

MILLIY XOSLANGAN LISONIY BIRLIKLARNI TIL KORPUSIDA TEGGLASHNING METODOLOGIK ASOSLARI

Yorkulov Zafarjon Bakirovich
mustaqil tadqiqodchi

Ёркулов Зафаржон Бакирович
независимый исследователь

Yorkulov Zafarjon Bakirovich
independent researcher

DOI: <https://doi.org/10.5281/zenodo.21935189>

Korpus tilning real kommunikativ amaliyotdagi holatini namoyon etishi lozim. Korpus representativ bo'lishi uchun u tilning turli registrari, janrlari, kommunikativ vaziyatlari va funksional uslublarini mutanosib ravishda qamrab olishi shart. Ya'ni korpusning representativlik darajasi tanlovning sifat ko'rsatkichlariga bevosita bog'liqdir – millionlab so'zlardan iborat korpus ham noto'g'ri tanlangan material asosida tuzilgan bo'lsa, representativ hisoblanmaydi. Aksincha, metodologik jihatdan puxta tanlangan nisbatan kichik korpus ishonchli natijalar berishi mumkin. Korpus tarkibiga kiritiladigan matnlar maqsadli tanlov, sathli tanlov, tasodifiy tanlov yoki kombinatsiyalashgan tanlov usullari asosida shakllantirilishi mumkin. Sathli tanlov usulo matnlarni oldindan belgilangan parametrlar janr, davr, muallif, auditoriya, kommunikativ vazifa, geografik hudud va boshqa mezonlar asosida mutanosib taqsimlash imkonini berishi statistik jihatdan aniq xulosalar shakllantirishda muhim xususiyat hisoblanadi. Korpus tarkibidagi turli tipdagi matnlarning nisbiy ulushi ilmiy asoslangan mezonlarga muvofiq taqsimlanishi uning muvozanatlilikini anglatadi. Shakllantirilgan bazada muayyan janr yoki uslub haddan tashqari ustunlik qilsa, ayrim leksik, grammatik yoki pragmatik birliklarning chastotasi sun'iy ravishda ortib ketadi va shu bois tadqiqot natijalari obyektivligini yo'qotadi. Muvozanatlilikni ta'minlashda quyidagi parametrlar hisobga olinishi lozim:

turli uslublar qamrovi (badiiy, ilmiy, publitsistik, rasmiy, so'zlashuv);

turli kommunikativ janrlar qamrovi;

yozma va og'zaki nutq nisbatida mutanosiblik;

matn yaratilgan davrlar qamrovi;

nutq mualliflarning ijtimoiy-demografik xususiyatlari (jins, yosh, kasb va boshqalar);

geografik hudud va dialektal variantlar qamrovi;

Takidlash joizki, korpusning sifat jihatdan to'g'ri shakllantirilishi uning hajmidan muhimroq. T. McEnery va A. Hardie korpusning ilmiy qiymati tadqiqot maqsadiga mos ravishda tanlangan materiallar tizimi bilan belgilanishini qayd etadilar. Shuning uchun maxsus korpuslarda hajm nisbatan kichik bo'lsa ham, representativlik va muvozanatlilik tamoyillari to'liq ta'minlangan taqdirda ular yuqori ilmiy samaradorlikka ega bo'lishi muqarrar. Milliy-madaniy xoslangan lisoniy birliklarni jamlovchi korpusni shakllantirishda ham ushbu metodologik tamoyillar hal qiluvchi ahamiyat kasb etadi. Zotan, ushbu korpus turli janr va diskurslarga mansub autentik matnlarni qamrab olishi, erkak va ayol nutqi namunalarining mutanosib taqsimlanishini ta'minlashi, shuningdek, yozma hamda og'zaki nutq materiallarini o'z ichiga olishi zarur. Korpus uchun til materialini to'plashda autentiklik, obyektivlik hamda reproduktivlik tamoyillariga ham tayanish shart. Ushbu tamoyillar korpusdagi til materialining ishonchliligini, tadqiqot natijalarining xolisligini va ularning boshqa tadqiqotchilar tomonidan qayta tekshirilish imkoniyatini ta'minlab, empirik tadqiqotlarning metodologik poydevori bo'ladi.

Milliy xoslangan leksemalarni korpusda teglashda yuqorida ko'rsatilgan representativlik, muvozanatlilik, autentiklik, obyektivlik va reproduktivlik tamoyillari bevosita teglashning o'ziga emas, balki leksemalarni tanlash, ularning milliy xoslanganligini aniqlash va tegni izchil qo'llash jarayonlariga xizmat qiladi. Buni quyidagicha tizimlashtirish mumkin.

Representativlik tamoyili milliy xoslangan leksemalarning korpusda turli mavzu, uslub, janr, davr va kommunikativ vaziyatlarda yetarli darajada aks etishini ta'minlaydi. Masalan, do'ppi leksemasi faqat badiiy matnlardan emas, publitsistik, ilmiy-ommabop, tarixiy va boshqa matnlardan ham tanlanishi kerak. Shunda uning milliy-madaniy mazmuni va qo'llanish xususiyatlari to'liqroq namoyon bo'ladi.

Muvozanatlilik milliy xoslangan leksemalarning turli semantik-madaniy guruhlarida korpusda mutanosib aks etishini nazarda tutadi. Ya'ni korpus faqat milliy taom nomlaridan iborat bo'lib qolmasligi,

unda milliy kiyim-kechak, marosim, qarindoshlik, uy-ro‘zg‘or, kasb-hunar, diniy tushunchalar va boshqa ma‘no guruhlar ham mavjud bo‘lishi kerak.

Masalan:

MILLIY_TAOM

MILLIY_KIYIM

MAROSIM

URF_ODAT

QARINDOSHLIK

UY_RO‘ZG‘OR

KASB_HUNAR

DINIY_QADRIYAT

Korpusni shakllantirishda autentiklik eng muhim mezonlardan biri. Milliy xoslangan leksema sun‘iy tuzilgan gap asosida emas, real nutqdagi qo‘llanishi asosida teglanishi shart. Obyektivlik tamoyili milliy xoslangan leksemalarni subyektiv taxmin asosida emas, oldindan belgilangan mezonlar asosidateglashni talab qiladi. Masalan, har qanday o‘zbekcha leksemani milliy xoslangan deb belgilash mumkin emas. Uy, non, bola, suv kabi birliklar o‘zbek tilida faol qo‘llansa-da, ularning o‘zi muntazam milliy xoslangan leksema bo‘lmaydi, Milliy xoslanishni belgilashda leksema → milliy referent → milliy-madaniy sema → madaniy funksiya → kontekst tekshirilishi lozim. Masalan: iroqi → o‘zbek milliy bosh kiyimi → milliy-madaniy sema mavjud → kiyim-kechak madaniyatiga mansub → MILLIY_KIYIM.

5. Reprodukativlik tamoyili teglashning bir xil holatlarda bir xil natija berishini ta‘minlaydi. Ya‘ni bir annotator do‘ppiga MILLIY_KIYIM tegini bersa, boshqa annotator ham ayni mezon asosida shu tegni berishi kerak.

Buning uchun teglashning quyidagi yo‘riqnomasiga amal qilinadi. Unda:

milliy xoslangan leksema nima;

qaysi birliklar teglanadi;

qaysi birliklar teglanmaydi;

har bir tegning chegarasi;

bir leksema bir nechta kategoriyaga tegishli bo‘lsa, qaysi teg tanlanishi;

kontekstda milliy ma‘no qanday aniqlanishi

aniq ko‘rsatiladi.

Ma‘lumki, ayrim leksemalar nutqda umumiy ma‘noda ham, milliy-madaniy xoslangan ma‘noda ham qo‘llanishi mumkin. Agar leksema boshqa kontekstda umumlashgan yoki ko‘chma ma‘noda ishlatilsa, uning milliy xoslanganlik darajasi qayta tekshiriladi. Bu yondashuv kontekstuallik tamoyili bilan shartlangan.

Shunday qilib, tamoyillar asosida korpusda birlikni teglashda quyidagi ketma-ketlik shakllanadi:

Dasturxonga palov tortildi.

Teg strukturasi	Natija
Leksema	palov
Milliy xoslanganlik	mavjud
Referent	milliy taom
Madaniy soha	oziq-ovqat / milliy taom
Kontekst	to‘y marosimi
Asosiy teg	MILLIY_TAOM
Qo‘shimcha teg	MAROSIM_TO‘Y

Milliy xoslangan birliklarning real nutqdagi qo‘llanish xususiyatlarini obyektiv aks ettirish uchun korpus materiallari turli funksional uslub va kommunikativ janrlarni qamrab olishi zarur, degan xulosaga kelindi. Bunda badiiy, publitsistik, rasmiy, ilmiy-ommabop va so‘zlashuvga xos materiallarning mavjudligi birliklarning turli kontekstdagi qo‘llanishini qiyoslash imkonini beradi. Yozma va og‘zaki nutq, turli davrlar hamda hududiy variantlarning hisobga olinishi milliy xoslangan birliklarning qo‘llanish doirasini to‘liqroq ifodalashga xizmat qiladi. Muayyan lisoniy birlikning faqat boshqa tillarda muqobili mavjud yoki mavjud emasligi sababli milliy xoslanganlikni belgilashda birlikning o‘zbek madaniyati bilan bog‘liq referenti, tarkibidagi milliy-madaniy semantik komponent va muayyan ijtimoiy-madaniy tajribani aks

ettirish darajasi kompleks ravishda hisobga olinishi lozim. Milliy xoslangan birlikni faqat lugʻaviy shakli asosida belgilash ayrim hollarda notoʻgʻri natijaga olib keladi. Birlikning milliy-madaniy mazmuni uning konkret matndagi qoʻllanishi orqali aniqlanishi sababli kontekst teglashning majburiy komponenti sifatida belgilanadi. Shu asnoda bir xil lisoniy shaklning umumiy maʼnoda yoki milliy-madaniy maʼnoda qoʻllangan holatlarini farqlash imkonini beruvchi kontekstual mezon ishlab chiqildi. Lisoniy birliklarni marosim va urf-odatlar, ijtimoiy munosabatlar, qarindoshlik, milliy taomlar, kiyim-kechak, uy-roʻzgʻor, kasb-hunar, diniy tushunchalar va qadriyatlar kabi mazmuniy guruhlariga ajratish maqsadga movofiq. Bunday tasnif keyingi bosqichlarda muayyan kategoriyalarning chastotasi va funksional taqsimotini aniqlash imkonini beradi.

Foydalanilgan adabiyotlar roʻyxati

1. Abjalova M. Oʻzbek tili milliy korpusida soʻzshaklni leksikografik baza asosida qidiruv imkoniyatlari // “Kompyuter lingvistikasi: muammolar, yechim, istiqbollar” Respublika I ilmiy-texnikaviy konferensiyasi. – 2021. – B. 12–17.
2. Biber D. Representativeness in Corpus Design // Literary and Linguistic Computing. – 1993. – Vol. 8, No. 4. – P. 243–257.
3. McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. – Cambridge: Cambridge University Press, 2012. – 294 p.
4. Poʻlatov A. Kompyuter lingvistikasi. – Toshkent: Akademnashr, 2011. – 160 b.
5. Sinclair J. Corpus, Concordance, Collocation. – Oxford: Oxford University Press, 1991. – 179 p.
6. Zaxarov V., Mengliyev B., Xamroyeva Sh., Xoliyorov Oʻ. Korpus lingvistikasi: oʻquv qoʻllanma. – Toshkent, 2021. – 164 b.