

TALABALAR YOZMA NUTQI ASOSIDA TA'LIMY LEKSIKOGRAFIK MODEL YARATISH

Kuvondikova Gavhar Isomiddinovna,
mustaqil izlanuvchisi
kuvondikovag@gmail.com
O'zMU

Annotatsiya. Mazkur maqolada nofilologik yo'nalish talabalari tomonidan yozilgan esse matnlari asosida ta'limiy korpus yaratish va ushbu korpus materiallari asosida leksikografik model ishlab chiqish masalalari ko'rib chiqiladi. Tadqiqotda esse matni statistik va leksik-semantik tahlil qilinib, ta'limiy lug'at yaratish uchun metodologik model taklif etiladi.

Kalit so'zlar. *Korpus lingvistikasi, ta'limiy korpus, leksikografik model, esse matnlari, chastota tahlili, leksikografik model.*

Abstract. This article examines the creation of an educational corpus based on essays written by non-philological students and the development of a lexicographic model using corpus data. A corpus of essays was statistically and lexically analyzed. The study proposes a methodological model for developing an educational dictionary based on real written language data.

Keywords: *Corpus linguistics, educational corpus, lexicographic model, student essays, frequency analysis, lexicographic model.*

Kirish. Zamonaviy tilshunoslikda korpus lingvistikasi tilni empirik materiallar asosida o'rganishga imkon beruvchi muhim ilmiy yo'nalishlardan biri sifatida qaratiladi. Korpus lingvistikasi til birliklarini nutq materiallari asosida statistik va funksional jihatdan tahlil qilish imkonini beradi. Shu sababli korpusga asoslangan tadqiqotlar zamonaviy tilshunoslikda alohida metodologik ahamiyat kasb etmoqda.

Korpus lingvistikasining nazariy asoslari haqida T.McEnery va A.Wilson shunday ta'kidlaydi: korpus lingvistikasi til birliklarini katta hajmdagi elektron matnlar majmuasi asosida o'rganishga imkon beruvchi ilmiy metodlar tizimidir [1]. Shuningdek, mashhur tilshunos D. Biber korpus tahlilining muhim jihatlaridan biri sifatida statistik metodlarning qo'llanishini alohida qayd etadi. Uning fikriga ko'ra, korpus tahlili til birliklarining real qo'llanish chastotasini aniqlash imkonini beradi [2]. Korpus lingvistikasi leksikografiya bilan ham bevosita bog'liqdir. Zamonaviy lug'atlar ko'pincha korpus materiallari asosida yaratiladi. Atkins va Rundellning ta'kidlashicha, korpusga asoslangan leksikografiya lug'at tuzishda real nutq materiallaridan foydalanishni ta'minlaydi [3].

Mazkur tadqiqotda nofilologik yo'nalish talabalari tomonidan yozilgan esse matnlari asosida ta'limiy korpus uchun leksikografik model yaratish masalalari ko'rib chiqiladi.

Tadqiqot metodologiyasi

Mazkur tadqiqot korpus lingvistikasi metodologiyasiga asoslangan holda amalga oshirildi. Tadqiqot materiallari sifatida turli yo'nalishlarda tahsil olayotgan nofilologik yo'nalish talabalari tomonidan yozilgan esse matnlari tanlab olindi. Tadqiqot korpusini shakllantirish jarayonida jami 1300 nafar talaba tomonidan yozilgan yozma ishlar yig'ildi. Ushbu esse matnlari turli mavzularni qamrab olgan bo'lib, ular talabalar nutqida qo'llaniladigan faol leksik birliklarni aniqlash imkonini beradi.

Tadqiqotning dastlabki bosqichida yig'ilgan esse matnlari raqamlashtiriladi va yagona elektron formatga keltiriladi. Matnlarni qayta ishlash jarayonida ortiqcha belgilar, grafik xatoliklar hamda texnik kamchiliklar bartaraf etiladi. Shundan so'ng barcha matnlar maxsus korpus bazasiga joylashtiriladi. 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.

Korpusni shakllantirish jarayoni quyidagi bosqichlarda amalga oshiriladi:

1. Talabalarning esse matnlarini yig'ish.
2. Matnlarni skanerlash va raqamlashtirish.
3. Matnlarni tahrirlash va tozalash (cleaning).
4. Elektron matn bazasini yaratish.
5. Korpus formatini shakllantirish.

Mazkur jarayon natijasida talabalar yozma nutqini aks ettiruvchi ta'limiy korpus yaratiladi. Korpus shakllantirilgandan so'ng matnlar statistik jihatdan tahlil qilindi. Statistik tahlil til birliklarining matnda qanchalik tez-tez qo'llanishini aniqlashga imkon beradi. Tilshunoslikda bunday tahlil chastota tahlili deb yuritiladi. Matnlardagi leksik birliklarni aniqlash va ularning chastota ko'rsatkichlarini hisoblash jarayonida AntConc korpus tahlili dasturidan foydalaniladi. Ushbu dastur matnlar ichidagi so'zlarning takrorlanish chastotasini aniqlash, konkordanslarni ko'rsatish hamda kollokatsion bog'lanishlarni tahlil qilish imkonini beradi.

Ushbu metodologiya korpus lingvistikasi va amaliy leksikografiya tamoyillariga asoslanadi hamda quyidagi bosqichlarni o'z ichiga oladi.

- a) matn korpusini shakllantirish;
- b) matnlarni raqamlashtirish;
- c) so'zlar chastotasini aniqlash;
- d) leksik birliklarni tanlash;
- e) lug'at maqolalarini tuzish;
- f) xatolarni qayd etish.

Mazkur metodologiya talabalar yozma nutqi korpusiga asoslangan lug'at yaratish uchun ilmiy asos bo'lib xizmat qiladi. Esselar asosida yaratilgan lug'at

talabalar nutqida faol qo'llaniladigan leksik birliklarni aniqlash hamda ularning nutqdagi qo'llanishini ko'rsatish imkonini beradi

Statistik tahlil natijasida aniqlangan leksik birliklar keyingi bosqichda leksik-semantik tahlil qilindi. Ushbu tahlil orqali so'zlarning ma'no guruhlari, tematik maydonlari hamda nutqdagi funksional xususiyatlari aniqlanadi.

Tadqiqotning yakuniy bosqichida statistik va semantik tahlil natijalariga asoslanib, ta'limiy leksikografik model ishlab chiqiladi. Ushbu model ta'limiy korpus asosida yaratiladigan lug'atning struktur elementlarini belgilaydi. Ushbu model talabalar yozma nutqi asosida yaratiladigan ta'limiy lug'atlar uchun metodologik asos bo'lib xizmat qiladi

Natija

Mazkur tadqiqotning dastlabki bosqichida nofilologik yo'nalish talabalari tomonidan yozilgan esse matnlari asosida ta'limiy leksikografik modelni sinovdan o'tkazish maqsadida kichik hajmdagi eksperimental materialdan foydalanildi. Tajriba materiali sifatida 30 ta esse matni tanlab olindi va ular asosida leksik birliklarning dastlabki chastota tahlili amalga oshirildi. Mazkur bosqich pilot tadqiqot xarakteriga ega bo'lib, uning asosiy maqsadi talabalar yozma nutqida faol qo'llanadigan leksik birliklarni aniqlash hamda ular asosida ta'limiy lug'at modelini ishlab chiqishdan iborat.

Shuni alohida ta'kidlash lozimki, ushbu mini lug'at faqatgina 30 ta esse matni asosida yaratilgan eksperimental model hisoblanadi. Tadqiqotning keyingi bosqichlarida esa 1300 ta esse matnidan iborat keng ko'lamli korpus shakllantirilib, uning asosida yanada mukammal ta'limiy lug'at yaratish rejalashtirilgan. Mazkur yondashuv korpus lingvistikasida keng qo'llaniladigan bosqichma-bosqich modellashtirish tamoyiliga asoslanadi. Dastlab kichik hajmdagi matnlar asosida model sinovdan o'tkaziladi, keyinchalik esa u katta hajmdagi korpus materiallari asosida takomillashtiriladi. Ushbu mini lug'at quyidagi vazifalarni bajaradi:

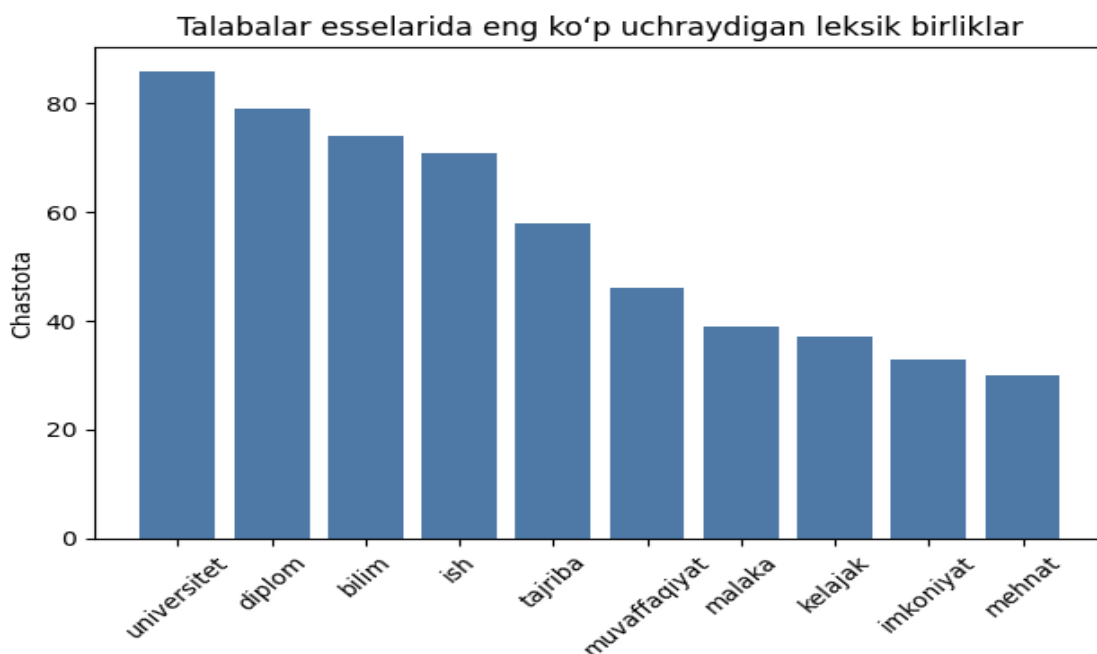
- ta'limiy korpus asosida lug'at yaratish metodologiyasini sinovdan o'tkazish;
- talabalar yozma nutqida faol qo'llanadigan leksik birliklarni aniqlash;
- kelajakda yaratiladigan keng qamrovli ta'limiy lug'at uchun model ishlab chiqish.

Korpus lingvistikasi tadqiqotlarida lug'at yaratish jarayoni ko'pincha yozma nutq materiallari asosida amalga oshiriladi. Bunday yondashuv leksik birliklarning tabiiy qo'llanishini aniqlash va ularning semantik hamda funksional xususiyatlarini tahlil qilish imkonini beradi. Shu sababli zamonaviy leksikografiya korpusga asoslangan metodlar muhim metodologik vosita sifatida qoraladi. Mazkur tadqiqotda ham talabalar yozma nutqi asosida ta'limiy leksikografik model ishlab chiqish maqsad qilindi. Leksikografik model lug'at maqolalarining struktur elementlarini belgilash hamda leksik birliklar haqidagi lingvistik ma'lumotlarni tizimli ravishda taqdim etishga xizmat qiladi.

Tadqiqot natijalariga ko'ra ta'limiy lug'at uchun quyidagi struktur model taklif etildi: lemma → grammatik tavsif → semantik izoh → kontekst misoli → chastota ko'rsatkichi → leksik xatolar. Mazkur model korpus lingvistikasida keng qo'llaniladigan chastota tahlili hamda kontekstual tahlil tamoyillariga asoslanadi va ta'limiy lug'at yaratishda tizimli yondashuvni ta'minlaydi.

Tahlil natijasida talabalar esse matnlarida eng ko'p uchraydigan leksik birliklar aniqlanib, ularning statistik ko'rsatkichlari hisoblandi.

1-jadval. Talabalar esselarida eng ko'p uchraydigan leksik birliklar



Tahlil natijalari shuni ko'rsatadiki, talabalar yozma nutqida ta'lim va kasbiy rivojlanish bilan bog'liq leksik birliklar ustunlik qiladi. Ushbu mini lug'at modeli nofilologik yo'nalish talabalari esselari asosida shakllantirilgan bo'lib, u leksik birliklarning struktur-semantik, funksional hamda chastotaviy xususiyatlarini tizimli tarzda aks ettirishga xizmat qiladi. Lug'at maqolasi korpus lingvistikasining asosiy tamoyillari – kontekstual tahlil, chastota tahlili va kollokatsion tahlil asosida tuzildi. Quyida “universitet” leksemasi asosida eksperimental lug'at maqolasi keltiriladi:

2-jadval. Lug'at modeli

Komponent	Tavsif
Lemma	Universitet
So'z turkumi	Ot
Semantik izoh	Oliy ta'lim beruvchi ilmiy va ta'lim muassasasi
Korpusdan misol	Universitetda o'qish insonning kelajagini belgilaydi
Kontekstual tavsif	Talabalar essellarida “universitet” leksemasi asosan “bilim olish”, “kasbiy rivojlanish” hamda “kelajak” tushunchalari bilan semantik jihatdan bog'liq holda qo'llanadi.

Kollokatsiyalar	universitetga kirish, universitetda o'qish, universitetni tamomlash
Tipik xato	Universitetda oqish
Normativ shakl	Universitetda o'qish
Tematik guruh	ta'lim tizimi
Chastota	86

Mazkur lug'at modeli talabalar yozma nutqida qo'llaniladigan leksik birliklarning semantik xususiyatlarini hamda ularning kontekstda qo'llanishini ko'rsatishga xizmat qiladi. Taklif etilayotgan leksikografik modelning ahamiyati shundaki, taklif etilgan model quyidagi ilmiy va amaliy vazifalarni bajaradi:

- talabalar yozma nutqining leksik tarkibini aniqlash;
- ta'lim jarayonida qo'llaniladigan faol leksik birliklarni tizimlashtirish;
- ta'limiy korpus asosida lug'at yaratish metodologiyasini ishlab chiqish;
- kelajakda yaratiladigan keng qamrovli ta'limiy lug'at uchun konseptual asos yaratish.

Shu tariqa 30 ta esse asosida yaratilgan mini lug'at modeli kelajakda 1300 ta esse matni asosida shakllantiriladigan katta hajmdagi ta'limiy lug'at uchun metodologik poydevor vazifasini bajaradi.

Xulosa

Talabalar yozma matnlari asosida yaratilgan korpus ta'limiy lug'atlar yaratishda muhim empirik manba hisoblanadi. Bunday korpuslar real nutq materiallarini aks ettirgani sababli ularga asoslangan leksikografik modellar ham amaliy ahamiyatga ega. Tadqiqot natijasida quyidagi xulosalarga kelindi:

- talabalar esse matnlari ta'limiy korpus yaratish uchun muhim lingvistik material hisoblanadi;
- korpus tahlili yordamida eng ko'p uchraydigan leksik birliklarni aniqlash mumkin;
- ushbu birliklar asosida ta'limiy leksikografik model yaratish mumkin.

Foydalanilgan adabiyotlar ro'yxati

1. McEnery T., Wilson A. Corpus Linguistics. Edinburgh University Press, 2001.
2. Biber D. Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press, 1998.
3. Atkins B., Rundell M. The Oxford Guide to Practical Lexicography. Oxford University Press, 2008.
4. 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.
5. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. InProceedings of the International Conference on Language Technologies for All (LT4All), 2019.
6. Abdullah Alfaifi, Eric Atw, Hedeya Ibrahim. Arabic Learner Corpus (ALC) v2. A New Written and Spoken Corpus of Arabic Learner, 2014, Saudia Arabia.
7. D. Stewart, S. Bernardini, G. Aston Introduction: Ten years of TaLC / Corpora and language learners, Vol 17, John Benjamins Publishing Company Amsterdam/Philadelphia, 2004. - Pp. 3.
8. Larisa Goulart Da Silva Academic vocabulary: a corpus linguistics study on how Brazilian students write academic English, 2016, Brazilia.
9. Abdurakhmonova N., Shakirovich I. A., Khudayberganov N. S. Morphological analyzer (morfoAnalyse) Python package for Turkic language. Science and Education, 3(9), 2022. Pp. 146-156.
10. Granger S. The international corpus of learner English: A new resource for foreign language learning and teaching and second language acquisition research. TESOL Quarterly, 37(3), 2003. Pp. 538-546.



11. Hammarberg B. Introduction to the ASU corpus, a longitudinal oral and written text corpus of adult learners' Swedish with a corresponding part from native Swedes. Stockholm University: Department of Linguistics, 2010.
12. Thompson P. Spoken language corpora. In M. Wynne (Ed.), Developing Linguistic Corpora: a Guide to Good Practice (pp. 59-70). Oxford: Oxbow Books, 2005.
13. Biber D. Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press, 1998.
14. Atkins B., Rundell M. The Oxford Guide to Practical Lexicography. Oxford University Press, 2008.