

**BOG'DORCHILIK TERMINLARINING RAQAMLI KORPUSINI
YARATISH: O'ZBEK TILI MISOLIDA**

Diyarov Akmal To'lqin o'g'li

*Kattaqo'rg'on davlat pedagogika instituti o'qituvchisi,
filologiya fanlari bo'yicha falsafa doktori (PhD)*

Annotatsiya. Maqolada o'zbek tilidagi bog'dorchilik terminlarining raqamli korpusini yaratish tamoyillari va bosqichlari yoritiladi. Korpus uchun manbalarni tanlash, matnlarni qayta ishlash (normallashtirish, tokenizatsiya, lemmatizatsiya), terminlarni annotatsiyalash sxemasi, ma'lumotlarni saqlash formatlari va sifat nazorati usullari tavsiflanadi. O'zbek tilining agglutinativ tabiati, imlo variantlari va shevaviy xilma-xillik korpus yaratishdagi asosiy muammolar sifatida ko'rsatiladi. Taklif etilgan yondashuv terminografiya, tarjima va til texnologiyalari uchun amaliy asos bo'lishi mumkin.

Kalit so'zlar: raqamli korpus, bog'dorchilik terminlari, annotatsiya, lemmatizatsiya, terminografiya, o'zbek tili, korpus lingvistikasi.

Bog'dorchilik o'zbek xalqining qadimiy xo'jalik tarmog'i bo'lib, bugungi kunda u agrotexnologiyalar, eksport va ilmiy tadqiqotlar bilan bog'liq holda jadal rivojlanmoqda. Bu jarayon sohaga oid terminologiyaning kengayishini, yangi terminlarning paydo bo'lishini va mavjud terminlarning ma'no jihatidan o'zgarishini taqozo etadi. Terminlar hozircha asosan bosma lug'atlar va tarqoq o'quv adabiyotlarida saqlanadi, bu esa ularni tizimli qidirish, chastotasini aniqlash va qo'llanilish doirasini kuzatishni qiyinlashtiradi. Shu bois bog'dorchilik terminlarining raqamli korpusini yaratish dolzarb vazifa hisoblanadi.

Korpus deganda ma'lum maqsad uchun maxsus tanlangan, mashina o'qiy oladigan va tilshunoslik ma'lumotlari bilan boyitilgan matnlar to'plami tushuniladi (Sinclair, 1991; McEnery va Hardie, 2012). Terminologik korpusning oddiy matnlar to'plamidan farqi shundaki, unda faqat so'zlar emas, balki terminlar, ularning ta'riflari, sinonimlari va qo'llanilish kontekstlari tizimli tarzda aks etadi. Bunday korpus bir vaqtning o'zida lug'at, qidiruv tizimi va tadqiqot vositasi vazifasini bajaradi. Maqolaning maqsadi – o'zbek tilidagi bog'dorchilik terminlari korpusini yaratishning metodik asoslarini ishlab chiqish va uning tuzilmasini taklif etishdir.

Korpusni yaratishda birinchi qadam manbalarni tanlashdir. Terminologik korpus uchun manbalar rang-barang va ishonchli bo'lishi lozim. Taklif etilayotgan tarkibga quyidagi turdagi matnlar kiradi: izohli va maxsus terminologik lug'atlar, qishloq xo'jaligi va bog'dorchilikka oid darsliklar va o'quv qo'llanmalari, ilmiy maqolalar va dissertatsiya avtoreferatlari, me'yoriy hujjatlar va standartlar, ommabop va ixtisoslashgan matbuot, shuningdek fermerlarga mo'ljallangan yo'riqnomalar va internet-resurslar. Har bir manba uchun nashr yili, muallif,

“DUNYO TA'LIMI SIFATINI OSHIRISHNING ILG'OR USULLARI ILMIY JURNALI”

5-Oktyabr, 2026-yil

janr, uslub va hudud kabi metama'lumotlar qayd etilishi kerak. Bu ma'lumotlar korpusni davr, uslub yoki mintaqa bo'yicha filtrlash imkonini beradi va terminlarning tarixiy hamda hududiy farqlarini kuzatishga xizmat qiladi. Shevalardagi bog'dorchilik terminlarini qamrab olish uchun dala yozuvlari va og'zaki nutq transkriptlarini ham qo'shish maqsadga muvofiq.

Ikkinchi qadam matnlarni qayta ishlashdir. O'zbek tili uchun bu bosqich bir qancha o'ziga xos qiyinchiliklarga ega. Birinchidan, imloda *o'* va *g'* harflari turli belgilar bilan yoziladi (to'g'ri apostrof, qiyshiq apostrof, ' yoki ' belgilari), natijada bir xil so'z kompyuter uchun turli shakllarda ko'rinadi. Shuning uchun matnlar yagona yozuv me'yoriga keltirilishi, ya'ni normallashtirilishi lozim. Ikkinchidan, ayrim eski manbalar kirill yozuvida bo'lib, ularni lotin yoziviga o'tkazish va natijani qo'lda tekshirish kerak. Uchinchidan, o'zbek tili agglutinatив til bo'lgani uchun bitta termin juda ko'p shaklda uchraydi: *daraxt*, *daraxtlar*, *daraxtlarimizning*, *daraxtzor* kabi shakllar bir leksemaga yoki unga yaqin birliklarga tegishli. Bunday sharoitda faqat so'z shakliga tayanib qidirish natijasi to'liq bo'lmaydi, shu sababli har bir so'z shakli o'z lemmasiga keltiriladi (lemmatizatsiya) va qo'shimchalar morfologik jihatdan ajratiladi. Bog'dorchilikda ko'p uchraydigan *bog'dorchilik*, *payvandlamoq*, *ko'chatzor* kabi yasama so'zlar uchun yasovchi qo'shimchalarni ham alohida ko'rsatish terminlarning so'z yasaliş modellarini o'rganishga imkon beradi.

Uchinchii qadam terminlarni annotatsiyalash, ya'ni har bir birlikni tilshunoslik ma'lumotlari bilan boyitishdir. Terminologik korpus uchun quyidagi annotatsiya sxemasini taklif etamiz:

Maydon	Mazmuni	Misol (<i>payvand</i>)
Lemma	termin asosiy shakli	<i>payvand</i>
So'z turkumi	grammatik toifa	ot
Morfologik tuzilish	asos va qo'shimchalar	<i>payvand</i> + <i>-la-</i> + <i>-moq</i>
Etimologik qatlam	kelib chiqish manbasi	fors-tojik
Tematik guruh	sohaviy tasnif	agrotexnik amal
Ta'rif	qisqa izoh	o'simlikning bir qismini boshqasiga biriktirish
Ma'no turi	to'g'ri yoki ko'chma	to'g'ri; ko'chma: yangi g'oyani qo'shish
Variant	imlo yoki shevaviy shakl	alohida qayd etiladi

Bunday sxema termini bir vaqtning o'zida lingvistik, tarixiy va tematik jihatdan tavsiflaydi. Tematik guruhlar sifatida o'simlik qismlari, meva va sabzavot navlari, agrotexnik amallar, asbob-uskunalar, kasallik va zararkunandalar hamda yangi texnologiyalar bilan bog'liq terminlar ajratilishi mumkin. Etimologik qatlam maydoni tarixiy-etimologik tadqiqotlar uchun alohida ahamiyatga ega: u turkiy, fors-tojik, arab va rus-baynalmilal

qatlamlar nisbatini avtomatik hisoblashga imkon beradi. Birlamchi lug'at bazasi sifatida *O'zbek tilining izohli lug'atidagi bog'dorchilik leksikasidan* foydalanish, uni maxsus lug'atlar va matn korpusidan olingan yangi terminlar bilan to'ldirish tavsiya etiladi.

Ma'lumotlarni saqlash uchun xalqaro standartlarga tayanish korpusning uzoq muddat foydalanilishini ta'minlaydi. Matnlar va metama'lumotlar TEI XML formatida (TEI Consortium) saqlanishi mumkin, gap darajasidagi morfologik va sintaktik annotatsiya uchun esa Universal Dependencies loyihasida qabul qilingan CoNLL-U formati qulay (Nivre va boshq., 2020). Terminlar ro'yxati va ularning annotatsiyasi esa jadval ko'rinishida (CSV yoki JSON) alohida saqlanib, mavjud lug'at va qidiruv tizimlariga oson ulanishi mumkin. Korpusni qidirish va tahlil qilish uchun Sketch Engine kabi korpus menejerlaridan yoki ochiq kodli muqobillaridan foydalanish mumkin; ular konkordans, chastota ro'yxati va so'z hamkorligi (kollokatsiya) tahlilini taqdim etadi (Kilgarrieff va boshq., 2014).

Sifat nazorati korpusning ishonchliligini belgilaydi. Avtomatik ravishda bajarilgan lemmatizatsiya va annotatsiya natijalari tilshunoslar tomonidan tanlab tekshirilishi, murakkab holatlar esa ikki mustaqil annotator tomonidan belgilanib, ularning kelishuvi statistik ko'rsatkich orqali baholanishi kerak. Ko'p ma'noli terminlar (masalan, *ildiz*, *shox*, *payvand*) uchun to'g'ri va ko'chma ma'nolarni ajratish ayniqsa muhim, aks holda chastota ko'rsatkichlari noto'g'ri talqin qilinadi. Shuningdek, matnlarni korpusga kiritishda mualliflik huquqi masalalari hisobga olinishi, ochiq litsenziyali va ruxsat berilgan manbalarga ustuvorlik berilishi zarur.

Tayyor korpus bir qancha amaliy vazifalarni hal etishga imkon beradi. Birinchidan, terminlarning chastotasi va qo'llanilish doirasini aniqlab, ularni ta'lim va lug'atlar uchun tanlash mumkin. Ikkinchidan, bir tushunchaning bir nechta nomi (masalan, o'zbekcha va o'zlashgan variantlar) o'rtasidagi raqobatni kuzatib, standartlashtirish bo'yicha asosli tavsiyalar berish mumkin. Uchinchidan, yangi terminlar va neologizmlarni vaqt bo'yicha kuzatib borish mumkin. To'rtinchidan, korpus terminlarning ko'chma ma'nolari va metaforik qo'llanilishini (masalan, *ildiz otmoq*, *meva bermoq*) chuqur o'rganish uchun material beradi. Beshinchidan, bu ma'lumotlar mashina tarjimai va til texnologiyalari uchun o'quv resursi sifatida xizmat qilishi mumkin.

Shu bilan birga, korpus yaratish jarayonida bir qancha muammolar mavjud: o'zbek tili uchun morfologik analizatorlar va izohlangan resurslarning cheklanganligi, shevalar va og'zaki nutq materialining kamligi, imlo me'yorlari o'zgarishi va manbalarning bir xil emasligi. Ularni hal etish uchun bosqichma-bosqich yondashuv tavsiya etiladi: avval kichik, lekin sifatli namunaviy korpus yaratish, uning asosida annotatsiya qoidalarini sinab ko'rish, so'ngra korpusni kengaytirish.

Xulosa qilib aytganda, bog'dorchilik terminlarining raqamli korpusi terminografiya, tarixiy-etimologik tadqiqotlar, tarjima va til texnologiyalari uchun yagona ma'lumot bazasini

yaratadi. Taklif etilgan manba tanlash, qayta ishlash, annotatsiya va sifat nazorati tamoyillari boshqa tarmoq terminologiyalari uchun ham qo'llanilishi mumkin.

FOYDALANILGAN ADABIYOTLAR:

1. Kilgarriff, A., Baisa, V., Bušta, J. va boshq. The Sketch Engine: ten years on. *Lexicography*, 1(1), 2014, pp. 7–36.
2. McEnery, T., Hardie, A. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, 2012.
3. Nivre, J. va boshq. Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. *Proceedings of LREC 2020*. Marseille: ELRA, 2020.
4. Sinclair, J. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.
5. TEI Consortium. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. <https://tei-c.org/guidelines/p5/>
6. *O'zbek tilining izohli lug'ati*. 5 jildlik. Toshkent: O'zbekiston milliy ensiklopediyasi, 2006–2008.