UDC 81.33
DOI https://doi.org/10.24919/2308-4863/93-2-46

Maryna TARNAVSKA,
orcid.org/0000-0002-5476-911X
Candidate of Philological Sciences,
Associate Professor at the Department of Translation, Applied and General Linguistics
Volodymyr Vynnychenko Central Ukrainian State University
(Kropyvnytsky, Ukraine) tarnavsky78@gmail.com

ENHANCING TRANSLATION PRECISION THROUGH CORPUS-BASED ANALYSIS

The article offers an extensive examination of the systematic development of translation competence grounded in corpus technologies and the analysis of terminological structures in specialized texts, highlighting this process as a multidimensional and interrelated component of professional training. The research focuses on the gradual transition from working with individual lexical items and recurrent collocations to addressing more complex translation issues involving neologisms, low-frequency units, and multi-component compounds characteristic of specialized discourse. The study emphasizes the need to redirect translators away from intuitive equivalence selection toward an analytical, evidence-based approach supported by corpus data, parallel texts, and specialized terminological resources. Particular emphasis is placed on procedural aspects that simulate real-world translation practice and highlight the importance of mastering corpus tools. These include research-oriented activities, construction of personalized CQL queries in Sketch Engine, the interpretation of concordance lines, and context-based comparison of language units followed by the justification of translation choices. The article clearly demonstrates that such procedures promote the development of critical thinking, strengthen the ability to justify translation solutions, and contribute to shaping the translator's professional identity as a linguistic researcher and language analyst. A significant component is a corpus-based project involving the creation of a mini-glossaries, which enable the integration of previously acquired skills and support the development of a comprehensive strategy for handling terminology within narrow-specialized fields. The practical value of such an activity is emphasized, particularly its role in ensuring terminological consistency, enhancing translation accuracy, and facilitating more effective work with specialized texts.

The research proves the effectiveness of a corpus-oriented approach in translator training and its capacity to foster autonomy, research competence, and evidence-based translation decisions – the features that define the standards of contemporary professional translation.

Key words: *corpus linguistics, linguistic analysis, specialized translation, terminological collocations CQL queries, translation competence.*

Марина ТАРНАВСЬКА,
orcid.org/0000-0002-5476-911X
кандидат філологічних наук,
доцент кафедри перекладу, прикладної та загальної лінгвістики
Центральноукраїнського державного університету імені Володимира Винниченка
(Кропивницький, Україна) tarnavsky78@gmail.com

КОРПУСНИЙ АНАЛІЗ ЯК ІНСТРУМЕНТ ПІДВИЩЕННЯ ТОЧНОСТІ ФАХОВОГО ПЕРЕКЛАДУ

Стаття пропонує розгорнутий аналіз поетапного формування перекладацької компетентності на основі корпусних технологій і дослідження термінологічних структур у фахових текстах, розглядаючи цей процес як багатовимірний та взаємопов'язаний компонент професійної підготовки. У центрі уваги перебуває поступовий перехід від роботи з окремими лексемами та типовими колокаціями до розв'язання складніших перекладацьких завдань, пов'язаних із неологізмами, рідковживаними одиницями та багатокомпонентними конструкціями, характерними для фахового дискурсу. Обґрунтовується доцільність переорієнтації перекладача від інтуїтивного добору відповідників до аналітичного, доказового підходу, що спирається на корпусні дані, паралельні тексти та спеціалізовані термінологічні джерела. Особливу увагу приділено практичним аспектам роботи, які моделюють реальні умови перекладацької діяльності та актуалізують необхідність методичного оволодіння корпусними інструментами. Серед них – пошуково-аналітична діяльність, конструювання персоналізованих CQL-запитів у Sketch Engine, інтерпретація конкордансів і контекстуальне зіставлення мовних одиниць з подальшим обґрунтуванням вибору перекладацького рішення. Висвітлено, яким чином такі інструменти сприяють розвиткові критичного мислення, формують здатність до аргументованого ухвалення рішень і підтримують становлення професійної ідентичності перекладача як дослідника мови та лінгвіста-аналітика. Важливою складовою такого процесу є створення глосаріїв на основі корпусних даних, які забезпечують інтеграцію раніше

набутих умінь і сприяють формуванню комплексної стратегії роботи з термінологією у вузькоспеціалізованих сферах. Підкреслюється практична значущість такого продукту як засобу підтримання термінологічної узгодженості, підвищення точності перекладу та оптимізації подальшої роботи зі спеціалізованими текстами.

Дослідження демонструє ефективність корпусоорієнтованого підходу в підготовці перекладачів та його потенціал у формуванні автономності, дослідницьких навичок і доказовості перекладацьких рішень – характеристик, що визначають рівень сучасного професійного перекладу.

Ключові слова: *корпусна лінгвістика, лінгвістичний аналіз, фаховий переклад, термінологічні колокації, CQL-запити, перекладацька компетентність.*

The Issue at Hand. The key problem lies in the complexity of working with specialized texts, which requires not only expert knowledge of both languages but also understanding of the guidelines that shape professional terminology, structure, and grammar. In this context, introducing corpus methods raises a natural question: can corpus analysis make an already demanding process even more complicated? The answer depends greatly on how corpus tools are incorporated. While experienced users can quickly benefit from corpus data to refine terminology and resolve lexical uncertainties (Akkoyunlu & Kilimci, 2017: 370–371), translators who are unfamiliar with these tools may struggle without having a clearly organized multi-stage procedure. For this reason, a structured and gradual systematic framework is essential – one that views corpus analysis as a practical aid in solving specialized texts translation problems rather than as an additional burden.

Latest Research Analysis. Current research shows a growing interest in using corpus linguistics to analyse specialized discourse and support translation. The works by scholars, such as Norbert Schmitt and Stefan Th. Gries, have shown the value of frequency data, collocational patterns, and distributional behaviour for understanding specialized vocabulary (Gries, 2021). In translation-focused studies, researchers, including Abdurrahman Kilimci, Aslı Nur Akkoyunlu, Ying Lyu, and Ziman Han, have observed improvements in translators' accuracy and decision-making when corpus-based techniques were used (Lyu & Han, 2023; Akkoyunlu & Kilimci, 2017). Additional studies on data-driven learning by Amel Lusta, Özcan Demirel, and Behbood Mohammadzadeh confirm that corpus methods can enhance consistency and provide evidence for making right translation choices (Lusta et al., 2023). Ukrainian researchers – T. Anokhina, V. Babych, I. Kobyakova, N. Lemish, S. Matvieieva, S. Shvachko, and A. Zernetska – have also contributed to expanding corpus approaches in linguistic and translation research (Anokhina et al., 2023; Matvieieva et al., 2022; Lemish et al., 2020). Together, these works point to a clear trend toward data-driven methods that help translators base their decisions on authentic language use rather than on intuition alone.

The article aims at presenting a practical model for integrating corpus-based procedures into the overall process of specialized translation. It outlines a multi-stage system that helps translators move from basic lexical analysis to solving complex terminological issues using evidence from real texts. By examining how each stage functions in practice, the article highlights the strategies and tools that make corpus data useful for refining terminology, checking patterns, and supporting well-reasoned translation decisions. The broader goal is to offer a coherent procedural structure that links corpus evidence with professional translation tasks and can be adapted to different specialized fields.

The main body of the article. The systematic progression from mastering individual terms to analysing terminological collocations provides translators with a comprehensive framework for approaching the lexical core of specialized texts. Nevertheless, the decisive test of translation competence frequently occurs at the lexical periphery, where the translator deals with the elements that are void of straightforward one-to-one equivalents. Consequently, the next analytically structured stage within this corpus-supported translation process focuses on the examination and resolution of complex terminological problems. These might involve low-frequency items, neologisms, compound terms as well as hyphenated compounds. The primary objective of this phase extends far beyond the immediate task of finding an equivalent; it aims to adjust how conscious translators perceive lexical challenges. The goal is to encourage them to reconsider an unfamiliar or problematic term not as an impassable obstacle but as a solvable professional problem that requires a systematic, evidence-based approach (Pietrzak & Kornacki, 2021).

To foster this cognitive shift, the analytical process with problem-oriented tasks might be designed to reflect authentic translation challenges. The central activity at this stage is problem-solving translation when translators work with genuine specialized texts excerpts intentionally selected to contain complex terminology such as field-related neologisms, rare terminological expressions, or multi-element hyphenated lexical units. The essence of the tasks is not to guess the correct equivalent but to conduct systematic

research. Using a carefully selected set of resources, including pre-compiled specialized texts corpora in Sketch Engine for investigating contextual patterns, as well as parallel specialized texts to identify established terminological usage, and specialized online dictionaries or glossaries, translators are expected to research independently or collaboratively and propose a translation equivalent supported by explicit evidence. This approach transforms translation from a passive reception of meanings into an active, analytic, corpus-supported procedure (Lu & Xu, 2023: 241–243). The problem-solving tasks are followed by an essential comparative multiple discussion. This stage functions as the consolidation point for more advanced analytical reasoning, when translators present and justify their choices, providing background behind each choice in reference to the empirical data collected. The collaborative evaluation examines why one translational option succeeds while another may not. The analysis focuses on register suitability: whether the proposed term aligns with the norms and stylistic expectations of specialized discourse. It also considers semantic precision: whether the chosen equivalent accurately reflects the conceptual scope of the source unit. In a much similar way, it assesses the degree of naturalness: whether the translation sounds authentic in the target language or reads as a literal calque. This type of comparative, data-supported discussion enhances collective critical awareness and strengthens translators' ability to articulate and defend their decisions, which is one of the essential components of professional translator competence (Akkoyunlu & Kilimci, 2017: 386).

The analytical value of this stage is threefold. First, it systematically questions the idea of «untranslatability», illustrating that even the most vague term can be decoded using appropriate corpus tools and strategies, which significantly reduces hesitation while dealing with complex units. Second, it contributes to the formation of a resilient, resourceful professional identity: the translator assumes the role of a linguistic investigator rather than a transmitter of established notions. Third, and most importantly, it reinforces the principle of justified, evidence-based translation. By grounding the decisions in corpus data and verified parallel sources instead of unexamined intuition, translators learn to produce choices that are not only linguistically accurate but professionally validated. This marks a shift from a beginner's reliance on a single «correct» dictionary entry to an expert capacity to manage ambiguity and construct contextually appropriate solutions.

Building on this foundation, the translation mastery process shifts to the practical application of Sketch Engine tools for solving real-world translation problems independently. The primary aim of this stage is to provide the transition from gathering pre-selected examples to independently exploring linguistic patterns in corpora. Conceptually, this phase functions as a corpus-analysis stage within translation studies and demonstrates how translators use corpus evidence to inform decision-making. This work begins with practical demonstrations of constructing basic CQL-queries to locate linguistic structures that frequently complicate the process of translation. One example is the query «[tag="N.*"]{}», which retrieves sequences of two nouns and is especially beneficial for identifying terminological noun clusters characteristic of specialized discourse. These findings are not abstract descriptions, each step including actual corpus results accompanied by brief analysis and allowing translators to see how query construction directly contributes to understanding usage patterns relevant to translation.

The next and the most crucial step involves independent corpus-based inquiries. Translators are given a problematic phrase or term whose translation raises uncertainty. Their task is to formulate and carry out an appropriate corpus query, examine the retrieved concordance lines, compare usage across different contexts, and propose an equivalent in the target language that is both semantically precise and stylistically adequate. This process strengthens critical reasoning, enhances familiarity with corpus tools, and transforms the translator being a passive receiver of information into an active researcher capable of validating hypotheses (Yan & Wang, 2022: 8). The value of this stage lies in developing the translator's sustained ability to independently resolve complex lexical issues. Instead of depending on pre-existing dictionary entries or external authority, translators acquire the whole set of tools enabling them to verify assumptions through authentic data. Fundamentally, this training is not just about mastering query syntax; it establishes a professional working method based on linguistic evidence rather than intuition. This approach builds confidence and equips translators to handle future challenges autonomously.

On dealing with the analytical and practical corpus-based modules, the workflow comes to a comprehensive project designed to synthesize all previously developed competencies through collaborative, corpus-supported research. This final stage, which can be defined as an extensive group-based task, marks the transition from guided exercises to independent project-oriented application, thereby consolidating professional skills and preparing translators for real-world practices. The main objective is for translators

to create a specialized mini-glossary on a selected topic, encouraging deeper engagement with the lexical features of a given specialized field and moving from understanding toward the production of structured terminological knowledge. The project is developed as a multi-stage collaborative activity within the groups. Each group might be assigned a narrow, specific topic with the focused scope ensuring depth and requiring each group to operate as subject-matter specialists, reflecting the division of tasks common in professional localization environments. The structural core of the project lies in its research encouraging procedure. Rather than receiving ready-made data, each group uses all previously acquired tools: frequency lists, collocational profiles, and especially independent iterative queries conducted in Sketch Engine. These steps are indispensable for building up glossary entries that go beyond simple «term-equivalent» correspondence and include corpus-based illustrative sentences, translator's notes on style and usage, as well as observations on grammatical peculiarities derived from actual corpus evidence.

The final product of this multi-session research is a consolidated, collaboratively developed glossary. It serves not merely as an exercise but as a practical compilation of corpus-verified terminology. The broader value of this phase is significant, as it has high long-term benefits. First, it reinforces computational research skills through repeated use of corpus tools to solve complex terminological problems. Second, it develops essential collaborative and organizational skills typical of professional translation and localization processes. Third, and most important, it results in a corpus-based resource, a glossary grounded in empirical evidence, which can be of multiple use and translators may refer to that in their future professional work. This complex collaborative corpus product ensures that the acquired knowledge becomes a lasting, functional tool to support translation accuracy, efficiency, and terminological consistency (Pietrzak & Kornacki, 2021).

And just as important is that the entire process, from the initial core lexicon to the collaborative glossary, is validated through a structured assessment system designed not to test rote memorization but to evaluate translators' ability to function as competent, analytical professionals. The assessment might include a fixed benchmark component that tests recognition of key terms and collocations and the ability to provide accurate equivalents in the target language. In spite of its foundational character, this stage confirms mastery of the essential lexical core usage that enables faster analysis and understanding of specialized texts to be translated.

Conclusions and further research prospects. In conclusion, the multi-stage corpus-supported system presented here demonstrates its effectiveness in fostering a data-driven, analytically grounded model of translator competence, particularly in addressing complex and low-frequency lexical items in specialized texts. It ensures that the students shift from intuition-based choices to evidence-supported decision-making and go on to create a corpus-verified glossary that, apart from its skills-improving value, could function as a durable professional resource. As for future prospects, the same structured approach could be applied to various specialized fields, such as financial, medical, or technical translation with a view to determining whether its efficiency varies across different domains and what field-specific adjustments might be required. Further research could also compare fruitfulness of the procedure applied to different corpus platforms or investigate the development or retaining of the analytical translation skills acquired through this approach. Ultimately, extending this model to collaborative, project-based environments, such as complete student-built corpora of various specialized fields, can help to assess how corpus-driven training can be scaled and adapted to authentic professional workflows.

BIBLIOGRAPHY

1. Akkoyunlu A., Kilimci A. Application of Corpus to Translation Teaching: Practice and Perceptions. International Online Journal of Education and Teaching. 2017. Vol. 4. pp. 369–396.

2. Gries S. T. Analyzing Linguistic Data: A Practical Introduction to Statistics Using R (2nd ed.). Cambridge University Press. 2021. 374 pages.

3. Lyu Y., Han Z. Applying data-driven learning in self-translation of academic discourse: A case study of a Chinese medical student. Frontiers in Psychology. 2023. 14:1071123. pp. 1–10. https://doi: 10.3389/fpsyg.2023.1071123

4. Lusta A., Demirel Ö., Mohammadzadeh B. Language Corpus and Data Driven Learning (DDL) in Language Classrooms: A Systematic Review. Heliyon. 2023. Vol. 9. e22731. 10.1016/j.heliyon.2023.e22731.

5. Anokhina T., Kobyakova I., Schvachko S. Innovative Methodology for Teaching European Studies Using a Corpus Approach. Philological Treatises. 2023. Vol. 15. No. 2. pp. 7–16.

6. Matvieieva S. A., Lemish N. Ye., Zernetska A. A., Babych V. I., Torgovets M. S. English-Ukrainian Parallel Corpus: Prerequisites for Building and Practical Use in Translation Studies. Studies about Languages. 2022. Vol. 1. pp. 61–74.

7. Lemish N. Ye., Aleksieieva O. M., Denysova S. P., Matvieieva S. A., Zernetska A. A. Linguistic Corpora Technology as a Didactic Tool in Training Future Translators. Information Technologies and Learning Tools. 2020. Vol. 79. No. 5. pp. 242–259.

8. Pietrzak P., Kornacki M. Focus on the Translation Trainee. Research in Language. 2021. 19(2). pp. 107–115. https://doi.org/10.18778/1731-7533.19.2.01

9. Lu J., Xu Y. Practical Research on Corpus-based Translation Teaching Modes of College English. International Journal of Education and Humanities. 2023. 8(2). pp. 241–245. https://doi.org/10.54097/ijeh.v8i2.7822

10. Yan D., Wang J. Teaching data science to undergraduate translation tragroupinees: Pilot evaluation of a task-based course. Frontiers in Psychology. 2022. 13:939689. pp. 1–18. https://doi: 10.3389/fpsyg.2022.939689

REFERENCES

1. Akkoyunlu A., Kilimci A. (2017). Application of corpus to translation teaching: Practice and perceptions. International Online Journal of Education and Teaching. 4. 369–396.

2. Gries S. T. (2021). Analyzing linguistic data: A practical introduction to statistics using R (2nd ed.). Cambridge University Press.

3. Lyu Y., Han Z. (2023). Applying data-driven learning in self-translation of academic discourse: A case study of a Chinese medical student. Frontiers in Psychology. 14. 1071123. https://doi.org/10.3389/fpsyg.2023.1071123

4. Lusta A., Demirel Ö., Mohammadzadeh B. (2023). Language corpus and data driven learning (DDL) in language classrooms: A systematic review. Heliyon. 9. e22731. 1–10. https://doi.org/10.1016/j.heliyon.2023.e22731

5. Anokhina T., Kobyakova I., Schvachko S. (2023). Innovative methodology for teaching European studies using a corpus approach. Philological Treatises. 15(2). 7–16.

6. Matvieieva S. A., Lemish N. Ye., Zernetska A. A., Babych V. I., Torgovets M. S. (2022). English-Ukrainian parallel corpus: Prerequisites for building and practical use in translation studies. Studies about Languages. 1. 61–74.

7. Lemish N. Ye., Aleksieieva O. M., Denysova S. P., Matvieieva S. A., Zernetska A. A. (2020). Linguistic corpora technology as a didactic tool in training future translators. Information Technologies and Learning Tools. 79(5). 242–259.

8. Pietrzak P., Kornacki M. (2021). Focus on the translation trainee. Research in Language. 19 (2). 107–115. https://doi.org/10.18778/1731-7533.19.2.01

9. Lu J., Xu Y. (2023). Practical research on corpus-based translation teaching modes of college English. International Journal of Education and Humanities. 8(2). 241–245. https://doi.org/10.54097/ijeh.v8i2.7822

10. Yan D., Wang J. (2022). Teaching data science to undergraduate translation trainees: Pilot evaluation of a task-based course. Frontiers in Psychology. 13. 939689. 1–18. https://doi.org/10.3389/fpsyg.2022.939689

Дата першого надходження рукопису до видання: 27.11.2025
Дата прийнятого до друку рукопису після рецензування: 10.12.2025
Дата публікації: 30.12.2025