

TERM-COMBINATION IN THE CONTEXT OF CORPUS-APPLIED TRANSLATION STUDIES

Yuliya Demyanchuk

Ph.D., Assistant Professor, Lviv State University of Life Safety, Ukraine

e-mail: y.demianchuk@gmail.com, orcid.org/0000-0001-8722-5568

Summary

This research presents an investigation of term-combination as a corpus-applied translation phenomenon, which constitutes a distinct branch of applied linguistics, merging corpus analysis methods with the study of fundamental translation doctrines. The aim of the study was to examine the theoretical postulates of collocational analysis and paradigmatic transformation of term-combinations in corpus-applied translation studies, determine the quantitative feature of reproducing term-combinations in the corpus of texts, characterize the corpus feature of contextual adequacy, and demonstrate the corpus feature of grammatical concord. The research results demonstrate that the reinforcement of logical determinism of the linguistic phenomenon characterizes an integrated hierarchical approach to the classification of official-business texts, in which collocational relationships between linguistic units are traced, contributing to the expansion of the sphere of terminological combination realization in context. Overall, official-business texts fulfill a dual role: informative and suggestive. The realization of these functions can be both implicit, encoded in the internal form of term-combinations, and explicitly expressed. The pragmatic orientation of official-business texts manifests through corpus-applied translation means. Official-business documents construct the official-business discourse, which we understand as a specific environment of linguistic communication aimed at the realization of lexical-semantic, derivational, grammatical, and stylistic levels. The system of approaches applied in our research constitutes a coherent corpus-applied translation formation, structuring the manner of organization of translation methods directed at the analysis of official-business terminology.

Key words: term-combination, corpus-applied translation studies, text corpus, features of terminological combination reproduction.

DOI <https://doi.org/10.23856/5902>

1. Introduction

The history of studying the concept of “term-combination” as a linguistic and later corpus concept has been long-standing. Until recently, in the most general sense, a term-combination was understood as a “linguistic construction consisting of two or more terms combined together to denote a specific concept, object, or phenomenon” (*Masamitsu, 2005*). A term-combination has its own properties that distinguish it from simple collocations. First and foremost, a term-combination has its specificity in linguistic studies, linguistic experiments, language teaching platforms, and language localization. Hence arises the two-tier ontological nature of a term-combination: it is a stably constructed and structurally fixed monolithic construction. The specification of a term-combination is determined by syntactic norms of its usage, specific grammatical structure, and pragmatic relationships between morphological units and categories. The formation of the concept of “term-combination” can be traced back to the emergence of specialized terminology in various domains of science. The notion of the first patterns of

term-combinations, according to S. Seljan & E. Gašpar (2012), appeared in ancient Greek texts, such as the philosophical works of Aristotle. Such word combinations helped to accurately convey complex concepts and ideas. In the medieval period, term-combinations gained popularity, especially in scientific texts, medicine, jurisprudence, and religious texts.

In the context of European epistemology (19th–20th century), the development of science and technology, the expansion of scientific knowledge, and the emergence of new disciplines like physics, chemistry, and biology led to the active formation of non-trivial term-combinations to describe new concepts, processes, and phenomena. As a result of the broadening of the scope of permanent terminological transformation in philological studies, the investigated phenomenon was officially designated as “term-combination” (Felbaum, 2015).

2. Methodology

The methodology of investigating term-combinations as a corpus-applied translation phenomenon in this article is based on a comprehensive approach that combines the analysis of paradigmatic transformation, quantitative assessment of reproduction, contextual adequacy, and grammatical concordance of term-combinations in the corpus of texts.

The methodological perspective of the research encompasses the following stages:

The first stage of the research involves collocational analysis of the term-combination, which entails understanding typical word or phrase combinations for the purpose of speech synthesis. In our study, special attention is given to the analysis of paradigmatic transformation of term-combinations within the scope of its formal change. Typically, this involves changes in number, gender, case, or substitution of one term with another that carries an identical semantic meaning. An indicator of paradigmatic transformation also includes the objectification of the term-combination in the context of official-business texts, which reveals its potential, as well as causal relationships and changes in pertinent meanings.

The concluding stage of the research involves identifying corpus-based features of reproduction of term-combinations in corpus-applied translation studies. The utilization of these features contributes to emphasizing the logical determinism of the linguistic phenomenon and ensures the consistency and precision of translating epistemological terminology.

3. Results and discussion

The concentrated theoretical ideas of modern study of term-combinations have found manifestation in the work “General problems of translation of phraseological units” by Uzbek researcher Shakhlo Mukhtorivna Yusufjanova (2022). According to the scholar’s understanding, term-combinations serve as tools for determining concepts and anthropocentrism of speech; the expressive aspect of term-combinations encompasses not only the lexemes within them but also morphemes, and even inflections.

One of the peculiarities of term-combinations is their specific semantics, which can differ from the literal meanings of individual words. Terms combined in expressions often interact and complement each other, forming a new concept. In Ch. Fellbaum’s interpretation, the dominant and thus the cross-cultural perception of term-combinations serves as the foundation for comprehending a particular scientific field, which unifies its terminological apparatus through the creation of systematic and consistent terminology (2019). This trend has revealed the complexity of the axiological concept of “term-combination” and expanded the heuristic possibilities of its study.

In the context of corpus, linguistic, and paradigmatic transformations, the term-combination is actualized based on the principles of corpus-applied translation studies, while preserving its unambiguous and standardized lexical meaning. Therefore, the integrity of the term-combination is ensured through the combination of its formalized and intersubjective identifying features. Corpus linguistics and linguistic transformations influence the development and application of term-combinations in corpus-applied translation studies.

The differentiation in modern science between simple word combinations and term-combinations as different aspects of specific and combined meaning branches out discrete characteristics of lexical collocation and term-combination. This correlation within the system of lexical collocation/term-combination characterizes the corresponding function of the term-combination in corpus-applied translation studies. The foundation of contemporary corpus-applied principles, arising under the influence of corpus-applied translation studies, is laid upon teleological universals – term-combinations.

Considering that the term-combination is a dominant entity in the source and target languages (*Seljan & Gašpar, 2012*), in our view, it is the pragmatic perspective of translation studies that has led to a rational division of methodological principles into relevant clusters:

Collocational analysis of term-combinations in corpus-applied translation studies is an essential tool for studying and understanding specialized vocabulary in the context of translation. Collocations are stable linguistic associations between words or word combinations that often occur together in texts and have specific meanings (*Jaworska, 2017*).

In corpus-applied translation studies, collocational analysis of term-combinations is used to identify, reproduce, and investigate typical word combinations or phrases in the process of speech synthesis identification. Such analysis involves context beyond a single system since the act of speech is materialized through modeling collocational relationships within the text. In our research, for conducting collocational analysis of term-combinations in corpus-applied translation studies, both blocks of factors are important: the use of a text corpus that aggregates a vast amount of textual material from various sources, such as scientific articles, official-business documents, translations, generating collocations and the expansion of the context realization sphere within which term-combinations emerge.

The selectivity and structured nature of collocational analysis of term-combinations are reflected in its speech representation, in the identification of close connections between terms and common lexicon encountered in specialized texts. The text segmentation aimed at highlighting marked official-business discourse units also has an epistemological basis: it facilitates the perception of context, the expression of precise lexical-semantic meaning of terminological concepts, and focuses attention on statistical collocation extraction methods, linguistic models of term-combinations, and natural language processing tools, which actualize the concept of term-combinations in corpus-applied translation studies.

Paradigmatic transformation of term-combinations in corpus-applied translation studies refers to the process of changing or varying the usage of terms or term-combinations in a specific context based on corpus analysis. This process typically aims to enhance the formalization and structuring of translation. By analyzing texts from official-business discourse, researchers can identify various patterns of term or term-combination usage in different contexts, as well as their combinations with other words.

The most significant aspect of paradigmatic transformation of term-combinations lies within the change of their form. Variants of usage often demonstrate changes in the forms of terms, such as numbers, genders, cases, or the substitution of one term with another that has an identical subject-meaning. Objectification of term-combinations within the context

of official-business texts reveals their cause-and-effect potential, which manifests in the alteration of nuances in their current meaning. Thus, a term-combination may have several synonymous variants, each carrying its own shades of meaning. A characteristic feature of paradigmatic transformation is its pragmatic orientation towards the adequate reflection of translation variants. It usually involves breaking down the term-combination into separate words or altering the word order to better reproduce the represented expression in the target language.

Corpus features of reproducing term-combinations in corpus-applied translation studies. The comprehension of translation in the field of computational language processing starts at an algorithmic level. As crucial components of algorithmic models, we consider software products and databases, which serve as efficient archivers among other translator's tools.

Condensed within corpus linguistics, text corpora are designed to address various tasks, involving the extraction of inherent characteristics of linguistic constructions for qualitative linguistic analysis (*Bednarek, 2021*). In a more generalized sense, the classification distribution of text corpora into bilingual and multilingual (*Fantinuoli, 2022*), representing numerous original texts with corresponding translations, allows for the interpretation of their linguistic content. It's important to also consider corpora where all languages are equivalent categories (*Pystejovsky, 2004*): 1) corpora created based on official-business texts, such as those from international organizations like NATO, UN, WTO, or the European Community; 2) pseudo-parallel or comparable corpora gathered under similar criteria, combining texts from the same thematic group written in two or more languages, comprising original and translated texts.

The capabilities of corpus linguistics in operational processes, such as forming text corpora, studying linguistic content within the corpus, revealing mechanisms of language interaction within contextual content reflection, contribute to linguistic processing of this content reflection. It's in this way that the identification of term-combinations belonging to words with a particular function of pragmatic realism transmits information about official-business texts of international organizations.

The initiation of studies on the extraction of term-combinations from official-business texts was spurred by the research efforts of various scholars in different domains. S. Picciotto's work on financial terminology (*2011*), I. Simonnæs's investigation into legal terminology (*2019*), S. Flusberg, T. Matlock & P. Tibodeau's study of war metaphors in public discourse (*2018*), J. Griffin's analysis of UN terminology in the field of human rights (*2008*), and W. Goode's work on trade terminology (*2008*) collectively delved into the meaning and attributes of attributive terms. These attributive terms, distinct from common lexicon, are organized, interconnected, and complexly structured determiners.

Renowned theorist of official-business terminology, Thierry Fontenelle, believes that linguistic advancements contribute to the expansion of corpus studies in the direction of collocations and term-combinations. Fontenelle introduced the terms "lexical" and "grammatical collocation," reflecting the interdependence of quantum states between two different linguistic constructions (*1994*). The collaborative work of K. R. McKeown & D. R. Radev in "Collocations: A handbook of natural language processing" (*2000*) elucidates crucial linguistic components used for various natural language processing tasks, encompassing natural language generation, machine translation, information retrieval, sentiment analysis, and language learning. The researchers conclude that computational systems will influence collocation search, involving "three dimensions: frequency, mutual information of two random variables, and χ^2 test" (*McKeown, Radev 2000*). Following the modern scientific paradigm of corpus linguistics,

“an automatic system for extracting and investigating collocations forms its types of term-combinations, including those of official-business nature” (Ibid.).

Despite the application of statistical and automated methods for extracting term-combinations from collocations (such as Sketch Engine, association methods, strength of association criteria), these methods primarily characterize high-frequency collocations that serve as an intermediate level between single terms and idiomatic expressions. However, within this scope, certain lexical variants are still utilized, expanding the capabilities of the statistical approach and enhancing the likelihood of collocational relationships based on given keywords. In some cases, automated systems may conflate general vocabulary with term-combinations, which can impact the communicative accuracy of specialized texts.

The adequate selection of optimal features for representing term-combinations in corpus-applied translation studies can be attributed to traditional linguistic analyses and linguistic studies. Within the framework of the traditional approach to reproducing term-combinations, the following features are employed: structural integrity, which does not involve separate components within the term-combination; lexical precision; stylistic unification; and adequate syntactic structure of reproducing the term-combination in context.

In corpus-applied translation studies, terminologization, which involves enriching the terminological vocabulary by extracting linguistic units from a text corpus, prompts the choice of optimal features for reproducing term-combinations in translation.

Quantitative features for reproducing term-combinations within a text corpus also illustrate the use of mathematical modeling methods (characteristic method, principal component analysis), statistical association measures (log-likelihood, MI, t-score), and the TF-IDF measure (product of two factors) reflecting the integration aspect in the process of extracting term-combinations from a set of collocations. This results in a constant multiplier applied to each word, logical contextual activation through the KWIC format, which assists in developing a program for reading a text file and an algorithm for number extraction within a string, and the utilization of software tools like Sketch Engine and Orange, which play a crucial role in the high-quality corpus analysis of official-business texts from international organizations.

Taking into account the representativeness of official-business texts in both the source and target languages, the frequency of occurrence of a term-combination in context allows for its pluralization in translation. The quantitative feature ensures consistency of the analyzed phenomenon and contributes to the precision of translating epistemological terminology.

The **corpus feature of contextual adequacy** involves contextual analysis in which the term-combination is used. The linguistic format (linguistic representation) of an official-business text is oriented towards emphasizing the importance of reproducing the term-combination in various contexts. This is done to understand the semantic nuances of language elements and to find an appropriate translation that takes these patterns into account. Representing term-combinations in official-business texts through the lens of corpus-applied translation studies has its specificity.

Nominalist models of designations of international organizations, *well-known names of secretaries and political figures, military strategic objects, combat zones, military operations, international capital movement, trade activities, resolutions of the UN General Assembly, international migration* form the basis for corpus research into modern terminology. This feature is based on the idea that achieving contextual adequacy requires considering an extensive context, including the context in the source language, the context in the recipient language, and the reproduction of cultural specificity in both languages.

Therefore, when selecting a term-combination to work with in a text corpus, we have in mind databases that enhance the reception of contextual elements in both the source and target languages, represent intertextual relationships, and demonstrate contextual adequacy in corpus-applied translation studies.

The enrichment of the conceptual framework of corpus studies enhances the overall linguistic format of the text corpus related to the analysis of official-business issues. In the context of official-business discourse, the emphasis on corpus terminological parameters is particularly prominent. Specifically, the *corpus feature of grammatical agreement* demonstrates evaluative models, grammatical guidelines, and priorities for grammatical correspondence between the original and translation in corpus-applied translation studies. The essence of the grammatical agreement feature, as a corpus-determined phenomenon that is formed and manifested in official-business texts, takes into account the grammatical correctness of translation and is an important aspect of its quality. In this regard, grammatical norms and rules of language, both in the original and the target language, appeal to precedent nominalizations.

The influence of grammatical forms and syntactic structures is highly persistent, which is reflected in a deep understanding of syntactic peculiarities in both the source and target languages. Additionally, the use of grammatically balanced phrases, functional invariable parts of speech, prepositions, pronouns, and other grammatical elements is an important aspect of grammatical agreement of term-combinations in the text corpus. Demonstrating adequate grammatical constructions contributes to preserving the integral structure and semantics of the text during translation.

In the context of our research, the corpus feature of grammatical agreement enables the description of the usage of tenses, verb forms, numerals, and other grammatical categories that ensure accurate information transmission. The linguistic approach directs its toolkit towards analyzing terminological combinations in official-business texts. The research value of the grammatical agreement feature lies in reproducing complex grammatical structures in the source language, through the corpus of texts, reflecting similar cases in the target language and substantiating the choice of an appropriate translation.

Thus, analyzing a corpus of official-business texts using an integrated hierarchical approach to document classification, utilizing the quantitative feature of term-combination reproduction in the text corpus, the feature of contextual adequacy, and the corpus feature of grammatical agreement becomes a significant trigger for quality research of official-business texts. The use of these features emphasizes the logical determinism of the linguistic phenomenon and ensures consistency and precision in translating epistemological terminology.

4. Conclusions

Therefore, the influence of the text corpus, linguistic transformations, and translation pragmaticism on the analysis and reproduction of term-combinations in corpus-applied translation studies underscores the dominant importance of term-combinations in the context of the functional orientation of the source and target languages. The new system of knowledge in collocation analysis of term-combinations initiates research that offers a theoretical perspective on an important toolkit for studying and understanding specialized vocabulary in the context of translation. Techniques and methods of reproducing and researching typical word combinations or phrases in the process of speech synthesis, involving context and using the text corpus in corpus-applied translation studies, contribute to forming collocational relationships and expanding the scope of context realization in which term-combinations emerge. At the same time, the

procedure of outlining paradigmatic transformation of term-combinations in corpus-applied translation studies tends toward changing and varying the way term-combinations are identified in the context of official-business texts based on corpus data analysis. In this regard, we interpret the component complex of this phenomenon as a formalized and structured unity of two or more pragmatically related terminological units.

References

1. Aston, G. & Burnard L. (1998). *The BNC handbook: exploring the British National Corpus with SARA*. p. 256. Edinburgh: Edinburgh University Press.
2. Bednarek, M. (2021). *Using corpus linguistics to study indexicality in Indigenous-authored television drama: Keyword analysis and lexical profiling*. p. 1–21. *Discourse, Dialogue and Characterisation in TV Series*. Granada: Editorial Comares.
3. Fantinuoli, C. (2022). *Specialized Corpora from the Web and Term Extraction for Simultaneous Interpreters*. *Working Papers on the Web as Corpus*. p. 173–190.
4. Fellbaum Ch. (2019). *How flexible are idioms? A corpus-based study*. *Linguistics*, Vol. 57, No. 4, p. 735–767.
5. Fellbaum, Ch. (2015). *The treatment of multi-word units*. *Oxford Handbook of Lexicography*. p. 411–425. Oxford: Oxford University Press.
6. Flusberg, S., Matlock T. & Thibodeau P. (2018). *War metaphors in public discourse*. *Metaphor and Symbol*. Vol. 33. p. 1–18. 10.1080/10926488.2018.1407992.
7. Fontenelle, T. (1994). *What on earth are collocations? English today: the international review*. Vol. 10 (40), No. 4. p. 42–48.
8. Goode, W. (2007). *Dictionary of trade policy terms*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511910050.
9. Griffin, J. (2008). *On human rights*. Oxford: Oxford University Press.
10. Jaworska, S. (2017). *Corpora and corpus linguistics approaches to studying business language*. *Handbooks of applied linguistics*. p. 583–606. Berlin: De Gruyter. ISBN 9781614514862
11. McKeown, K. R. & Radev D. R. (2000). *Collocations. A handbook of natural language processing*. p. 507–523.
12. Masamitsu, I. (2005). *Quantitative linguistics in Japan*. p. 82–95. Berlin–New York: de Gruyter.
13. Picciotto, S. (2011). *Regulating global corporate capitalism*. Cambridge: Cambridge University Press. DOI:10.1017/CBO9780511792625.
14. Pystejovsky, J. (2004). *Lexical semantic techniques of corpus analysis brandies*. p. 413. Oxford: Oxford University Press.
15. Seljan, S. & Gašpar, A. (2012). *First steps in term and collocation extraction from English-Croatian corpus*. *Computational language analysis, computer-assisted translation and e-language learning*. p. 149–156. Zagreb: Zavod za informacijske studije.
16. Shakhlo, M. Y. (2022). *General Problems of translation of Phraseological Units*. *Current research journal of philological science*, Vol. 3, No. 02, p. 30–35. <https://doi.org/10.37547/philological-crjps-03-02-07>
17. Simonncæs, I. & Kristiansen M. (2019). *Legal Translation: Current issues and challenges in research, methods and applications*. p. 410. Berlin.