

MATNLARNI LINGVOSTATISTIK TAHLIL QILISHDA KORPUS USULLARIDAN FOYDALANISH

Nuritdinov Abrorbek Sayfiddin o'g'li

University of Business and Science

Til va adabiyot ta'limi kafedrası o'qituvchisi

E-mail: nuritdinovabrorbek@gmail.com

<https://doi.org/10.5281/zenodo.15581991>

Annotatsiya: Ushbu maqolada matnlarni xronologik tahlil qilishda zamonaviy korpus lingvistikasi usullaridan foydalanish masalalari ko'rib chiqiladi. Korpus usullarining matnlar vaqt bo'yicha rivojlanishini aniqlashdagi o'rni, lingvistik birliklarning tarixiy o'zgarishini aniqlash va matnlarni davriy guruhlariga ajratish imkoniyatlari tahlil qilinadi. Muallif lingvistik korpuslar asosida tanlangan matnlar ustida xronologik tahlil olib borib, natijalarga empirik yondashuv asosida ilmiy izohlar beradi. Tadqiqot davomida elektron korpus vositalari, avtomatlashtirilgan tahlil dasturlari va statistik metodlar qo'llanildi.

Kalit so'zlar: korpus lingvistikasi, xronologik tahlil, lingvistik birlik, tarixiy o'zgarish, matn strukturasi, elektron korpus, avtomatlashtirilgan tahlil.

Аннотация. В статье рассматриваются вопросы применения современных методов корпусной лингвистики при хронологическом анализе текстов. Анализируется роль корпусных методов в выявлении динамики развития текстов во времени, в определении исторических изменений лингвистических единиц и в классификации текстов по временным периодам. Автор проводит хронологический анализ отобранных текстов на основе лингвистических корпусов и предлагает научные интерпретации на основе эмпирических данных. В исследовании использовались электронные корпусные инструменты, автоматизированные программы анализа и статистические методы.

Ключевые слова: корпусная лингвистика, хронологический анализ, лингвистическая единица, исторические изменения, структура текста, электронный корпус, автоматизированный анализ.

Abstract. This article explores the application of modern corpus linguistic methods in the chronological analysis of texts. It examines the role of corpus tools in identifying the temporal development of texts, tracking historical changes in linguistic units, and categorizing texts into chronological groups. The author conducts a chronological study of selected texts based on linguistic corpora and provides scientific interpretations grounded in empirical data. The research employs electronic corpus tools, automated analysis software, and statistical techniques to ensure accuracy and objectivity.

Keywords: corpus linguistics, chronological analysis, linguistic unit, historical change, text structure, electronic corpus, automated analysis.

Zamonaviy lingvistik tadqiqotlarda korpus lingvistikasi usullarining qo'llanilishi tahliliy va statistik asoslangan yondashuvni shakllantirishga xizmat qilmoqda. Xususan, matnlarni xronologik nuqtai nazardan o'rganish orqali til birliklarining vaqt mobaynidagi o'zgarishlarini, ularning qo'llanish chastotasi va kontekstual ko'rinishlarini aniqroq baholash imkoniyati paydo bo'lmoqda[1]. Xronologik tahlil til birliklarining tarixiy rivojlanishini kuzatish, davriy til o'zgarishlarini aniqlash, matnlardagi semantik o'zgarishlarni

sistemalashtirishda muhim vosita bo'lib xizmat qiladi[2]. Ushbu yondashuv ayniqsa diaxron lingvistika, stilistika, matnshunoslik va tarixiy semantika sohalarida samarali qo'llanilmoqda. Shu bilan birga, korpus asosida avtomatlashtirilgan vositalar yordamida katta hajmdagi matnlar ustida tahlil olib borish imkoniyati ilmiy izlanishlarda aniqlik va takrorlanuvchanlik tamoyillarini ta'minlaydi[3].

Zamonaviy elektron korpuslar (masalan, British National Corpus, Russian National Corpus, yoki O'zbek tilining Milliy korpusi) xronologik parametrlar asosida matnlarni ajratish, so'z va iboralar tarqalishini aniqlash, til birliklarining tarixiy kontekstini aniqlash imkonini beradi[4]. Matnlarning vaqt bo'yicha strukturaviy va semantik o'zgarishlarini tahlil qilish nafaqat til taraqqiyotini tushunishga, balki madaniy va ijtimoiy jarayonlarning tilga ta'sirini ham yoritishga xizmat qiladi. Shu nuqtai nazardan, ushbu maqolada korpus lingvistikasining xronologik tahlilga oid metodlari tahlil qilinib, tanlangan matnlar misolida amaliy tahlil natijalari taqdim etiladi. Ushbu yondashuv asosida til birliklarining tarixiy dinamikasi, ularning matnlardagi tarqalishi va o'zgarish tendensiyalari ochib beriladi.

So'nggi yillarda korpus lingvistikasi lingvistik tadqiqotlarning asosiy metodologik poydevorlaridan biriga aylandi. Bu yondashuv, ayniqsa, tarixiy-mukammal tahlilni talab qiluvchi lingvistik sohalarida keng qo'llanilmoqda. McEnery va Hardie (2012) korpus lingvistikasini nafaqat tildagi faktlarni aniqlash, balki ularning vaqt doirasida qanday o'zgarishini kuzatish vositasi sifatida ham ko'rsatib o'tadi[5]. Sinclair (1991) esa korpus asosida tuzilgan konkordanslar yordamida lingvistik birliklarning kontekstual ko'rinishlarini aniqlash imkonini beruvchi yondashuvni ishlab chiqqan bo'lib, bu metod xronologik tahlil uchun alohida ahamiyatga ega[6]. Bunday yondashuvni Nemeth (2000) til birliklarining semantik taraqqiyotini o'rganishda muvaffaqiyatli qo'llab, tarixiy matnlar ustida amalga oshirilgan korpus tahlillarini taklif etadi[7].

O'zbek tilshunosligida esa korpus asosida tadqiqot olib borish tajribalari so'nggi yillarda shakllanmoqda. Isamuxamedova (2021) O'zbek tilining Milliy Korpusini yaratish va undan tahliliy maqsadlarda foydalanishning ilmiy-me'yoriy asoslarini yoritgan[8]. Shu bilan birga, Eshonqulova (2022) korpus asosida matnlarda so'z shakllari va ularning tarixiy o'zgarishini tahlil qilishda avtomatlashtirilgan dasturlar samaradorligini ko'rsatgan[9]. Shunday qilib, xorijiy va mahalliy adabiyotlar xronologik tahlilning metodik jihatlari bilan birga, amaliy qo'llanish shakllarini ham ochib beradi. Ammo bu boradagi tadqiqotlar ko'proq individual til birliklariga qaratilgan bo'lib, matnlarning davriy guruhlar bo'yicha sistematik tahlili hali yetarlicha ishlab chiqilmagan.

Tadqiqotda sifat va miqdoriy tahlil usullari uyg'unlashtirildi. Xronologik tahlil asosida matnlar davrlar bo'yicha (XX asr boshlaridan XXI asrgacha) guruhlariga ajratildi va har bir guruh uchun lingvistik birliklar korpus usullarida tahlil qilindi.

Ma'lumot manbasi sifatida quyidagi elektron korpuslardan foydalanildi:

- **British National Corpus (BNC)** – ingliz tilidagi tarixiy matnlar uchun;
- **Russian National Corpus (RNC)** – rus tilidagi tarixiy matnlar uchun;
- **O'zbek tilining Milliy Korpusi (OTMK)** – o'zbek matnlari uchun;

Matnlar **Sketch Engine**, **AntConc**, va **Voyant Tools** kabi avtomatlashtirilgan korpus tahlil dasturlari yordamida tahlil qilindi. So'zlar va iboralarining **chastotasi**, **kollokatsiyasi**, **kontekstual tarqalishi** hamda **stilistik o'zgarishlari** xronologik mezonlar asosida aniqlandi.

Tahlil qadamlarining ketma-ketligi quyidagicha belgilandi:

Matnlar tarixiy davrlar asosida tanlab olindi (1900–1930, 1931–1960, 1961–1990, 1991–2020).

Har bir davr uchun 50 ming soʻzdan iborat matnlar korpusga joylandi.

Soʻz shakllari, semantik birliklar va sintaktik konstruksiyalar aniqlanib, ularning oʻzgarish tendensiyalari kuzatildi.

Statistik va vizual grafik tahlillar amalga oshirildi.

Olib borilgan tadqiqot natijalari shuni koʻrsatadiki, korpus lingvistikasi asosida amalga oshirilgan xronologik tahlil matnlar tarkibida til birliklarining vaqt mobaynidagi oʻzgarish tendensiyalarini aniqlashda yuqori darajadagi aniqlikni taʼminlaydi. Har bir tarixiy davr matnlarida semantik, morfologik va sintaktik birliklarning qoʻllanish chastotasida sezilarli farqlar kuzatildi.

Masalan, XX asr boshlariga oid oʻzbek matnlarida arabcha va forsa komponentlar yuqori chastotada uchragan boʻlsa, keyingi davrlarda bu birliklar asta-sekin qisqarib, oʻrnini milliy leksik birliklar egallaganligi aniqlandi[10]. Bu tendensiya Oʻzbek tilining Milliy korpusida 1920-2020 yillarga oid matnlar asosida tekshirilganda aniqlangan boʻlib, soʻz boyligi va morfologik shakllarning soddalashuvi bilan kechgan[11]. Shuningdek, stilistik jihatdan qaralganda, XX asr oʻrtalariga tegishli matnlarda rasmiy va siyosiy uslubning ustunligi, XXI asrga kelib esa publitsistik va badiiy uslubdagi xilma-xillik kuchaygani kuzatildi. Bu holat ayniqsa rus tilidagi matnlarda ham kuzatiladi: RNC (Russian National Corpus) asosida oʻrganilgan matnlarda Stalinizm davridagi siyosiy ritorika va keyingi davrlarda sodda, xalqona tildan foydalanish keskin farq qiladi[12].

Ingliz tilidagi korpus (BNC) asosida olib borilgan tahlil esa boshqa bir tendensiyaning aniqlanishini koʻrsatadi: 1960-yillardan boshlab ingliz tilidagi matnlarda yozma nutq bilan ogʻzaki nutq unsurlari oʻrtasidagi chegaraning sezilarli darajada qisqarganligi koʻzga tashlanadi. Bu esa matnlarning stilistik diversifikatsiyasiga olib kelgan[13]. Korpus tahlili vositasida kollokatsiyalar, konkordanslar va kontekstual variantlar aniqlanganda, har bir davr uchun muayyan leksik birliklar va ifoda shakllari ustunligi aniqlangan. Misol uchun, “taraqqiyot”, “sotsializm”, “mehnat” singari soʻzlar 1950-1980 yillardagi oʻzbek matnlarida yuqori chastotaga ega boʻlgan boʻlsa, XXI asr matnlarida “innovatsiya”, “raqamlashtirish”, “ijtimoiy tarmoq” kabi terminlarning faol ishlatilgani qayd etildi[14].

Tadqiqot davomida matnlarni xronologik tahlil qilishda korpus lingvistikasi usullarini qoʻllash orqali bir nechta muhim ilmiy va amaliy natijalarga erishildi. Quyida asosiy natijalar tizimli tarzda bayon etiladi:

Matnlar tarkibida til birliklarining davrlar boʻyicha oʻzgaruvchanligi aniqlandi.

Oʻzbek tiliga oid korpusda (OTMK) 1920-2020 yillar oraligʻida tuzilgan matnlar asosida olib borilgan tahlil soʻz boyligi va leksik birliklarning tarixiy davrlar boʻyicha sezilarli farqlanishini koʻrsatdi. Jumladan, XX asr boshidagi matnlarda arabcha-forsa oʻzlashmalar chastotasi yuqori boʻlsa, XXI asr matnlarida globalizatsiyaga oid yangi soʻzlar va inglizcha terminlar koʻp uchrashini qayd etildi.

Stilistik va uslubiy jihatdan sezilarli oʻzgarishlar aniqlangan. XX asr oʻrtalarida siyosiy va ideologik rangga ega boʻlgan iboralar (masalan, “kommunistik qurilish”, “sotsialistik mehnat”) koʻp ishlatilgan boʻlsa, XXI asrga oid matnlarda yangilik, raqamli texnologiyalar va axborot jamiyati bilan bogʻliq leksik birliklar (“innovatsiya”, “IT sohasi”, “ijtimoiy tarmoq”) ustunlik qilgan.

Soʻzlar va iboralarining kollokatsion xususiyatlari ham tarixiy davrga bogʻliq holda oʻzgargani kuzatildi. Sketch Engine dasturida tuzilgan konkordanslar orqali aniqlandi-ki, bir soʻz turli davrlarda turlicha semantik kontekstda qoʻllanilgan. Masalan, “taraqqiyot” soʻzi 1950-yillardagi matnlarda koʻproq sotsialistik kontekstda ishlatilgan boʻlsa, 2000-yillardan boshlab texnologik va iqtisodiy rivojlanish maʼnosida faol qoʻllanilgan.

Morfologik shakllarda soddalashuv tendensiyasi aniqlandi. Tahlil shuni koʻrsatdiki, oʻzbek tilidagi matnlarda murakkab ot yasovchi vositalarning qoʻllanish chastotasi pasayib, oʻrniga qisqa va sodd shakllar kirib kelgan. Bu holat ogʻzaki til unsurlarining yozma matnlarda koʻproq ishlatilayotganini tasdiqlaydi.

Avtomatlashtirilgan korpus tahlil vositalari (Sketch Engine, AntConc) yordamida yuqori aniqlikdagi statistik maʼlumotlar olindi. Har bir davr uchun alohida kollokatsiyalar roʻyxati tuzildi va ularning semantik-dinamik harakati grafik va diagrammalar vositasida vizual koʻrsatildi. Bu yondashuv orqali oʻzbek tilidagi tarixiy til oʻzgarishlarini ilmiy asoslangan tarzda tasniflash imkoni yaratildi.

Mazkur tadqiqotda matnlarni xronologik tahlil qilishda korpus lingvistikasi yondashuvlarining nazariy va amaliy imkoniyatlari chuqur oʻrganildi. Oʻzbek tilidagi matnlarning turli tarixiy davrlardagi lingvistik xususiyatlari korpus vositalari yordamida tahlil qilinar ekan, til birliklarining oʻzgaruvchan tabiati, ularning davriy dinamikasi va uslubiy oʻziga xosliklari aniqlandi. Korpus asosida olib borilgan xronologik tahlil uslubi matnshunoslikda yangi ilmiy yondashuvlar uchun zamin yaratdi. Tadqiqot jarayonida leksik birliklar, kollokatsiyalar, morfologik shakllar va uslubiy strukturalar vaqti oʻtishi bilan qanday oʻzgarishga uchrashi aniqlanib, til taraqqiyotining obyektiv manzarasi yaratildi. Korpus platformalari yordamida olingan statistik maʼlumotlar esa bu oʻzgarishlarni raqamli asosda isbotlash imkonini berdi.

Shunday qilib, korpus lingvistikasi vositalari zamonaviy matnshunoslik va tarixiy tilshunoslikda ishonchli, tizimli va takrorlanadigan tadqiqotlar olib borish imkoniyatini taqdim etadi. Ushbu yondashuv til taraqqiyotining empirik asoslangan xronologiyasini yaratish, matnlar tahlilini chuqurlashtirish va lingvistik holatni kompleks yoritish uchun istiqbolli ilmiy metod hisoblanadi.

References:

Используемая литература:

Foydalanilgan adabiyotlar:

1. McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
2. Бондарко, А.В. (1990). Функциональная грамматика. – Л.: Наука, с. 34–39.
3. Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
4. Исамухамедова, Ш. (2021). Ўзбек тили миллий корпусини яратишда лингвистик мезонлар. // *Филология масалалари*. – №4. – Б. 25–29.
5. McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
6. Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
7. Nemeth, M. (2000). Historical Semantics and Corpus Linguistics. *Journal of Historical Linguistics*, 15(2), 134–145.

8. Исамухамедова Ш. (2021). Ўзбек тили миллий корпусини яратишда лингвистик мезонлар. // *Филология масалалари*. – №4. – Б. 25–29.
9. Eshonqulova, M. (2022). Korpus lingvistikasi: O'zbek matnlari asosida amaliy tahlil. // *O'zbek tili va adabiyoti*, №2. – Б. 41–46.
10. Исамухамедова, Ш. (2021). Ўзбек тили миллий корпусини яратишда лингвистик мезонлар. // *Филология масалалари*, №4, Б. 25–29.
11. Eshonqulova, M. (2022). Korpus lingvistikasi: O'zbek matnlari asosida amaliy tahlil. // *O'zbek tili va adabiyoti*, №2, Б. 41–46.
12. Крейдлин, Г.Е. (2014). Структура и стиль советской публицистики 1930-х годов. // *Вопросы языкознания*, №1.
13. McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
14. Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

INNOVATIVE
ACADEMY