoslanishiga ko'ra lingvistik annotatsiyalash masalasi xususida". *O'zbekiston: til va madaniyat. Amaliy filologiya.* 2 (5): 84-96.

korpus tarkibidagi matnlar internet manzillari orqali (ma'lum bir ishlovdan so'ng) avtomatik ravishda korpusga qabul qilinadi. Shu sababli milliy korpus tuzuvchilari korpusda turli uslub va nutq shakllariga oid matnlar proporsiyasi, ya'ni milliy tilning barcha qirralarini ko'rsatuvchi reprezentativlik masalasiga alohida e'tibor berishlari talab qilinadi.

*Annotatsiya yoki razmetka tushunchalari va uning korpus tuzishdagi ahamiyati.* Turli lingvistik muammolarni hal qilish uchun faqat matnlar majmuasiga ega bo'lish yetarli emas. Matnlarda turli qo'shimcha lingvistik va ekstralingvistik ma'lumotlar talab qilinadi. Shu ehtiyojdan kelib chiqib, korpus tilshunosligida annotatsiyalangan korpus tushunchasi paydo bo'ldi. Annotatsiyalash (teglash, izohlash, razmetkalash; inglizcha tagging, annotation) matn va uning tarkibiy qismlariga maxsus teg (razmetka, izohlar) birliktirishdan iborat.

Uning quyidagi ko'rinishlari mavjud:

1) tashqi, ekstralingvistik (muallif haqidagi ma'lumot, matn haqidagi ma'lumot: muallif, unvon, nashr etilgan yil va joy, janr, mavzu; muallif haqidagi ma'lumotlar nafaqat uning ismini, balki yoshi, jinsi, hayot yillari va boshqalarni ham o'z ichiga olishi mumkin. Ushbu ma'lumotni kodlash metateglash yoki metarazmetka deb ataladi);

2) tarkibiy (bob, paragraf, jumla, so'z shakli) va lingvistik tavsiflovchi, matn elementlarining grammatik va boshqa xususiyatlarini ko'rsatuvchi.

### **Asosiy qism**

Lisoniy birliklarni uslubiy xosligiga ko'ra razmetkalash annotatsiyalashning lingvistik tavsiflovchi turiga mansub bo'lib, quyida fe'l grammatik shakllarini uslubiy xosligiga ko'ra teglash masalalari xususida fikr yuritimiz.

*Mayl shakllarining uslubiy xoslanishini ifodalash xususida.* Fe'l mayllari o'ziga xos shakllarga ega bo'lib, turli xil modal ma'nolarni ifodalaydi. Shunga qaramay, fe'l mayllarining biri o'rnida ikkinchisini qo'llash mumkin. Ana shu holat ulardagi keng uslubiy imkoniyatlarni ro'yobga chiqaradi. Fe'lning ijro maylida modallik, ta'sirchanlik deyarli bo'lmaydi. Bunda faqat harakatning bajarilish-bajarilmasligi haqida xabar berish ma'nosi anglashiladi. Ammo bu mayl ba'zan buyruq yoki shart mayllari o'rnida kelib, ular anglatgan ma'nolarni ifodalashi mumkin: *uylarni yig'ishtirasan, oshxonani tozalaysan, so'ngra kinoga borasan (yig'ishtir, tozala, bor buyruq fe'llari o'rnida kelgan)* [Qilichev 1992, 61]. Bunday holatda mayl

shaklini bildiruvchi qo'shimchani uslubiy xoslanishini hozircha ko'rsatib bo'lmaydi. Chunki uslubiy xoslanganlik haqidagi ma'lumot bilan teglashtirishda faqat lisoniy birlikning shakliga asoslanamiz. Yuqorida ko'rsatilgan holatni ham teglash mumkin, ammo buning uchun korpusda semantik analizator ishlashi kerak: ya'ni morfologik shakl bilan aniqlanmaydigan birliklar semantik analizator funksiyasi ishga tushganda avtomatik tahlil qilinishi mumkin.

Fe'ning buyruq-istak mayli buyurish, so'rash, qistash, iltimos kabi ma'nolarni bildiradi. Bu ma'nolar ikkinchi va uchinchi shaxslarga qaratilgan bo'ladi. Ammo buyruq maylining birinchi shaxs ko'rinishi mavjud emas. Chunki so'zlovchi o'ziga o'zi buyurishining imkoni yo'q. Shunga qaramay, badiiy asarlarda personajlar nutqida ma'lum uslubiy talab bilan o'ziga o'zi undash uchraydi. Bunda so'zlovchi o'zini ikkinchi shaxs o'rniga qo'yib gapiradi. Buning o'ziga xos semantik-uslubiy xususiyatlari bor. Jumladan, buyruq bevosita so'zlovchining o'ziga tegishli bo'ladi: *Rajabboy, bu ishni qiling* ("men bu ishni qilay" ma'nosida, so'zlovchi birinchi shaxs). Bu tarzda qo'llanish **so'zlashuv** uslubiga xos bo'lib, o'zini ovsarlikka olish yoki g'ururlanish ma'nosida keladi [Qilichev 1992, 62]. Bu kabi holatlar haqida ham aynan yuqoridagi muammoni aytish mumkin.

E.Qilichevning ta'kidlashicha, fe'ning buyruq-istak mayl shakllari **badiiy uslubda, publitsistikada** keng qo'llanadi. Shart mayli fe'l o'zaklariga -sa affiksini qo'shish bilan hosil qilinadi. -sa shaklli fe'llarda modallik, hissiy-ta'sirchanlik ma'nolari kuchli seziladi. -sa shakli orqali nafaqat shart ma'nosi, balki orzu, xohish, iltimos, sabab ma'nolari ham ifodalanadi: *ishlasa, borsam, borib kelsang. Menga so'z bersangiz edi* kabi [Qilichev 1992, 62].

Ko'rinyaptiki, fe'ning mayl shakllari ichida **so'zlashuv, badiiy, publitsistik uslubga** xoslanish ilmiy jihatdan asoslangan, ammo bunday xoslanish fe'ning ma'lum bir uslubida maxsus morfologik shakl bilan farqlanmaydi, ya'ni bir morfologik shakl ma'lum kontekstda uslubiy xoslanishni ifodalasa, boshqa kontekstda uslubiy betaraf bo'ladi. Bunday holatlarni avtomatik tahlil qilishning hozircha imkoni yo'q. O'zbek kompyuter lingvistikasining keyingi bosqichlarida ushbu tahlilni amalga oshiruvchi analizatorlar yaratilsa, kontekstga qarab uslubiy xoslanishni aniqlash imkoni paydo bo'ladi.

*Fe'l zamon shakllarining uslubiy xususiyatini avtomatik aniqlash.* Fe'ning zamon kategoriyasi – obyektiv voqelikning tilda aks etishining bir ko'rinishi. Tilda zamon tushunchasi fe'ning zamon shakli orqali ifodalanadi. Fe'l zamonlarining har biri o'z shakliga

ega. Bu shakllar o'zaro ichki sinonimiya hosil qilishi mumkin yoki bir zamonga xos shakl o'zi anglatishi kerak bo'lgan ma'noni emas, balki boshqa zamon ma'nosini bildirishi ham mumkin. Ya'ni shakl ma'nosida ko'chish yuz beradi. Bularning hammasi ma'lum bir uslubiy effektini yuzaga chiqaradi [Qilichev 1992, 62].

*O'tgan zamon fe'li.* Bu zamon ma'nosi *-di, -gan, -gan edi, -(i) b edi, -(a)r edi* shakllari vositasida yuzaga chiqadi. Bu shakllar nutqda harakatning o'tgan zamonda amalga oshganligini ifodalash ma'nosida o'zaro sinonim bo'lib kelsa-da, yana har biri o'ziga xos modal ottenkalariga ega. Jumladan, *-di* shaklli fe'llarda harakatning o'tgan zamonda aniq bajarilganligi ifodalansa, *-(i)b edi* shaklli fe'llarda o'tgan zamonda bajarilgan harakat hikoya qilish, eslash orqali bayon qilinadi. O'tgan zamon shakllari nutq uslublariga munosabati jihatidan ham farqlanadi. Masalan, *-(i)b edi, -guvchi edi* shakllari **so'zlashuv** nutqiga, *-moqda edi* shakli esa **rasmiy uslubga** xos. Nutqni jonli qilish, u yoki bu harakatni tinglovchining ko'z o'ngida yoki yaqin vaqt ichida aniq bajarilganligini ifodalash uchun *-di* shaklli o'tgan zamon fe'li kelasi zamon ma'nosi uchun ishlatiladi: *Qani ketdik. Ertaga boy meni qoziga chaqirtiradi, yorimni tortib oladi, endi men nima qilaman? (Ertakdan).* Yuqoridagi ma'lumotlardan kelib chiqib, **so'zlashuv va rasmiy uslubga** xosligi ko'rsatilgan shakllarga shunday axborotni beruvchi teglar biriktiramiz.

Zamon	z	Zamon shakllari	
		Uslubiy betaraf	Uslubiy xoslangan
O'tgan zamon	z1	-di, -gan/kan/qan, -gan edi/kan edi/qan edi, -gan ekan/-kan ekan/-qan ekan, -gan emish /-kan emish/-qan emish, -b edi/-ib edi, -uvdi/-ovdi, -b(shaxs-son) ekan/-ib(shaxs-son) ekan, -r edi/+ar edi, -rdi/+ardi, -yotgan edi/-ayotgan edi, -yotgandi/-ayotgandi, -yotib edi/+ayotib edi, -guvchi edi, -ydigan edi/-adigan edi, -moqda edi, -moqchi edi, -b/+ib	- gandi/-kandi/-qandi ( <b>so'zl.</b> ) -ganakan/ -kanakan/-qanakan ( <b>so'zl.</b> ) -ganmish/ -kanmish/-qanmish ( <b>so'zl.</b> )

*Hozirgi zamon fe'li.* Bu zamon ma'nosi *-yap, -yotir, -moqda* shakllari vositasida yuzaga chiqadi. Bu formalar o'zaro sinonimlik hosil qiladi: *boryapman / borayotirman / bormoqdaman; ishlayapman / ishlayotirman / ishlamoqdaman* kabi. Hozirgi zamon ma'nosini

*yot, tur, o'tir, yur* fe'llari yordamida analitik yo'l bilan ham hosil qilish mumkin. Hozirgi zamon fe'li shakllaridan *-yap* nutqiy betaraf bo'lib, faol qo'llanadi; *-yotir* shakli she'riy nutqda ko'proq uchraydi; *-moqda* shakli kitobiy uslubga xos bo'lib, obrazli ifodalarda, tantanali publitsistik nutqlarda ko'p ishlatiladi. Shuningdek, manbalarda hozirgi zamon shakllari ma'lum uslubga xoslangani bilan o'zaro farqlanishi ko'rsatiladi: *-yapti* uslubiy betaraf (neytral) shakl bo'lsa, *-moqda, -yotir* yozma nutq uslubiga xos [Nurmonov, Sobirov, Yusupova 2012, 161].

Hozirgi zamon shakllari boshqa zamon ma'nosida ham ishlatilishi mumkin. Jumladan, hozirgi zamon shakllari o'tgan zamon ma'nosida qo'llanganda, badiiylik oshiriladi, ta'sirchanlik kuchaytiriladi. So'zlovchi harakatning bevosita kuzatuvchisi sifatida o'tgan zamon voqealarini hozirgi zamon holat fe'llari orqali tasvirlaydi. Bo'lib o'tgan voqealarni o'quvchi ko'zi oldida yangidan gavdalantiradi: *Birpasdan keyin Farmonqul, uning ketidan xotini qizini yetaklab kirdi. U mehmondan qochmadi. Hammamiz o'tirmoqdamiz, qani endi birontamizdan sado chiqsa.* (A.Q.) E.Qilichevning ko'rsatishicha, agar o'tirmoqdamiz fe'li kontekstdagi boshqa o'tgan zamon fe'llari bilan bir xil shaklda berilsa, ma'noda o'zgarish bo'lmaydi. Lekin matndagi tasviriylik yo'qolib, voqea oddiy xabarga aylanib qoladi. Hozirgi zamon fe'l shakllari kelasi zamon o'rnida ham qo'llanadi. Bunda ma'no ta'sirchanligi bo'rttiriladi: *Ertaga dam olishga ketyapmap.* Fe'l zamon shakllarining bu kabi uslubiy o'ttenka hosil qilish imkoniyatlari shaklga emas, ma'noga xoslangani sababli ularni avtomatik tahlilga tortib bo'lmaydi. Ko'rsatilgan qo'shimchalarning uslubiy xoslanishi nazariy asosga ega. Zamon shakllarining uslubiy xoslanishi shakliy jihatdan farqlangan holatlarning bazada aks ettirilishini quyida ko'rsatamiz.

Zamon	z	Zamon shakllari	
		Uslubiy betaraf	Uslubiy xoslangan
Hozirgi zamon	z2	-yotibdi ekan/+ayotibdi ekan, -yotib (shaxs-son) ekan/+ayotib (shaxs-son) ekan, -yotgan ekan/+ayotgan ekan, -yotgan emish/+ayotgan emish, -yotgan emish/+ayotganmish,-yap, -yapti ekan, -yapti emish, -yap+(shaxs-son)+ekan (shaxs-son), -yap+(shaxs-son)+emish, -yap+(shaxs-son)+mish, -yap+(shaxs-son)+emish (shaxs-son), -yap+(shaxs-son)+mish (shaxs-son) -yotib/+ayotib	-yaptikan (so'zl.) -yaptimish (so'zl.) -yap+(shaxs-son)+ekan -yotir/+ayotir (yozma nutq) -yotir ekan/+ayotir ekan <b>(yozma nutq)</b> -moqda (yozma nutq) -moqda emish (so'zl.) -moqdamish (so'zl.) -moqda ekan (so'zl.) -ydigan emish/+adigan emish (so'zl.) -ydiganmish/+adiganmish (so'zl.)



*Kelasi zamon fe'li.* Kelasi zamon shakllari nutq so'zlanayotgan paytdan keyin bajariladigan harakatni bildiradi. Bu zamon *-ajak(-yajak), -moqchi, -digan* affikslari yordamida yasaladi. Kelasi zamon yasovchi bunday affikslar o'zaro sinonimlik munosabatni tashkil qiladi; *borajakman – bormoqchiman – boradiganman; yozajaksan – yozmoqchisan – yozadigansan.* Kelasi zamon shakllari ham o'rni bilan boshqa zamon ma'nolarini ifodalash uchun ishlatiladi. Kelasi zamon shakllarining uslubiy xoslanishi fe'lning *-a* ravishdosh shakliga *-di, -jak*; fe'l asoslariga *-ar, -ur, -gusi* qo'shimchalarini qo'shish orqali hosil bo'ladi. Masalan, *keladi, kelajak, kelar (ur), kelgusi* kabi. Bulardan *-adi, -ar* uslubiy betaraf shakllar bo'lsa, *-jak, -gay, -ur, -gusi* shakllari badiiy va publitsistik uslubga xos. Uslubiy betaraflik belgisiga ko'ra *-adi* va *-ar* shakllari bir guruhni tashkil qilsa ham, ular ifodalanayotgan harakatga gumon-taxmin ma'nosini bildirish-bildirmasligiga ko'ra bir-biridan farq qiladi. Birinchisi bunday ma'noga betaraf bo'lsa, ikkinchi shakl gumon-taxmin ma'nosiga ega.

Ushbu ma'lumotlardan kelib chiqqan holda fe'l zamon shakllarining uslubiy xoslanishi MBda quyidagicha aks ettiramiz:

Zamon	z	Zamon shakllari	
		Uslubiy betaraf	Uslubiy xoslangan
Kelasi zamon	z3	-y/+a, -ydi ekan/-adi ekan, -ydi emish/-adi emish, -r/-ar, -r emish/-ar emish, //-rmish/ armish, -r ekan/-ar ekan, -moqchi, -moqchi emish, -moqchi ekan, -ydigan/-adigan, -ydigan ekan/-adigan ekan, -y/-ay dedi	-yajak/+ajak ( <b>badiiy / publitsistik</b> ) -gay/+kay/+qay ( <b>badiiy / publitsistik</b> ) -gu/+ku/+qu(shaxs-son) dir ( <b>badiiy</b> ) -rkan/+arkan ( <b>badiiy / publitsistik</b> ) -moqchimish ( <b>so'zl.</b> ) -ur ( <b>badiiy / publitsistik</b> ) -gusi ( <b>badiiy / publitsistik</b> )

*Fe'l nisbat shakllarining uslubiy xoslanishini ko'rsatuvchi teglar.* Hozirgi o'zbek tilida fe'lning besh xil daraja shakli mavjud bo'lib, ular barcha nutq uslublarida uchraydi. Ayrim daraja shakllari ma'lum bir nutq uslubiga xoslangan ham bo'ladi.

*Orttirma daraja shakllari* semantik-grammatik planda bitta vazifani bajarsa ham, nutqda ishlatilishi jihatidan farqlanadi. *-t* affiksi *-giz (-g'iz, -qiz)* shakllari bilan sinonimlik hosil qiladi: *kirit – kirgiz, yurit – yurgiz, to'lat – to'lg'az* kabi. Xuddi shuningdek, turli shakldagi orttirma daraja yasovchi quyidagi affikslar ham o'zaro

o'rin almashtira oladi: *-giz, -dir (kiygiz, kiydir), -iz, -dir, (emiz, emdir), -t, -tir, -g'iz (to'lat, to'ldir, to'lg'az)* va h.k. Bir leksemaga orttirma daraja yasovchi bir morfemaning turli fonetik variantlari qo'shilishi mumkin: *ko'rgaz – ko'rgaz, ko'rg'az, turgiz – turg'iz – turg'az* kabi. Bunday fonetik variantlar qisman **dialektal** (sheva) xususiyatlari bilan bog'liq bo'lib, emotsional-ta'sirchanlikni kuchaytirish, personajni nutqiy tavsiflash kabi uslubiy vazifalarni bajaradi. Ko'rinib turibdiki, fe'l nisbat shakllari uslubiy vazifa bajarsa-da, shakliy jihatdan uslubiy xoslanganligini ajratib bo'lmaydi. Bunday holatda avtomatik ishlov berish hozircha imkonsiz.

Fe'lning orttirma darajasini hosil qiluvchi shakllar faol-nofaolligiga ko'ra ham farqlanadi. *-tir (-dir), -gaz (-qaz, -kaz, -giz, -g'iz, -qiz)* affikslari *sinonimiyasida -giz (-kiz, -qiz)* forma yasovchilar faol emas. *-gaz, -g'az, -g'iz* affikslari esa ko'proq dialektal yoki individual nutqqa xos. *-kaz, -qaz* fonetik variantlari ba'zan ma'no farqlash uchun ham ishlatiladi. Chog'ishtiring: *o'tkaz* – biror voqeani boshdan kechirmoq; *o'tqaz* – ko'chat o'tqazish (qurish) ma'nosida. Bu leksik ma'noni farqlamaslik uslubiy nuqsonga olib keladi [Qilichev 1992, 56].

Ba'zan bir nisbat ma'nosi o'rnida boshqa nisbat shakllaridan foydalaniladi. Xususan, bosh nisbat ma'nosida birgalik nisbat shakllari ishlatilishi mumkin. Masalan, *keldi – kelishdi, kelyapti – kelishyapti*. Shuningdek, bosh nisbat ma'nosida majhul nisbat shakllaridan foydalaniladi: *qildim o'rniga qilindi, bajardim o'rniga bajarildi, kiring o'rniga kirilsin* kabi. Bunday holat ma'lum uslubiy emotsional ma'no talabi bilan ro'y beradi. Ko'pincha keyingi holat hurmat, kamtarlik ma'nolarini ifodalaydi. Ko'rib o'tilganidek, fe'l nisbat shakllari shakl jihatdan uslubiy betaraf. Boshqa uslubiy ma'nolar esa faqat kontekstda anglashiladi. Shu sababli nisbat ko'rsatkichlarini ma'lumotlar bazasida quyidagicha aks ettiramiz:

I.2	Nisbat	n	Uslubiy betaraf
A	Aniq	n1	+∅
B	O'zlik	n2	1) +n/+in, 2) +l/+il
C	Majhul	n3	1) +n/+in, 2) +l/+il
D	Orttirma	n4	1) +ar 2) +ir, 3) +giz/+kiz/+qiz/+g'iz/+gaz/+kaz/+qaz, 4) +iz, 5) +dir/+tir, 6) +t/+it
E	Birgalik	n5	+sh/+ish

*Fe'lning funksional shakllarining uslubiy xususiyatini avtomatik aniqlash imkoniyati.* Fe'lning funksional shakllari asosiy

xususiyatlariga ko'ra ikki turga bo'linadi:

- 1) fe'ning ma'lum bir vazifa uchun xoslangan shakllari;
- 2) harakatning turli tavsifini ko'rsatuvchi va modal ma'no ifodalovchi shakllar.

Fe'ning ma'lum bir vazifa uchun xoslangan shakllariga *harakat nomi, sifatdosh va ravishdosh* kiradi.

*Harakat nomi.* Fe'ning bu shakli *-(i) sh, -(u)v, -moq, -maslik, -ganlik* affikslari yordamida yasaladi. Bu shakllar ma'no jihatidan bir-biriga yaqin turadi. Ammo ular nutq uslublarida qo'llanishi, faol-nofaolligiga ko'ra o'zaro farqlanadi. Jumladan, *-(i)sh* affiksi bilan harakat nomi yasalishi faol bo'lsa, *-(u)v* shakli bilan harakat nomi yasalishi nofaol. *Toshkentga o'qishga keldim* deyish mumkin. Ammo *Toshkentga o'quvga keldim* deb bo'lmaydi. Chunki *-(u)v* affiksi bilan yasaluvchi shakl ko'proq ot ma'nosida qo'llanadi. *-moq* affiksi bilan harakat nomi yasash ham nisbatan chegaralangandir. Uning *-mak* varianti she'riy nutqqa xos. Klassik adabiyotda *-mak* shakli faol qo'llangan. Jonli so'zlashuv nutqida harakat nomi gapda ega vazifasida ko'proq qo'llanadi. Egasi harakat nomi bilan ifodalangan gaplar ta'sirchanlik xususiyatga ega bo'ladi. Chog'ishtiring: *chekish zararli. Yugurish foydali. Ayb emasdir sevmog, sevilmog (U.)*.

Ushbu ma'lumotlardan kelib chiqqan holda harakat nomiga oid axborotni quyidagicha annotatsiyalaymiz:

	<b>Fe'ning vazifa shakli</b>	<b>V</b>	<b>Uslubiy betaraf</b>	<b>Uslubiy xoslangan</b>	<b>Uslub turi</b>
a	harakat nomi shakli	V1	+sh/+ish +маслик +v/+uv	+moq +moqlik -мак +shlik/+ishlik	kitobiy uslub badiiy uslub badiiy uslub yuqori uslub

*Sifatdosh.* Fe'ning bu shakli *-gan (-kan, -qan), -ar, -r, -gusi (-g'usi, -qusi), -jak, -mish* affikslari yordamida yasaladi. *-gusi (-g'usi), -jak* affikslari kammahsul, *-mish* shakli esa arxaiklashgan. Bu shakllar she'riy nutqda ishlatiladi. *-gan, -ar* sifatdosh shakllari aniq zamonni ifodalamaganda harakat nomi shakllari *-ish, -moq* bilan sinonim bo'lib kela oladi: *toza havoda dam olganga nima yetsin* (olishga). *Keldi ochilur chog'ing, o'zliging namoyon qil* (H. H.) (ochilish). *Adashgan ayb emas. - Adashish ayb emas* kabi. *-gan* shakli aniq zamon ifodalab, aniqlovchi bo'lib kelganda hamda *-gan* shakli so'z otlashganda *-ish* shakli so'zlar bilan o'rin almashtira olmaydi: *O'qiyotgan bola keldi: Qo'rqqanga qo'sh ko'rinar* (maqol.) kabi. Ushbu ma'lumotlardan kelib chiqqan holda MBda sifatdosh yasovchi

shakllarni quyidagicha teglash mumkin:

I.3	Fe'ning vazifa shakli	V	Uslubiy betaraf	Uslubiy xoslangan
b	sifatdosh shakli	V2	+gan/+kan/+qan, +ganlik/+kanlik/+qanlik, +ganda edi/+kanda edi/+qanda edi, +gandaydi/ +kandaydi/+qandaydi, +gan/+kan/+qan (egalik) bo'lsin, +gan/+kan/+qan bo'ldi, +yotgan/+ayotgan, +ydigan/+adigan, +ydigan edi/+adigan edi	-r/+ar, +guvchi/+uvchi/+vchi, +gulik/+g'lik/+rlik, +gur/+kur/+qur/+g'ur

*Ravishdosh.* Fe'ning bu shakli *-(i)b, -a, -y, -gali, -guncha (-kuncha, -quncha), -gach, -gani (-kani, -qani)* affikslari yordamida hosil qilinadi. *-(i)b* va *-a, -y* ravishdosh yasovchilari ma'no jihatdan bir-biridan ancha farqlansa ham, ayrim o'rinlarda sinonim bo'lib keladi: *o'qib biladi – o'qiy biladi; yozib oladi – yoza oladi* kabi. Ravishdoshning *-may* shakli she'riyatda *-mayin* shaklida ham uchraydi. *Bilmayin bosdim tikanni, tortadurman dardini.*

*-gani, -gali* shakllari hozirgi nutqda ko'p ishlatilmaydi, ko'proq she'riy nutqqa xosdir:

*Shiypon oldi gul hovuz,*

*Gul tergani kelganmiz.*

*Gulni bahona qilib,*

*Yor ko'rgani kelganmiz.* (Xalq og'zaki ijodi)

Nozu istig'no bilan ko'rsatur ming g'amza jon qilgali shaydo meni (U.)

Harakatning bajarilish holatini ifodalashda *-gan -cha, -guncha shakllari -gan, -g'an* holda konstruksiyalari bilan ham sinonimik qator hosil qiladi: *Anvar eshik kesagasigidan ushlagan cha goh Alimardonga, goh Muqaddamga qarab turar edi.* (U.H.) (Ushlagan holda,) Ko'rinadiki, ravishdosh yasovchi shakllarda uslubiy imkoniyat katta. Ulardan maqsadga muvofiq foydalanish nutqni ko'rkamlashtiradi. Ravishdosh shakllari uslubiy xoslanishini MBda quyidagicha belgilash mumkin:

I.3	Fe'ning vazifa shakli	V	Uslubiy betaraf	Uslubiy xoslangan
c	ravishdosh shakli	V3	1) +b/+ib, 2) +gancho/+kancho/+qancho, 3) +gudek/+kudek/+qudek, 4) +gach/+kach/+qach, 5) +guncha/+kuncha/+quncha	+gani/+kani/+qani/ +gali +gali

*Harakatning turli tavsifini ko'rsatuvchi va modal ma'no ifodalovchi shakllar.* Bu shakllar ko'rinishiga ko'ra ikkiga bo'linadi:

- 1) sintetik shakllar,
- 2) analitik shakllar.

Sintetik shakllar harakatning davomiyligi, takroriyiligini ifodalaydi. Bu *-a, -la, -kila (-gila, -qila, -g'ila)* shakllari orqali yuzaga chiqadi: *bur-bura, sava-savala, turt-turtkila, cho'z-cho'zg'ila, tep-tepkila, qashi - qashila* kabi. Harakatning kuchsiz darajasi esa *-(i) nqira, -(i)msira* shakllari vositasida yuzaga keladi: *qo'rqinqira, kulimsira, yig'lamsira* kabi. Ushbu shakllar, asosan, badiiy uslubda ishlatilib, personajlarning ruhiy holati, ichki kechinmalarini yaqqol gavdalantrish uchun xizmat qiladi: *Uning yanog'i pir-pir uchar, qovog'i shishinqiraganday ko'rinar edi* (A.Q.). *Zohid ishqin desaki, qilgay fosh, yig'lamsiraru ko'ziga kelmas yosh* (N.). Anglashiladiki, harakatning turli tavsifini ko'rsatuvchi va modal ma'no ifodalovchi sintetik shakllar badiiy uslubga xos. Shu sababli MBda ushbu shakllarning barchasiga uning uslubiy xoslanganligi, badiiy uslubga tegishlilik haqida ma'lumot biriktiriladi:

	<i>Harakatning modal ma'nosini ifodalovchi shakllar</i>	<b>Uslubiy betaraf</b>	<b>Uslubiy xoslangan</b>
	Sintetik shakl		<i>-a, -la, -kila (-gila, -qila, -g'ila)</i>

Analitik yo'l bilan modal ma'no ko'makchi fe'llar va to'liqsiz fe'l yordamida hosil qilinadi. Har bir ko'makchi fe'l yetakchi fe'l bilan birikib modal ma'nolar ifodalaydi. Masalan, ishning endi boshlanganligi (kuylay boshladi), davomlilik (ishlab yuribdi), tugallanganligi (o'qib chiqdi) kabi ma'nolar boshla, yur, chiq ko'makchi fe'llari orqali ifodalangan. Modal ma'no ifodalash jihatidan ham ayrim ko'makchi fe'llar sinonimlik tashkil qiladi. Masalan, biror harakatni bajara olish imkoniyati *bil, ol* yordamchi fe'llarida teng ifodalanadi. Chog'ishtiring: *o'qiy biladi – o'qiy oladi, kela biladi – kela oladi* kabi. Shunga qaramay, ayrim ko'makchi fe'llar ba'zi fe'llar bilan birika olmaydi. Masalan, yoz ko'makchi fe'li quvonmoq, tekshirmoq, o'smoq qabi fe'llar bilan birika olmaydi. Og'zaki va yozma nutqda bunga e'tibor qilmaslik natijasida uslubiy xatolarga yo'l qo'yiladi.

Ko'makchi fe'lli so'z qo'shilmalari uslubiy vazifa bajarsada, ularning shakli uslubiy ma'nosini hozircha avtomatik aniqlash imkonini bermaydi.

*Edi, ekan, emish, erur* to'liqsiz fe'llarda aslida e o'zagidan hosil bo'lgan. To'liqsiz fe'l ot, sifat, son, olmosh kabi so'z turkumlari bilan

birikib, ularga fe'llik tusini beradi. Fe'llar bilan birikib kelganda esa harakatning aniq bajarilganligi, harakatning bajarilish ehtimoli borligi, payti, maqsadi kabi modal ma'nolarni ifodalab keladi. Erur to'liqsiz fe'li tarixiy-badiiy asarlarda davr ruhini berish uchun ishlatiladi. To'liqsiz fe'ning ekan, *-kan* shakllari so'zlashuv nutqi uslubida ko'proq qo'llanadi. Yuqoridagi ma'lumotlar asosida to'liqsiz fe'llarni quyidagicha teglash mumkin:

	To'liqsiz fe'llar	Uslubiy betaraf	Uslubiy xoslangan
		edi, emish	ekan, -kan (so'zl.) erur (tar.)

### Xulosa

Xulosa sifatida aytish mumkinki, til korpuslari jahon amaliyotida keng qo'llanilmoqda va an'anaviy ta'lim texnologiyalariga samarali innovatsion qo'shimchalar sifatida namoyon bo'lmoqda. O'zbek tili milliy korpusi bazasidagi lisoniy birliklarni uslubiy belgilariga ko'ra annotatsiyalash kelgusida uslubiy analizatorni yaratishda muhim manba bo'lib xizmat qiladi.

### Adabiyotlar

- Дубовик А.Р. 2017. Автоматическое определение стилистической принадлежности текстов по их статистическим параметрам // Компьютерная лингвистика и вычислительные онтологии. Выпуск 1.
- Захаров В.П., Богданова С.Ю. 2011. Корпусная лингвистика. Иркутск: ИГЛУ.
- Муҳаммедова С. 2006. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма. Тошкент, 166 б.
- Национальный корпус русского языка // Вопросы языкознания. 2006. Москва: Наука.
- Недошивина Е.В. 2006. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. Санкт-Петербург.
- Қўнғуров Р. 1990. Нутқ маданияти ва услубият асослари. Тошкент.
- Гулямова Ш. 2022. Ўзбек тили семантик анализаторининг лингвистик асослари: Филология фан. д-ри. дисс. автореф. Фарғона.
- Хамроева Ш. 2021. Ўзбек тили морфологик анализаторининг лингвистик таъминоти: филол. фан докт. диссертацияси автореф. Фарғона, – 83 б.
- Синклер Д. Предисловие к книге “Как использовать корпуса в преподавании иностранного языка”/ Д. Синклер [Электронный ресурс]. – Режим доступа: <http://www/ruscorpora.ru/corpora-info.html>, свободный.

- Қиличев Э. 1992. Ўзбек тилининг амалий стилистикаси (грамматика).  
Ўқув қўлланма. Тошкент: Ўқитувчи.
- Nurmonov A., Sobirov A., Yusupova Sh. 2012. Hozirgi o'zbek adabiy tili.  
Toshkent: Ilm-ziyo.
- Sayfullayeva R., Mengliyev B., Boqiyeva G., Qurbonova M., Yunusova M.,  
Abuzalova M. 2009. Hozirgi o'zbek adabiy tili. Toshkent: Fan  
va texnologiyalar.

## **Style of grammatical forms of the verb in particular on the problem of linguistic annotation according to customization**

Dilrabo Elova<sup>1</sup>

### **Abstract:**

In the field of computer linguistics, which is rapidly developing in world linguistics, improving the quality of natural language processing, asking and receiving answers, using corpora in linguistic research, linguodidactics, and lexicography has become an urgent issue. In linguistics, in particular, in the field of computer linguistics, the problem of selecting corpus units, their linguistic, including morphological, semantic, syntactic and stylistic tagging, developing a system of stylistic tags, and creating linguistic support for distinguishing stylistic features that includes special information for the language corpus remains relevant.

The verb word group is a word group rich in grammatical categories, and these forms, like lexical units, mean different stylistic meanings. A single verb expresses a number of semantic and grammatical aspects such as the process of action and its description, whether the action is fast or slow, whether it is repeated, whether it is performed or not, the number of people who perform the action, and when it is performed.

This article discusses the principles of annotating lexical and syntactic form-forming adverbs of the verb according to their stylistic features, their system of tags, the importance of automatic processing according to stylistic features in the national corpus of the Uzbek language, and search parameters for stylistic features.

---

<sup>1</sup>Elova Dilrabo Qudratilloevna - Alisher Navoi Tashkent State University of Uzbek Language and Literature, Senior Lecturer, Department of Applied Philology, Doctor of Philology (PhD)

**E-mail:** elova@navoiy-uni.uz

**ORCID:** 0000-0002-2329-1811

**For reference:** Elova D. 2022. "Style of grammatical forms of the verb in particular on the problem of linguistic annotation according to customization". *Uzbekistan: language and Culture. Applied philology*. 2 (5): 84-96.

**Key words:** *verb, lexical form maker, syntactic form maker, inflectional forms of the verb, tense forms of the verb, task forms of the verb, relative forms of the verb, linguistic annotation, speech styles, stylistic tagging.*

**References:**

- Dubovik A.R. 2017. Avtomaticheskoe opredelenie stilisticheskoy prinadlejnosti tekstov po ix statisticheskim parametram // Kompyuternaya lingvistika i vichislitelnie ontologii. Vipusk 1.
- Zaxarov V.P., Bogdanova S.Yu. 2011. Korpusnaya lingvistika. – Irkutsk: IGLU.
- Muhammedova S. 2006. Harakat fe'llari asosida kompyuter dasturlari uchun lingvistik ta'min yaratish. Metodik qo'llanma. – Toshkent, p 166.
- Natsionalniy korpus russkogo yazyka // Voprosi yazykoznanija. 2006. Moskva: Nauka.
- Nyedoshivina V. 2006. Programmy dlya raboti s korpusami tekstov: obzor osnovnix korpusnyx menedjerov. Uchebno-metodicheskoe posobie. Sankt-Peterburg.
- Qo'ng'urov R. 1990. Nutq madaniyati va uslubiyat asoslari. Toshkent.
- Gulyamova Sh. 2022. O'zbek tili semantik analizatorining lingvistik asoslari: Filologiya fan. d-ri. diss. avtoref. Farg'ona.
- Hamroyeva Sh. 2021. O'zbek tili morfologik analizatorining lingvistik ta'minoti: filol. fan dokt. dissertatsiyasi avtoref. Farg'ona, – p 83.
- Sinkler D. Predisloviye k knige "Kak ispolzovat korpusa v prepodavanii inostrannogo yazyka"/ D. Sinkler [Elektronnyy resurs]. – Rejim dostupa: <http://www.ruscorpora.ru/corpora-info.html>, svobodniy.
- Qilichev E. 1992. O'zbek tilining amaliy stilistikasi (grammatika). O'quv qo'llanma. Toshkent: O'qituvchi.
- Nurmonov A., Sobirov A., Yusupova Sh. 2012. Hozirgi o'zbek adabiy tili. Toshkent: Ilm-ziyo.
- Sayfullayeva R., Mengliyev B., Boqiyeva G., Qurbonova M., Yunusova M., Abuzalova M. 2009. Hozirgi o'zbek adabiy tili. Toshkent: Fan va texnologiyalar.



## O'zbek-ingliz tillarining teglangan parallel korpusini yaratish bosqichlari

Botir Elov<sup>1</sup>  
Ma'rufjon Amirqulov<sup>2</sup>

### Abstrakt

Maqolada kompyuter lingvistikasining asosiy yo'nalishlaridan biri va ko'plab masalalari yechilishi lozim bo'lgan korpus lingvistikasi, monolingual korpuslar hamda parallel korpuslar haqida, bundan tashqari parallel korpus sohasida jahon tajribasiga tayangan holda o'zbek-ingliz tillarining parallel korpusini yaratish bosqichlari haqida so'z yuritiladi. O'zbek-ingliz tillari parallel korpusining dasturiy va lingvistik tamoyillarini asoslash, tanlangan birliklarni lingvistik va ekstralingvistik teglash, parallel korpus tuzish algoritmini ishlab chiqish kabi ustuvor vazifalar haqida ma'lumotlar beriladi. Parallel korpusga qanday ma'lumotlar tanlanishi, ma'lumotlarga qo'yiladigan talablar va o'zbek-ingliz parallel korpusining yaratilishi tadqiqotchi va foydalanuvchilarga qanday imkoniyatlar taqdim etishi to'g'risida fikrlar yuritiladi. Ushbu jarayonda material tanlash kabi lingvistik va uslubiy muammolar bilan bir qatorda parallel korpusni yaratishdagi dasturiy qiyinchiliklar xususida xulosalar aks etadi.

**Kalit so'zlar:** *Korpus, teglash, tenglashtirish, parallel korpus, Pos teglash, XML kodlash*

### Kirish

So'ngi yillarda korpus lingvistikasi jadallik bilan rivojlanib bormoqda. Parallel korpuslarning qo'llanishi amaliy tilshunoslik tadqiqotlari ko'lamini yanada kengaytiribgina qolmay, chet tilini o'rgatish bo'yicha olib boriladigan tadqiqotlar uchun yangi falsafiy

---

<sup>1</sup>*Elov Botir Boltayevich* – texnika fanlari bo'yicha falsafa doktori (PhD), dotsent. Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti Kompyuter lingvistikasi va raqamli texnologiyalar kafedrasini mudiri.

**E-pochta:** elov@navoiy-uni.uz

**ORCID:** 0000-0001-5032-6648

<sup>2</sup>*Amirqulov Ma'rufjon Aliqul o'g'li* - Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti Kompyuter lingvistikasi mutaxassisligi 2-kurs magistranti.

**E-pochta:** amirkulovmaruf@navoiy-uni.uz

**ORCID:** 0000-0002-4025-8466

**Iqtibos uchun:** Elov B., Amirqulov M. 2022. "O'zbek-ingliz tillarining teglangan parallel korpusini yaratish bosqichlari". *O'zbekiston: til va madaniyat. Amaliy filologiya*. 2 (5): 97-109.

qarashlarni shakllantirmoqda [Wen, Wang, Liang, 2005]. Parallel korpus atamasi odatda lingvistik doiralarda bir-birining tarjimai bo'lgan matnlarga murojaat qilish uchun ishlatiladi. Bir tilli hisoblanmish monolingual korpuslarni butun dunyo bo'ylab keng ko'lamda uchratish mumkin. Ingliz tili monolingual korpusi, ayniqsa, eng katta hajmi egallaydi. Monolingual korpuslarni yaratishdagi muvaffaqiyat bilan taqqoslanganda, parallel korpuslar hajmi, ayniqsa o'zbeklar ishtirok etgan parallel korpuslar, unchalik ko'p emas va bu o'zaro tillarda axborotni qayta ishlash, til o'rgatish, tadqiqot hamda ikki tilli lug'at tuzish kabi jarayonlarning istiqboliga sezilarli darajada to'sqinlik qiladi. Dunyo kompyuter lingvistikasida oxirgi 30 yil davomida parallel korpuslarni avtomatik tarjima platformasi, tezaurus, elektron lug'at sifatida qo'llash va ularni ilmiy-nazariy jihatdan o'rganish harakati avj oldi. Shu jarayonda o'zbek-ingliz, ingliz-o'zbek parallel korpusini tuzishning dasturiy va lingvistik tamoyillarini asoslash, tanlangan birliklarni lingvistik va ekstralingvistik teglash, parallel korpus tuzish algoritmini ishlab chiqish kabi ustuvor vazifalar morfologiya, sintaksis, tarjimashunoslik sohaslarida uzusning formal va noformal registrini aniqlash, faqat tilshunos intuisiyasiga tayanilgan, subyektivlikka moyil tadqiqotlarning ishonchiligi va obyektivligini ta'minlash, yangi avlod korpus lug'atlari va korpus grammatikalarini yaratishga keng imkon yaratadi [Karimov, 2022].

Yuqori sifatli tarjimaga ega parallel korpusni yaratish kompyuter lingvistikasining mashina tarjimai sohasidagi dolzarb masalalardan biriga aylanib bormoqda. Afsuski, bunday yuqori sifatli parallel korpusni yaratish juda ko'p sabablarga ko'ra qiyin masala hisoblanadi. Sifatli tarjimalarning mualliflik huquqini sotib olish, parallel matnlarni, gaplarni va so'zlarni muqobil tenglashtiruvchi dasturlarni topishdagi qiyinchiliklar, korpusni yaratishdagi yuqori sarf-xarajatlar shular jumlasidandir [Park, Lee, Eo, Seo, Lim, 2021]. Parallel korpus uchun manba tanlashdagi eng ishonchli ma'lumotlar sifatida inson tomonidan bajarilgan tarjimalarni olishimiz mumkin [Hutchins, 2001, 5; Rojo, 2018, 257]. Inson omili orqali tarjima qilingan matnlar orqali korpusning aniqlik darajasi ortadi, mashina tarjimai uchun esa sifatli ma'lumotlar bazasi sifatida xizmat qiladi hamda foydalanuvchilarga mavjud avtomatik tarjima dasturlar bilan solishtirganda ancha yuqori sifatdagi tarjimaga ega bo'lish imkoniyatini yaratadi. Bu jarayon esa korpusdagi to'g'ri va sifatli materiallar qamrovi ortgani sababli yanada yaxshilanib borishi, parallel korpuslarni yaratish kompyuter lingvistikasidagi nechog'lik

muhim masalalardan biri ekanligini namoyon qiladi.

### **Asosiy qism**

Eng keng tarqalgan fikr va qarashlarga ko'ra "korpus" (lotinchadan so'zma-so'z "tana" ma'nosida) yozma matnlar, ayniqsa muayyan bir yozuvchining muayyan mavzularidagi asarlarini qamrab olgan jamlanma sifatida qaralgan [Saad, 2015]. Korpus lingvistikasini tilshunoslikda foydalanish uchun tuzilgan, matnlarning katta hajmdagi bazasi orqali tilni o'rganish metodologiyasi sifatida ko'rish mumkin [Leech, 2010, 103]. Korpus va korpus lingvistikasiga doir dastlabki tadqiqot hamda rivojlanish jarayonlari asosan ingliz tilshunosligi doirasida amalga oshirilgan. 1960-yillarda yaratilgan Braun korpusi birinchi mashinada o'qiladigan korpus sifatida tarixda iz qoldirgan bo'lsa, undan so'ng yaratilgan Lankaster-Oslo/Bergen (LOB) korpusi [Svartvik, 1992, 7] haqida ham shunday deyish mumkin. Korpusning mavjudligi katta hajmdagi lingvistik ma'lumotlarga ega bo'lishga imkon yaratibgina qolmay, ko'plab kvantativ va variatsion tadqiqotlar uchun ham dasturulamal bo'lib xizmat qildi. Dastlabki ikkita korpus (Braun va LOB) bir xil dizayn mezonlari bo'yicha tuzilgan edi, bu esa bir xil tuzilgan korpuslardagi ingliz tilining ikki xil variantini qiyoslash g'oyasini olib keldi. Natijada korpus metodologiyasiga tamoman begona bo'lgan qiyoslash nuqtayi nazari kirib keldi. Biroq, ko'p yillar davomida korpuslar bir tilli bo'lib qoldi, bu esa qiyosiy tadqiqotlar imkoniyatini cheklovchi sabab edi. Qiyoslash faqatgina boshqa tillarda yaratilgan bir tilli korpuslar orqali amalga oshirilardi. Bitta korpusda ikki va undan ortiq tillarning parallel ma'lumotlarini topish korpus lingvistikasi rivojlanish bosqichining ilk davrlarida yetishib bo'lmas jarayon edi. Ko'p tilli korpuslarni yaratish haqidagi fikrlar yaqin yillarga borib taqaladi. 1990-yillar boshlarida olimlar Stig Yoxansson va Knut Xofland ingliz-norveg tillari parallel korpus (ENPC)ni uchun harakatlarni boshlashgan [Johansson, Hofland, 1994, 25]. Yaratilishi mo'ljallangan korpus ingliz va norveg tillarining asl va tarjima matnlarini o'z ichiga olishi hamda unda ma'lumotlarni tenglashtirish va qidiruv tizimining yangi texnologiyalari yaratilishini talab qilar edi. Stig Yoxanssonning boshqa loyihalari singari, ingliz-norveg tillari parallel korpusi uning ingliz-shved tillari parallel korpus (ESPC)ni mualliflari hisoblanmish shved hamkasblarining yordami bilan yaratilgan edi.

Ma'lumotlarni tenglashtirish uchun foydalaniladigan dasturiy ta'minot Knut Xofland tomonidan yaratilgan bo'lsa, Jarle Ebel tomonidan parallel konkordans uchun tizim yaratildi. Ingliz-shved

tillari parallel korpusidagi asl va tarjima matnlar gaplar kesimida tenglashtirilgan bo'lib, har bir gap yoki gaplar birligi boshqa tildagi xuddi o'sha gap yoki gaplar birligiga o'ziga xos xususiyatga ega identifikatsiya tegi bilan birlashtirilgan edi. Quyidagi misolda ushbu korpusdagi teglangan va tenglashtirilgan gapdan namuna keltiramiz:

(1a) <s id=ABR1.1.1.s326 corresp=ABR1T.1.1.s325>But how come you speak the language so fluently?" </s>(ABR1)

(1b) <s id=ABR1T.1.1.s325 corresp=ABR1.1.1.s326>Men hvordan har det seg at De snakker språket så flytende?" (ABR1T)

Yuqoridagi kabi ikki va ko'p tilli korpuslarning turli jabhalardagi o'rni beqiyos. Qiyosiy tahlil jarayonida ham parallel korpuslarning nechog'lik muhimligi quyida keltiriladigan fikrlar orqali dalillanadi. Qiyosiy tahlil – bu ikki yoki undan ortiq tillarni maqsadli ravishda ularning farqli va o'xshash tomonlarini tizimli taqqoslash hamda tavsiflash demakdir [Johansson, 2007]. Parallel korpuslarning qiyosiy tahlil jarayonidagi ustunliklari xususida olimlar Aijmer va Altenberglar tomonidan tushuntirib berilgan:

- Ular monolingual korpuslarda erishib bo'lmaydigan ustunlik, ya'ni tillarni taqqolash imkonini beradi;
- Ular bir qator qiyosiy maqsadlarda ishlatilishi va bizning tilga xos tipologik va madaniy farqlar, shuningdek, umumiy o'ziga xos xususiyatlar haqidagi tushunchamizni oshirishi mumkin;
- Ular manba va tarjima, mahalliy va nomahalliy matnlar o'rtasidagi farqlarni o'rganishda ishlatiladi;
- Ular bir qator amaliy dasturlar, masalan, leksikografiya, tilni o'rgatish va tarjima sohasida ishlatilishi mumkin.

Bir tilli korpuslar singari, parallel korpuslar leksika, leksika – grammatika va nutqiy xususiyatlarni o'rganish uchun juda mos keladi. Korpusning teglangan versiyasi esa o'sha grammatik konstruksiyalarni o'rganishni yanada osonlashtiradi.

Korpusni yaratishga doir ko'plab adabiyotlar ma'lum bo'lib, ulardan birining sharti korpus muvozanatlangan bo'lishini taqozo etadi. Bunda turli soha va janrlarga tegishli matnlar korpusdan nisbatan teng hajmni egallashi kerak, ana shunda o'sha tilning bor jozibasini ko'rish mumkin. Biroq, parallel korpuslarni, xususan, o'zbek-ingliz tillari parallel korpusini tuzishda yuqoridagi talabga amal qilish qiyin. Chunki ushbu tillarning elektron bilingual matnlarini topish anchayin mushkul masala. Hozirda bir necha o'zbek asarlarining ingliz tiliga, ingliz asarlarining o'zbek tiliga o'girilgan variantlarini topish mumkin, ammo bu ham ikki tilning jozibasini to'laqonli ochib berolmaydi va hajman parallel korpus

uchun kam miqdor hisoblanadi. Shunday bo'lsada, mamlakatimizda korpus lingvistikasi doirasida parallel korpuslarni yaratishdagi ilk qadamlar sifatida o'zaro yaxshi tarjima qilingan matnlarning elektron shakli qancha miqdorda topilsa, ularning barchasidan unumli foydalanish maqsadga muvofiq bo'ladi. Demak, o'zbek-ingliz tili juftligidagi matnlarni internet orqali, masalan, HTML fayl shaklida yoki turli elektron hujjatlar shaklida yig'ish mumkin. Ushbu holatda keraksiz HTML teglari, chizmalar, jadvallar fayllardan olib tashlanadi va korpusga kiritish uchun faqat matndan iborat shaklga keltiriladi.

Ko'plab olimlarning tadqiqotlariga binoan, korpusga, xususan o'zbek-ingliz tillari parallel korpusiga kiritilishidan oldin, har qanday matn quyidagi jarayonlardan o'tmog'i lozim:



**Matnni "tozalash".** Parallel korpus uchun manba tanlashda ko'plab matnlar internet tarmog'i orqali jamlanadi va shu sabab ularning ko'pi HTML fayllardan tashkil topadi. HTML fayllarda matnlardan tashqari ko'plab teglar, ortiqcha havolalar mavjud bo'lishini hisobga olgan holda, daslabki bosqichda parallel matnlar ushbu keraksiz belgilardan holi qilinadi.

**Ekstralingvistik teglash.** Ushbu bosqichda matnlarga ekstralingvistik teglar biriktiriladi. Masalan, matnning qaysi sohaga tegishli ekanligi, kim tomonidan yozilganligi yoki tarjima qilinganligi,

davri, qaysi janrga tegishli ekanligi ekstralingvistik teglash doirasida amalga oshiriladi.

**Parallel tenglashtirish.** Parallel matnlar paragraflar va gaplar kesimida tenglashtiriladi. Ushbu bosqichni amalga oshirish uchun ko'plab avtomatik dasturlar yaratilgan bo'lib, ular bir-biridan aniqlik darajasi, ma'lum bir til yoki til oilasiga moslashtirilganligi va yana boshqa xususiyatlariga ko'ra farq qiladi. Jahondagi ko'plab parallel korpuslar Gale va Church [Gale, Church 1993, 75] taklif etganidek, gaplar kesimida tenglashtirilgan. Bunda parallel matnlardagi gaplar uzunligiga qarab bir-biriga tenglashtiriladi.

**Parallel tenglikni tekshirish.** Yuqorida ta'kidlanganidek parallel matnlarni tenglashtirishda bu jarayonni avtomatik amalga oshiruvchi dasturlardan foydalaniladi. Biroq bu dasturlar 100% aniqlikda matnlarni tenglashtiradi deb hisoblay olmaymiz. Chunki, ingliz tilidagi bitta gap bilan ifodalangan matn bo'lagi o'zbek tilida ikkita gap bilan ifodalanishi mumkin yoki aksincha. Bu holatda avtomatik tenglashtiruvchi dastur parallel matnlarni faqat 1:1 nisbatda tenglashtiradi. Birga ko'p va ko'pga bir tenglashtirish uchun esa inson omili kerak bo'ladi. Shu sabab ushbu bosqichda qandaydir xatolar inson omili orqali tekshiriladi va to'g'rilanadi. LF Aligner, E Align va shu kabi avtomatik tenglashtiruvchi dasturlar yaratilgan bo'lib, ular ingliz tili va boshqa tillar matnlarini gap hamda paragraf kesimida tenglashtiradi.

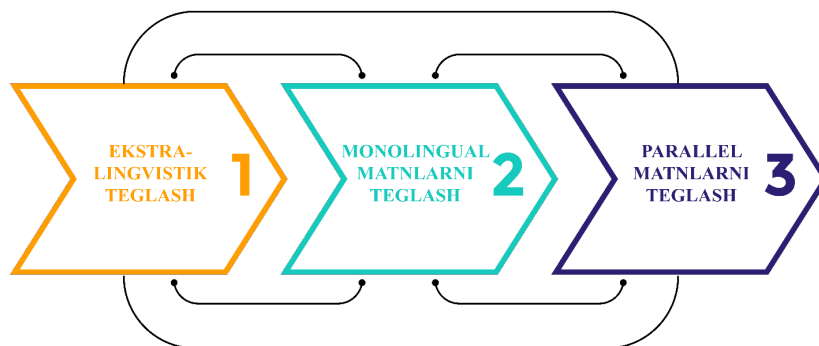
**Bilingual kodlash (xml).** Ushbu bosqichda tasdiqlangan matnlar XML formatda kodlanadi. Kodlash avtomatik XML kodlovchi dasturlar orqali amalga oshiriladi.

**Lemmatizatsiya va pos teglash.** Bugungi kunda parallel korpus sifatini oshirish hamda foydalilik ko'lamini oshirish uchun lemmatizatsiya va Pos teglash muhim omil bo'lib xizmat qiladi. Ushbu jarayonlarni amalga oshirish katta mehnat talab qilsada, parallel korpusning ko'plab xususiyatlarga ega bo'lishiga samarali hissa qo'shadi.

**Matnlarni indekslash.** Oxirgi bosqichda hamma tekshirishlardan o'tgan matnlar indekslanadi. Indekslash jarayonining amalga oshirilishi qidiruv tizimining optimallashtirishiga yordam beradi. Ya'ni katta matnli ma'lumotlar kichik-kichik qismlarga bo'linib indekslanadi va bu esa qidirilayotgan natijaga erishish vaqtini sezilarli darajada tezlashtiradi.

**Parallel korpusni teglash masalasi.** Korpus faqatgina ma'lumotlari teglangandan so'nggina foydali bo'lishi mumkin. Masalan, turli sathlarda teglangan parallel korpusgina **tarjima**

**xotirasi** bo'lib xizmat qiladi. Parallel korpus uchun eng muhim teglash bu **tenglashtirish**dir. Gaplar parallel korpuslarning eng muhim tenglashtirish mahsuloti hisoblanadi. Monolingual korpuslarga tegishli teglarni parallel korpuslarga ham qo'llash mumkin. Misol uchun Pos (so'z turkumi) va sintaktik teglash shular jumlasidandir. Biroq, bu jarayonlarni amalga oshirish katta mehnat va sabrni talab qilishini unutmaslik lozim. Jahon tajribasi hamda bir nechta mashhur parallel korpuslarga tayangan holda korpusni teglashning 3 ta turini aniqladik:



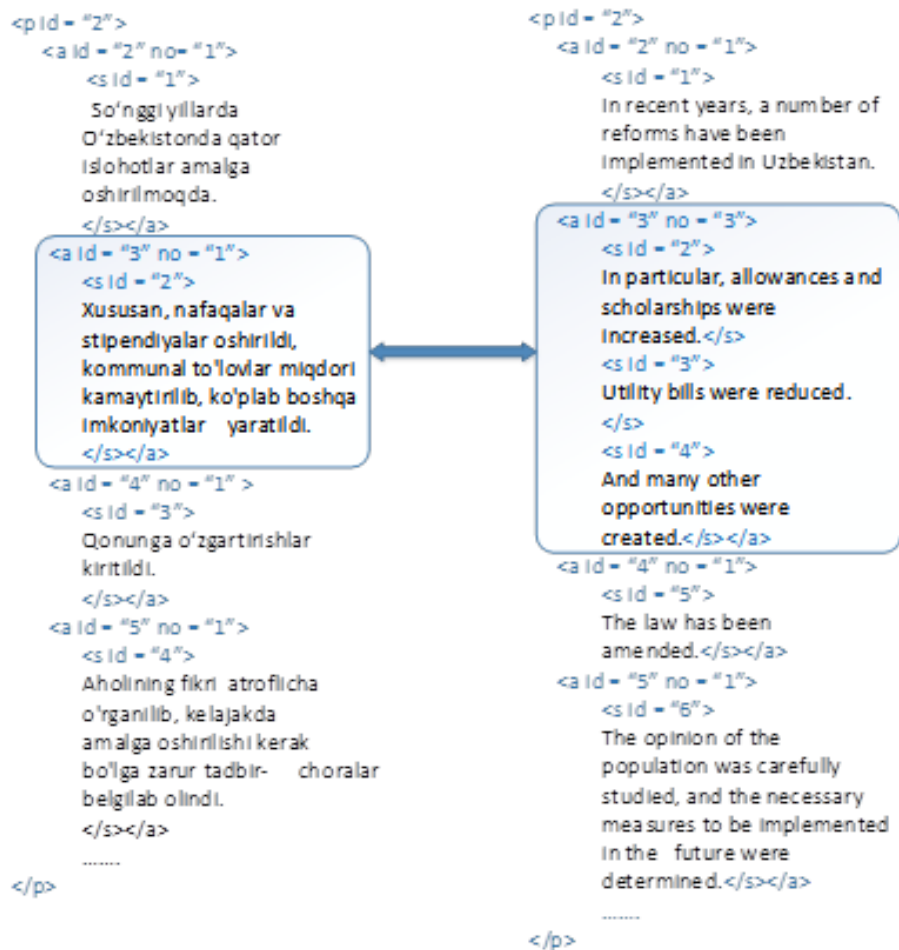
**Ekstralingvistik teglash.** Ekstralingvistik teglash o'z ichiga parallel korpusga kiritilgan to'liq bo'lgan matnlar haqidagi ekstralingvistik ma'lumotlarni qamrab oladi. Masalan, yuqorida ta'kidlangani kabi matnning muallifi, tarjimoni, sohasi, yozma yoki ovoz shaklida ekanligi, qaysi davrga tegishli ekanligi va sarlavhasi shular jumlasidandir. Parallel korpusdagi ekstralingvistik teglangan ma'lumotlar ko'plab tadqiqotchilar uchun qo'l kelishi mumkin. Masalan, tadqiqotchilar faqatgina bir sohaga, bir davrga yoki bir yozuvchiga tegishli bo'lgan matnlarga qiziqayotgan bo'lsa ekstralingvistik teglangan ma'lumotlar qidiruvi orqali o'zi izlayotgan natijaga soniyalar ichida erishishi mumkin [5].

**Monolingual matnlarni teglash.** Ushbu teglash jarayoni parallel korpusdagi ikki til uchun alohida amalga oshiriladi. Matnlar, paragraf, gap va so'zlar kesimida annotatsiyalanadi. Bundan tashqari o'zbek-ingliz tillari parallel korpusida o'zbek tilidagi so'zlar ustida ularning lemmasini biriktirish hamda Pos (so'z turkumlari) teglash jarayoni amalga oshiriladi. Ingliz tilidagi so'zlar uchun ham shu jarayon qo'llanadi.

**Parallel matnlarni teglash.** Bu bosqichda ikki til o'rtasida "ko'prik" o'rnatiladi. Original matn va ularning tarjimalari paragraf va gaplar kesimida tenglashtiriladi. So'zlar kesimida tenglashtirishni amalga oshirish hozirgi vaqtda qiyin vazifa hisoblanadi. Birinchidan,

so'z kesimida tenglashtiruvchi dasturlarning yetishmasligi bo'lsa, ikkinchidan, bu jarayon uchun katta mehnat talab etiladi.

**Bilingual kodlash (XML) masalasi.** Yaratilgan korpusni internet orqali ishga tushirishga qulay bo'lishi uchun barcha matnlar bir xil kodlanishi zarur [Baobao, 2004]. Shu sabab, XML ga asoslangan freymvork tanlanishi maqbul tanlovdir. Bunda barcha o'zbek va ingliz matnlari alohida kodlanadi va ushbu matnlar tilidan qat'iy nazar matnning "bosh"i va "tana"sidan tashkil topadi. Barcha ekstralingvistik ma'lumotlar bosh qismiga, monolingual teglar, lingvistik ma'lumotlar hamda matnning o'zi tana qismiga joylashtiriladi. Parallel matnlarni tenglashtirishlar esa ikki tilning tana qismidagi id raqamlari orqali amalga oshiriladi. Quyida o'zbek-ingliz tillaridagi matnning bir qismi tenglashtirilishining taxminiy ko'rinishini keltiramiz:



**Segmentlash va PoS teglash masalasi.** Korpusni yaratish ko'plab mashaqqat va vaqtni talab qiluvchi jarayondir. Avtomatik dasturlarning yordamisiz korpusni tuzish amrimahol. Ingliz tili



so'zlarini PoS teglovchi dasturlarning bir necha turi allaqachon yaratilgan bo'lsa, o'zbek tili xususida bunday deb ayta olmaymiz. Agar o'zbek tili uchun ham PoS tegger yaratilsa o'zbek tili so'zlari quyidagicha PoS teglanadi va segmentlarga ajratiladi:

```
<TEXT>
<TEXT HEAD>
  <FIELD>Hayvonot</FIELD><STYLE>Ilmiy</STYLE>
  <PERIOD>2022</PERIOD><U_TITLE>Sut emizuvchi hayvonlar</U_TITLE>
</TEXT HEAD>
<TEXT_BODY>
  <p id = "1">
    <a id = "1" no = "1">
      <s id = "1">
        <U_TITLE><w pos = "t">Hayvonlar</w>
          <w pos = "NOUN">Arslon</w>
          <w pos = "CON">va</w>
          <w pos = "NOUN">sirtlonlar</w>
          <w pos = "NOUN">sut emizuvchilar</w>
          <w pos = "VERB">hisoblanadi.</w></U_TITLE></s></a></p>
    <p id = "2">
```

Korpusdagi so'zlarni morfologik teglashda ko'pgina tillar uchun majburiy bo'lgan 12 kategoriya mavjud [Zeroual, Lakhouaja, 2022, 61]:

Verbs	Fe'llar	VERB
Nouns	Otlar	NOUN
Pronouns	Olmoshlar	PRON
Adjectives	Sifatlar	ADJ
Adverbs	Ravishlar	ADV
Numerals	Sonlar	NUM
Determiners (determiners and articles)	Aniqlovchi (mas., artikllar)	DET
Adpositions (prepositions and postpositions)	Qo'shimchalar (old va keyingi)	ADP
Particles	Qismlar	PRT
Conjunctions	Bog'lovchilar	CON
Other (typos, abbreviations...)	Boshqa (xatolar, qisqartmalar)	X
Punctuation marks	Tinish belgilari	SENT

Bundan tashqari parallel korpusga lemmatizatsiya va tokenizatsiya funksiyasini qo'shsak, korpusning foydalilik darajasi yanada ortadi. Quyida rus tili milliy korpusidan namuna keltiramiz:

"book"

Найдено: 455 документов, 5 320 вхождений.

Поискать в других корпусах: [основном](#), [параллельном \(все языковые пары\)](#), [многоязычном](#).

Страницы: 1 2 3 4 5 6 ... 11 ... 46 [следующая страница](#)

1. André Aciman. Find Me (2019) | Андре Асима. Найди меня (Наталья Рашковская, 2020) [омонимия не снята] [Всё примеры \(18\)](#)

eng	For a moment she looked so totally forlorn that, while staring at my open <b>book</b> , I caught myself struggling to come up with something to say, if only about to erupt in our little corner at the very end of the car. [André Aciman. Find Me (2019)   Андре Асима. Найди меня (Наталья Рашковская, 2020)]	<b>book</b>	Лемма	book (см. перевод)	ЛИМЯ И
rus	На мгновение она показалась мне настолько несчастной, что я, продолжая наш разговор в хвосте вагона. [André Aciman. Find Me (2019)   Андре Асима. Найди меня (Наталья Рашковская, 2020)]		Грамматика	сущ. им. ед. глаг. изъяв. наст. глаг. инф. прил.	КЛМЯ Б
eng	Once again I made a motion to pick up my <b>book</b> , convinced we were done talking. [André Aciman. Find Me (2019)   Андре Асима. Найди меня (Наталья Рашковская, 2020)]		Доп. признаки	book, bcooma, posred, numred, casered, timed, poem	И (Ната Рашковская, 2020)
rus	Я снова протянул руку к книге, уверившись, что на этот раз разговор окончен. [André Aciman. Find Me (2019)   Андре Асима. Найди меня (Наталья Рашковская, 2020)]				Сообщить об ошибке
eng	A bit later, with my <b>book</b> still open, I started looking out at the rolling Tuscan landscape and thinking about the time I spent in the hills of Tuscany. [André Aciman. Find Me (2019)   Андре Асима. Найди меня (Наталья Рашковская, 2020)]				
rus	Чуть позже, не закрывая ее, я принялся смотреть на пробегающие за окном тосканские пейзажи и задумался. [André Aciman. Find Me (2019)   Андре Асима. Найди меня (Наталья Рашковская, 2020)]				

Yuqoridagi ilovada rus tili milliy korpusi tarkibidagi ingliz-rus parallel korpusidan "book" so'zi qidirilmoqda. Avvalo, "book" so'zi qatnashgan parallel matnlar aks ettirilgan. Ko'rib turganimizdek, so'zning lemmasi "book", grammatik jihatdan ot va fe'l so'z turkumi, birlikda ekanligi ko'rsatilgan. Bu nafaqat "book" so'zi uchun, balki qidiruv natijasida paydo bo'lgan matnning istalgan, u xoh rus tili, xoh ingliz tili so'zining ustiga bosish orqali ko'rinadigan ma'lumotlardir. Demak korpusdagi so'zlarning har biriga lingvistik izoh berilib teglangan, bundan tashqari matnlarga ekstralingvistik teglar birlashtirilgan. Qidiruv jarayonida qidirilayotgan so'zning nechta hujjatda va nechta o'rinda kelayotganligi ham aks etgan. Bir so'z bilan aytganda, rus tili milliy korpusidagi parallel korpuslarni mukammal teglangan deb sanash mumkin. Nafaqat lingvistik, balki yuqori sifatli ekstralingvistik teglanganlik xususiyati ham rus tili milliy korpusi ustida qanchalik ter to'kilganligini ko'rsatadi.

**Korpusni indekslash masalasi.** Korpusdan qidirish jarayoni vaqt talab etuvchi jarayon bo'lishi mumkin. Ammo, bu jarayonni tezlashtirish imkoni mavjud. Buning uchun qilinadigan ish ushbu korpusdagi matnlarning indekslanishidir. Bunda teskari indekslash usulidan foydalaniladi. Matnlar qismlarga bo'lib indekslanadi. Shunda matnning biror qismidagi ma'lumot qidirilayotganda, butun matn emas uning ozgina qismigagina murojaat qilinadi va qidiruv natijasi soniyalar ichida tezda namoyon bo'ladi. Agar yangi qo'shilayotgan matnlar ustida to'g'ri XML kodlash amalga oshirilsa, butun korpusni qayta kodlashni amalga oshirishga zarurat bo'lmaydi. Demak, indekslash yaratilishi kutilayotgan parallel korpusning mukammal qidiruv tizimini tashkil etuvchi omillaridan biri sanalmish tezkorlik talabini bajarishni o'z zimmasiga olar ekan, tatqiqotchilar o'zbek-ingliz parallel korpusini yaratishda ushbu jihatga alohida e'tibor qaratishlari maqsadga muvofiqdir.

## **Xulosa**

Korpus lingvistikasi jadallik bilan rivojlanib bormoqda. Bu esa o'z navbatida yurtimizda ham ushbu soha taraqqiyoti uchun harakatlarni ildamlashtirish masalalarini hal etishni taqozo qilmoqda. Buning uchun tashlanadigan ilk va istiqbolli qadamlardan biri – bu o'zbek hamda jahon tillari juftligidagi parallel korpuslarni yaratish. Xususan, o'zbek-ingliz tillari parallel korpusining yaratilishi tillararo ma'lumotlarni qayta ishlash, shuningdek, o'zbek-ingliz tilli leksikografiya, til tadqiqoti va tilni o'qitish uchun muhim manba bo'lib xizmat qilishi shubhasiz, albatta. Ammo, bunday korpusni qanday yaratish, undan qanday qilib samarali ravishda foydalanish, unda qanday ma'lumotlarni jamlash kabi vazifalar dolzarb masala bo'lib qolmoqda. Ushbu maqola korpus sohasida jahon tajribasiga tayangan holda, o'zbek-ingliz parallel korpusini yaratish bosqichari to'g'risida nazariy ma'lumotlar beradi. Umid qilamizki, ushbu ish yurtimizdagi korpus sohasiga qiziquvchilar uchun dasturulamal bo'lib xizmat qiladi va soha rivojiga munosib hissasini qo'shadi.

## **Adabiyotlar:**

- Chanjun Park, Seolhwa Lee, Hyeonseok Moon Sugyeong Eo, Jaehyeong Seo and Heuseok Lim. How should human translation coexist with NMT? Efficient tool for building high quality parallel corpus. 35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia
- John Hutchins. Machine translation and human translation: in competition or in complementation. *International Journal of Translation*, 13(1-2):5–20, 2001.
- Jorge Leiva Rojo. Aspects of human translation: the current situation and an emerging trend. *Hermeneus: Revista de la Facultad de Traducción e Interpretación de Soria*, (20):257–294, 2018.
- Q.F.Wen, L.F.Wang and M.C.Liang: Spoken and Written English Corpus of Chinese Learners[M], Beijing: Foreign Language Teaching and Research Press, 2005.(In Chinese)
- CHANG Baobao. Chinese-English Parallel Corpus Construction and its Application. PACLIC 18, December 8th-10th, 2004, Waseda University, Tokyo.
- Gale, W. A., & Church, K. W. 1993. A program for aligning sentences in bilingual corpora. *Computational Linguistics*, 19(3), 75-102.
- Karimov Rustam. O'zbek-ingliz parallel korpusini tuzishning lingvistik va dasturiy masalalari. Dissertasiya. Buxoro – 2022
- Imad Zeroual, Abdelhak Lakhouaja. MulTed: a multilingual aligned and tagged parallel corpus. *Applied Computing and Informatics* Vol. 18 No. 1/2, 2022 pp. 61-73
- Alkahtani, Saad. Building and verifying parallel corpora between Arabic and English. DOCTOR OF PHILOSOPHY. Bangor University,

2015.

- Leech, G.N. 2010. Corpus linguistics. In K. Malmkjaer (ed.), *The Routledge Linguistics Encyclopedia, Third Edition*. London: Routledge, 103-113.
- Svartvik, J. 1992. Corpus linguistics comes of age. In J. Svartvik (ed.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*. Berlin: Mouton de Gruyter, 7-13.
- Johansson, S. & Hofland, K. 1994. Towards an English-Norwegian parallel corpus. In U. Fries, G. Tottie and P. Schneider (eds), *Creating and Using English Language Corpora*. Amsterdam: Rodopi, 25-37.
- Johansson, S. 2007. *Seeing through Multilingual Corpora*. Amsterdam: Benjamins.

## **Steps of creating a tagged parallel corpus of the uzbek-english languages**

Botir Elov<sup>1</sup>  
Ma'rufjon Amirqulov<sup>2</sup>

### **Abstract:**

The article talks about corpus linguistics, which is one of the main directions of computer linguistics, monolingual corpora and parallel corpora, and also about the stages of creating a parallel corpus of Uzbek-English languages based on world experience in the field of parallel corpora is maintained. In addition, information is provided about priority tasks such as establishing the programming and linguistic principles of the parallel corpus of Uzbek-English languages, linguistic and extralinguistic tagging of selected units, and developing an algorithm for creating a parallel corpus. Considerations are given on how to select data for the parallel corpus, the requirements for the data, and what opportunities the creation of the Uzbek-English parallel corpus provides to researchers and users.

---

<sup>1</sup>*Elov Botir Boltayevich* – doctor of philosophy in technical sciences (PhD), associate professor. Head of the Department of Computer Linguistics and Digital Technologies of Tashkent State University of Uzbek Language and Literature named after Alisher Navoi.

**E-mail:** elov@navoiy-uni.uz

**ORCID:** 0000-0001-5032-6648

<sup>2</sup>*Amirqulov Ma'rufjon Aliqul o'g'li* – 2nd year master's degree in Computer Linguistics at Alisher Navoi Tashkent State University of Uzbek Language and Literature.

**E-mail:** amirkulovmaruf@navoiy-uni.uz

**ORCID:** 0000-0002-4025-8466

**For reference:** Elov B., Amirqulov M.. 2022. "Steps of creating a tagged parallel corpus of the uzbek-english languages". *Uzbekistan: language and culture. Applied philology*. 2 (5): 97-109.

In this process, linguistic and methodological problems, such as material selection, as well as programmatic difficulties in creating a parallel corpus are reflected.

**Key words:** *Corpus, tagging, alignment, parallel corpora, POS tagging, XML encoding.*

**References:**

- Chanjun Park, Seolhwa Lee, Hyeonseok Moon Sugyeong Eo, Jaehyeong Seo and Heuseok Lim. How should human translation coexist with NMT? Efficient tool for building high quality parallel corpus. 35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia
- John Hutchins. Machine translation and human translation: in competition or in complementation. *International Journal of Translation*, 13(1-2):5-20, 2001.
- Jorge Leiva Rojo. Aspects of human translation: the current situation and an emerging trend. *Hermeneus: Revista de la Facultad de Traducción e Interpretación de Soria*, (20):257-294, 2018.
- Q.F.Wen, L.F.Wang and M.C.Liang: Spoken and Written English Corpus of Chinese Learners[M], Beijing: Foreign Language Teaching and Research Press, 2005.(In Chinese)
- CHANG Baobao. Chinese-English Parallel Corpus Construction and its Application. PACLIC 18, December 8th-10th, 2004, Waseda University, Tokyo.
- Gale, W. A., & Church, K. W. 1993. A program for aligning sentences in bilingual corpora. *Computational Linguistics*, 19(3), 75-102.
- Karimov Rustam. O'zbek-ingliz parallel korpusini tuzishning lingvistik va dasturiy masalalari. Dissertasiya. Buxoro – 2022
- Imad Zeroual, Abdelhak Lakhouaja. MulTed: a multilingual aligned and tagged parallel corpus. *Applied Computing and Informatics* Vol. 18 No. 1/2, 2022 pp. 61-73
- Alkahtani, Saad. Building and verifying parallel corpora between Arabic and English. DOCTOR OF PHILOSOPHY. Bangor University, 2015.
- Leech, G.N. 2010. Corpus linguistics. In K. Malmkjaer (ed.), *The Routledge Linguistics Encyclopedia*, Third Edition. London: Routledge, 103-113.
- Svartvik, J. 1992. Corpus linguistics comes of age. In J. Svartvik (ed.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*. Berlin: Mouton de Gruyter, 7-13.
- Johansson, S. & Hofland, K. 1994. Towards an English-Norwegian parallel corpus. In U. Fries, G. Tottie and P. Schneider (eds), *Creating and Using English Language Corpora*. Amsterdam: Rodopi, 25-37.
- Johansson, S. 2007. *Seeing through Multilingual Corpora*. Amsterdam: Benjamins.

## Matnlarda NER obyektlarini aniqlashdagi muammolar

Madina Samatboyeva<sup>1</sup>

### Abstrakt:

Nomlangan obyektning tanib olish axborotni boshqarishning ko'plab boshqa sohalar uchun asos bo'lib xizmat qiladi. Tabiiy tilni qayta ishlashda nomli obyektlarni tanib olish (NER) matndagi obyektlarning muayyan turlarini tanib olish va ajratib olish muammosidir. Haqiqatan ham, ismga ega bo'lgan har qanday aniq "narsa" NER obyekti sifatida ajratiladi. Ushbu maqolada mashinaviy o'qitish yordamida NER obyektlarini aniqlashda yuzaga keladigan muammolar ingliz tili va o'zbek tili matnlari tarkibida solishtirildi va tadqiq qilindi.

**Kalit so'zlar:** *tabiiy tilni qayta ishlash, NLP, mashinaviy o'qitish, nomlangan obyekt, NER, matn.*

### Kirish

Axborotni olish (IE) kompyuter lingvistikasining asosiy kichik vazifalaridan biri bo'lib, tizimli ma'lumotlarni tizimli bo'lmagan yoki yarim tuzilmali mashina tomonidan o'qilishi mumkin bo'lgan hujjatlardan avtomatik ravishda ajratib olishga qaratilgan. Umatndagi barcha kompaniya nomlarini topishdan tortib, voqeaning barcha obyektlarini topishgacha bo'lgan keng ko'lamli kichik vazifalarni o'z ichiga oladi. Masalan, kim qaysi kompaniyaga ishga kirgan yoki kim qaysi bo'limda ishlayapti. Bunday imkoniyatlar foydalanuvchiga kerakli ma'lumotlarni topish uchun katta hajmdagi onlayn matnlarni saralash uchun tobora muhim ahamiyat kasb etmoqda.

Nomlangan obyektlar (NER)ni tanib olish va uning turlarini avtomatik identifikatsiya qilish vazifasi axborot olishning eng intensiv o'rganilgan usullaridan biridir. Til tahlilining taqdimoti odatda lug'atdagi so'zlarni qidirish va ularni ot, fe'l, sifat va boshqalar sifatida aniqlashdan boshlanadi. Lekin ko'pchilik matnlar juda ko'p

---

<sup>1</sup>Samatboyeva Madina To'liqinjon qizi - Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti Kompyuter lingvistikasi mutaxassisligi 1-kurs magistranti.

**E-pochta:** samatboyevamadina@navoiy-uni.uz

**ORCID:** 0000-0002-5146-4484

**Iqtibos uchun:** Samatboyeva M. 2022. "Matnlarda NER obyektlarini aniqlashdagi muammolar". *O'zbekiston: til va madaniyat. Amaliy filologiya*. 2 (5): 110-118.

ismlarni o'z ichiga oladi. Agar tizim ularni lug'atdan topa olmasa, ularni aniqlay olmaydi. Bu o'rinda matnni lingvistik tahlil qilish qiyin. Shunday qilib, NER ko'plab tabiiy tillarni qayta ishlash (NLP) vazifalarida (masalan, ma'lumot olish (IR) yoki mashina tarjimasini (MT)) muhim ahamiyatga ega. [Simon, 2013]

### **Asosiy qism**

Matn tarkibidan NER obyektlarini avtomatik aniqlashda bir nechta muammolar mavjud. Bu muammolar matn tarkibidagi NER obyektlarini avtomatik tanib olish, ularga mos modellar yaratish, teglash, obyekt turini hamda uning matn tarkibidagi chegarasini belgilashda yuzaga keladi. Dastlab, biz ikki tushunchani o'rganib olishimiz zarur. Bular: *NER va nomlangan obyekt*.

Nomlangan obyektning tanib olish (NER) - *obyektning identifikatsiyalash* yoki *obyektning ajratib olish* deb ham ataladi - matndagi nomlangan obyektlarni avtomatik ravishda aniqlaydigan va ularni oldindan belgilangan toifalarga tasniflaydigan tabiiy tilni qayta ishlash (NLP) usuli. Matndagi NER obyektlari shaxslar, odamlar, tashkilotlar, joylar, vaqtlar, miqdorlar, pul qiymatlari, foizlar va boshqalarning nomlari bo'lishi mumkin. [ <https://monkeylearn.com> ]

**Nomli obyekt** - tegishli nom bilan ko'rsatilgan yoki ko'rsatilishi mumkin bo'lgan haqiqiy yoki xayoliy har qanday narsa. Nomlangan obyekt nima ekanligini bilish uchun NER modeli obyektning tashkil etuvchi so'z yoki so'zlar qatorini (masalan, London shahri) aniqlay olishi va u qaysi obyekt toifasiga tegishli ekanligini (London - Shahar (city)) bilishi kerak. *Nomlangan obyektlarni tanib olish* ularning *havolarini topish!* *nazarda tutilgan obyekt yoki shaxsni aniqlash!* *vazifasi nafaqat matnda, shuningdek, unga toifa berishni!* o'z ichiga oladi.

(NER), ehtimol, ma'lumot olish yo'lidagi birinchi qadam bo'lib, matndagi nomli obyektlarni oldindan belgilangan toifalarga, masalan, shaxslar, tashkilotlar, joylar, vaqt ifodalari, miqdorlar, pul qiymatlari, foizlar va boshqalarni aniqlash va tasniflashga intiladi. [Susan Li. 2018] SpaCy kutubxonasidan foydalangan holatda ajratib olingan NER obyektlari bilan tanishamiz [ <https://towardsdatascience.com> ]:

No	TYPE	DESCRIPTION
1	PERSON	People, including fictional.
2	NORP	Nationalities or religious or political groups.

3	FAC	Buildings, airports, highways, bridges, etc.
4	ORG	Companies, agencies, Institutions, etc.
5	GPE	Countries, cities, states.
6	LOC	Non-GPE locations, mountain ranges, bodies of water.
7	PRODUCT	Objects, vehicles, foods, etc. (Not services)
8	EVENT	Named hurricanes, battles, wars, sports events, etc.
9	WORK OF ART	Titles of books, songs, etc.
10	LAW	Named documents made into laws.
11	LANGUAGE	Any named language.
12	DATE	Absolute or relative dates or periods.
13	TIME	Times smaller than a day.
14	PERCENT	Percentage, including «%»
15	MONEY	Monetary values, including unit.
16	QUANTITY	Measurements, as of weight or distance.
17	ORDINAL	«first», «second», etc.
18	CARDINAL	Numerals that do not fall under another type.

### 1-jadval. SpaCy kutubxonasidagi NER obyektlari

No	Type	Tag	Sample Categories	Example sentences
1	Person	PER	people, characters	Turing is a giant of computer science.
2	Organization	ORG	companies, sports teams	The IPCC warned about the cyclone.
3	Location	LOC	regions, mountains, seas	The Mt. Sanitas loop is in Sunshine Canyon.
4	Geo-Political Entity	GPE	countries, states, provinces	Palo Alto is raising the fees for parking.
5	Facility	FAC	bridges, buildings, airports	Consider the Golden Gate Bridge.
6	Vehicles	VEH	planes, trains, automobiles	It was a classic Ford Falcon.

### 2-jadval. NER obyektlarining teglangan va kategoriyalangan ko'inishi [<https://medium.com>]

Shunday qilib, birinchi navbatda, biz nom, joylashuv, voqea, tashkilot va boshqalar kabi obyektlar toifalarini yaratishimiz va NER modeliga tegishli ma'lumotlarni ta'minlashimiz kerak. Keyin, ba'zi so'z va ibora namunalarini tegishli obyektlar bilan belgilash orqali biz NER modeliga obyektlarning o'zini qanday aniqlashni o'rgatamiz. [<https://monkeylearn.com>]

Endi esa ushbu NER obyektlari aniqlangan dasturdan misol taqdim etamiz [<https://towardsdatascience.com>]:





### 3-rasm. NER obyektlarini aniqlovchi dastur ko'rinishi

Ushbu 3-rasm-jadvalga e'tibor qaratsak, uning tarkibida barcha NER obyektlari maxsus ranglar bilan ajratilgan va obyektlarning barchasiga toifa (obyekt turi nomi) belgilangan. Bundan tashqari, NER hujjat tahlilining bir turi hisoblanadi. U matnda qaysi obyektlar - shaxslar, joylar, tashkilotlar, sanalar, manzillar va hokazolar eslatib o'tilganligini va semantik tahlil orqali xulosa chiqarish mumkin bo'lgan shaxsning atributlarini belgilaydi. Masalan, shaxslar olmoshlarda va nomli eslatmalarga ishora qiluvchi qisqaroq shakllarda ham belgilanadi. Ushbu usul bilan nomni tanib olishning bunday turi anaforikdir, chunki obyektlar anaforalar orqali tan olinadi. Quyidagi misolga e'tibor qarating:

Michael Jordan was one of the best basketball players of all time. Scoring was Jordan's stand-out skill, but he still holds a defensive NBA record, with eight steals in a half.

### 4-rasm. <https://docs.expert.ai/nlapi/latest/guide/entity-recognition/>

Ushbu misolda Maykl Jordan haqida ikkita eslatma mavjud:

To'liq ismi: Michael Jordan

Anoforalar: Jordan (Jordan's), u – olmoshi (he – pronoun)

Ushbu anaforik usul NER obyektlarini matn tarkibidan aniq ajratib olish uchun to'liq samarali usul bo'la olmaydi. Matnni qayta ishlab, uning tarkibidagi obyektlarni avtomatik aniqlash jarayonida ham bir qancha muammolar yuzaga keladi. Quyida ushbu NER obyektlarini aniqlashdagi muammolar bilan tanishamiz.

**Segmentatsiyaning noaniqligi.** Bunda so'zlar semantik jihatdan tahlil qilinib, ular orasidagi chegara to'g'ri ajratilishi talab

etiladi. Masalan: Buyuk Britaniya (yoki: O‘zbekiston Respublikasi) – bunda so‘zlar qismi 2 ta, ammo ma’no bitta. Demak ular bitta obyekt, bitta NER.

Nyu York – bunda esa qismlar bitta ma’noga egami yoki ular alohida ma’noga ega so‘zlarmi? Ya’ni birikmali so‘zlar kelganda ular bitta obyekt deb qaraladimi yoki ular alohida semaga ega ikkita obyektmi? Mana shunday muammoli vaziyatlar bizda dastlabki yechimlarni taqdim etishni talab etadi.

**Teglarni belgilashdagi noaniqlik.** NER obyektlarini avtomatik aniqlashda, avvalo, teglash va modellashtirish jarayoni amalga oshiriladi. Teglash jarayoni esa tildagi har bir sintaktik bo‘lak uchun mos teglarni talab etadi. Teglash jarayonida teglanayotgan obyekt bitta tegga bo‘ysunsa yaxshi, ammo teglanish uchun tanlangan obyekt bir nom ostida bir nechta ma’noni hosil qilsa, bu holatda teglash jarayonida ham noaniqliklar yuzaga keladi. Masalan, quyidagi matnga e’tibor qaratamiz [<https://medium.com>]:

No	Name	Possible Categories
1	Washigton	Person, Location, Political Entity, Organization, Vehicle
2	Downing St.	Location, Organization
3	IRA	Person, Organization, Monetary Instrument
4	Louis Vuitton	Person, Organization, Commercial Product

*5-jadval. Bitta NER nomi ostida turli obyektlarning ko‘rinishi*  
[PER Washigton] was born into slavery on the farm of James Burroughs.

[ORG Washington] went up 2 games to 1 in the four-game series.

Blair arrived in [LOC Washington] for what may well be his last state visit. In June, [GPE Washington] passed a primary seatbelt law.

The [VEH Washington] had proved to be a leaky ship, every passage I made... [<https://medium.com>]

Yuqoridagi ingliz tilidagi matn tarkibida bitta “Vashington” obyekt ostida bir nechta teglarni talab etuvchi NER turlari (inson ismi (PER), tashkilot (ORG), yo‘nalish (LOC), siyosiy tuzilma (GPE), transport (VEH)) keltirilgan. Bu o‘rinda matn qayta ishlanganda dastur NER obyekt qaysi obyekt turi aynan shu matn tarkibida kelayotganini to‘g‘ri ajrata olishi kerak. Bunday muammo, nafaqat ingliz tili matnlari uchun, balki o‘zbek tilidagi matnlarni qayta ishlashda ham mavjuddir. Quyida bitta “Alisher Navoiy” NER obyektiga e’tibor qaratamiz:

1. Alisher Navoiy – buyuk shoir. (Shaxs nomi)
2. Men “Alisher Navoiy” mahallasida yashayman. (Mahalla nomi)
3. U kecha “Alisher Navoiy” romanini o‘qidi. (Roman)
4. “Alisher Navoiy” fondi matnshunoslik bilan shug‘ullanadi. (Fond nomi)
5. Shahrimiz markazida “Alisher Navoiy” kutubxonasi joylashgan. (Kutubxona)
6. “Alisher Navoiy” bog‘i yurtimiz ko‘rkiga ko‘rk qo‘shmoqda. (Bog‘)
7. Davron Navoiy viloyatida yashaydi. (Viloyat)
8. Alisher Navoiy nomidagi davlat mukofoti ta‘sis etildi. (Davlat mukofoti)
9. Kecha televizorda “Alisher Navoiy” filmi namoyish etildi. (Film nomi)
10. O‘zRFA Alisher Navoiy nomidagi Til va adabiyot instituti (Institut nomi)
11. Alisher Navoiy nomidagi opera va balet akademik teatri ochildi. (Teatr nomi)
12. Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti (Universitet)
13. Alisher Navoiy qishloq vrachlik punkti (Samarqand, Bulung‘ur tumanidagi qishloq vrachlik punkti)
14. Alisher Navoiy QFY (Samarqand, Bulung‘ur tumanidagi QFY)
15. Alisher Navoiy nomidagi fermer xo‘jaligi (Samarqand, Narpay tumanidagi fermer xo‘jaligi)
16. Alisher Navoiy haykallari (Toshkent, Navoiy, Samarqand, Boku, Tokio, Moskva, O‘sh, Shanxayda mavjud haykallari)
17. Alisher Navoiy nomidagi Merkuriydagi krater (Astronomik krater)

**Chegaralarning noaniqligi.** Ushbu vaziyatda matn tarkibidagi NER obyektlari chegarasi aniq bo‘linishni talab etadi. Quyida ikki misolga e‘tibor qaratamiz:

1. Bog‘da *qizil* va *oq* gullar ochilib turibdi.
2. Tarixda Angliyada “*Qizil va oq gullar*” qo‘zg‘oloni bo‘lib o‘tgan.

1-gapda ikkita NER obyekti mavjud, ya‘ni “qizil” va “oq” (**COL**).  
2-gap tarkibida esa bitta NER obyekti mavjud – “Qizil va oq gullar” (**voqea nomi – VEH**)

Ushbu vaziyatda dastur matnini qayta ishlashda obyekt chegarasini to'g'ri aniqlay olishi zarur. Bunda, albatta, kompyuter uchun biz taqdim etgan modellar ishlaydi. Ya'ni bunda *B* – *boshi* (*begin*), *E* – *oxiri* (*end*), *I* – *oraliq* (*intermediate*), *S* – *bitta so'z* (*singular*) modellaridan foydalanishimiz mumkin.

**Bosh harflar bilan bog'liq noaniqlik.** Juda ko'p hollarda, nomlangan obyektlar matn tarkibida bosh harf bilan beriladi. Bu ularning "nomlanganlik" belgisi bilan bog'liq. Qayta nomlangan obyekt, o'z sifatiga, ko'rinishiga, holatiga qarab boshqa nom bilan ataladi. Bu o'rinda atab qo'yilgan nom matn tarkibida (gapning qaysi o'rnida kelishidan qat'iy nazar) bosh harf bilan beriladi. Ammo NER obyektlarini aniqlashda har doim ham bosh harflar qoidasiga tayanish biz kutgan ijobiy samarani bermaydi. Masalan:

"Davlat rahbarining topshirig'iga asosan, davlat oliy ta'lim muassasalariga o'qishga qabul qilish jarayonlarini muvofiqlashtirish bo'yicha komissiyaning tegishli bayoni bilan Qrim va Sevastopol hududlaridagi OTMda tahsil olayotgan O'zbekiston fuqarolarining ham o'qishini O'zbekistondagi davlat OTMga istisno tariqasida ko'chirishga ruxsat berildi. Bu qonun ko'pgina talabalarga yengillik taqdim etdi" [<https://kun.uz/uz/>].

Ushbu gapda qizil bilan belgilab ko'rsatilgan obyektlar NER obyektlaridir va ular gap tarkibida bosh harfda kelmoqda. Ammo nuqtadan so'ng punktuatsion qoidaga bo'ysungan holatda bosh harf bilan yoziluvchi so'zlar (davlat, bu) NER obyektidir emas.

Boshqa tillar doirasida tadqiq qilganimizda ham, masalan, ingliz va rus tillarida nomlar bosh harfda yozilishi ham yozilmasligi ham mumkin. Bunga qarama-qarshi tarzda, nemis tilida esa barcha nomlar to'liq bosh harfda ifodalanadi. Ammo arab va xitoy tilida esa bosh harf tushunchasi mavjud emas. Bundan ko'rinadiki, NER obyektlarini avtomatik aniqlashda bosh harflar har doim biz kutgan natijani bermaydi.

**Sonlar va yillar bilan bog'liq noaniqliklar.** Bunda raqamlar bilan ifodalangan yillar aniqlikni beradi. Masalan, "*Biz 2010-yil Toshkentga borgan edik*" – ushbu gapda 2010-yil birligini NER obyektidir sifatida qabul qilish mumkin. Ammo "*Biz may oyida Toshkentga borgan edik*" – gapida *may oyi* birligini NER obyektidir sifatida qabul qilib bo'lmaydi. Chunki bunda vaqtning aniqligi yo'q. Ya'ni may oyi qaysi yilga tegishli ekanligi noma'lum. Bundan ko'rinadiki, vaqt bu - NER degan tushunchaga kelmasligimiz kerak.

### **Xulosa**

Axborot yaratish tezligi yillar davomida o'sib bormoqda,

ayniqsa hujjatlar va hisobotlar kabi matnga asoslangan ma'lumotlar juda ko'p. Ushbu ma'lumotlardan foydalanib, biz kuchli algoritmlar yarata olamiz. Biroq, bu algoritmlar yoki kompyuter dasturlarini noldan yaratish, ayniqsa, matnga asoslangan ma'lumotlardan axborot olish uchun keng ko'lamlı tajribalarnı talab qiladi. Shuning uchun biz matnga asoslangan hujjatlardan muhim obyektlarnı aniqlash va chiqarish uchun NLP (Tabiiy tilni qayta ishlash) usullariga tayanamiz.

Nomlangan obyektlarnı tanib olish (NER) obyektlarini aniqlash matnnı avtomatik qayta ishlashga mo'ljallangan NLP dasturlari uchun muhim qism hisoblanadi. Maqolada, nomlangan obyektı tanib olish asoslari va undan foydalanish holatlari yoritildi. NER obyektlarnı aniqlashdagi muammoli vaziyatlar tahlil qilindi va o'rganildi. Shuni alohida aytib o'tishimiz mumkinki, SpaCy - maxsus NER modellarini yaratish va ulardan foydalanish uchun yaxshi kutubxona. Demak, ushbu kutubxona yordamida yuqorida biz taqdim etgan muammoli vaziyatlarga ham yechim topa olishimiz mumkin.

#### **Adabiyotlar:**

- Simon 2013 Simon, Eszter. 2013. Approaches to Hungarian Named Entity Recognition. PhD. Thesis. [http://www.cogsci.bme.hu/~ktkuser/PHD\\_iskola/dissertations/20131011\\_Simon\\_Eszter/ertekezes.pdf](http://www.cogsci.bme.hu/~ktkuser/PHD_iskola/dissertations/20131011_Simon_Eszter/ertekezes.pdf)
- [https://monkeylearn.com/blog/named-entity-recognition/#:~:text=Named%20entity%20recognition%20\(NER\)%20helps,to%20deal%20with%20large%20datasets](https://monkeylearn.com/blog/named-entity-recognition/#:~:text=Named%20entity%20recognition%20(NER)%20helps,to%20deal%20with%20large%20datasets)
- Susan Li. 2018 Changing the world, one post at a time. Sr Data Scientist, Toronto Canada. <https://www.linkedin.com/in/susanli/>.  
<https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da>
- [https://monkeylearn.com/blog/named-entity-recognition/#:~:text=Named%20entity%20recognition%20\(NER\)%20helps,to%20deal%20with%20large%20datasets](https://monkeylearn.com/blog/named-entity-recognition/#:~:text=Named%20entity%20recognition%20(NER)%20helps,to%20deal%20with%20large%20datasets)
- <https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da>
- Mehal Gupta. 2020. <https://medium.com/data-science-in-your-pocket/named-entity-recognition-ner-using-conditional-random-fields-in-nlp-3660df22e95c>  
<https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da>
- <https://kun.uz/uz/news/2022/11/09/qrin-va-sevastopol-hududlaridagi-talabalar-ham-istisno-tariqasida-oqishini-kochirishi-mumkin>

# Problems of NER object identification in texts

Madina Samatboeva<sup>1</sup>

## Abstract:

Named entity recognition serves as the basis for many other areas of information management. In natural language processing, named entity recognition (NER) is the problem of recognizing and extracting certain types of entities in text. In fact, any particular “thing” that has a name stands out as a NER object. This article compared and studied the problems that arise when identifying NER objects using machine learning, as part of English and Uzbek texts.

**Key words:** *natural language processing, NLP, machine learning, named object, NER, text.*

## References:

- Simon 2013 Simon, Eszter. 2013. Approaches to Hungarian Named Entity Recognition. PhD. Thesis. [http://www.cogsci.bme.hu/~ktkuser/PHD\\_iskola/dissertations/20131011\\_Simon\\_Eszter/ertekezes.pdf](http://www.cogsci.bme.hu/~ktkuser/PHD_iskola/dissertations/20131011_Simon_Eszter/ertekezes.pdf)
- [https://monkeylearn.com/blog/named-entity-recognition/#:~:text=Named%20entity%20recognition%20\(NER\)%20helps,to%20deal%20with%20large%20datasets](https://monkeylearn.com/blog/named-entity-recognition/#:~:text=Named%20entity%20recognition%20(NER)%20helps,to%20deal%20with%20large%20datasets)
- Susan Li. 2018 Changing the world, one post at a time. Sr Data Scientist, Toronto Canada. <https://www.linkedin.com/in/susanli/>. <https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da>
- [https://monkeylearn.com/blog/named-entity-recognition/#:~:text=Named%20entity%20recognition%20\(NER\)%20helps,to%20deal%20with%20large%20datasets](https://monkeylearn.com/blog/named-entity-recognition/#:~:text=Named%20entity%20recognition%20(NER)%20helps,to%20deal%20with%20large%20datasets)
- <https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da>
- Mehal Gupta. 2020. <https://medium.com/data-science-in-your-pocket/named-entity-recognition-ner-using-conditional-random-fields-in-nlp-3660df22e95c>
- <https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da>
- <https://kun.uz/uz/news/2022/11/09/qrim-va-sevastopol-hududlaridagi-talabalar-ham-istisno-tariqasida-oqishini-kochirishi-mumkin>

---

<sup>1</sup>Samatboyeva Madina To'liqinjon qizi - 1st year master's degree in Computer Linguistics at Alisher Navoi Tashkent State University of Uzbek Language and Literature.

**E-mail:** samatboyevamadina@navoiy-uni.uz

**ORCID:** 0000-0002-5146-4484

**For reference:** Samatboyeva M. 2022. “Problems of NER object identification in texts”. *Uzbekistan: language and culture. Applied philology.* 2 (5): 110-118.

## MAQOLA TAQDIM QILISH TALABLARI

O'zbekiston: til va madaniyat (O'zTM) – zamonaviy O'zbekiston (sobiq Turkiston) bilan bog'liq bevosita Markaziy Osiyo mintaqasini birlashtiradigan til, tarix, san'at, etnografiya, madaniyat va ijtimoiy fanlar sohalarini qamrab olgan ilmiy jurnaldir. O'zTM munozarali, zamonaviy, innovatsion, konseptual jihatdan qiziqarli, original mavzudagi ilmiy tadqiqotlarni nashr qiladi. Jurnal lingvistika, adabiyotshunoslik, tarjimashunoslik, din, falsafa, ilohiyot, fan, ta'lim, metodika, sotsiologiya, psixologiya, tarix, madaniyat, san'at, etnologiya, etnografiya, antropologiyaga oid ilmiy yo'nalishdagi maqolalar va taqrizlar hamda konferensiya hisobotlarini qabul qiladi.

### I. Maqola taqdim etish uchun umumiy talablar

Qo'lyozmalar o'zbek, ingliz, rus, fors, shuningdek, boshqa turkiy tillarda ham qabul qilinadi. Agar muallif o'z maqolasini jurnalning muayyan sonida nashr ettirmoqchi bo'lsa, unda qo'lyozma jurnal nashridan kamida besh oy oldin taqdim etilishi lozim.

Qo'lyozmalar MS Word (.doc) formatida (uzlangcult@gmail.com) elektron pochtasiga yuboriladi. Iqtiboslar va izohlar uchun MS Word menejerini qo'llash mumkin.

Barcha qo'lyozmalar tahririyatga muallif (mualliflar) haqidagi qisqacha ma'lumot bilan taqdim etiladi.

Asosiy matn *Times New Roman* shrifti, 14 hajm, satr oralig'i 1 interval, hoshiyalar chapdan 3 sm, o'ngdan 1,5 sm, yuqori va pastdan 2 sm bo'lishi kerak.

Maqolalar *The Chicago Manual of Style, 16<sup>th</sup> Edition* formatida shakllantiriladi. Maqola matni 3 000–5 000 so'zdan iborat bo'lishi kerak.

O'zbek va ingliz tillarida 100–150 so'zdan iborat abstrakt (Abstrakt) va 5–10 so'zdan kam bo'lmagan kalit so'zlar (o'zbek va ingliz tillarida). Abstraktda maqolaning qisqacha mazmuni va dolzarbligi, tadqiqot natijalari aks etishi lozim.

Adabiyotlar ro'yxati 5 sahifadan oshmasligi kerak.

Kitobga taqriz (ingliz yoki boshqa tillarda bo'lishi mumkin) 1500 so'zdan oshmasligi talab etiladi.

Taqriz formati: 1) sarlavha: kitob nomi, muallif (mualliflar), nashr qilingan shahar: nashriyot nomi, nashr yili, sahifasi soni. Narxi, ISBN raqami, (qattiq/yumshoq muqova); 2) taqriz so'ngida: taqrizchining F.I.O., ish joyi, pochta manzili.

### II. Maqola bo'limlarini rasmiylashtirish

Maqola nomi – normal harflarda, to'q bo'yoqda, 16 hajm.

Maqola nomi o'zbek va ingliz tillarida (agar maqola boshqa tilda yozilgan bo'lsa, maqola yozilgan til va ingliz tilida) beriladi.

Maqola kirish, asosiy qism bo'limlari va xulosadan tashkil topadi.

Maqola bo'limlari sarlavhasi – to'q bo'yoqda, 14 hajm.

### III. Maqolada tarjimalardan foydalanish

Boshqa tillardagi matn yoki boshqa manbalar tarjimoni aniq ko'rsatilishi kerak. Agar matn maqola muallifi tomonidan tarjima qilingan bo'lsa, u holda "tarjima muallifniki" shaklida beriladi.

Rasmiy nashrdan olingan tarjima-matn tahrir qilinmaydi.

Zarur holatda tarjima matnga sana, turli diakritik belgilar va boshqa elementlar kiritilishi mumkin.

Tarjima qilingan matn olingan manba nomi asl holicha beriladi. Zarur deb topilsa, uning nomi qavs ichida berilishi mumkin.

Geografik nomlar tarjima qilinmaydi va asl shaklida beriladi.

Tashkilotlar nomi tarjima qilinmaydi va asl shaklida beriladi.

Davr nomi rasmiy qabul qilingan shaklda beriladi.

### IV. Ko'chirma va tarjima parchaning berilishi

Manbadan olingan ko'chirma parcha asosiy matndan 1 qator tashlab ajratiladi, satr oralig'i 1 interval, markazda, 12 hajmda yoziladi.

Ko'chirmaning tarjimasini qavs ichida ( ) satr boshidan yozilishi kerak. Bunday ko'chirma *Times New Roman* shrift, 12 hajm, normal yozuvda beriladi.

### V. Havola va izohlar berish

Manbaga havola matn ichida to'rtburchak qavsda [ ] beriladi. Havola qilingan manbalar bir nechta bo'lsa, ular nuqtali vergul (;) bilan ajratiladi.

Izohlar tegishli sahifa pastida, tartib raqami bilan joylashtiriladi.

### VI. Qo'lyozma (toshbosma) manbalar va nashr etilgan asarlar bibliografiyasi

Bibliografiyada muallif yoki asar nomi satr boshidan, boshqa barcha qatorlari xatboshidan yoziladi. Adabiyotlar *bibliografiyada* o'zbek lotin alifbosi tartibida ko'rsatiladi.

#### V.1. Qo'lyozma va toshbosma manbalar bibliografiyasi

Qo'lyozma yoki toshbosma manbalarni bibliografiyada o'zi yozilgan grafikada berish maqsadga muvofiq. Lotin alifbosidagi transliteratsiyasini berish ham mumkin. Ba'zan qo'lyozma asarning nomi muallif ismidan oldin yozilishi ham mumkin.

Muallif nomi. Ko'chirilgan asr (agar mavjud bo'lsa). Asar nomi. Qo'lyozma (toshbosma): saqlanayotgan joy, inventar raqam.

Xondamir. XV asr. Makorim ul-axloq. Qo'lyozma: O'zFASHI, № 742.

**Matnda qo'lyozma (toshbosma)ga havola berish:**

[Xondamir, Makorim, 17<sup>a</sup>]

#### VII.2. Kitoblar uchun

**Bibliografiyada:**

Familiya, ism. Nashr yili. *Kitob nomi*. Shahar: Nashriyot nomi.



Qudratullayev, Hasan. 2018. *Boburning adabiy-estetik olami*. Toshkent: Ma'naviyat.

**Matnda kitobga havola:**

[Familiya kitob nashr yili, sahifa raqami]

[Qudratullayev 2018, 99]

Agar bir muallifning bir yilda nashr qilingan kitoblaridan foydalanilgan bo'lsa, bibliografiyada kitobning nashr yili o'zbek lotin alifbosi harflari bilan ajratilib ko'rsatiladi.

Sirojiddinov, Shuhrat. 2011 (a). *Alisher Navoiy: manbalarning qiyosiy-tipologik, tekstologik tahlili*. Toshkent: Akademnashr.

Sirojiddinov, Shuhrat. 2011 (b). *O'zbek adabiyotining falsafiy sarchashmalari*. Toshkent: Akademnashr.

**Matnda kitobga havola:**

[Sirojiddinov 2011 (a), 99]

[Sirojiddinov 2011 (b), 67]

**Ikki muallif tomonidan yozilgan kitobni bibliografiyada berish tartibi:**

Familiya, Ism va Ism Familiya. Nashr yili. *Kitobning nomi*. Shahar: Nashriyot nomi.

Abdurahmonov, G'anjion, Alibek Rustamov. 1984. *Navoiy tilining grammatik xususiyatlari*. Toshkent: Fan.

**Matnda kitobga havola:**

[Familiya va Familiya nashr yili, sahifa raqami]

[Abdurahmonov, Rustamov 1984, 52]

Agar kitobning uch va undan ortiq muallifi bo'lsa, bibliografiyada barcha mualliflarning ismi to'liq yoziladi. Bunday kitobga havola qilinganda, birinchi muallifning ismi yozilib, davomida *va boshqalar* deb ko'rsatiladi:

[Familiya va boshqalar nashr yili, sahifa raqami]

[Vohidov va boshqalar 2010, 847]

**Kitob yoki to'plam maqolasini bibliografiyada berish tartibi:**

Familiya, ism. Nashr yili. "Maqola nomi." *Kitob yoki to'plam nomi*, Ism Familiya, Ism Familiya muharrirligida, maqola sahifasi raqamlari. Shahar: Nashriyot.

Abdug'afurov, Abdurashid. 2016. "Badoye' ul-bidoya"ning tuzilish sanasi". *XX asr o'zbek mumtoz adabiyotshunosligi*, Olim To'laboyev muharrirligida, 174–184. Toshkent: "O'zbekiston milliy ensiklopediyasi" Davlat ilmiy nashriyoti.

**Matnda kitob yoki to'plam maqolasiga havola:**

[Familiya nashr yili, sahifa raqami]

[Abdug'afurov 2016, 176]

**Elektron shaklda nashr qilingan kitoblar uchun:**

Elektron kitobning bir nechta formati bo'lsa, bibliografiyada foydalanilgan format ko'rsatiladi. Elektron kitobning internet manzili (URL) hamda shu manba olingan sana ko'rsatilishi lozim.

### **Elektron kitobni bibliografiyada berish:**

Familiya, Ism. Nashr yili. *Kitob nomi*. Shahar: Nashriyot nomi. URL. Foydalanilgan sana.

Mamatov, Ulug'bek. 2018. *O'zbekiston madaniyatida tarixiy janrdagi tasviriy san'at asarlari*. Toshkent: Mumtoz so'z. <https://kitobxon.com/uz/catalog/sanat/>. 12.03.2019.

### **Matnda elektron kitobga havola:**

[Familiya nashr yili, sahifa raqami]

[Mamatov 2018, 11]

### **Ikki mualliftomonidan yozilgan elektron kitobni bibliografiyada berish tartibi:**

Familiya, Ism va Ism Familiya. Nashr yili. *Kitobning nomi*. Shahar: Nashriyot nomi. Internet adres (URL).

Sirojiddinov, Shuhrat va Sohiba Umarova. 2017. *O'zbek matnshunosligi qirralari*. Chikago: Chikago universiteti nashriyoti. <http://press-pubs.uchicago.edu/founders/>.

### **Matnda elektron kitobga havola:**

[Familiya nashr yili, sahifa raqami]

[Sirojiddinov 2017, 19-hujjat]

## **VII.3. Jurnal maqolasi uchun**

### **Chop etilgan jurnal maqolasini bibliografiyada berish tartibi:**

Familiya, Ism. Nashr yili. "Maqola nomi". *Jurnal nomi* jurnal soni: maqola sahifalari.

Mahmudov, Nizomiddin. 2013. "Termin, badiiy so'z va metafora". *O'zbek tili va adabiyoti* 4: 3 – 8.

### **Matnda jurnal maqolasiga havola:**

[Familiya nashr yili, sahifa raqami]

[Mahmudov, 2013, 5]

### **Elektron jurnal uchun:**

Elektron jurnal uchun jurnalning DOI manzili ko'rsatiladi. Agar DOI manzili mavjud bo'lmasa, internet adresi ko'rsatilishi kerak (URL). DOI – bu o'zgarmas ID bo'lib, internet tarmoqlarining elektron adreslari tizimiga ulangan, ya'ni manbani boshqaruvchi <http://dx.doi.org/> manzildir.

### **Elektron jurnal maqolasini bibliografiyada berish:**

Familiya, Ism. Nashr yili. "Maqola nomi". *Jurnal nomi* jurnal soni: maqola sahifalari. DOI adres (yoki URL).

Aminov, Hasan. 2018. "O'zbekiston san'atida temuriylar siymosi". *O'zbekistonda xorijiy tillar* 2: 246 – 253. doi: 10.36078/1596780051.

### **Matnda maqolaga havola:**

[Familiya nashr yili, sahifa raqami]

[Aminov 2018, 248]

## **7.4. Gazeta yoki ilmiy-ommabop jurnal uchun**

Gazeta yoki ilmiy-ommabop jurnal maqolasiga havola matn shaklida beriladi (masalan, Muhammadjon Imomnazarovning 27.02.2005dagi

“O‘zbekiston adabiyoti va san’ati” gazetasida chop etilgan maqolasida aytilganidek...); odatda, bunday manbalar umumiy adabiyotlar ro‘yxatida keltirilmaydi. Agar keltirilsa, kitoblarga qo‘yiladigan talablarga asosan beriladi.

Agar onlayn maqolaga havola berilayotgan bo‘lsa, uning internet manzili (URL), maqola olingan sana ko‘rsatilishi kerak.

### **Gazeta yoki ilmiy-ommabop jurnal maqolasini bibliografiyada berish tartibi:**

Familiya, Ism. Nashr yili. “Maqola nomi.” *Gazeta-Jurnal nomi*, nashr sanasi. Imomnazarov, Muhammadjon. 2005. “Jomiy “Xamsa” yozganmi?.” *O‘zbekiston adabiyoti va san’ati*, January 25.

#### **Matnda maqolaga havola:**

[Familiya nashr yili, sahifa raqami]

[Imomnazarov 2005, 4]

### **Elektron gazeta yoki ilmiy-ommabop jurnal maqolasini bibliografiyada berish:**

Familiya, Ism. Nashr yili. “Maqola nomi.” *Jurnal nomi*, nashr sanasi. Internet adres.

Jabborov, Rustam. 2019. “Navoiyning Tabrizda yashagan xorazmlik kotibi”. *UZA: O‘zbekiston Milliy axborot agentligi*, 08.12. <https://uza.uz/uz>.

#### **Matnda maqolaga havola:**

[Familiya nashr yili, sahifa raqami]

[Jabborov 2010, 17]

Maqola so‘ngida foydalanilgan adabiyotlar o‘zbek lotin alifbosi tartibida beriladi. Adabiyotlar ro‘yxati ikki qismdan iborat bo‘lishi, birinchi qismda foydalanilgan adabiyot chop etilgan grafikada yuqorida ko‘rsatilgan shaklda rasmiylashtirilishi, ikkinchi qismda esa barcha foydalanilgan adabiyotlar o‘zbek lotin alifbosida berilishi talab qilinadi. Misol uchun:

#### **Adabiyotlar**

Баранов, Х.К. 1958. Арабско – русский словарь. Москва: Наука.

#### **Adabiyotlar**

Baranov, X.K. 1958. Arabsko – russkiy slovar. Moskva: Nauka.

Maqolani rasmiylashtirish talablarining ingliz tilidagi variantini “*The Chicago Manual of Style, 16<sup>th</sup> Edition*” qo‘llanmasi yoki <https://www.chicagomanualofstyle>. havolasidan ko‘rib olishingiz mumkin.

## **GUIDELINES FOR CONTRIBUTORS**

Uzbekistan: Language and Culture is an academic journal, publishing research in linguistics, history, literature, translation studies, arts, ethnography, philosophy, anthropology and social studies. We aim to publish cutting edge, innovative, conceptually interesting, original case studies and new research, which shape and lead debates in multifaceted studies. We do not publish economic analyses or policy papers. Any

opinions and views expressed in publications are the opinions and views of the authors, and the publishers are not responsible for the views/ reviews of the contributors.

The journal is published four times a year. The language of articles can be English, Russian and Uzbek. Other Turkic languages are also welcomed. In addition to research articles, the journal welcomes book reviews, literature overviews, conference reports and research project announcements.

### **1. General**

- Submission Guideline

1. Manuscripts may be submitted at any time during the year. However, if the author wishes to have his/her manuscript published in a certain issue of the journal, the submission should be made at least five months in advance of the proposed publication date.

2) Manuscripts should be submitted by email (uzlangcult@gmail.com) as an attachment in MS Word document (.doc) format and use MS Word Source.

3) All manuscripts should be submitted with a cover page including an email address, a mailing address and a short introduction about the author(s) /contributor(s)'.

### **2. Manuscript format**

1) The main texts should be written in Times New Roman font, 12 point, and single-spaced in 44 pagination with 1-inch margins.

2) Submissions must follow the author-date system of The Chicago Manual of Style, 16th Edition.

3) Quotations are given in brackets in the text.

4) A research article should normally be no more than 9,000 words in length, including the following contents:

- an abstract of 150-200 words (in English, Russian, and Uzbek) and seven to ten keywords;

- a list of references of no more than five (5) pages;

- tables and figures, if any.

5) A book review should generally be about 1,500 English words (or other languages) in length, and must include the heading and closing in the following format:

- Heading: Title of the Book. By Author's Name(s). City of Publication: Publisher Name, Year. pp. Price, ISBN:, (hardcover/paperback).

- Closing: Book reviewer's name, affiliation and postal address at the end.

6) Style Points Headings. Limit: Four levels.

Level 1. Title Style (e.g. the first letter of each word upper case, except prepositions), Bold, and 14 point.

Level 2. Title Style, Italics, 14 Point.

Level 3. Modified "down" style (first letter upper case, or first letter of first two words if the first word is an article), Bold, and 12 point.

Level 4. Modified down style, Bold, 11 point.

3. Style and Usage

## 1) Translation

- Translated excerpts from classical texts or non-English sources should be annotated with clarification of its original/published language and translator. Likewise, “Author’s own” translations of quoted texts should be noted as such.

- The author is expected to provide an English translation of key terms in the work, rather than a translator without expertise in the subject.

- Excerpts or quoted texts from published translation will not be edited. However, UzLC editors may query or modify translations of key terms or texts provided by the author.

- Where necessary, short supplementary information such as dates, an item in its original characters, or the Romanized form of a non-English item, may be included.

- Names of foreign publishers, and titles of sources published in a foreign language should primarily appear in Romanized form without translation. However, if necessary, a translation may be added in brackets ([ ]).

## 2) Names and Terms

- Place Names (foreign):

Designation for division of areas should be either translated or hyphenated after the given area name.

Designation for geographical/structure names are not hyphenated, and appear without the equivalent English term.

Institutional names are considered proper nouns. Their names should appear following the preference of the individual institutions.

3) The descriptive designation of a period is usually lowercase, except for proper names or traditionally capitalized terms.

## 4. Quotation

1) Block Quotations:

- A block quotation should start with double line spacing and an indentation from the left margin. From the second paragraph of the block quotation, additional paragraph indentation is needed.

Texts in block quotation should be written in Times New Roman 10 pts., and not be entirely italicized.

## 5. Others

1) There is one space after sentence punctuation and not two.

2) The end parenthesis, closing quotation mark, and footnote numbers come after the sentence punctuation.

3) For parentheses within parentheses, use brackets ([ ]).

## 6. Basic Citation Format

The following examples illustrate citations using the **author-date** system. Each example of a reference list entry is accompanied by an example of a corresponding parenthetical citation in the text. For more details and many more examples, see chapter 15 of *The Chicago Manual of Style*.

## BOOK

### Reference List (hanging indent):

Pollan, Michael. 2006. *The Omnivore's Dilemma: A Natural History of How Eating Has Evolved*. New York: Penguin.

**In Text Cite:**

[Pollan 2006, 99–100]

### Reference List (hanging indent):

Ward, Geoffrey C., and Ken Burns. 2007. *The War: An Intimate History, 1941–1945*. New York: Knopf.

**In Text Cite:**

[Ward and Burns 2007, 52]

For four or more authors, list all of the authors in the reference list; in the text, list only the first author, followed by et al. (“and others”):

[Barnes et al. 2010, 847]

### Reference List (hanging indent) book chapter:

Kelly, John D. 2010. “Seeing Red: Mao Fetishism, Pax Americana, and the Moral Economy of War.” In *Anthropology and Global Counterinsurgency*, edited by John D. Kelly, Beatrice Jauregui, Sean T. Mitchell, and Jeremy Walton, 67–83. Chicago: University of Chicago Press.

**In Text Cite:**

[Kelly 2010, 77]

Chapter of an edited volume originally published elsewhere (as in primary sources):

### Reference List (hanging indent) book originally published elsewhere:

Cicero, Quintus Tullius. 1986. “Handbook on Canvassing for the Consulship.” In *Rome: Late Republic and Principate*, edited by Walter Emil Kaegi Jr. and Peter White. Vol. 2 of University of Chicago Readings in Western Civilization, edited by John Boyer and Julius Kirshner, 33–46. Chicago: University of Chicago Press. Originally published in Evelyn S. Shuckburgh, trans., *The Letters of Cicero*, vol. 1 (London: George Bell & Sons, 1908).

**In Text Cite:**

[Cicero 1986, 35]

## BOOK PUBLISHED ELECTRONICALLY

If a book is available in more than one format, cite the version you consulted. For books consulted online, list a URL; include an access date only if one is required by your discipline. If no fixed page numbers are available, you can include a section title or a chapter or other number.

### Reference List (hanging indent):

Austen, Jane. 2007. *Pride and Prejudice: A Novel in Five Books*. New York: Penguin Classics. Kindle edition.

**In Text Cite:**

[Austen 2007, 101]

**Reference List (hanging indent):**

Kurland, Philip B., and Ralph Lerner, eds. 1987. *The Founders' Constitution*. Chicago: University of Chicago Press. <http://press-pubs.uchicago.edu/founders>

**In Text Cite:**

[Kurland and Lerner, chap. 10, doc. 19]

**JOURNAL ARTICLE****Article in a print journal**

In the text, list the specific page numbers consulted, if any. In the reference list entry, list the page range for the whole article.

**Reference List (hanging indent):**

Weinstein, Joshua I. 2009. "The Market in Plato's Republic." *Classical Philology* 104:439-58.

**In text cite:**

[Weinstein 2009, 440]

Article in an online journal

Include a DOI if the journal lists one. A DOI is a permanent ID that, when appended to <http://dx.doi.org/> in the address bar of an Internet browser, will lead to the source. If no DOI is available, list a URL. Include an access date only if one is required by your discipline.

**Reference List (hanging indent):**

Kossinets, Gueorgi, and Duncan J. Watts. 2009. "Origins of Homophily in an Evolving Social Network." *American Journal of Sociology* 115:405-50. doi:10.1086/599247.

**In text cite:**

[Kossinets and Watts 2009, 411]

**Article in a newspaper or popular magazine**

Newspaper and magazine articles may be cited in running text ("As Sheryl Stolberg and Robert Pear noted in a New York Times article on February 27, 2010..."); they are commonly omitted from a reference list. The following examples show more formal versions of the citations. If you consulted the article online, include a URL; include an access date only if your discipline requires one. If no author is identified, begin the citation with the article title.

**Reference List (hanging indent):**

Mendelsohn, Daniel. 2010. "But Enough about Me." *New Yorker*, January 25.

**In text cite:**

[Mendelsohn 2010, 68]

**Reference List (hanging indent):**

Stolberg, Sheryl Gay, and Robert Pear. 2010. "Wary Centrists Posing Challenge in Health Care Vote." *New York Times*, February 27. <http://www.nytimes.com/2010/02/28/us/politics/28health.html>.

**In text cite:**

[Stolberg and Pear 2010, 12]

## WEBSITE

A citation to website content can often be limited to a mention in the text (“As of July 19, 2008, the McDonald’s Corporation listed on its website . . .”). If a more formal citation is desired, it may be cited as in the examples below. Because such content is subject to change, include an access date or, if available, a date that the site was last modified. In the absence of a date of publication, use the access date or last-modified date as the basis of the citation.

### **Bibliography (hanging indent):**

Google. 2009. “Google Privacy Policy.” Last modified March 11. <http://www.google.com/intl/en/privacypolicy.html>.

### **In text cite:**

[Google 2009]

### **Reference List (hanging indent):**

McDonald’s Corporation. 2008. “McDonald’s Happy Meal Toy Safety Facts.” <http://www.mcdonalds.com/corp/about/factsheets.html>.

### **In text cite:**

[McDonald’s 2008]



## **IZOHLAR**

Jurnal 2017-yil 26-oktyabrda O'zbekiston Respublikasi Matbuot va axborot agentligi tomonidan № 0936 raqam bilan ro'yxatdan o'tgan. Tahririyatga kelgan maqolalar mualliflarga qaytarilmaydi.

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

**Manzil:** Toshkent shahri, Yakkasaroy tumani, Yusuf Xos  
Hojib ko'chasi 103-uy.  
Telefonlar: +99871 281-45-11, +99871 281-41-93.  
Website: [www.uzlc.navoiy-uni.uz](http://www.uzlc.navoiy-uni.uz)  
E-mail: [uzlangcult@gmail.com](mailto:uzlangcult@gmail.com)

Bosishga 30.06.2022-yilda ruxsat etildi.  
Bichimi 70x100 1/16, Ofset bosma. "Cambria" garniturasi.  
Shartli b.t. 7,51. Nashr b.t. 7,62.

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