Ilmiy tadqiqotlar va ularning yechimlari jurnali
Aoka guvohnoma № 323759 ISSN: 3060-4818

UO‘K 811.111'373.46:81'322

INGLIZ TERMINOLOGIYA MAKTABI VA KORPUS-ASOSLI METODOLOGIYANING SUN’IY INTELLEKT ATAMALARINI TADQIQ ETISHDAGI O‘RNI

Solijonova Mahbubaxon Bo‘riboy qizi
mustaqil izlanuvchi,
Andijon davlat universiteti filologiya fakulteti
xorijiy tillar filologiyasi kafedrasi, Andijon
ORCID iD: 0009-0000-0618-9172;
e-mail: makhbuba.solijanova@gmail.com
https://doi.org/10.5281/zenodo.22161574

Annotatsiya. Maqolada ingliz terminologiya maktabining shakllanish bosqichlari va uning korpus-asosli metodologiyasi tahlil qilinadi. H.Picht va J.Draskau, J.C.Sager, J.Pearson, J.Sinclair hamda L.Bowker ishlarining ketma-ketligi asosida ingliz an’anasida terminni lug‘aviy izolyatsiyada emas, real matn muhitida o‘rganish tamoyilining qanday shakllangani ko‘rsatiladi. Tadqiqotda ADU-SI Corpus (2 275 000 so‘z) materialida korpus-asosli yondashuvning sun’iy intellekt terminologiyasini o‘rganishdagi amaliy samaradorligi baholanadi. Natijalar shuni ko‘rsatdiki, ta’rifli kontekstlarni (definitional contexts) avtomatik ajratish usuli 150 ta atamaning 82 foizi uchun ishonchli semantik tavsif berish imkonini beradi.

Kalit so‘zlar: ingliz terminologiya maktabi, korpus lingvistikasi, ta’rifli kontekst, terminografiya, sun’iy intellekt atamalari, kollokatsiya, matn muhiti.

АНГЛИЙСКАЯ ТЕРМИНОЛОГИЧЕСКАЯ ШКОЛА И РОЛЬ КОРПУСНОЙ МЕТОДОЛОГИИ В ИССЛЕДОВАНИИ ТЕРМИНОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация. В статье анализируются этапы формирования английской терминологической школы и её корпусная методология. На основе последовательности работ Х.Пихта и Й.Драскау, Дж.Сейгера, Дж.Пирсон, Дж.Синклера и Л.Боукер показано, каким образом в английской традиции сложился принцип изучения термина не в словарной изоляции, а в реальной текстовой среде. Оценивается практическая эффективность корпусного подхода на материале ADU-SI Corpus объёмом 2 275 000 слов. Результаты показали, что метод автоматического выделения дефиниционных контекстов позволяет получить надёжное семантическое описание для 82 % из 150 терминов.

Ключевые слова: английская терминологическая школа, корпусная лингвистика, дефиниционный контекст, терминография, термины искусственного интеллекта, коллокация, текстовая среда.

THE ENGLISH SCHOOL OF TERMINOLOGY AND THE ROLE OF CORPUS-BASED METHODOLOGY IN STUDYING ARTIFICIAL INTELLIGENCE TERMS

Abstract. The article analyses the formative stages of the English school of terminology and its corpus-based methodology. Drawing on the sequence of works by Picht and Draskau, Sager, Pearson, Sinclair and Bowker, it demonstrates how the principle of studying the term in a real textual environment rather than in lexicographic isolation took shape in the English tradition. The practical

effectiveness of the corpus-based approach is assessed using the ADU-SI Corpus (2,275,000 words). The results show that automatic extraction of definitional contexts yields a reliable semantic description for 82 per cent of the 150 terms studied.

Key words: English school of terminology, corpus linguistics, definitional context, terminography, artificial intelligence terms, collocation, textual environment.

Terminologiya fanining rivojlanishida kontinental (Vena, Praga, Moskva) va anglofon an'analar o'rtasidagi metodologik farq muhim ahamiyatga ega. Agar kontinental maktablar terminni tushunchadan kelib chiqib, onomasiologik yo'nalishda o'rgansa, ingliz an'anasi termini real matn qo'llanishidan kelib chiqib, semasiologik yo'nalishda tahlil qilishga ustunlik beradi.

Mazkur farq sun'iy intellekt (SI) terminologiyasini o'rganishda hal qiluvchi ahamiyat kasb etadi. SI sohasi atamalari rasmiy standartlarda qayd etilishidan ancha oldin ilmiy maqolalar, texnik hujjatlar va dasturlash resurslarida faol qo'llanila boshlaydi. Shu sababli terminni faqat lug'at ta'rifi orqali o'rganish sohaning haqiqiy holatini aks ettirmaydi.

Maqolaning maqsadi — ingliz terminologiya maktabining shakllanish bosqichlarini tavsiflash va uning korpus-asosli metodologiyasi SI terminologiyasini o'rganishda qanday amaliy natija berishini empirik material asosida ko'rsatishdan iborat. Tadqiqot vazifalari: (1) ingliz an'anasining nazariy asoslarini aniqlash; (2) korpus-asosli usullar tizimini tavsiflash; (3) ularning SI materialidagi samaradorligini o'lchash.

Ilmiy yangilik shundan iboratki, ingliz terminologiya maktabining metodologik ketma-ketligi birinchi marta o'zbek tilshunosligi kontekstida tizimli tavsiflanadi va uning tatbiqi ikki tilli SI korpusi materialida miqdoriy jihatdan baholanadi.

Ingliz terminologiya maktabining shakllanishida H.Picht va J.Draskauning qo'llanmasi burilish nuqtasi bo'ldi — mualliflar Wüsterning nemis tilidagi Umumiy terminologiya nazariyasini birinchi marta ingliz tilida sistemli tarzda taqdim etib, kontinental an'anani anglofon ilmiy muhitiga olib kirdilar [1, 34-b.].

J.C.Sagerning "A Practical Course in Terminology Processing" asari qiyosiy terminologik tahlilning metodologik asoslarini belgilab berdi. Sager terminologik ishni uch bosqichga ajratdi: terminlarni aniqlash, ularni tavsiflash va me'yorlashtirish [2, 89-b.]. Uning yondashuvida terminning tabiiyligi (naturalness) prinsipi alohida ajratiladi.

J.Pearson "Terms in Context" tadqiqotida korpus lingvistikasi metodologiyasini terminografiyaga ilk bor tizimli tatbiq etib, terminlarni lug'aviy izolyatsiyada emas, balki real matn muhitida o'rganish tamoyilini asosladi [3, 42-b.]. Pearson ta'rifli kontekstlarni (definitional contexts) avtomatik aniqlash uchun leksik-sintaktik shablonlar tizimini ishlab chiqdi.

J.Sinclairning korpus lingvistikasi sohasidagi ishlari kollokatsiya va kolligatsiya tushunchalarini terminologik tahlilga olib kirdi [4, 110-b.]. Sinclairning "idiom principle" g'oyasiga ko'ra, til birliklari erkin tanlanmaydi, balki oldindan shakllangan birikmalar sifatida qo'llaniladi — bu tamoyil terminologik birikmalarni aniqlashda samarali vosita bo'lib xizmat qiladi.

L.Bowker terminologiya va tarjima kesishmasidagi masalalarni o'rgangan holda, korpus asosidagi terminologik ishning amaliy jihatlarini tavsifladi [5, 308-b.]. Uning ishlarida ikki tilli korpuslardan foydalanish metodikasi batafsil bayon etilgan.

Zamonaviy bosqichda P.Drouin va M.-C.L'Homme terminologik ma'lumotlarni yarim avtomatik ajratish usullarini rivojlantirdilar [6, 71-b.]. Ularning ishlarida statistik va lingvistik mezonlarni birlashtirgan gibrid yondashuv taklif etiladi.

B.Biber va S.Conradning registr nazariyasi matn turlariga qarab leksik xususiyatlarning o'zgarishini tavsiflaydi [7, 156-b.]. Bu nazariya SI atamalarining turli diskurslardagi xatti-harakatini o'rganishda metodologik asos bo'lib xizmat qiladi.

O‘zbek tilshunosligida korpus-asosli terminologik tadqiqotlar hozircha boshlang‘ich bosqichda. A.Nurmonovning nazariy ishlarida korpus yondashuvining istiqbollari qayd etilgan bo‘lsa-da [8, 176-b.], SI sohasiga tatbiq etilgan maxsus tadqiqot mavjud emas.

Tadqiqot ADU-SI Corpus materialida o‘tkazildi. Korpus hajmi 2 275 000 so‘z bo‘lib, undan 1 565 000 so‘zi ingliz tilida, 475 000 so‘zi o‘zbek tilida va 235 000 so‘zi ikki tilli manbalardan iborat. Korpus 2018–2024-yillar oralig‘idagi sakkizta segmentni qamrab oladi.

Pearson metodologiyasiga muvofiq ta’rifli kontekstlarni ajratish uchun quyidagi leksik-sintaktik shablonlar qo‘llanildi: X is a Y that…, X refers to…, X is defined as…, X, also known as…, by X we mean…. O‘zbek tilidagi segment uchun mos shablonlar ishlab chiqildi: X — bu Y, X deb… ataladi, X quyidagicha ta’riflanadi.

Kollokatsion tahlil Sinclair metodologiyasi asosida amalga oshirildi. Har bir atama uchun besh so‘zlik oyna (±5) doirasida statistik ahamiyatli qo‘shni so‘zlar aniqlandi. Ahamiyat mezoni sifatida o‘zaro axborot ko‘rsatkichi (Mutual Information) qo‘llanildi.

Terminlarni avtomatik ajratishda Drouin va L’Homme taklif etgan gibrid mezon ishlatildi: nomzod birlik terminologik maqomga ega deb hisoblanadi, agar u (a) ixtisoslashgan segmentda umumiy segmentga nisbatan kamida uch barobar ko‘p uchrasa va (b) barqaror morfosintaktik shaklga ega bo‘lsa.

Baholash uchun tanlanma 150 ta yadro atamadan iborat bo‘ldi. Har bir atama uchun ta’rifli kontekst topilgan-topilmagani, kollokatsion profilning barqarorligi va tavsifning to‘liqligi qayd etildi. Natijalarning ishonchliligi test-qayta-test usulida tekshirildi (intra-rater $\kappa = 0{,}82$).

Korpus-asosli metodologiyaning SI terminologiyasiga tatbiq etilishi quyidagi natijalarni berdi.

1-jadval

Ta’rifli kontekstlarni ajratish samaradorligi (n = 150)

Segment	Til	Ta’rif topilgan	Foiz
K1 Ilmiy maqolalar	ingliz	141	94,0 %
K2 Texnik hujjatlar	ingliz	128	85,3 %
K3 Ommaviy nashrlar	ingliz	96	64,0 %
K4 Akademik nutq	o‘zbek	89	59,3 %
K5 Ommabop matnlar	o‘zbek	54	36,0 %
Umumiy ko‘rsatkich	—	123	82,0 %

Jadvaldan ko‘rinadiki, ta’rifli kontekstlarni ajratish usuli ingliz tilidagi ilmiy segmentda eng yuqori samaradorlikni ko‘rsatdi (94,0 foiz). Bu natija Pearson metodologiyasi akademik janrga yo‘naltirilgani bilan izohlanadi: ilmiy maqolalarda yangi atama kiritilganda uning ta’rifi keltirilishi janr me’yori hisoblanadi.

O‘zbek tilidagi segmentlarda ko‘rsatkich sezilarli past (59,3 va 36,0 foiz). Bizning fikrimizcha, bu ikki sababga ega. Birinchidan, o‘zbek tilidagi SI matnlarida atama ko‘pincha inglizcha shaklda, ta’rifsiz keltiriladi. Ikkinchidan, o‘zbek ilmiy uslubida ta’rif berish shablonlari hali to‘liq barqarorlashmagan.

Kollokatsion tahlil natijalari Sinclairning idiom prinsipini tasdiqladi. Korpusdagi 150 ta atamadan 118 tasi (78,7 foiz) barqaror kollokatsion profilga ega bo‘lib chiqdi. Masalan, neural network atamasi deep, convolutional, recurrent, train, architecture so‘zlari bilan statistik ahamiyatli birikma hosil qiladi.

Muhim kuzatuv shundan iboratki, o‘zbek tilidagi ekvivalentlarning kollokatsion profili inglizcha asl shaklnikidan sezilarli farqlanadi. “Neyron tarmoq” birikmasi o‘zbek korpusida asosan “sun’iy”, “model”, “o‘qitish” so‘zlari bilan kelsa, inglizcha neural network ancha keng va texnik jihatdan differensiallashgan profilga ega. Bu farq o‘zbek SI terminologiyasining hali to‘liq shakllanmaganidan dalolat beradi.

Terminlarni avtomatik ajratishning gibrid mezoni qanoatlanarli natija berdi: nomzod birliklarning 71,4 foizi qo‘lda tekshiruvda terminologik maqomni tasdiqladi. Xatolarning asosiy manbai — umumiy leksikaga oid, lekin SI matnlarida ko‘p uchraydigan so‘zlar (model, system, data).

Olingan natijalarni umumlashtirib aytish mumkinki, ingliz terminologiya maktabining korpus-asosli metodologiyasi SI terminologiyasini o‘rganish uchun samarali vosita bo‘lib xizmat qiladi, biroq uning o‘zbek tili materialiga tatbiqi qo‘shimcha moslashtirishni talab etadi. Xususan, o‘zbek tili uchun ta’rif shablonlari tizimi kengaytirilishi va agglutinativ morfologiyani hisobga oluvchi normalizatsiya bosqichi kiritilishi lozim.

Tadqiqot ingliz terminologiya maktabining Picht–Draskau, Sager, Pearson, Sinclair va Bowker chizig‘i bo‘yicha shakllangan metodologik an’anasini tavsifladi. Mazkur an’ananing markaziy tamoyili — terminni real matn muhitida o‘rganish — SI terminologiyasi kabi shakllanish bosqichidagi sohalar uchun ayniqsa samarali.

Empirik tahlil natijalari korpus-asosli yondashuvning amaliy samaradorligini tasdiqladi: ta’rifli kontekstlar 150 ta atamaning 82 foizi uchun aniqlandi, 78,7 foiz atama esa barqaror kollokatsion profilga ega bo‘lib chiqdi. Ayni paytda o‘zbek tilidagi segmentlarda ko‘rsatkichlar sezilarli past bo‘lib, bu o‘zbek SI terminologiyasining me’yorlashuv darajasi pastligini aks ettiradi.

Amaliy tavsiyalar quyidagilardan iborat. Birinchidan, o‘zbek tili uchun ta’rifli kontekstlarni ajratuvchi leksik-sintaktik shablonlar tizimini kengaytirish. Ikkinchidan, agglutinativ morfologiyani hisobga oluvchi lemmalash bosqichini terminologik ish oqimiga kiritish. Uchinchidan, o‘zbek ilmiy uslubida yangi atama kiritilganda uning ta’rifini keltirish amaliyotini me’yoriy tavsiya darajasiga ko‘tarish.

Kelgusi tadqiqotlar uchun istiqbolli yo‘nalish — o‘zbek tili uchun maxsus terminologik ajratish vositasini (term extractor) ishlab chiqish va uni SI sohasi korpusida sinovdan o‘tkazish.

FOYDALANILGAN ADABIYOTLAR RO‘YXATI:

1. Picht H., Draskau J. Terminology: An Introduction. – Guildford: University of Surrey, 1985. – 265 p.
2. Sager J.C. A Practical Course in Terminology Processing. – Amsterdam: John Benjamins, 1990. – 254 p.
3. Pearson J. Terms in Context. – Amsterdam; Philadelphia: John Benjamins, 1998. – 246 p.
4. Sinclair J. Corpus, Concordance, Collocation. – Oxford: Oxford University Press, 1991. – 179 p.
5. Bowker L. Terminology and Translation // Handbook of Terminology. Vol. 1 / Ed. by H.J. Kockaert, F. Steurs. – Amsterdam: John Benjamins, 2015. – P. 304–323.
6. Drouin P., L’Homme M.-C. Semi-automatic Methods for Extracting Terminological Data: Corpus-Based Approaches // Terminology. – 2022. – Vol. 28, № 1. – P. 68–95.
7. Biber D., Conrad S. Register, Genre, and Style. 2nd ed. – Cambridge: Cambridge University Press, 2019. – 406 p.
8. Nurmonov A. O‘zbek tilshunosligining nazariy asoslari. – Toshkent: O‘zbekiston, 2012. – 312 b.
9. L’Homme M.-C. Lexical Semantics for Terminology: An Introduction. – Amsterdam: John Benjamins, 2020. – 263 p.
10. Ahmad K., Rogers M. Corpus Linguistics and Terminology Extraction // Handbook of Terminology Management. – Amsterdam: John Benjamins, 2001. – P. 725–760.

11. Mahlberg M. Corpus Stylistics and Dickens's Fiction. – New York; London: Routledge, 2013. – 194 p.
12. Pecman M., Kübler N. Investigating Sublanguages in Specialized Discourse: A Corpus-Based Approach // Procedia – Social and Behavioral Sciences. – 2015. – Vol. 198. – P. 422–431.
13. Daille B., Romary L. Terminology-Based Knowledge Representation: Corpus-Driven Approaches // Terminology. – 2021. – Vol. 27, № 2. – P. 142–165.
14. IEEE Std 2894-2022. IEEE Guide for an Architectural Framework for Explainable Artificial Intelligence. – New York: IEEE, 2022. – 34 p.
15. Cabré M.T. Terminology: Theory, Methods and Applications / Transl. by J.C. Sager. – Amsterdam; Philadelphia: John Benjamins, 1999. – 248 p.