

O'ZBEK TILI KORPUSINING IMKONIYATLARI

Mohichehra Abramatova Abdurazzoq qizi

Termiz davlat universiteti o'zbek filologiyasi fakulteti
 abramatovamohichehra@gmail.com

Annotatsia: Maqolada solishtirish mumkin bo'lgan ikki tilli korpusdagi matnlarni moslashtirish va taqqoslanadigan korpusdan ajratib olingan terminologik birliklarning tarjima ekvivalentlarini o'rnatish muammosi muhokama qilinadi. Potentsial atamalar ro'yxati korpus menejeri dasturidan foydalangan holda o'zbek va ingliz subkurslaridan ajratib olinadi, so'ng olingan terminologik birliklarning leksik tarkibi o'rganiladi va tarjima ekvivalentlarini o'rnatish usullari taklif etiladi. Grammatikani o'rganish tilning tuzilishini, shu jumladan morfologiya, sintaksis va so'zlarning boshqa xususiyatlarini, masalan, ularning grammatik sinflarini (fe'llar, ismlar va boshqalarni) tushunishni o'z ichiga oladi. Leksikografiyadan farqli o'laroq, grammatika uzoq yillik empirik tadqiqot an'analari ega emas. So'nggi paytgacha ona tilida so'zlashuvchilar o'z tillarining grammatik boyliklaridan qanday foydalanganliklarini o'rganishga kam e'tibor berilardi.

Kalit so'zlar: terminologiya, atamalar ta'rifi, matnlar korpusi, taqqoslanadigan matnlar korpusi, tarjima, ikki tilli leksikografiya, korpus leksikografiyasi.

Korporativ tilshunoslik - kompyuter texnologiyalaridan foydalangan holda lingvistik korpuslarni (matnlar korpuslarini) qurish va ulardan foydalanishning umumiy tamoyillarini ishlab chiqadigan hisoblash tilshunosligining bir tarmog'i. Matnlarning lingvistik yoki lingvistik korpusi deganda ma'lum lingvistik muammolarni hal qilish uchun mo'ljallangan, mashinada o'qiladigan, birlashtirilgan, tuzilgan, belgilangan, filologik jihatdan malakali til ma'lumotlari to'plami tushuniladi. Hozirgi vaqtda "korpus" tushunchasining ko'plab ta'riflari mavjud. Masalan, E.Finegan darsligida berilgan ta'rifda shunday deyilgan: "Korpus odatda mashinada o'qiladigan formatdagi matnlarning vakillik to'plami bo'lib, matn yaratilgan vaziyat haqidagi ma'lumotlarni, masalan, axborotni o'z ichiga oladi. ma'ruzachi, muallif, qabul qiluvchi yoki auditoriya haqida" [4]. Vikipediya korpusni statistik tahlil va gipotezalarni tekshirish, muayyan holatlarni tekshirish yoki ma'lum sohalar uchun til qoidalarini tasdiqlash uchun foydalaniladigan katta va tuzilgan matnlar to'plami (hozirda odatda elektron shaklda) deb ta'riflaydi [12]. T.MakEneri va E.Vilsonlar quyidagi ta'rifni beradilar: korpus – til modeli sifatida foydalanish uchun aniq til mezonlariga muvofiq tanlangan til qismlari to'plamidir [9]. V.V. Rikov matnlar korpusini mantiqiy kontseptsiyaga asoslangan matnlar to'plami, bu matnlarni birlashtiradigan va matnlarni korpus, algoritmi va matnlar korpusini tahlil qilish dasturi, mafkuraviy tahlil qilish qoidalarida o'z ichiga olgan mantiqiy g'oya sifatida belgilaydi. , va u bilan bog'liq metodologiya [10].

Ta'riflarda zamonaviy matnlar korpusining asosiy xususiyatlari - maqsad ("mantiqiy g'oya"), mashinada o'qiladigan format, maxsus tanlash protsedurasi natijasida reprezentativlik, metallingvistik ma'lumotlarning mavjudligi ta'kidlangan. Mashina muhitida og'zaki materialning standartlashtirilgan tasviri uni qayta ishlash uchun standart dasturlardan foydalanishga imkon beradi.

Ishlarni yaratish imkoniyati va ulardan foydalanish ma'nosi quyidagi shartlar bilan belgilanadi:

1) korpusning etarlicha katta (vakillik) hajmi ma'lumotlarning tipikligini kafolatlaydi va lingvistik hodisalarning butun spektrini taqdim etishning to'liqligini ta'minlaydi;

2) har xil turdagi ma'lumotlar korpusda tabiiy kontekstual shaklda bo'lib, ularni har tomonlama va ob'ektiv o'rganish imkoniyatini yaratadi;

3) bir marta yaratilgan va tayyorlangan ma'lumotlar to'plami turli tadqiqotchilar tomonidan va turli maqsadlarda bir necha marta ishlatilishi mumkin.

"Matn korpusi" tushunchasi matn va lingvistik ma'lumotlarni boshqarish tizimini ham o'z ichiga oladi, bu so'nggi paytlarda ko'pincha ingliz korpus menejeri (korpus menejeri) deb ataladi. Bu korpusdagi ma'lumotlarni qidirish, statistik ma'lumotlarni olish va natijalarni foydalanuvchiga qulay shaklda taqdim etish uchun dasturiy ta'minotni o'z ichiga olgan maxsus qidiruv tizimi. Korpus qidiruvi har qanday so'z uchun moslikni yaratish imkonini beradi – manbaga havolalar bilan kontekstda ushbu so'zning barcha uchraydigan ro'yxati. Korpuslardan til va nutq birliklari haqida turli ma'lumot materiallari va statistik ma'lumotlarni olish uchun foydalanish mumkin. Xususan, korpus asosida so'z shakllari, leksemalar, grammatik kategoriyalarning chastotasi haqida ma'lumotlar olish, chastota va kontekstlarning turli davrlardagi o'zgarishini kuzatish, leksik birliklarning qo'shma kelishi haqidagi ma'lumotlarni olish va hokazo.

Muayyan davrdagi lingvistik ma'lumotlarning reprezentativ majmuasi filning leksik tarkibidagi o'zgarish jarayonlari dinamikasini o'rganish, turli janrlar va mualliflarning leksik va grammatik xususiyatlarini tahlil qilish imkonini beradi. Korpora, shuningdek, turli tarixiy va zamonaviy lug'atlarni tayyorlash uchun ko'p o'lchovli leksikografik ish uchun manba va vosita bo'lib xizmat qilish uchun mo'ljallangan. Korpus ma'lumotlari grammatikalarni qurish va takomillashtirish va tillarni o'rganish uchun ishlatilishi mumkin. An'anaviy tadqiqotlarda e'tibordan chetda qolgan sohalar korpus grammatikasini tadqiq qilishning kuchli tomoni ekanligi isbotlangan, bu grammatikaga so'z, jumla va nutq darajalarida qo'llanilishi mumkin. Bu erda morfologik belgilarning registrlar bo'yicha taqsimlanishini tahlil qilib, ulardan foydalanish va funksiyasi muammosini ko'rib chiqamiz. Korpusdan foydalanib, morfologik xususiyatning tarqalishini uni qo'llash konteksti bilan bog'lash va u bajaradigan funktsiyalarni yaxshiroq tushunish mumkin.

"Korpus lingvistikasi" darsligida ushbu muammoni hal qilish yo'li nominalizatsiya (tuzama otlar) uchta registrda taqsimlanishi bilan

ko'rsatilgan. Maqola ikki tilli solishtiriladigan yagona korpus - mavzu sohasining ikki tilli matn korpusining asosiy axborot birliklaridan olingan ma'lumotlarni taqqoslash uchun tarjima ekvivalentlarini tahlil qilish va o'rnatish usullariga bag'ishlangan; muammoni o'z vaqtida hal qilish. Korpus ma'lumotlarini o'rganishning maqsadi ma'lum bir mavzu bo'yicha mutaxassislar, uning moddiy va nomoddiy ob'ektlarini yaratuvchilar va iste'molchilari o'rtasida ikki tilli aloqani ta'minlashdir.

Matnlar korpusi deganda kompyuterda o'qiladigan formatdagi, dasturiy ta'minot yordamida qayta ishlash uchun mavjud bo'lgan, yetarli darajada katta va tadqiqot uchun vakili bo'lgan tabiiy tildagi matnlarning tuzilgan majmuasi tushuniladi. Bitta mavzuni qamrab oluvchi ikki tilli matn majmuasi maxsus muloqot doirasida mavzu sohasiga xos atamalardan foydalanish misollarini o'z ichiga oladi. Bu terminologiya tadqiqotida, ikki tilli terminologiya operatsiyalarini boshqarishda va ikki tilli terminologik lug'atlarda qo'llaniladigan tasdiqlangan dasturiy vosita sifatida tan olingan [7].

Hozirgi vaqtda keyingi tadqiqot va qayta ishlash uchun taqqoslanadigan ikki tilli korpuslardan ma'lumot olish muammosi mavjud. Ishchi tilda matnning teng uzunligi va hizalanishi ikki tilli korpusning reprezentativligi va taqqoslanuvchanligini ta'minlash uchun muhimdir. Ikki tilli korpuslar orasida parallel va taqqoslanadigan korpuslar mavjud. Agar parallel korpusda asl matnlar va jumlar bilan moslangan tarjima matnlari mavjud bo'lsa, taqqoslanadigan korpus bir xil mavzu sohasi, hajmi va tuzilishidagi matnlarni o'z ichiga oladi, lekin bir-birining tarjimasini emas [3]. Taqqoslanadigan korpusda ikkala kichik korpus matnlari o'rtasida terminologik kelishuv bo'lishi mumkin, bu matnlardan umumiy terminologik birliklar va iboralarni ajratib olish va ularni tarjima qilish muammosini hal qilishni nazarda tutadi [2].

Shuni alohida ta'kidlash kerakki, tushunish darajalarini tushunishning ketma-ket bosqichlari deb hisoblash mumkin emas. Tilni bilishning eng oddiy darajasidan boshqa har qanday darajasi uchun yuqoridagi har qanday darajani bajarish mumkin. Bularning barchasi "matn" va "o'quvchi" tizimlari o'rtasidagi munosabatlarga bog'liq. Chet tilini o'rgatish amaliyotida o'quvchilar o'qiganlarida eng yuqori tushunchaga ega bo'lgan matnlarni olishlarini ta'minlashga harakat qilish kerak. Shuning uchun matnlarni tanlashga, ya'ni darslik matnining shishishiga etarlicha e'tibor berish kerak. Aks holda kitobxonlik o'quvchiga zavq bag'ishlamaydi, balki og'ir akademik burchga aylanadi.

Aksariyat hollarda, birinchidan, o'rganilayotgan materialning lingvistik jihatlariga e'tibor qaratish mumkin bo'lib qoladi (masalan, o'quvchilarga o'xshash tuzilishga ega atamalar bilan javoblarni taklif qilish). talabalar (masalan, mutlaqo nomaqbul javoblarni kiritish orqali buni osonlashtiring va shu bilan variantlar doirasini toraytiring) [1]. Chastotali lug'atlarning ko'p birliklari terminologik emas. Non terminlar salbiy lug'atni tashkil qiladi va so'zlar ro'yxati hajmini kamaytirish uchun lug'atlardan olib tashlanishi kerak.

Inkor lug'at tarkibiga aniqlovchilar, yordamchi fe'llar, yuklamalar, bog'lovchilar, olmoshlar, tegishli otlar va ularning hosilalari kiradi. Bundan tashqari, korpus tarkibida uchraydigan bir leksik birlikning grammatik shakllari majmuasini asosiy shaklga keltirish, ya'ni lemmatizatsiyani amalga oshirish zarur.

Terminologik birliklarning lemmatizatsiyasi va qiyoslanishini soddalashtirish uchun bir ildizga ega bo'lgan hosila atamalarni alohida murakkab hosila birliklarga birlashtirish mumkin. Barcha hosila so'zlar eng katta so'z yasovchi birlik sifatida so'z yasovchi uyaga birikishi mumkin. Uya doirasida uni tashkil etuvchi murakkab so'z yasovchi birliklarni ajratib ko'rsatish mumkin: so'z yasovchi paradigmalar (ot paradigmalari, sifatlar paradigmalari) va so'z yasash zanjirlari [11]. Ikki tildagi so'z yasash uyalarini taqqoslash barcha hosilalarning ekvivalentlarini izlashni tezlashtiradi, tillardan birining terminologik tizimida atamalar yo'qligini aniqlaydi va atamalarning lemmatizatsiyalangan ro'yxatining leksik tarkibi haqida xulosa chiqaradi. .

Lug'atlarning notekis hajmda bo'lishining sabablari o'zbek tilining rivojlangan morfologik tizimi va o'zbek so'zlarining keng ko'lamli so'z yasovchi uya hosil qilish qobiliyatidadir. Masalan, ikki so'z yasovchi zanjir "semantika – sematizatsiya – sematizatsiya" va "semantika". - semantik - semantiklash - semantiklashtirish" kabi tamoyillar asosida shakllangan. Biroq, bu zanjirlarning barcha tarkibiy qismlari ishda taqdim etilmaydi. Masalan, o'zbek tilidagi sematizatsialash fe'li korpus tarkibida ifodalanmaydi, shuning uchun u tanlangan mavzu bo'yicha ilmiy nutqda ko'p uchramaydi. Ammo refleksiv fe'l "sematizatsiya qilish" bilan ifodalanadi, inglizcha so'zlar zanjirida uning o'xshashi yo'q, chunki ingliz tilida takrorlash g'oyasi maxsus fleksiyonlar bilan ifodalanmaydi.

Ingliz tilining subkorpusida o'zbek tilidagi "sematization lash – sematization lash" ga o'xshash leksik birliklarni ham kuzatmaymiz. Bundan tashqari, o'zbek tilining subkorpusida rus tilidan o'zlashtirilgan, so'z semantikasi bilan turtkilangan "psixosemantika, semantika-kognitiv va semantika-pragmatika" kabi ko'plab hosila birliklarni qayd etish mumkin. Ingliz subkorpusida bu nomlar va xususiyatlarning o'xshashlari yo'q. Yakka turtkili hosila so'zlar semantik bo'lib, ingliz subkorpusida faqat uchta bor: leksik-semantik n leksiko-semantik, ular sinonimlar va semantik bo'lmagan semantik - semantik. Lug'atlar tahlilidan xulosa qilishimiz mumkinki, yangi atamalarning asoslarni birlashtirish yo'li bilan shakllanishi o'zbek tilidagi ilmiy matnlarga ko'proq xosdir.

Ingliz tilidagi "semantik" so'zining yuqori chastotasini hisobga olsak, u ko'pincha terminologik iboralarda uchraydi va ingliz tili uchun "ikki mavzu sohasini birlashtirish" g'oyasini sintaktik havola orqali ifodalash odatiy holdir. Hosila so'zlarga yangi ma'no beruvchi o'zak va fleksiyalar tufayli lug'atning so'z yasovchi uyalarini kengroq bo'lib, chastotali lug'at hajmiga ta'sir qiladi. Bilim sohasini bildiruvchi leksik birliklar haqida gap ketganda, ularning yuqoriligini bog'lash mumkin. so'z yasash salohiyati va ulardan foydalanish

chastotasi va natijada ushbu bilim sohasining ma'lum bir filning ilmiy nutqiga aylanishi bilan keng so'z yasash uyalarini yaratish qobiliyati.

O'zbek tilida ikki asosli birliklarning ustunligi to'g'risidagi xulosalar yuqori hosilaviy qiymatga ega bo'lgan korpusning boshqa chastotali so'zlari uchun ham amal qiladi. Semantika-semantik ekvivalentlar juffligi sifatida so'z terminatori va termini ularning til subkorpusida eng ko'p uchraydigan narsalardan biridir. Ikkala so'z ham ulkan uyalar hosil qiluvchi ko'plab lotinlarni rag'batlantiradi. Terminning o'zbek korpusidagi so'z chastotasi uning ingliz subkorpusidagi ekvivalentidan birmuncha yuqoriroqdir. Ammo hosila poyasining hosila uyasi ko'p ikki asosli hosilalar tufayli "uya" atamasidan ancha katta. Tadqiqotning maqsadlaridan biri tarjima ekvivalentlarini yaratishda taqqoslanadigan korpuslarning imkoniyatlarini o'rganish bo'lganligi sababli, o'zbek va ingliz lug'atlarida eng ko'p uchraydigan ochiq sinf so'zlarini solishtirish qiziq.

Korpusdagi ochiq sinfnig ellikta eng tez-tez uchraydigan so'zlari katta ehtimollik bilan atamalar ekanligini hisobga olib, biz har bir subkorpusning chastota lug'atlarining dastlabki 50 ta so'zi o'rtasida tarjima ekvivalentlarini o'rnatish imkoniyatini ko'rib chiqamiz. Namuna ko'pincha umumiy derivatsion uyalarning elementlarini o'z ichiga oladi. Agar o'zbek tilidagi ma'lum bir so'z shakli uchun inglizcha ekvivalenti topilmagan bo'lsa, uni inglizcha ekvivalenti bilan bir xil generativ o'zakdan kelib chiqqan boshqa so'z shakllari bilan bog'lash mumkin.

Terminning ma'nosi kontekstda ochilib, ingliz tilidagi ot, jumladan, terminologik ma'nosi nominal iboraning o'zagini o'zgartiruvchi vazifasini bajarishga moyilligini hisobga olsak, terminologik birikmalarni iboraning leksik birliklari bilan solishtirishimiz mumkin. ekvivalentlari topilmadi. Shunisi e'tiborga loyiqki, o'zbek tilidagi deyarli barcha so'zlarning ekvivalentlari bor, ayrim birliklardan tashqari.

Mavjud atamalarni tasniflash usullariga muvofiq, masalan, nutq qismlari va qo'llanilish doirasi bo'yicha [6; 8], qolgan so'zlar quyidagicha ta'riflanishi mumkin:

- umumiy ilmiy atamalar-fe'llar (aniqlash, olib borish, o'zaro bog'lash, bog'lash);
- otning umumiy ilmiy atamalari (fan, belgi, ism, dunyo, ma'no, savol, ifoda, muammo, mavzu, komponent, shaxs);
- otlarning fanlararo atamalari (lug'at, botanika).

Ushbu so'zlarning korpusda mavjudligi alohida umumiy ilmiy atamalar bilan izohlanadi, ular bir vaqtning o'zida nafaqat turli fanlarning terminologik tizimlariga, balki turli xil maxsus maqsadli tillarga (ASL) tegishli bo'lishi mumkin [8]. ularning yuqori chastotasi. Namunada ushbu tadqiqot mavzusi bo'lmagan (botanika) terminologik tizimlarga tegishli atamalarni ajratib ko'rsatish mumkin. Buning sababi shundaki, korpusdagi matnlarning aksariyati terminologiyani o'rganishga bag'ishlangan bo'lib, boshqa fanlarning terminologik tizimlari esa matn mualliflari tomonidan o'rganish uchun material hisoblanadi.

Matnlarda mualliflarning izlanishlarini aks ettiruvchi misol sifatida boshqa fanlarning terminologiya tizimidagi atamalar keltirilgan. Ko'pgina ESPSda nominal atamalar og'zaki so'zlardan ustun turadi, amaliy fanlarga oid ba'zi matnlar, masalan, muhandislik, harbiy, geologiya [6] bundan mustasno. Turli korpuslarning leksik tarkibiga asoslanib, bitta mavzu sohasini tavsiflovchi turli tillarda terminologik tizimlarning rivojlanishi bir xil emas degan xulosaga kelish mumkin. Biroq, ikki tilli taqqoslanadigan korpusda ekvivalentlarni topish tartib-qoidalarining ishlab chiqilishi atamalar bo'yicha taqqoslanadigan korpusga moslashish mumkinligini ko'rsatdi.

Terminlarni qo'llanish sohasiga ko'ra tasniflash terminologik iboralarning terminologik ma'nosi haqida qo'shimcha xulosalar chiqarishga, ularning ta'rifiga ta'sir ko'rsatishga imkon beradi. Taqqoslanadigan korpusdan olingan atamalar terminologik tizimlarni va keyingi leksikografiyani tuzish va solishtirish uchun ishlatilishi mumkin.

ADABIYOTLAR:

1. Bakirova X.B. Nolingvistik universitetlarda xorijiy tilda terminologiyani o'qitish. Innovatsion nutq xalqaro jurnali. Integratsiya va ta'lim. Jildi: 01-son: 01. 2020. <http://summusjournals.uz/index.php/ijdiie>.

2. Belyaeva L.N. Zamonaviy tarmoq makonidagi lingvistik texnologiyalar: mahalliyashtirish sohasidagi tilshunos: monografiya. Kitob uyi, 2016. - S. 134.

3. Delpech E., Dyle B. Taqqoslanadigan korpuslardan olingan lug'at bilan ishlash: tekshirish va almashish // Terminologiya va bilim muhandisligi bo'yicha 9-konferentsiya materiallari. Fiontar: Dublin shahar universiteti, 2010, 229-223-betlar.

4. Finegan E. TILL: tuzilishi va ishlatilishi. - Nyu-York: Harcourt Brace College Publishers, 2004. Fung P. Parallel bo'lmagan korpuslardan terminologiyaning tarjimalarini topish. - URL: http://www.cs.columbia.edu/nlp/papers/1997/fung_mckeown_97.pdf.

5. Gerd A.S. Maxsus maqsadlar uchun til o'rganishga kirish: o'quv qo'llanma. Filologiya fakulteti, 2011. - B. 60.

6. Xurshid A., Rojers M. Terminologiyani boshqarish: korpus yondashuvi. URL: <http://www.mt-archive.info/90/Aslib-1992-Ahmad.pdf>.

7. Leichik V.M. Terminologiya: predmeti, usullari, tuzilishi. 3-nashr. M.: LKI nashriyoti, 2007. - S. 256.

8. Mc Enery T., Wilson A. Korpus tilshunosligi. - Edinburg: Edinburg universiteti nashriyoti, 2001.

9. Rikov V.V. Matnlar korpusi ob'ektga yo'naltirilgan paradigmaning amalga oshirilishi sifatida // Materiallar 9.

10. "Dialog-2002" xalqaro seminari. - M.: Nauka, 2002.

11. Tixonov A.N. Zamonaviy rus tilidagi leksik uyasi: monografiya. Yelets, 2006. 270 b. Vikipediya - <http://en.wikipedia.org/wiki/>.

12. Niyazova G.; Raupova L. Detektiv janrining mashhurligining asosiy belgilari va sabablari. OI 2021. 2. - B. 515-524.