European International Journal of Philological Sciences (ISSN: 2751-1715)

Received: 14 April 2026 | Accepted: 02 May 2026 | Published: 31 May 2026

DOI: 10.55640/eijps-06-05-28

RESEARCH ARTICLE

A Corpus-Based Study of Structural Models in English-Uzbek Cardiological Terminology: Addressing Asymmetry, Nominalization, And Terminological Chaos

Batirova Khurshida Egamberdiyevna

Uzbek State World Languages University, Translation Faculty, The English Translation Theory Department, Teacher, Uzbekistan

VOLUME: Vol.06 Issue05 2026

PAGE: 139-142

Abstract

Contemporary medical discourse is characterized by a significant transition from simple mono-lexical units to complex multi-word terms (MWTs) and dense noun phrases. This structural transformation presents substantial challenges for specialized translation and terminological standardization, particularly in low-resource language pairs such as English and Uzbek. This study establishes a corpus-based methodological framework for identifying and analyzing MWTs within the cardiology domain. By utilizing quantitative criteria such as frequency, keyness, and dispersion across academic and clinical subcorpora, the research evaluates structural-semantic patterns in English cardiological discourse and their Uzbek equivalents. The findings reveal a marked dominance of two-component nominal structures and a profound structural asymmetry between English source terms and Uzbek target expansions. This research provides a data-driven foundation for developing standardized English-Uzbek cardiological equivalents, bridging the gap between international Greco-Latin traditions and native linguistic resources.

KEYWORDS

Medical terminology, corpus linguistics, cardiological multi-word terms, structural models, English-Uzbek translation, terminological chaos, onomastic angst.

INTRODUCTION

Medical English is a highly specialized linguistic subsystem defined by its precision, systematicity, and heavy reliance on classical roots. Traditionally, the language of medicine was viewed as a stable system of communication among experts. However, the rapid expansion of medical science in the digital age has led to what researchers describe as "terminological chaos," where the need to promptly name new pathological conditions, diagnostic procedures, and surgical instruments often outpaces the development of standardized nomenclature. [6, 14]

In clinical settings, terminology serves as an indispensable tool for effective academic and professional communication. The concept of "genre" in medicine is key, encompassing both written research articles and spoken clinical interactions. Within these genres, a significant shift has been observed toward "nominalization" – the process by which complex concepts are packed into dense multi-word terms and multi-component noun phrases rather than simple words. In the specialized field of cardiology, this precision is a clinical requirement. Concepts such as acute myocardial infarction or congestive heart failure function as single semantic entities, yet their internal structural-semantic complexity poses significant hurdles for cross-linguistic transfer. [14]

A notable research gap exists regarding the authentic use of

medical language in English-Uzbek translation. Standard bilingual dictionaries often focus on monolexical units that fail to reflect the phrasal nature of modern cardiological discourse. Furthermore, the field experiences "onomastic angst" – the tension between traditional eponyms named after discoverers and the push for standardized descriptive terms. For the Uzbek language, which is increasingly participating in global medical discourse as a Turkic language with unique patterns of adaptation, the systematic alignment of these complex English structures remains under-researched. This study employs corpus linguistics to bridge this gap, identifying real-world usage patterns through objective statistical evidence rather than subjective selection. [2, 12]

METHODS

The study utilizes an interdisciplinary framework combining morphology, semantics, and a quantitative, corpus-driven methodology to extract and analyze cardiological terms from authentic professional discourse. This approach minimizes subjectivity and ensures that high-frequency terms, which are more likely to be encountered in clinical practice, are prioritized.

Subcorpus Composition The research data was elicited from a compiled cardiological subcorpus representing three distinct registers of medical communication:

Academic Publications: High-impact journals, such as The New England Journal of Medicine and the Journal of Medical Case Reports (JMCR), were selected to represent expert-to-expert scientific writing.

Clinical Documentation: Anonymized medical case reports (MCRs), discharge summaries, and specialist letters were included to reflect the "professional jargon" and authentic linguistic patterns used in active clinical management.

Standardized Reference Systems: The Unified Medical Language System (UMLS), SNOMED CT, and ICD-10 were used for semantic validation and concept normalization. [4]

Corpus Tools and Selection Criteria Primary processing was conducted using computational tools such as Sketch Engine and AntConc. These tools allowed for the generation of N-grams (sequences of 2 to 6 words) and concordances to observe terms in their natural "semantic environment." [4, 15]

Following the methodological model for corpus-derived medical term lists, candidates were evaluated based on several criteria: [15]

Frequency: A normalized measure of occurrences per million words (pmw) to distinguish established terms from rare occurrences.

Keyness: Identifying words with significantly higher relative frequency in the cardiological corpus compared to a general language corpus.

Dispersion: Ensuring the term appears consistently across diverse document types and authors rather than being concentrated in a single source.

Structural Modeling: Terms were categorized based on their constituent components, such as Noun + Noun or Adjective + Noun patterns.

Morphological Decomposition: Breaking complex terms into their constituent prefixes, roots, and suffixes to evaluate semantic motivation and improve learner retention.

RESULTS AND DISCUSSIONS

Analytical findings confirm that cardiological discourse is predominantly multi-component. The structural distribution of terms within the corpus highlights the necessity of phrase-level translation. [7, 13]

Structural Model Distribution and Examples Based on the analysis of the compiled subcorpus and comparison with established medical datasets, the following structural patterns were identified:

One-Component Terms: These consist of simple or derived words, often with clear Greco-Latin roots. In medical English, it is established that approximately 75% to 98% of core vocabulary is of Greek or Latin origin. [6, 7]

Examples include: aorta, ventricle, myocardium, and thrombosis. In Uzbek, these are often rendered via direct international borrowing or transliteration: aorta, ventrikula, miokard, tromboz.

Two-Component Terms: This is the most frequent model in clinical settings. Studies on medical corpora indicate that terms with two elements comprise approximately 38% of the terminological load. [7, 13]

Key examples in cardiology include:

Heart failure — yurak yetishmovchiligi (Noun+Noun).

Blood pressure — qon bosimi (Noun+Noun).

Mitral valve — mitral qopqoq (Adjective+Noun).

Cardiac arrest — yurak to'xtashi.

Three-Component and Multi-Word Terms (MWTs): These units often represent specific diagnoses or physiological measures. Research shows that while multi-word units are fewer in unique types, they carry a high functional weight, often accounting for 18% of the terminological variety in clinical texts. [15, 16]

Myocardial infarction — miokard infarkti.

Cardiac output — yurak chiqarish hajmi.

Differential diagnosis — differensial diagnoz.

Coronary artery disease — koronar arteriya kasalligi.

Acute coronary syndrome — o'tkir koronar sindrom.

Semantic Classification The terms were categorized into thematic groups based on their core semes, revealing the following distribution:

Anatomy (prevailing group): Names of internal organs and physiological processes such as left ventricle, pulmonary artery, and erythrocyte sedimentation rate. [8, 13]

Pathology/Disorders: Terms denoting diseases, such as thrombosis, fibrillation, and heart scarring.

Diagnosis/Procedures: Units related to clinical examination, including echocardiography, electrocardiogram (EKG), and coronary angioplasty.

Symptoms: Terms describing patient complaints, such as chest pain, shortness of breath (dyspnoea), and palpitations.

Translation Strategies and Asymmetry The analysis reveals a profound structural-semantic asymmetry in English-Uzbek alignment. English often utilizes compact Adjective + Noun or Noun + Noun structures that lack direct morphological equivalents in Uzbek, necessitating various translation mechanisms:

Descriptive Expansion: English terms are often expanded in Uzbek to maintain semantic transparency for laypersons. For example, infectious disease is rendered as yuqumli kasallik ("contagious disease"). [6, 11]

This is vital because studies show that nearly 30% of drugs are not taken correctly due to poor patient understanding of instructions and terminology. [6, 11]

Calque Translation: Preserving the structural logic of the source, such as airborne infection becoming havo-tomchi infeksiyasi. [6, 11]

Hybrid Borrowing: In high-tech medicine, neologisms such as stenting, bypass, and screening are often borrowed directly and adapted morphologically into Uzbek as stentlash and baypas. [4, 15]

Machine Translation and AI Risks The research evaluated the effectiveness of Neural Machine Translation (NMT) in cardiological discourse. While AI systems achieve high fluency, they frequently struggle with polysemy and MWTs. In studies involving low-resource languages, NMT systems achieve an accuracy rate of approximately 81% for unique terms, but performance drops significantly with complex multi-word units. [13]

A critical risk identified is "terminological chaos" caused by ambiguous abbreviations; for instance, the initialism "RS" can refer to four distinct syndromes (Rett, Reye, Raynaud, or Rumination) depending on the context, leading to potential clinical errors in AI-mediated translation.

Eponyms and "Onomastic Angst" Cardiology remains heavily populated by eponyms (e.g., Fallot's tetralogy, Purkinje fibers). While international bodies like the WHO advocate for descriptive standards to ensure safety, clinicians often persist in using eponyms for mnemonic convenience and historical value. In English-Uzbek translation, this creates a "double terminology" where the eponym (Fallot tetradasi) often requires a descriptive accompaniment in parentheses for the target audience. [13]

CONCLUSION

A corpus-based analysis confirms that cardiological terminology is inherently multi-word, dynamically evolving, and structurally dense. The structural dominance of two- and three-component terms necessitates a shift in translation strategy from simple word-matching toward phrase-level alignment.

For the dissertation topic, "A Corpus-Based Study of the Structural-Semantic Features of English Cardiological Terms and Their Translation into Uzbek," this study provides a robust methodological foundation. It highlights that the future of medical communication in Uzbek depends on the development of standardized, corpus-derived MWT lists that account for different registers of discourse. By aligning international Greco-Latin traditions with native linguistic resources and utilizing human expert oversight for AI-mediated translation,

researchers can significantly reduce medical misunderstandings and enhance the accuracy of professional communication in global healthcare.

REFERENCES

1. Abdullah, G. M. A. (2025). Translating Medical Terminology. British Journal of Applied Linguistics, 5(1), 60–70.
2. Abel, E. L. (2014). Medical Eponym Angst. Names, 62(2), 76–85.
3. Abrayeva, S. E. (2025). Features and Methodology of Studying Cardiological Terminology. ISSN 2181-3485, 15(09), 157–159. [4, 15]
4. Afzal, Z., et al. (2014). ContextD: an algorithm to identify contextual properties of medical terms. BMC Bioinformatics, 15:373.
5. Alshehri, A. A. (2024). Building the Medical Lexicon: A Corpus-Based Approach. Forum for Linguistic Studies, 6(6), 558–574.
6. Andreeva, M. I., & Shaekhova, R. R. (2024). Medical terminology of patients' medical cases: structural and semantic analysis. Vestnik of Samara University, 30(4), 228–234.
7. Džuganová, B. (2019). Medical Language – A Unique Linguistic Phenomenon. JAHR, 10/1(19), 129–145.
8. Ehsanzadeh, S. J., et al. (2026). Teaching medical terminology: comparing the effectiveness of methods. BMC Research Notes, 19:16.
9. Jabborova, S. A. (2026). Lexical-Semantic Features of Medical Terminology in English and Uzbek. World of Philology, 5(2).
10. Kasperė, R., et al. (2023). Medical terminology issues: machine translation in a low-resource language. SKASE Journal of Translation and Interpretation, 16(2), 5–22.
11. Lysanets, Y. V., & Bieliaieva, O. M. (2023). The use of eponyms in medical case reports. Journal of Medical Case Reports, 17:151.
12. Marciniak, M., & Mykowiecka, A. (2014). Terminology extraction from medical texts in Polish. Journal of Biomedical Semantics, 5:24.
13. Milosavljević, N., & Antić, Z. (2015). Medical English Genres - Indispensable Tool for Effective Communication. Acta Facultatis Medicae Naissensis, 32(1), 77–81.
14. Noll, R., et al. (2025). Assessing GPT and DeepL for terminology translation in the medical domain. BMC Medical Informatics and Decision Making, 25:237.
15. Shpak, Y. (2024). Structural and Semantic Aspects of Modern English Medical Terminology. Studia Neofilologiczne, 20, 133–147.
16. Vintar, Š. (2000). Using Parallel Corpora for Translation-Oriented Term Extraction. Original scientific paper, 143–152.
17. Werneck, A. L., & Batigália, F. (2011). Anatomical eponyms in Cardiology from the 60s to the 21st century. Rev Bras Cir Cardiovasc, 26(1), 98–106.
18. Zaripova, A. R. (2025). A Corpus-Based Analysis of Medical Vocabulary in Academic Journals. International Journal of Medical Sciences, 5(11), 1125–1128.