



*Nilufar ABDURAXMONOVA,*  
Professor, O'zbekiston Milliy universiteti, Toshkent, O'zbekiston  
E-mail: [Abduraxmonovan@gmail.com](mailto:Abduraxmonovan@gmail.com)

*Gavhar KUVONDIKOVA,*  
Mustaqil izlanuvchi, O'zbekiston Milliy universiteti, Toshkent, O'zbekiston  
E-mail: [kuvondikovag@gmail.com](mailto:kuvondikovag@gmail.com)

DSc, professor S.Zokirova taqrizi asosida

## STAGES OF PREPARING ESSAY MATERIALS FOR LINGUISTIC CORPUS CONSTRUCTION

Annotation

This article explores the formation of a written learner corpus based on essays produced by non-philology students. The study applies linguistic and statistical analysis to identify lexical features, semantic fields, and typical errors in students' written speech. The findings reveal the dominance of vocabulary related to education, professional activity, and competence. The results are significant for developing educational corpora and learner-oriented dictionaries.

**Keywords:** corpus linguistics, learner corpus, student essays, lexical analysis, semantic fields, written discourse, linguistic errors

## ЭТАПЫ ПОДГОТОВКИ МАТЕРИАЛОВ ЭССЕ ПРИ СОЗДАНИИ ЛИНГВИСТИЧЕСКОГО КОРПУСА

Аннотация

В данной статье рассматривается вопрос формирования корпуса письменной речи на основе эссе студентов нефилологических направлений. В исследовании проведён лингвистический и статистический анализ текстов, выявлены их лексические особенности, семантические поля, а также типичные ошибки письменной речи. Результаты анализа показали преобладание лексики, связанной с образованием, профессиональной деятельностью и компетенциями. Полученные данные имеют важное значение для создания учебных корпусов и разработки учебных словарей.

**Ключевые слова:** корпусная лингвистика, учебный корпус, студенческие эссе, лексический анализ, семантические поля, письменная речь, языковые ошибки

## LINGVISTIK KORPUS YARATISHDA ESSE MATERIALLARINI TAYYORLASH BOSQICHLARI

Annotatsiya

Mazkur maqolada nofilologik yo'nalish talabalari tomonidan yozilgan esselar asosida yozma nutq korpusini shakllantirish masalasi ko'rib chiqiladi. Tadqiqotda talabalar esselari lingvistik va statistik jihatdan tahlil qilinib, ularning leksik tarkibi, semantik maydonlari hamda yozma nutqda uchraydigan tipik xatolar tizimi aniqlangan. Tahlil natijalari talabalar nutqida ta'lim, kasbiy faoliyat va kompetensiyaga oid birliklar ustunligini ko'rsatdi. Tadqiqot natijalari ta'limiy korpus yaratish va o'quv lug'atlarini tuzishda muhim ahamiyatga ega.

**Kalit so'zlar:** korpus lingvistikasi, ta'limiy korpus, talabalar esselari, leksik tahlil, semantik maydon, yozma nutq, lingvistik xatolar

**Kirish.** So'nggi yillarda tilshunoslikda korpus lingvistikasi alohida ilmiy yo'nalish sifatida jadal rivojlanmoqda. Til birliklarini real nutq materiallari asosida o'rganish zamonaviy lingvistik tadqiqotlarning muhim metodologik tamoyillaridan biri hisoblanadi. Korpus lingvistikasi matnlarning katta hajmdagi elektron bazasini yaratish hamda ularni statistik va lingvistik jihatdan tahlil qilish imkonini beradi.

Bugungi kunda dunyo tilshunosligida turli xil til korpuslari yaratilgan bo'lib, ular tilning turli qatlamlarini o'rganish uchun muhim manba bo'lib xizmat qilmoqda. Jumladan, milliy til korpuslari, so'zlashuv nutqi korpuslari, ixtisoslashgan terminologik korpuslar va o'quvchilar nutqi korpuslari keng qo'llanilmoqda. Bunday korpuslar til tizimining real qo'llanishini aks ettirish, til birliklarining chastotasini aniqlash hamda lingvistik modellarni ishlab chiqishda muhim ahamiyatga ega.

**Mavzuga oid adabiyotlar tahlili.** Mazkur maqolada nofilologik yo'nalish talabalari tomonidan yozilgan esselar asosida talabalar yozma nutqi korpusini shakllantirish imkoniyatlari ko'rib chiqiladi hamda ushbu materiallarning lingvistik xususiyatlari tahlil qilinadi. Tadqiqotning maqsadi nofilologik yo'nalish talabalari esselari asosida talabalar yozma

nutqi korpusini shakllantirish imkoniyatlarini aniqlash hamda ushbu matnlarning leksik tarkibini lingvistik va statistik jihatdan tahlil qilishdan iborat.

Kompyuter lingvistikasida korpus lingvistikasi til birliklarini nutq materiallari asosida o'rganishga imkon beruvchi muhim ilmiy yo'nalish sifatida shakllangan. Korpus lingvistikasi til tizimini tabiiy nutq namunalari asoslangan holda tahlil qilishni nazarda tutadi. Shu sababli korpus lingvistikasi metodlari bugungi kunda tilshunoslikning ko'plab yo'nalishlarida o'llanilmoqda. Mashhur korpus lingvisti T.McEnery korpus lingvistikasi haqida shunday yozadi: "Corpus linguistics is the study of language based on examples of real life language use." (McEnery, T., Wilson, A., 2001) Ya'ni, korpus lingvistikasi tilni real nutq namunalari asoslangan holda o'rganishdir. Bu yondashuv til birliklarining haqiqiy qo'llanishini aniqlash imkonini beradi. Shuningdek, ingliz tilshunosi D.Biber korpus lingvistikasining ahamiyatini quyidagicha ta'kidlaydi: "Corpus-based studies provide empirical evidence about patterns of language use." (D.Biber., Conrad, S., Reppen, R., 1998) Ya'ni korpusga asoslangan tadqiqotlar til birliklarining qo'llanish qonuniyatlarini aniqlash uchun empirik dalillar beradi.

Korpus lingvistikasi tadqiqotlarida yozma nutq materiallari alohida ahamiyat kasb etadi. Chunki yozma nutq matnlari til birliklarining aniq va izchil qo'llanishini aks ettiradi. Talabalar esselari esa zamonaviy yosh avlod nutqining leksik xususiyatlarini o'rganish uchun muhim empirik manba hisoblanadi. Talabalar yozma nutqi korpusini yaratish orqali quyidagi imkoniyatlar yuzaga keladi:

talabalar nutqidagi leksik birliklarning chastotasini aniqlash

yo'zma nutqdagi tipik xatolarni aniqlash

zamonaviy yoshlar nutqining lingvistik xususiyatlarini o'rganish

til ta'limi jarayonini takomillashtirish uchun ilmiy asos yaratish

Shu jihatdan qaraganda, talabalar esselari lingvistik korpus yaratishda muhim empirik material sifatida xizmat qiladi.

**Tadqiqot metodologiyasi.** Talabalar yozma nutqi korpusini yaratish jarayoni muayyan ilmiy tamoyillarga asoslangan holda amalga oshirilishi lozim. Korpus lingvistikasi tadqiqotlarida matnlarni tanlash, ularni tizimlashtirish va tahlil qilish jarayonlari aniq metodologik prinsiplarga tayangan holda tashkil etiladi. Mazkur tadqiqot doirasida talabalar esselari asosida yozma nutq korpusini shakllantirishda quyidagi asosiy prinsiplarga amal qilindi.

**Reprezentativlik prinsipi.** Korpus lingvistikasi tadqiqotlarida muhim tamoyillardan biri representativlik hisoblanadi. Bu prinsip korpusga kiritilgan matnlar tilning muayyan qatlamini yetarli darajada aks ettirishi kerakligini anglatadi. Mazkur tadqiqotda representativlikni ta'minlash maqsadida nofilologik yo'nalish talabalari tomonidan 1300 ta esse yig'ildi. Mazkur maqolada esa ushbu materiallardan tanlab olingan 30 ta esse tahlil uchun foydalanildi.

**Matnlarning bir xilligi prinsipi.** Korpus yaratishda muhim tamoyillardan yana biri matnlarning mavzu va janr jihatdan o'xshash bo'lishidir. Bu prinsip lingvistik tahlil natijalarining aniqligini ta'minlaydi. Mazkur tadqiqotda barcha esselar bir xil mavzu asosida yozilgan: "Yaxshi universitet diplomi yaxshi ish topishni kafolatlaydimi?" Bunday yondashuv matnlar o'rtasidagi semantik yaqinlikni ta'minlaydi hamda leksik birliklarni taqqoslash imkonini beradi.

1-jadval. Leksik birliklar tahlili

Leksema	So'z turkumi	Leksik qatlam	Kontekst
Yaxshi	Sifat	umumiste'mol leksika	yaxshi universitet
Universitet	Ot	termin (ta'lim sohasi)	universitetda o'qish
o'qish	Ot	umumiste'mol leksika	universitetda o'qish
Inson	Ot	umumiste'mol leksika	insonning faoliyati
Kelajak	Ot	umumiste'mol leksika	kelajakdagi faoliyat
Faoliyat	Ot	kitobiy/terminologik leksika	kelajakdagi faoliyat
Katta	Sifat	umumiste'mol leksika	katta ta'sir
ta'sir	Ot	o'zlashma leksika (arabcha)	ta'sir ko'rsatadi
ko'rsatadi	fe'l	umumiste'mol leksika	ta'sir ko'rsatadi

Talabalar esselarini tahlil qilish jarayonida matnlarda eng ko'p takrorlangan leksik birliklar aniqlanib, ularning chastotasi hisoblandi. Tahlil natijasida ta'lim va kasbiy faoliyat bilan bog'liq so'zlar ustunlik qilishi kuzatildi. Quyida 10 ta eng ko'p qo'llangan leksemalar sonini keltirib o'tamiz.

2-jadval Leksemalar qo'llanish chastotasi

	Leksema	Soni
1	Universitet	86
2	Diplom	79
3	Bilim	74
4	Ish	71
5	Tajriba	58
6	Muvaffaqiyot	46
7	Malaka	39
8	Kelajak	37
9	Imkoniyat	33
10	Mehnat	30

Talabalar yozma nutqini kontekstual tahlil qilish natijasida matnlarda takroran uchraydigan leksik birliklar ma'lum semantik maydonlarga birlashishi kuzatildi. Ushbu semantik maydonlar talabalarning dunyoqarashi, qadriyatlarini va ta'limga nisbatan konseptual yondashuvini aks ettiruvchi asosiy kognitiv birliklar sifatida namoyon bo'ladi. Tahlil natijasida quyidagi to'rtta asosiy semantik maydon ajratildi:

Ta'lim tizimi semantik maydoni. Mazkur maydon doirasiga "universitet", "talaba", "o'qish", "dars", "ta'lim" kabi birliklar kiradi. Ushbu birliklar talabalar nutqida eng yuqori chastotaga ega bo'lib, esselarning asosiy tematik yadrosini tashkil etadi. Universitet talabalarning shaxsiy va ijtimoiy shakllanishida muhim bosqich sifatida talqin qilinadi.

Kasbiy faoliyat semantik maydoni. Ushbu maydon "ish", "ish beruvchi", "faoliyat", "kasb", "lavozim" kabi birliklarni o'z ichiga oladi. Talabalar nutqida kasbiy faoliyat ko'pincha ta'lim natijasi sifatida baholanadi va ijtimoiy maqom hamda moddiy barqarorlik bilan bog'lanadi.

– Kompetensiya semantik maydoni. Mazkur maydon "bilim", "tajriba", "malaka", "ko'nikma" kabi birliklarni

qamrab oladi. Ushbu birliklar shaxsiy rivojlanish va kasbiy muvaffaqiyatning asosiy omillari sifatida talqin qilinadi. Ayrim hollarda bilim va amaliyot o'rtasidagi tafovut ham qayd etiladi.

– Natija semantik maydoni. Bu maydon "muvaffaqiyat", "kelajak", "yutuq", "imkoniyat" kabi birliklardan tashkil topadi. Talabalar nutqida muvaffaqiyat tushunchasi yaxshi ish, yuqori daromad va barqaror hayot bilan bog'lanadi.

Talabalar esselarini tahlil qilish jarayonida nafaqat semantik maydonlar, balki yozma nutqning sifat ko'rsatkichlarini belgilovchi xatolar tizimi ham aniqlandi. Mazkur xatolar yozma nutqning lingvistik me'yorlarga qanchalik mos kelishini ko'rsatadi va ta'limiy korpus materiallarining leksikografik tavsifini yaratishda muhim manba bo'lib xizmat qiladi. Tahlil asosida xatolar bir necha asosiy guruhlarga ajratildi: imlo xatolari, ortiqcha so'z ishlatish, grammatik xatolar, punktuatsiya xatolari hamda uslubiy nomuvofiqlik.

3-jadval Xatolar tasnifi

Xato turi	Tavsifi	Misol
Imlo xatolari	So'zlarning adabiy imlo me'yorlariga zid yozilishi	oqish → o'qish
Ortiqcha so'z ishlatish	Bir xil ma'noli birliklarning ortiqcha takrorlanishi	menimcha mening fikrimcha
Grammatik xatolar	Sintaktik yoki morfologik me'yorlarning buzilishi	talaba harakat qilish kerak
Punktatsiya xatolari	Tinish belgilarining noto'g'ri qo'llanishi	Men universitetda o'qiyman u juda yaxshi
Uslubiy nomuvofiqlik	Nutq uslubiga mos kelmaydigan birliklar	juda zo'r universitet

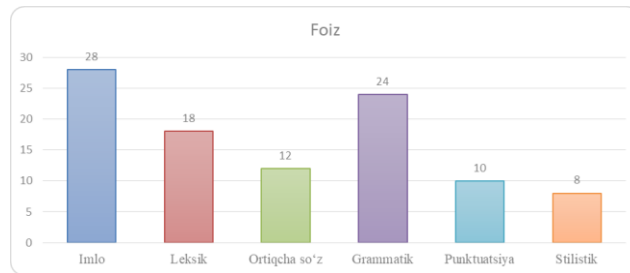
Imlo xatolari talabalar yozma nutqida eng ko'p uchraydigan kamchiliklardan biri sifatida kuzatildi. Bu turdagi xatolar so'zlarning adabiy imlo me'yorlariga zid yozilishi bilan xarakterlanadi. Masalan, oqish shaklining o'qish tarzida yozilishi lozim. Imlo xatolari matn mazmunini to'liq anglashga to'sqinlik qiladi, nutqning adabiy me'yoriyligini pasaytiradi. Ortiqcha so'z ishlatish esse matnlarida fikrning ixchamligi va aniqligiga salbiy ta'sir ko'rsatuvchi hodisa sifatida namoyon bo'ladi. Bu xatoda bir xil yoki o'zaro yaqin ma'noli birliklar takror qo'llanadi. Masalan menimcha, mening fikrimcha birikmasida ayni mazmun ikki marta ifodalangan. Bunday ortiqchalik nutqning mantiqiy zichligini susaytiradi, uslubiy soddalik va ravonlikka putur yetkazadi.

Grammatik xatolar sintaktik yoki morfologik me'yorlarning buzilishi bilan bog'liq bo'lib, ular gap qurilishi, so'z shakllari va bog'lanish vositalarida namoyon bo'ladi. Masalan, talaba harakat qilish kerak shaklida ega, kesim va modal konstruktsiya o'rtasidagi sintaktik moslik to'liq ta'minlanmagan; me'yoriy shakl talaba harakat qilish kerak tarzida bo'ladi. Grammatik xatolar nutqning struktur yaxlitligiga ta'sir etib, fikrning aniq va mantiqiy ifodalanishini qiyinlashtiradi. Punktatsiya xatolari tinish belgilarining noto'g'ri yoki yetarli qo'llanmasligi bilan tavsiflanadi.

Masalan, Men universitetda o'qiyman u juda yaxshi gapida ikki mustaqil predikativ birlik orasiga vergul yoki boshqa tegishli tinish belgisi qo'yilishi zarur. Punktatsion me'yorlarning buzilishi matnning sintaktik chegaralarini noaniqlashtiradi, o'quvchi tomonidan mazmunni anglash jarayonini murakkablashtiradi. Uslubiy nomuvofiqlik talabalarning yozma nutqida funksional uslublar aralashuvi natijasida yuzaga keladi. Masalan, juda zo'r universitet ifodasi so'zlashuv uslubiga xos bo'lib, akademik yoki rasmiy yozma matn talabalariga to'liq mos kelmaydi. Bunday holatlarda mazmun tushunarli bo'lsa-da, ifoda vositasi matnning janriy-uslubiy talabi bilan zid keladi.

Umuman olganda, mazkur xatolar tizimi talabalar esselarning lingvistik xususiyatlarini tavsiflashda muhim ahamiyatga ega. Ular bir tomondan yozma nutq malakasining real holatini ko'rsatsa, ikkinchi tomondan ta'limiy korpus asosida tuziladigan o'quv lug'ati, tavsifiy model va metodik tavsiyalar uchun amaliy material vazifasini bajaradi. Ayniqsa, xatolarni turkumlash asosida talabalar yozma nutqini normallashtirish, tipik nuqsonlarni aniqlash va ularni bartaraf etishga qaratilgan lingvodidaktik tavsiyalar ishlab chiqish mumkin bo'ladi.

4-jadval



**Xulosa.** Tadqiqot natijalari shuni ko'rsatadiki, nofilologik yo'nalish talabalari esselari lingvistik korpus yaratish uchun muhim empirik material hisoblanadi. Esselar talabalar yozma nutqining real holatini aks ettiradi hamda til birliklarining qo'llanish xususiyatlarini aniqlash imkonini

beradi. Talabalar esselari asosida yaratilgan yozma nutq korpusi zamonaviy yoshlar nutqining leksik xususiyatlarini o'rganish, yozma nutqdagi tipik xatolarni aniqlash hamda til ta'limini takomillashtirish uchun muhim ilmiy manba bo'lib xizmat qilishi mumkin.

#### ADABIYOTLAR

1. Abdurakhmonova N, Urdishev K. Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL). 2019;6(12019):131-7.
2. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. In Proceedings of the International Conference on Language Technologies for All (LT4All) 2019.
3. Abdullah Alfaifi, Eric Atw, Hedeya Ibrahim. Arabic Learner Corpus (ALC) v2. A New Written and Spoken Corpus of Arabic Learner 2014 Saudia Arabia
4. D. Stewart, S. Bernardini, G. Aston Introduction: Ten years of TaLC / Corpora and language learners, Vol 17, John Benjamins Publishing Company Amsterdam/Philadelphia, 2004, - P. 3
5. Larisa Goulart Da Silva Academic vocabulary: a corpus linguistics study on how Brazilian students write academic English. 2016-a. Brazilia
6. Abdurakhmonova, N., & Urdishev, K. (2019). Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL), 6(12019), 131-7.
7. Абдурахмонова, Н., & Абдувахобов, Г. (2021). О 'quv lug 'atini tuzishning nazariy metodologik asoslari. Международный журнал искусство слова, 4(6).
8. Abdurakhmonova, N. (2021). Formal-Functional Models of The Uzbek Electron Corpus. ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies, 10(8), 59-66.
9. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language
10. Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
11. Abdurakhmonova, N., Shakirovich, I. A., & O'G'Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. Science and Education, 3(9), 146-156.
12. Granger, S. (2003a). Error-tagged learner corpora and CALL: A promising synergy. CALICO Journal, 20(3): 465-480.
13. Granger, S. (2003b). The international corpus of learner English: A new resource for foreign language learning and teaching and second language acquisition research. TESOL Quarterly, 37(3): 538-546.