

DEVELOPING MEDICAL STUDENTS' TERMINOLOGICAL COMPETENCE THROUGH DATA-DRIVEN LEARNING

Isomiddinova Diyora Qutbiddin qizi

Tashkent State Medical University

Email: diyoraisomiddinova467@gmail.com

ORCID: <https://orcid.org/0009-0004-7861-8123>

Abstract

Corpus linguistics offers a principled way to select and teach specialized vocabulary from authentic disciplinary language. This conceptual paper discusses how frequency, dispersion, concordances, and collocations can inform medical English instruction and proposes a staged corpus-informed teaching framework. It argues that terminology teaching should integrate form, meaning, phraseology, and communicative use rather than rely on memorized bilingual word lists. The framework is a pedagogical proposal, not a report of an intervention.

Keywords: corpus linguistics; medical terminology; English for Specific Purposes; data-driven learning; concordance; medical education.

Introduction

Medical students encounter dense technical vocabulary in textbooks, patient records, clinical guidelines, and professional communication. Knowing a dictionary definition is insufficient when a term must be recognized in a case report, combined with conventional verbs, or explained to a patient. English for Specific Purposes (ESP) teaching therefore requires attention to how medical terms function in discourse. Corpus linguistics makes recurring patterns observable and offers a transparent basis for instructional choices.

Theoretical background

A corpus is a systematically assembled collection of electronic texts that can be queried to examine patterns of language use (McEnery & Hardie, 2012).

Concordance displays place a target expression in multiple authentic contexts; frequency lists identify recurrent forms; keyword analysis highlights items unusually prominent in a specialist corpus; and collocation measures reveal habitual combinations. Sinclair (1991) emphasizes the importance of phraseological patterning, while Johns (1991) introduced data-driven learning as a way for learners to investigate language evidence. Meta-analytic work by Boulton and Cobb (2017) suggests that corpus use can support language learning, although outcomes depend on instructional design and learner preparation.

Constructing a medical teaching corpus

A useful teaching corpus should reflect the target students' future communication needs. Teachers may compile appropriately licensed, de-identified materials from open-access clinical articles, educational case reports, patient-information leaflets, and professional guidelines. Texts should be tagged by genre and topic; metadata should record source, publication year, specialty, and intended audience. Patient-identifiable clinical records must not be added without proper authorization and safeguards. A small, carefully selected corpus can be more instructionally relevant than a large collection with mismatched genres.

Selecting a terminological minimum

Selection should combine frequency with dispersion across documents, disciplinary relevance, and curricular importance. High frequency alone can overrepresent general vocabulary or a narrow specialty. A candidate list can be checked against course objectives and reviewed by medical subject specialists. For instance, terms such as inflammation, lesion, diagnosis, and contraindication may be prioritized according to students' learning stage. The resulting list should include common word families, abbreviations, pronunciation challenges, and conventional multiword units.

From concordances to learning tasks

A scaffolded sequence begins with noticing: learners highlight a target term

in short concordance lines and infer its grammatical behavior. In guided analysis, they compare patterns such as “diagnosed with,” “risk of,” and “administer a dose,” checking which patterns are appropriate in specific genres. Next, they reconstruct authentic sentences, paraphrase specialist explanations for patients, and use selected expressions in simulated case summaries. The teacher verifies meanings and corrects misleading inferences, especially where clinical precision is essential.

Differentiation and assessment

Novice learners may need shortened concordance sets, glosses, and teacher-modeled searches; more proficient learners can compare genres or investigate synonym preferences. Assessment should distinguish recognition of terminology from productive use. A practical rubric can consider accurate meaning, collocation, grammatical form, register, and suitability for a communicative task. Pre- and post-instruction tasks can provide evidence of learning, but causal effectiveness requires a properly designed empirical study.

Limitations and ethical considerations

Corpus evidence describes usage within sampled texts; it does not itself establish clinical correctness. Poor corpus composition, unfamiliar interfaces, and excessive data can overwhelm students. Teachers must validate medical content against authoritative references and distinguish technical writing from patient-friendly communication. Copyright, source attribution, and privacy are essential when compiling and sharing materials.

Conclusion

Corpus-based terminology teaching can link lexical selection to authentic use and support active inquiry into medical English. A feasible model combines a purpose-built corpus, a reviewed terminological minimum, guided concordance work, communicative transfer, and explicit assessment. Future classroom research should test its effectiveness across medical specialties and proficiency levels.

References

- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393.
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *English Language Research Journal*, 4, 1–16.
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Anthony, L. (2022). AntConc (Version 4.0) [Computer software]. Waseda University. <https://www.laurenceanthony.net/software/antconc/>
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.
- Hyland, K. (2006). *English for Academic Purposes: An Advanced Resource Book*. Routledge.
- Flowerdew, L. (2012). *Corpora and Language Education*. Palgrave Macmillan.
- Isomiddinova, D. Q. (2026). A corpus-based approach to teaching English-Uzbek-Russian medical terminology. *Filologiya va uni o'qitish: muammolar va innovatsion yechimlar*, 1, 455.
- Isomiddinova, D. Q. (2026). Corpus linguistics in English for Specific Purposes (ESP): A focus on medical discourse. *Conference on Global Research Perspectives*, 5(1), 216–219.
- Isomiddinova, D. Q. (2026). Socially responsive pedagogical implications in language learning. *Global Research and Academic Innovations*, 2(4), 274–275.
- Isomiddinova, D. Q. (2026). Zamonaviy tibbiyotda ingliz tilining global ilmiy til sifatida o'rni. *Filologiya va Pedagogika ilmiy-metodik elektron jurnali*, 7(26), 18.
- Kilgariff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004). The Sketch Engine. *Proceedings of EURALEX 2004*, 105–116.