

MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
KYIV NATIONAL LINGUISTIC UNIVERSITY

FACULTY OF GERMANIC PHILOLOGY AND TRANSLATION

Department of Theory and Practice of Translation from the English Language

TERM PAPER

IN TRANSLATION STUDIES

**USING THE OPUS TEXT CORPUS TO SEARCH FOR TERMINOLOGICAL
EQUIVALENTS WHEN TRANSLATING TEXTS FROM ENGLISH INTO
UKRAINIAN**

Voitovych Karina

ІІа 04-20

Educational Programme:

English and a Second

Foreign Language:

Oral and Written Translation

Majoring 035 Philology

Research supervisor:

T. O. Anokhina

Kyiv 2024

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Київський національний лінгвістичний університет
Факультет германської філології і перекладу
Кафедра теорії і практики перекладу з
англійської мови

Представлено на кафедру _____
(дата, підпис секретаря кафедри)

Рецензування _____
(кількість балів, «до захисту» («на
доопрацювання»), дата, підпис керівника курсової
роботи)

Захист _____
(кількість балів, дата, підпис викладача)

Підсумкова оцінка _____
(кількість балів, оцінка за 4-х бальною системою,
дата, підпис викладача)

КУРСОВА РОБОТА

з перекладу

ВИКОРИСТАННЯ КОРПУСУ ТЕКСТІВ OPUS ДЛЯ ПОШУКУ
ТЕРМІНОЛОГІЧНИХ ВІДПОВІДНИКІВ ПІД ЧАС ПЕРЕКЛАДУ ТЕКСТІВ З
АНГЛІЙСЬКОЇ НА УКРАЇНСЬКУ МОВУ

Войтович Каріна

Студентка групи Па 04-20

Керівник курсової роботи _____
(підпис)

професор кафедри теорії та практики
перекладу з англійської мови Анохіна
Тетяна Олександрівна

Київ – 2024

CONTENTS

INTRODUCTION.....	1
CHAPTER 1.....	4
KEY ASPECTS OF USING A CORPUS OF TEXTS FOR SEARCHING TERMINOLOGICAL EQUIVALENTS WHILE TRANSLATING TEXTS	4
1.1 Fundamentals of corpus linguistics, the concept of a corpus of texts and their classification.....	4
1.2 Features of the translation of scientific and technical terminology from English into Ukrainian.....	9
1.3 General description of the OPUS corpus toolkit and their resource capabilities	13
CHAPTER 2.....	18
TECHNICAL ASPECTS OF USING A OPUS CORPUS FOR SEARCHING TERMINOLOGICAL EQUIVALENTS WHILE TRANSLATING TEXTS	18
2.1 Compilation of an English-Ukrainian glossary of terminology.....	18
2.2 Transformation analysis in the translation of scientific and technical terms of military discourse from English into Ukrainian.....	23
2.3 Summarizing the results of the study of the use of the OPUS corpus of texts in the translation of terminology	32
CONCLUSION	38
BIBLIOGRAPHY	40
LIST OF REFERENCE SOURCES	42
LIST OF DATA SOURCES	42
ANNEX.....	43
SUMMARY.....	54

INTRODUCTION

Corpus linguistics provides a tool, for studying language patterns and usage through the analysis of text collections known as corpora. In the field of translation corpus linguistics plays a role in aiding the translation of terminology. By examining and studying corpora translators can uncover usage patterns and variations in terminology across languages. This allows them to make choices when translating technical terms ensuring precision and consistency in their translations.

An important advantage of utilizing corpora for translation is the capability to search for term matches. Translators can search the corpus to locate instances where specific terms or phrases have been consistently translated within context. This not assists translators in finding equivalents for technical terminology but also helps maintain uniformity across translations. Furthermore, corpus linguistics enables translators to recognize collocations and usage patterns associated with terms offering insights into language nuances in various contexts. In essence corpus linguistics acts as a tool, for translators empowering them to create high quality translations that accurately convey the texts intended meaning.

The study is relevant because of the growing importance of translation in the contemporary world, particularly the need for precise and effective translations. Using the OPUS corpus of texts makes it possible to identify the best terminological counterparts, which enhances translation quality and advances linguistic diversity.

The object of the study is English-language texts of military discourse into Ukrainian and the use of the OPUS corpus of texts to find terminological equivalents.

The research is built upon an in-depth analysis of translation theories, techniques, and methodologies. Furthermore, it employs advanced techniques using corpora to uncover terminological parallels. The theoretical foundation of the study is based on the writings of eminent experts in the fields of computational linguistics and translation studies.

The aim of this research is to describe and establish the frequency and types of transformations utilized in the translation of scientific and technical terms within

military discourse from English into Ukrainian. Additionally, it seeks to find the most common types of transformations used and their respective percentages. Furthermore, it aims to determine the effectiveness of each transformation technique in maintaining accuracy and clarity in the translated texts. Lastly, it seeks to investigate the role of corpus linguistics tools, such as the OPUS corpus and the AntConc Corpus Manager AI tool, in facilitating the translation process, providing insights into their practical application.

The investigation subject is the OPUS text corpus, utilized in conjunction with the AntConc Corpus Manager AI tool, for the analysis of translation transformations in military discourse texts from English into Ukrainian.

The object of the study is the scientific and technical terminological vocabulary of military discourse.

Data sources are texts of military discourse using scientific and technical terminology.

Methods used in the research:

Theoretical: Conducting a theoretical analysis of phraseology theory and studying different approaches to studying phraseological units in newspaper headlines.

Compilation and Selection: Compiling and selecting texts from the OPUS corpus relevant to the study of military discourse.

Identification of Terminology: Identifying terminology related to military discourse within the compiled texts.

Data Analysis: Analyzing the identified terminology to determine the most frequently used lexical items and their contextual equivalents.

Translation Process: Translating the selected military discourse texts from English into Ukrainian using various transformation techniques.

Transformation Analysis: Analyzing the transformations employed during the translation process, including modulation, transposition, morphological replacement, etc.

Comparison and Evaluation: Comparing the original English texts with their translated versions in Ukrainian and evaluating the effectiveness of the translation strategies used.

Glossary Compilation: Compiling a glossary of translated terms along with an analysis of the transformation techniques applied.

Statistical Analysis: Conducting a statistical analysis of the frequency of each transformation type used in the translation process.

The study provides theoretical insights into the application of transformation techniques in the translation of scientific and technical terminology in military discourse by furthering our understanding of language processing mechanisms. Furthermore, it offers linguists and translators helpful, practical guidance on how to use the OPUS corpus and pinpoints efficient techniques for producing precise, contextually relevant translations in specific disciplines.

CHAPTER 1

KEY ASPECTS OF USING A CORPUS OF TEXTS FOR SEARCHING TERMINOLOGICAL EQUIVALENTS WHILE TRANSLATING TEXTS

1.1 Fundamentals of corpus linguistics, the concept of a corpus of texts and their classification

As a subfield of applied linguistics, corpus linguistics develops basic guidelines for building linguistic corpora using state-of-the-art computer technologies. The 1960 saw a sharp rise in this profession due to the quick advancement of computer technology (Barlow, 2000, p. 53–55). Only until the ICAME conference in 1983 did the word “corpus linguistics” itself come into common usage in the latter decades of the 20th century. According to scientific study, corpus linguistics is a discipline that seeks to provide an objective account of the language system.

The text corpus, which serves as both the basis for linguistic analysis and the end product of corpus linguistics, is the primary focus of research in this area. Corpus linguistics specifies the theoretical underpinnings and useful procedures for building and utilizing language corpora. The primary goal is to provide a thorough and organized representation of linguistic communication (Neumann, Freiwald & Heilmann, 2000, p. 98–129).

This branch of linguistics is unique in that it emphasizes communicative processes rather than an abstract language system. With this method, the process of meaningful communication will be precisely described in a language that can be applied to the solution of technical and scientific computer science problems (Johansson, 1998, 3–24).

O. M. Demska-Kulchytska (2003) claims that structuralism serves as the theoretical cornerstone of corpus linguistics, which exclusively examines language in written and spoken texts. The corpus of speech realization of a language, which is a sample of language material chosen and arranged in accordance with specific criteria

for the study and description of the language system, is the central idea in this discipline (p. 42).

Therefore, the corpus's presentation format (orality, written, or mixed) can be used to show texts; the corpus's linguistic tasks, which are determined by data collection regimes, can be limited/unlimited, monitoring or both. Multilingual and monolingual, annotated and unannotated corpora are separated based on whether words in the corpus texts have unique labels that indicate different linguistic categories. Using the concepts of total accountability or data selection makes working with the corpus easier.

A scientist's "corpus" might refer to a variety of ideas that they have offered. Specifically, a corpus is a collection of texts that include information on the circumstances surrounding their development, or it can be understood as a substantial array of linguistic data provided in electronic form (Baker, 1993, p. 223). Additionally, there are alternatives for categorizing corpora based on other standards, like annotations, language (monolingual or multilingual), etc.

The following qualities define corpus analysis: 1) an empirical approach to the study of language data (actual samples of language realisation in natural texts are analysed); 2) the analysis of large, structured corpora of natural texts is based on these corpora; 3) computer technologies are actively used to process language material; 4) the frequency of use of language units is studied using both qualitative and quantitative analytical methods, including statistical studies of compatibility (Коянджан, 2012, p. 48–53). Because actual language information is prioritized over linguistic intuition, corpus-based research allows researchers to eschew subjectivity and instead rely on objectivity (Biber, 1993, p. 243–257). This modifies the conventional method of language learning, and the outcomes help to re-evaluate existing linguistic theories.

Making the distinction between unaltered and annotated texts is a crucial criterion in corpus classification. Annotated corpora offer an extra layer of organization and let the user select which data to process. For markup, the Standard

Common Markup Language is typically utilized (Tiedemann, 2012, p. 2214–2218). Additionally, corpora can be categorized as monolingual and multilingual, synchronous and diachronic, static and dynamic, etc. While parallel and comparative corpora offer content for cross-linguistic comparisons, instructional corpora are designed to enhance foreign language learning. Moreover, depending on their intended use, corpora might be categorized as research or illustrative. The foundational corpus for corpus-based language study is the reference corpus, which was put forth by J. Sinclair. experimental corpora designed to analyze spoken and written language, respectively, such as literary and speech corpora (Бідненко, 2014, с. 243).

In several domains of linguistic study, including stylistics, dialectology, lexicography, sociolinguistics, and more, corpora are a common tool (Kenny, 2001, p. 429). Corpus linguistics is useful for dictionary creation, lexical and grammatical relation analysis, and language acquisition methods.

Text corpora's limited emphasis was a key characteristic in the early phases of corpus linguistics development. There should be annotations and, occasionally, a set of data based on the assigned responsibilities for every corpus of texts. The corpus should include information on the features of the word (part of speech, gender, number, case for nouns, type, tense, transitivity for verbs, etc.) for morphological study, for instance (Зімомп, 2010, с. 26–33).

A field for connecting words is necessary to examine expressions, and author and subject area specifications are necessary to build classifiers. A text corpus needs to contain a lot of data, particularly when it comes to machine learning, quality control of other algorithms and methods for automatic text analysis, or evaluating the performance of other algorithms for machine learning (Корунець, 2017, с. 448). Implementing a text corpus involves a number of time-consuming subtasks, including:

- gathering data on many subjects and from various sources in order to put the representativeness principle into practice;

- organizing and processing the gathered data;
- evaluating the organized and processed data;
- producing markup for the text corpus (Sinclair, 1991, p.179).

Since a corpus is nearly always formed for a specific job, the peculiarities of linguistic corpora and the challenges associated with their development lead to an issue of inconsistency of corpora once tasks are completed. This issue usually arises with corpora that are smaller than national ones. These corpora are nearly ubiquitous because to their extensive mark-up arsenal, but they lack a software interface or the capacity to get vast volumes of texts for software tools (Жулавська & Дорогокупля, 2013, с. 45). Furthermore, new applications for computational linguistics techniques are developing more quickly than new texts.

Texts are analyzed from several angles when tackling difficult tasks, such as whether texts on social media are fabricated, have an emotional bent, etc. Creating unique corpus markup is necessary to solve such jobs, particularly with the aid of machine learning (Карпенко, 2006, с. 29–33). Despite the fact that a great deal of work has been done to execute it, the corpora are not as popular as they once were once the duties are finished. The creation of a set of tools for building and annotating text corpora with the intention of completing different tasks could be the answer to this issue (Chesterman, 2004, p. 95).

Data for corpora can come from a variety of electronic libraries, encyclopedias, collections of texts on a certain subject, news sources, publicly available social media data, etc. Text retrieval with predefined properties (e.g., texts by topic or author) established by academics is a critical prerequisite for automatic corpus construction and markup (Коянджан & Репан, 2012, с. 48–53). There may be a variety of these attributes, thus in order to ensure that preexisting texts may be utilized, rules for extensible corpus markup must be created. Additionally, a tool that enables you to automatically retrieve the relevant texts is required.

Generally speaking, corpus linguistics is always changing, its accomplishments are utilized in many different domains, and the variety of corpora enables researchers to examine language from several angles.

1.2 Features of the translation of scientific and technical terminology from English into Ukrainian

The ultimate adoption of English as the language of modern scientific and technical literature (STL) is a defining trait of the evolution of modern science. The outcome is a translation of STL into Ukrainian from English (Коваленко, 2003, с. 165).

The language demands a high calibre of expertise from the specialized translator, including familiarity with the grammatical, stylistic, and linguistic peculiarities of technical and scientific writings. The study of the linguistic characteristics of scientific and technical translation during Ukraine's international establishment and the growth of its scientific and technical ties with other nations is closely linked to the issues surrounding terminology and is becoming more and more significant in all walks of life (Жуковська, 2013, с. 142).

Terminological units are used frequently in scientific and technical writing. There are hundreds of thousands of words and word-combinations (word-groups) in the legal, sports, scientific, technological, and other terminological systems. People outside of the specific specialty neither utilize nor comprehend these linguistic units. Every scientific discipline or activity has its own lexicon. Similar specialized terms exist for chemistry, physics, power engineering, economics, building construction, aviation, and many more fields. These include the medical field (Абабілова & Білокамінська, 2015, с. 126–128).

Linguistic analysis of the language used in scientific and technical writing demonstrates that these texts all adhere to a particular functional language style that has distinctive traits. The most challenging part of translating scientific and technical texts is accurately conveying the lexical units (terms), which include common and general scientific words. The domain of vocabulary accounts for a significant portion of translation errors, as it is the most complicated language in science and technology. As a result of its ongoing evolution, new terminology and meanings for

common words also emerge. There are many examples of poorly translated terminology and broad scientific words, as well as words that are frequently used.

Scientific and technical literature (STL) vocabulary analysis reveals that: ÷75% of STL vocabulary is made up of commonly used and general scientific words, i.e. words that serve different scientific fields, e.g., analysis, balance sheet, program, component (Коптілов, 2003, с. 165). The share of sector-specific terms, i.e. terms that are used only in certain fields of science (e.g., computer science terms tech: processor, browser), does not exceed 25%. As a result, translation of common and broad scientific terms as well as industry-specific terms requires specialization.

We believe that this review suggests vast majority of scientific work illustrate how this work has been done, what profound actions had been taken and we see many different transformations can be concluded accordingly to the research. We aim at the Ukrainian school of translation thus our research is primarily illustrating the national observations and national schooling (Максімов, 2006).

Translators must simultaneously handle a wide variety of tasks when translating scientific and technical texts. This calls for not just flawless language skills but also certain industry-specific information and the translator's utmost effort.

We are for sense translation in conveying the hidden meaning in the text when translating scientific and technical literature, it is necessary to strive for a special kind of terminological accuracy. It should be remembered that different mosques may not necessarily have identical technical characteristics with regard to volume. The information substance, rationality, objectivity, and accuracy of the scientific and technical language, as well as the clarity and comprehensibility that result from these attributes, are its defining characteristics. Furthermore, the texts that fit the bill might include these elements in varying degrees. Nonetheless, there is a common language use in all of these writings that satisfies the requirements of this communication domain. The usage of a lot of technical terms—words or phrases that define technical concepts – characterizes the language of technical literature. Because many words in the target language have many meanings, it is impossible to draw a clear separation

between terms. Words are no longer part of daily speech, where technology now plays a secondary (subordinate) function. However, when they carry the principal (main) relevant load, the technical sense (Кальниченко & Черноватий, 2020, с. 560).

There are a number of contentious issues in current linguistics that relate to nomenclature. The first is the confusing question of whether a term loses meaning when it becomes widely used. These days, it happens a lot because different forms of communication media (TV, radio, popular magazines, science fiction, the Internet, etc.) provide people with information about a wide range of topics, including science, technology, social life, commerce, law, sports, and the arts. Terms enter the mainstream literary language as a result of technical development and general education.

Given the circumstances, it makes sense that many terminological units enter common usage while maintaining their affiliation with their respective domains.

Linguists who hold the first point of view contend that words may only be considered terms if they have maintained their exclusivity and are not recognized, utilized, or known outside of their particular domain. The opposing viewpoint holds that regardless of their exclusivity, every terminological system should contain all linguistic units that transmit notions unique to a certain field of knowledge. The current body of study on diverse terminological systems has demonstrated and demonstrated that there is no unbreakable barrier separating terminology from the broader language system. However, for all intents and purposes, terminologies appear to adhere to the same set of rules and regulations as linguistic units. Therefore, it would be incorrect to think of a word as something “special” and isolated. Instead, interchange between terminological systems and the “common” lexicon is a rather than typical phenomena (Шпак, 2007, с. 310).

Polysemy and synonymy of terminological units are two more contentious issues in terminology. Some linguists believe that a phrase that is «ideal» should only have one meaning, or be monosemantic. Misunderstandings resulting from

polysemantic terminology can be extremely problematic in both professional and special discourse translation. On the one hand, it appears that this criterion is sensible. However, the language's facts do not support it. Polysemantic concepts have many different terminological systems. The context is the only factor that determines how well they are translated (Сухенко, 2000, с. 124).

Technical term translations completely rely on the translator's familiarity with the subject area of the original text. Translators should make a concerted effort to become well-versed in the terminology used in the relevant field and utilize online resources, terminological dictionaries, and other reference books. It is important to keep in mind that a phrase is often translated using its equivalent in the target language. Translation techniques like descriptive translation, synonymization, and analogues are only employed when no equivalent terms exist. Although terms are largely context-free linguistic entities, context frequently aids in defining the precise field to which a term belongs.

Inadequate translation techniques and incorrectly allocating a term to a certain group can also lead to errors made by translators. Word-for-word translation of an English term or one of its components that is equivalent to a Ukrainian term but has a different meaning is another error made by translators (Латишев, 2008, с. 320). When translating, it's important to keep in mind the term's meaning within the specific context and circumstance. Additionally, if a term is used in plural, its meaning may alter, therefore it must be considered.

1.3 General description of the OPUS corpus toolkit and their resource capabilities

High-quality language resources are essential for the development of the rapidly expanding discipline of Natural Language Processing (NLP). Parallel corpora, which are sets of sentences aligned in two or more languages, are essential for multilingual natural language processing (NLP) research projects like sentiment analysis, machine translation, and cross-lingual information retrieval. In this context, the OPUS corpus toolkit becomes an indispensable resource, providing an extensive set of tools and data created especially for creating and utilizing multilingual parallel corpora. The OPUS corpus toolkit's features are examined in this study, along with its main advantages, processing powers, and overall influence on multilingual NLP research (Baker, 2006, p. 206).

The vast array of parallel corpora that make up the OPUS corpus toolbox is its main asset. Each sentence in one language has a corresponding sentence in another language, and these corpora contain a wealth of aligned textual data. With the help of this framework, researchers can comprehend and evaluate a matching statement in a different language by utilizing one language's advantages (rich annotations, for example). The two main components of the OPUS collection are what give it its value:

Multilingual Breadth: OPUS has an impressive breadth of languages, in contrast to certain comparable corpora resources that concentrate mostly on major European languages. The collection includes languages that are commonly spoken, making it suitable for scholars conducting NLP research with dominant languages. Additionally, by expanding its scope to encompass a sizable number of languages with fewer resources, OPUS promotes the creation of NLP applications for varied language groups as well as the investigation of linguistic phenomena that have received less attention.

Domain Diversity: The OPUS collection's strength extends beyond its linguistic breadth. Because the corpora cover a wide variety of subjects, researchers

can customize their analysis to focus on certain areas of interest. News stories, excellent resources for researching current affairs, and sentiment analysis are examples of common areas. Scientific papers provide an insight into the specific vocabulary and writing styles used in different academic fields. Legal papers offer distinct challenges and prospects for scholars who are keen on exploring the nuances of legal language in various languages. Furthermore, web data—which includes product descriptions, forum debates, and social media posts—reflects the fluid and informal character of online communication. Researchers can investigate how language usage differs in various circumstances and communication styles thanks to the diversity of these domains (Tognini-Bonelli, 2001, p. 1).

The OPUS toolbox offers much more than just easy access to a large number of parallel corpora. It provides researchers with a set of processing tools made especially for modifying and getting these corpora ready for additional examination. These tools streamline the research workflow by addressing a number of issues related to working with parallel corpora.

Preprocessing Powerhouse: To guarantee data consistency and quality, preprocessing procedures are frequently necessary in the early phases of dealing with parallel corpora. Numerous tools are available in the OPUS toolkit to help with these preprocessing chores. The toolbox facilitates sentence alignment, an important step in determining the one-to-one correlation between sentences in different languages. The toolkit's tokenization capabilities also address tokenization, which is the process of dissecting text into discrete words or meaningful pieces.

The problem of choosing data is further complicated by the size of the OPUS corpora collection. Access to the complete corpus may not be necessary for researchers to address their particular research questions. With the filtering capabilities in the OPUS toolbox, researchers can choose particular data subsets according to a range of criteria. These standards might be language pairings, interest-specific topics, or even time intervals for diachronic research on language evolution.

Researchers can extract focused datasets that are suited to their research objectives thanks to this filtering feature, which makes for more effective and focused studies.

In their research activities, researchers frequently make use of a variety of NLP tools and procedures. The significance of data interoperability across various software and platform versions is acknowledged by the OPUS toolbox. Supporting numerous formats, including well-known ones, it makes integration with current NLP workflows and selected analytic tools easier (Tiedemann, 2009, p. 237–248).

The relevance of the OPUS corpus toolbox goes beyond providing processing tools and parallel corpora. With additional tools aimed at empowering academics and facilitating their investigation of multilingual NLP, the initiative cultivates a collaborative environment.

Ensuring the quality of parallel corpora is crucial while working with them. Tools for assessing parallel corpora's quality based on alignment precision and other factors are available in the OPUS. With the aid of these assessment metrics, researchers are able to determine the validity of the data they are utilizing and choose the best corpus to answer their particular research questions (Johansson, 1998, p. 3–24).

Any large-scale NLP resource's ability to succeed frequently depends on cultivating a lively and cooperative community. This kind of community is actively fostered by the OPUS initiative, which offers an abundance of resources. Researchers are guided through the accessible corpora and toolkit functions by means of extensive documentation. Because tutorials provide detailed instructions for particular activities, researchers with different degrees of experience can use the toolkit. Furthermore, forums help users ask questions, share best practices, and contribute to the toolkit's continuing development. They also encourage knowledge exchange. This cooperative setting encourages creativity and guarantees that the OPUS corpus toolbox will always be useful in the rapidly changing field of natural language processing.

OPUS is an open-source, free case system that is updated continuously and includes corpora of texts from L1 and L2 to L3...Ln from different online sources. Every text is transformed and ranked using the corpus linguistics methodology. By removing the letters C and R from the English word CORPUS, the corpus resource OPUS was created (Kennedy, 1998, p. 309).

If OPUS has more than 90 European and non-European languages, then multilingualism is the first feature. Parallelism is the second, as OPUS has a significant number of parallel text corpora. We are currently working on developing unique processing programs for each of the languages featured in OPUS because the corpus is multilingual and requires language-specific processing of its content (Halverson, 1998, p. 494–514).

J. Tiedemann reports in his paper “Parallel Data, Tools and Interfaces in OPUS” that there are over 3800 language pairings in OPUS, which translates to over 40 billion lexemes in 2.7 billion parallel units. It is also important to highlight that OPUS is a unique resource for researchers of all stripes because it offers tools for the parallel processing of monolingual L1 data in addition to a number of data retrieval methods.

J. Tiedemann highlights the potential of OPUS, which is to supplement the data with dependence information. The investigator is certain that it is worthwhile to rely on for this goal on statistical analyzers created from publicly available data. In order to do this, bits of speech must also be tagged, which has previously been done for a few languages and corpus segments (Tiedemann, 2009, p. 237–248).

This line will be based on previously researched models for several languages, as well as on contemporary tools like hunpos and MaltParser. Thus, by utilizing over 3800 language pairings, OPUS’s multilingual and parallel resource capabilities can guarantee the effectiveness of IPR implementation.

Conclusion to chapter 1

The study of language through the examination of sizable sets of authentic texts is known as corpus linguistics. These corpora can be categorized according to a number of factors, including language (monolingual or multilingual), annotation, and text format (written, spoken, or mixed). Because corpus linguistics is based on objective data analysis and provides insights into real language use, it has advantages over previous methods.

Translation into scientific and technical fields requires knowledge of the target domain as well as the target language. Translators face difficulties like correctly expressing technical jargon and taking into consideration how scientific language is always changing.

An extensive array of parallel corpora covering a variety of languages and domains is offered by the OPUS corpus toolbox. These corpora are useful for natural language processing (NLP) activities such as cross-lingual information retrieval, sentiment analysis, and machine translation.

OPUS is useful for NLP research on a variety of languages and themes because of its multilingual breadth and domain diversity. In addition to data access, OPUS offers processing tools for operations such as data filtering, tokenization, and sentence alignment. These characteristics improve data quality and expedite research operations.

With tools including tutorials, forums, and documentation, the OPUS effort promotes teamwork. With NLP always changing, this community-driven approach guarantees the toolkit's ongoing development and relevance.

CHAPTER 2

TECHNICAL ASPECTS OF USING A OPUS CORPUS FOR SEARCHING TERMINOLOGICAL EQUIVALENTS WHILE TRANSLATING TEXTS

2.1 Compilation of an English-Ukrainian glossary of terminology

We follow how the AntConc Corpus Manager's functioning, with a particular emphasis on the Option Word List Tab.

The free artificial intelligence program AntConc Corpus Manager (henceforth referred to as AntConc) is a multipurpose tool for statistical analysis of texts in world languages that belong to various discourses (philosophical, scientific, religious, etc.). (Kotyurova, 2020, p. 37). Dr. Laurence Anthony, Director of the English Language Learning Center in Science and Technology at Waseda University (Japan) School of Science and Technology, created AntConc interface (see Figure 1).

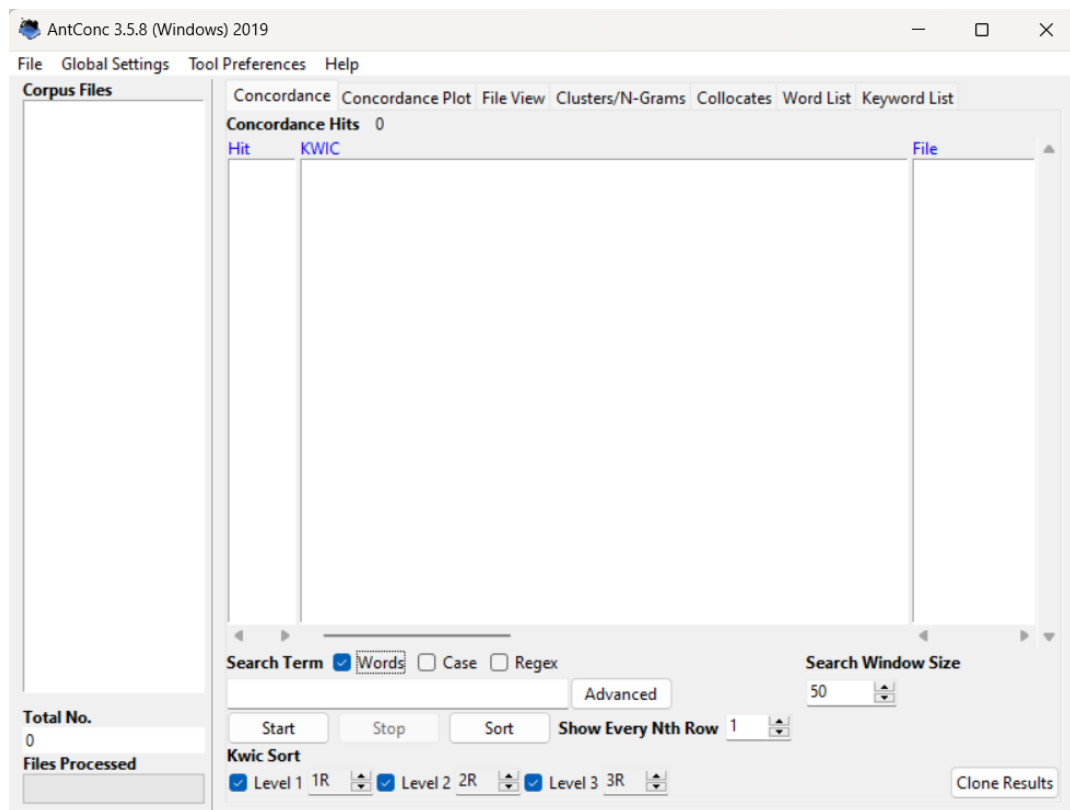


Fig. 1. AntConc Corpus Manager Interface

AntConc can be downloaded for Windows and Mac from the Internet. It has seven tabs, each of which corresponds to a different analytical tool. By clicking on

one, you can update it (you can also use the keys F1 through F7). The following are the primary instruments that I. A. Kotyurova (2020) highlights:

1. A *concordance*, also known as Key Words in Context – KWIC – is a tool that enables one to locate every instance of a term or phrase in the text;

2. A *concordance* Plot shows the researched words or phrases as a barcode within the text, enabling a visual assessment of the object's frequency and location within the text.

3. *File View* presents the chosen file's text in its original format;

4. *Words Clusters* is a tool for choosing a set of words or phrases that have a specified amount of elements to the left and right of a specified word or phrase;

5. *Collocates* function may instantly perform a statistical analysis of the words and phrases to the left and/or right of the search element;

6. *Word List* is a tool counts and displays every word and phrase that appears in the corpus as an ordered list;

7. A method called *Essential Word List* determines which terms and phrases in the corpus have a higher or lower frequency of significance than others.

Military discourse text with terminological lexical items in English was selected for the study on the use of the OPUS corpus for finding terminological matches when translating texts from English into Ukrainian. First of all, AntConc Corpus Manager is used to identify the most frequently used words and find terminological units.

The procedures below should be followed in order to choose keywords from the article « The President's Inbox Recap: The War in Gaza (Lindsay, 2024). To begin, convert the text document from.doc to.txt, upload it to AntConc, select the most often occurring keywords, utilize *the 6th Word List tab*, and hit the "Start" key without altering the program's "by default" settings. Note that the Word List tab is linked to *the Key Word List tab*; that is, *the Word List tab's* automatic renewal is not carried out when stop words, preference words, and/or lemma lists are changed. The scales are computed based on the previous frequency list, which does not take

modifications into consideration. Restarting *the Word List tab* is necessary to prevent this. Thus, the results are as follows. The following words are used most often in this text:

Defence (8), system (6), military (5), NATO (5), Patriot (5), model (4), support (4), exercise (4).

The following examples will help us illustrate the intertextuality based on the abovementioned keywords. The word *defence* appears 8 times throughout the text and it is the most frequently used token among other keywords. For example: *The Dutch Defence Minister, Kajsa Ollongren, emphasized her nation's commitment to Baltic security. has been ongoing concern regarding the deployment of air defence systems to the Baltic region. The former Lithuanian Minister of National Defence urged NATO to quicken the pace of deployment. Lithuania's defence budget experienced significant growth between 2020 and 2024. Despite the months of inactivity in air defence, the Netherlands provided substantial support to the Baltic republic in other areas. NATO has continued to increase its troop presence in Eastern Europe as a deterrent, particularly focusing on defence assets reaching the eastern flank quickly* (Lindsay, 2024).

So, it is see, the word defence is used both independently and in phrases such as: *The Dutch Defence Minister, air defence, Lithuanian Minister of National Defence, and defence assets.*

System is the second most frequently used keyword in the text. It is used 6 times in such sentences: *One of four Dutch MIM-104 Patriot surface-to-air missile defence systems will be deployed to Lithuania for an exercise in the summer of 2024. Arvydas Anušauskas urged NATO to quicken the pace of deployment for air defence systems in the region. The Rotational Air Defence Model, established at NATO's Vilnius summit, aims to deploy air defence systems to the Baltic region. The Patriot is by far the most expensive single weapon system supplied to Ukraine by the US. The intention behind the exercise is to train no-notice redeployment of air defence units and systems* (Lindsay, 2024).

Most often this word is used in phrases and collocations such as: (air) *defence systems* and *weapon system*.

And one more word is *Patriot* which is used 5 times in such sentences: *According to the Washington-based think tank, the Center for Strategic and International Studies, the **Patriot** is by far the most expensive single weapon system that the US has supplied to Ukraine. One of four Dutch MIM-104 **Patriot** surface-to-air missile defence systems will be deployed to Lithuania for an exercise in the summer of 2024. Assets such as the **Patriot** must reach the eastern flank quickly in the event of threats occurring in Eastern Europe* (Lindsay, 2024).

Thus, using AntConc, we identified the lexical core of the text and found the most frequently used words, and in conclusion, we can say that all these words are terminological units. However, other terminology was also identified using this tool, such as: *surface-to-air missile defense systems, rotational Air Defense Model, air defense tasks, air defense systems, field hospitals, medevac, logistics, military electronics, subsystems, GlobalData intelligence, think tank, weapon system, air policing, deterrent, reinforcements, Surface-to-air missile (SAM), Air defense system, Missile defense system, Airborne early warning (AEW), Electronic warfare (EW), Ballistic missile defense (BMD), Anti-aircraft artillery (AAA), Radar-guided missile, Tactical nuclear weapon, Anti-tank guided missile (ATGM), Intercontinental ballistic missile (ICBM), Cruise missile, Multi-role fighter aircraft, Armored fighting vehicle (AFV), Attack helicopter, Naval gun, Submarine-launched ballistic missile (SLBM), Anti-submarine warfare (ASW), Close air support (CAS), Unmanned aerial vehicle (UAV), Surface-to-surface missile, Guided bomb, Chemical weapon, Biological weapon, Amphibious assault vehicle (AAV), Mine-resistant ambush-protected vehicle (MRAP), Tactical ballistic missile (TBM), Rocket artillery, Anti-ship missile, Countermeasure system* (Lindsay, 2024).

The effective text analysis tool AntConc assisted in locating terminological units inside the text. The most common terms and phrases that are significant in the context of military discourse were found by using AntConc's functions. Key terms

utilized in the military setting, such as electronics, military logistics, air defense systems, and combat systems, can be identified by examining these terminological units. With AntConc, you may quickly scan through a lot of material and find the most important terms used in a certain situation. The ability to recognize major terms and concepts in the text has made this tool crucial for comprehending the primary ideas and themes of the military discourse that is being studied.

The next stage in language analysis is to use the OPUS corpus to discover the best relevant translations after important terms have been successfully identified with AntConc. The vast collection of literature from the OPUS corpus spans several languages and genres. Because words and expressions are examined in authentic language situations, using such a corpus enables you to obtain translations that are more accurate and contextually relevant. To view and analyze the options we follow the interface of OPUS (Figure 2.2).

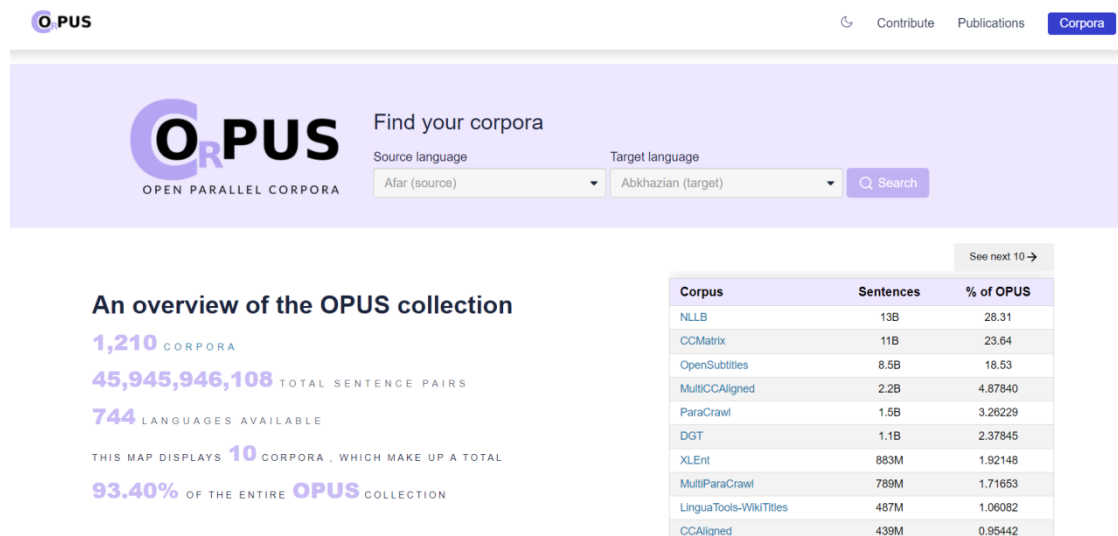


Fig. 2. 2 OPUS interface and parallel collection of texts

The lexical elements found by AntConc are first entered into the OPUS text corpus search engine to start the translation search process. With the help of this tool, translations that are contextually appropriate may be located, accounting for the variations in word usage across many contexts and languages.

Therefore, for the purpose of the study, all the terminological lexical items found using the AntCon program were placed in the OPUS text corpus, which played a key role in finding the translation of the word that was most relevant to our context.

So, before the start of selecting the contextual translation of the word, we chose the Corpus query tool (CWB), which is the main tool for finding matches.

The first step was to select the language pair to be involved, in our case, English-Ukrainian. The next step was to select a thematic corpus, in our case, tatoeba corpus. After that, proceeding to the main task, the selection and selection of translations of scientific and technical terminological lexical units. To do this, all the previously written out units were loaded into the program one by one, and translations were selected for each of them. The final result was a list of 50 terms in context in English and their contextual translation into Ukrainian.

Thus, the OPUS corpus of texts with the previously used artificial intelligence program AntConc Corpus Manager played a key role in, first of all, selecting terminology from the text of military discourse, finding the most frequently used lexical items, the lexical core of the text. In turn, with the help of text correlation, contextual equivalents of scientific and technical terminology were found, which greatly simplified the process of translation and selection of Ukrainian equivalents and ensured the correctness and accuracy of the translation.

2.2 Transformation analysis in the translation of scientific and technical terms of military discourse from English into Ukrainian

One sort of unique translation that stands out is military translation, which is characterized by a high number of terminology, clear and succinct presentation, and a dearth of artistic expression.

Military materials also include fiction, political texts, and technical documentation as well as acts of management and political literature. The correctness of a translation is crucial when translating military documents because the content can be used as a foundation for military operations, critical decision-making, etc.

A vast array of lexical, grammatical, and stylistic changes must be made for the translation of the original military text to reach communicative equivalency. To accurately express the meaning of a statement, lexical and semantic replacements must be made when translating military writings. As a result, when working with military discourse texts, translators need employ a variety of translation transformations.

There are many idioms and constructs in military literature that need to have lexical transformations applied in order to translate them into Ukrainian. When there is no dictionary match or when it cannot be used because of specific circumstances, lexical transformations are taken into consideration as a means of translating different military-technical originals. It is noteworthy that one crucial aspect of a high-quality translation is the skillful application of translation modifications.

When considering translation techniques, the classification of translation transformations by C. E. Maksimov defines translation transformations as lexical and semantic (generalization, differentiation, specification, semantic development, antonymic translation, compensation and complete rearrangement of text segments), as well as grammatical transformations (movement, i.e., changing the order of words and phrases, grammatical substitutions, additions and omissions

Let's take a look at the use of transformations that were used in the translation of terminology that was selected and translated using the OPUS corpus.

First of all, let's look at such a type of translation transformation as lexical. In the course of this study, only one type of lexical transformation was identified as being used, and that is transcoding. This is a method of translation when the sound and/or graphic form of a source language word is transmitted by means of the alphabet of the target language, cf.: *Cyber warfare* – *кібервійна*; *ballistic trajectory* – *балістична траєкторія*; *Rocket artillery* – *ракетна артилерія*; *Tactical ballistic missiles (tbms)* – *Тактичні балістичні ракети (ТБР)*; *Amphibious assault vehicles (aavs)* – *Амфібійні бойові машини (АБМ)*; *Guided bombs* – *керовані бомби*; *Intercontinental ballistic missile (ICBM)* – *міжконтинентальної балістичної*

ракету (МБР); Tactical nuclear weapons – Тактична ядерна зброя; Ballistic missile defence (BMD) – балістична ракетна оборона; tactical maneuver – тактичний маневр; subsystems – підсистеми; military electronics – військова електроніка; medevac – медевакуація.

After analyzing, we can understand that in each of these, the transition of the graphic and phonetic form of the word and the addition of morphemes inherent in the Ukrainian language to correctly display parts of speech.

The next significant group of transformations is LEXICAL-SEMANTIC TRANSFORMATIONS. All three types of these transformations were found in the analyzed text: concretization, generalization and modulation.

Therefore, concretization is formed by changing the generic name to a specific one, thereby narrowing the semantic field of the word being translated, cf.: *The leaders are airborne and ranger qualified* (Lindsay, 2024). – *Командири всіх рівнів проходять парашютно-десантну підготовку і курси рейнджерів.* In this sentence, the word *leaders*, which has a much broader concept, is translated by a word *командири* that has a narrower semantic field and is used in a more specialized context.

Air Policing missions aim to maintain airspace integrity... (Lindsay, 2024) – *Місії повітряного патрулювання спрямовані на збереження цілісності...* *Electronic warfare (EW) capabilities disrupted enemy communications and radar systems* (Lindsay, 2024). – *Можливості радіоелектронної боротьби (РЕБ) перешкоджали ворожим зв'язкам та радіолокаційним системам.* *Anti-aircraft artillery (AAA) units provided ground-based air defence against aerial threats* (Lindsay, 2024). – *Підрозділи протилітакової артилерії (ПЛА) забезпечували повітряну оборону на землі проти повітряних загроз.* *Anti-aircraft* is translated as *протилітакової*, but the connotation of this word is much narrower than in the target language.

The next type of transformation is generalization that is opposite to concretization. The meaning is to replace the species name with the generic name:

Attack helicopters conducted aerial reconnaissance missions to gather intelligence on enemy movements (Lindsay, 2024). – **Бойові вертольоти** виконували повітряні розвідувальні місії для збору інформації про рух ворожих військ. The word *attack* has narrower meaning than translated equivalent *бойові*.

The naval fleet is equipped with advanced Naval guns for naval warfare operations (Lindsay, 2024). – Морський флот обладнаний передовими **морськими гарматами** для морських бойових операцій. *The naval fleet is equipped with Cruise missiles for precision strikes against hostile naval assets* (Lindsay, 2024). – Морський флот обладнаний **крейсерськими ракетами**. для точних ударів по ворожим морським об'єктам. Here the word *missiles* was translated as *ракети*, but Ukrainian equivalent is *снаряд*. Since the first concept is much narrower, we can call this transformation concretization.

The most common lexical-semantic transformation during the studied text of the military discourse during the translation of scientific and technical terminological units was modulation, the meaning of which is to replace the dictionary equivalent with a contextual one that is logically related to the original word: *...team identified the target reference point for the upcoming artillery strike...* (Lindsay, 2024) – *...група визначила орієнтир для майбутнього артилерійського удару...* In the example used, modulation is manifested through the replacement of terms or phrases with similar but contextually appropriate ones. For example, Engl. “*target reference point*» was replaced with Ukr. “*орієнтир для майбутнього артилерійського удару*” which retains the meaning and context of the term but uses a more understandable and specific wording for the Ukrainian audience. This modulation helps to ensure that the translation is clear and accurate, taking into account the specifics of the language and audience.

...a prominent think tank to formulate policies on national security (Lindsay, 2024) – *...видатний аналітичний центр, щоб сформулювати політику у сфері національної безпеки*. In this example, the modulation consists of replacing the term «*think tank*» with «*аналітичний центр*», which is an adequate translation given the

context and specifics of the Ukrainian language. Such a substitution ensures that the text is clear and understandable for an audience unfamiliar with English terminology and conveys the same meaning or concept.

Airborne early warning (AEW) aircraft played a crucial role in detecting and tracking hostile aircraft (Lindsay, 2024). – **Літаки дальнього радіолокаційного стеження (ДРЛС) відіграли важливу роль у виявленні та відстеженні ворожих літаків.** In this case, the modulation consists of replacing the term “*Airborne early warning (AEW) aircraft*» with «*літаки дальнього радіолокаційного стеження (ДРЛС)*”, which is an adequate translation given the context and specifics of the Ukrainian language. This substitution ensures clarity and comprehensibility of the text for the Ukrainian reader and conveys the same meaning or concept.

Multi-role fighter aircraft excel in performing a wide range of combat missions with versatility (Lindsay, 2024). – **Багатоцільові винищувачі відзначаються виконанням широкого спектру бойових місій з універсальністю.** In this case, the modulation was done by replacing the term Engl. “*Multi-role fighter aircraft*” with Ukr. “*багатоцільові винищувачі*”, which is an adequate translation given the specifics of the Ukrainian language. This substitution allows us to preserve the semantic content and accuracy of the translation, conveying information about the ability of aircraft to perform various combat missions with great efficiency and flexibility.

The deployment of Armored fighting vehicles (AFVs) provides ground forces with enhanced protection and mobility (Lindsay, 2024). – **Розгортання броньованих машина піхоти (БМП) забезпечує сухопутним військам підвищену захищеність та мобільність.** In this case, the modulation consists of replacing the term Engl. “*Armored fighting vehicles (AFVs)*” with Ukr. “*броньовані машини піхоти (БМП)*”, which corresponds to the Ukrainian term and conveys the same meaning. This substitution allows us to maintain translation accuracy and

clarity of the text, describing the ability of these vehicles to provide ground troops with increased protection and mobility.

Lexical and grammatical transformations make up the next category of transformations and are crucial for translating English-language international legal papers into Ukrainian. These changes are required to accommodate for linguistic and cultural differences and to guarantee that the text is adequate and understandable for the Ukrainian audience.

Cf.: *The military base is equipped with advanced **surface-to-air missile defence systems** that are...* (Lindsay, 2024) – *Військова база обладнана передовими **системами протиповітряної ракетної оборони**, які...* The term surface-to-air missile defense systems is translated using, first of all, such lexical and grammatical transformation as transposition. In this example, the transposition transformation is used, which consists in replacing one word or phrase with another while maintaining the same grammatical category. The word *система* has changed its position in the phrase and now stands at the beginning.

Unmanned aerial vehicles (UAVs) play a pivotal role in conducting reconnaissance and surveillance missions (Lindsay, 2024). – *Безпілотні літальні апарати (БЛА) відіграють ключову роль у проведенні розвідувальних та спостережних місій.* *Aerial – літальні* it is replaced by a more specific concept, since the first meaning of the word is this *aerial – повітряний*.

As a rule, this transformation is used when translating phrases and phrase combinations. This can be done to better approximate the normal word order of the Ukrainian language or to emphasize the importance of specific information. This transposition helps to understand the sentence faster and more efficiently by providing a more natural word order in the translation.

Let's look at this transposition on the examples of other sentences: *...the implementation of the **rotational Air Defense Model** requires careful planning...* (Lindsay, 2024) – *...впровадження **ротаційної моделі повітряної оборони** потребує ретельного планування..., in this case, the position of the word model has*

changed. In this case, the word *model* is *transposed*. A fairly large number of terminological phrases and collocations consisting of 2 or more words are translated using this type of transformation, cf.: *air defence tasks* – *систем повітряної оборони*; *GlobalData intelligence* – *розвідки GlobalData*; *rotational Air Defence Model* – *Ротаційної моделі повітряної оборони*; *Countermeasure system* – *система протидії*.

Another lexical and grammatical transformation used in the study of terminology translation is substitution (morphological replacement), which consists in replacing a part of the original language during translation, which is often necessary due to lexical and grammatical differences between the two languages. Let's take a look at the example of a sentence: *Close air support (CAS) missions provide air assistance to ground forces engaged in combat operations* (Lindsay, 2024). – *Місії ближньої повітряної підтримки (БПП) забезпечують повітряну допомогу сухопутним військам, що знаходяться в бойових операціях*.

So, in this case, the word *air*, which in English plays the role of a noun, was translated into Ukrainian as an adjective *повітряної* during the translation process. This was necessary to ensure the adequacy of the translation and to correctly convey the meaning of the phrase. There are other examples of the use of digital transformation: *radar jamming* – *радіолокаційне спостереження*; *anti-tank guided missiles (ATGMs)* – *протитанкові керовані ракети (ПТКР)*; *anti-ship missiles* – *проти-корабельні ракети*; *stealth aircraft* – *літак-невидимка*. It should be noted that the change from one part of speech to another during translation can be different, such as from verb to noun, noun to adjective, etc.

Another lexical and grammatical transformation identified and used is antonymic translation: *The military employed advanced missile defense systems to intercept incoming projectiles* (Lindsay, 2024) – *Військові використовували передові системи протиракетної оборони, щоб перехопити вхідні снаряди*.

The phrase *missile defence* was translated as *протитанкова оборона*, thus the negative prefix *проти-* was added to the Ukrainian translation.

The main function of this transformation is to replace the form of a word or phrase with the opposite one (positive with negative and vice versa). The meaning of the translated unit remains basically the same. Lexical substitution can lead to structural changes. The variations in the language implementation of the statement's communicative structure are one of the justifications for the use of antonymic translation. The asymmetry of the lexical and semantic systems, which shows itself as the inability to articulate an idea in a single language and intersects throughout the translation process, is another justification for employing antonymic translation. In this instance, it definitely helps to produce appropriate perception.

Descriptive translation is the most common technique for translating terms; it uses extra lexical resources to guarantee that the reader fully understands the meaning and context of the target phrase. It is a lexico-grammatical transformation in which a phrase that explicates the meaning of a lexical item in the source language is substituted for the original word, providing a more or less comprehensive definition of the meaning in the target language.

The following examples of this type of transformation were found in the analyzed text: *The Radar-guided missile system accurately tracked and engaged hostile aircraft.* – *Система керованих ракет з радіолокаційним наведенням точно виявляла та атакувала ворожі літаки.*

The next example: *The launch of a Submarine-launched ballistic missile (SLBM) poses a serious threat to maritime security* (Lindsay, 2024) – *Запуск балістичної ракети, запущеної з підводного човна (БРПЧ), становить серйозну загрозу морській безпеці.* *The deployment of Mine-resistant ambush-protected vehicles (MRAPs) ensures the safety of military personnel in hostile environments* – *Розгортання броньованих машин зі захистом від пасток і засідок забезпечує безпеку військового персоналу в умовах ворожого середовища.* *Frontline soldiers are the backbone of any military operation, facing the brunt of combat head-on* (Lindsay, 2024) – *Військовослужбовці передових*

підрозділів – це головний складник будь-якої військової операції, які стикаються з головним ударом бойових дій безпосередньо.

In each case of using this transformation, auxiliary words are given, such as prepositions, adverbs, or even nominal parts of speech, such as nouns, verbs, and adjectives.

In each case of using this transformation, auxiliary words are given, such as prepositions, adverbs, or even nominal parts of speech, such as nouns, verbs, and adjectives. It is important to note the difference between a descriptive translation and a transformation such as addition. An explanation is created by adding lexical items to convey a potentially unclear term to the reader. In turn, addition is used to correctly convey a lexical and grammatical form, such as: *The military employed **Surface-to-surface missiles** for precise strikes against enemy targets* (Lindsay, 2024) – *Військові використовували **ракету типу земля-земля** для точних ударів по цілях ворога.*

In this case, such a lexical element as *типу*, which is absent in the original, was introduced into the translation in order to correctly convey the meaning of the sentence (original) being translated and/or to comply with the speech and language norms existing in the culture of the target language.

Several conclusions are drawn from the examination of changes made when translating scientific and technical words used in military discourse from English into Ukrainian. To begin with, one of the most often utilized transformations in the translation of these phrases is modulation. By selecting counterparts that are appropriate for the context and making sure the content is clear to the Ukrainian reader, this enables us to maintain the meaning and accuracy of the translation.

It should be mentioned that transcription and transliteration are employed when a term has little or no counterpart in Ukrainian or when maintaining terminological uniformity necessitates keeping the original spelling.

Furthermore, through the use of appropriate terms and phrases, these translations aid in the adaptation of English terminology to the Ukrainian language system, providing Ukrainian readers with convenience and clarity.

Overall, the analysis demonstrates that in order to produce a high-quality and intelligible text for the audience, translation processes and strategies pertaining to scientific and technical vocabulary used in military discourse need to be improved.

2.3 Summarizing the results of the study of the use of the OPUS corpus of texts in the translation of terminology

The way scientific and technical phrases are translated in military discourse matters a lot when it comes to providing clear and high-quality cross-linguistic information transfer. They are an essential instrument for modifying the text to account for linguistic, cultural, and terminological variances between languages.

Firstly, transformations make it easier for the reader to comprehend the text by helping you locate precise, intelligible synonyms for difficult-to-understand terminology. Modulation, for instance, makes it possible to substitute a difficult-to-understand phrase for a simpler analogue, which facilitates comprehension of the content.

Secondly, by tailoring the source language to the needs and quirks of the target language, transformations aid in maintaining the precision and specificity of the original language. As a result, they guarantee the translation will be very reliable and professional without sacrificing its technical and scientific integrity.

Furthermore, the industry's development of a single language standard is aided by the application of transformations in translation. This facilitates the preservation of terminology uniformity and stability across linguistic contexts, which is critical for cross-national specialist communication and information sharing.

In the course of the analysis, a total of 50 examples of technical scientific terminology lexical items were found and selected, which were analyzed for the use of various translation transformations. Thus, the results of this study, according to the

three types of transformation according to their functioning, are as follows: lexical (13), lexical semantic (12), lexical and grammatical transformations (25). So it is 26%, 24% and 50% respectively. Let's take a look on the Figure 2.1.

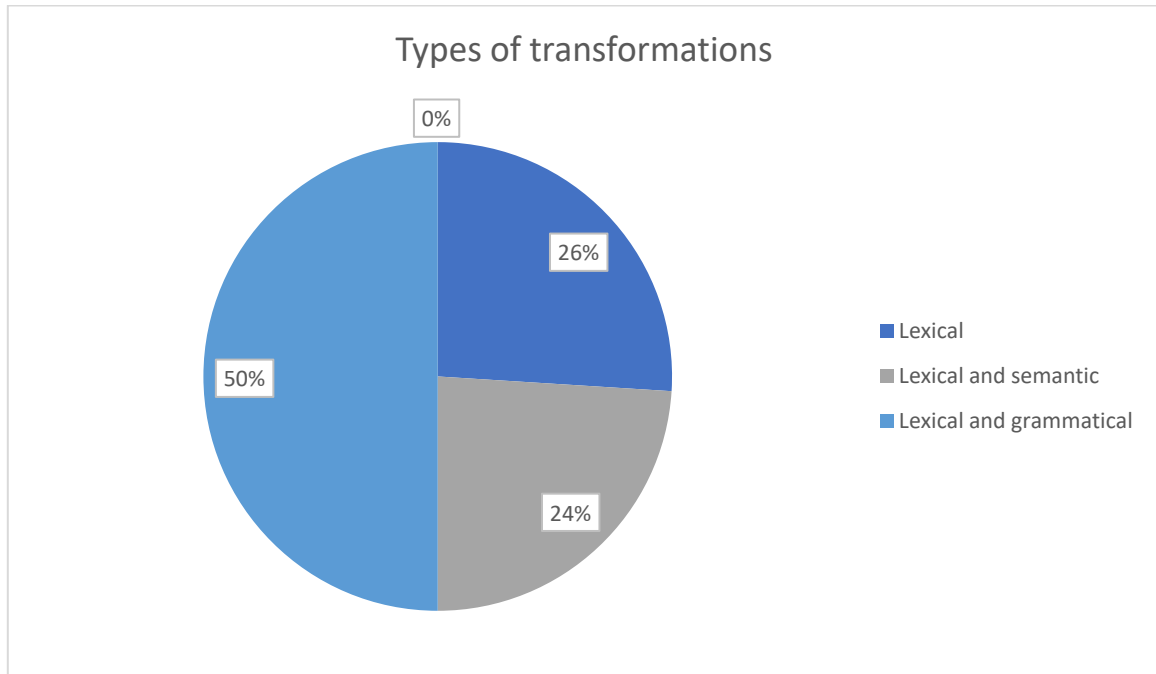


Figure 2.1 The use of groups of transformations in the translation in the analysis of scientific and technical terminology

We can infer from the study's findings that lexical-grammatical transformations are the most often employed kind. There could be multiple reasons for this.

First off, a lot of technical phrases and specialized vocabulary – many of which have unique grammatical characteristics – are present in military discourse. Therefore, lexical and grammatical transformation is frequently necessary to ensure the grammatical accuracy and logical flow of the translation in order to ensure an adequate translation of such phrases.

Second, it could be necessary to swap out a word for one that more accurately captures the meaning and context of the original due to the ambiguity and numerous interpretations of technical words in a military setting. This is frequently shown in the way that modulation and other lexical changes are used.

Furthermore, new phrases and concepts are continuously arising due to the military's quick development and rapid technology advancements, necessitating ongoing translation and terminology adaptation. We can ensure the relevance of the translation by accounting for these changes through grammatical and lexical transformation.

Due to the requirement for grammatical precision, contextual relevance, and quick changes in the language and technology domains, lexical and grammatical transformation ends up being the most often employed type of transformation in the translation of military speech.

Conversely, the least common sort of modification in translating military discourse is lexical-semantic. There are multiple explanations for this.

First of all, a high degree of specialized vocabulary and terminology with a precise meaning in relation to military tactics and operations is typically what defines military discourse. When terms or idioms with different semantic meanings are used in place of such exact and precise language, translation accuracy and clarity may suffer.

Second, technical phrases and idioms pertaining to particular weapons, equipment, and technical procedures are frequently used in military discourse. It is typically not advisable to redefine or replace such terminology because doing so may result in a loss of precision and clarity.

According to the study, the most commonly used type of transformation is Transcoding, which was used 13 times. This type of transformation involves replacing characters or codes of one alphabet with characters or codes of another alphabet. This approach is important in military translation because it ensures the clarity of technical terms and abbreviations.

The second most frequent type of transformation is Transposition, which was used 10 times. This type of transposition, which accounts for about 20% of the total, consists of replacing one word or phrase with another while maintaining the same

grammatical category. This approach helps to understand the sentence faster and more efficiently, providing a more natural word order in the translation.

The following types of transformations are used in descending order of frequency:

- Morphological replacement - used 7 times, which is also approximately 14%
- Modulation - was used 5 times, or 10%.
- Concretization was also used 5 times, which is also 10%.
- Addition - it was used 3 times, which is about 6% of the total.
- Descriptive translation was used 4 times, which is about 8%.
- Generalization - used 2 times, or 4%.
- Antonimic translation – used 1 time, or 2 %

Based on the results of the analysis, we can conclude that the least used type of transformation is Antonymic translation, cf. Figure 2.2.

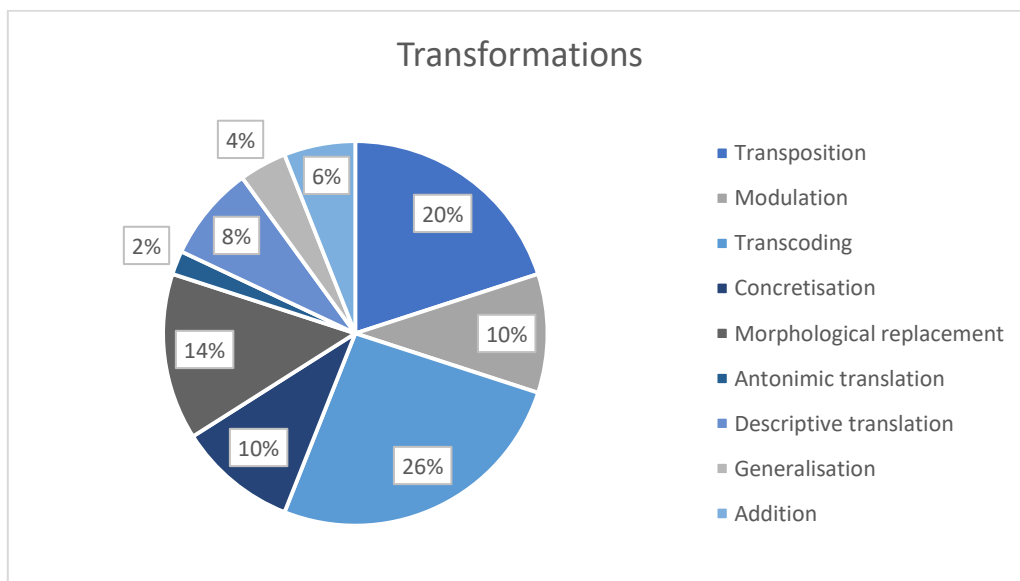


Figure 2.2 The use of different types of transformations in translation in the analysis of scientific and technical terminology

Important findings and intriguing patterns have been found by examining the usage of transformations in the translation of scientific and technical terms used in military discourse from English into Ukrainian. Lexico – grammatical transformations are the most often employed kind, indicating a strong need to modify

language to improve comprehension of the context. This kind of conversion turned out to be the most successful in maintaining the translation's grammatical accuracy and logical word relationships.

Modifying the proportion of various transformation types was also crucial. For instance, even though modulation is often a more popular transformation technique, transcoding was employed more frequently than it was. This could be due to the quirks of translating military jargon, where certain ideas would need to be changed more drastically to be accurate.

The least common sort of transformation was antonymic translation, which can signify that translators avoided using this type of transformation since it could result in a loss of accuracy or tried to accurately express the original meaning.

Overall, this study highlights how difficult it may be to translate military jargon and how crucial it is to employ a variety of transformation techniques in order to provide the most accurate and comprehensible translation.

Conclusion to chapter 2

The AntConc Corpus Manager AI tool and the OPUS corpus were important to our study, especially when it came to choosing terms from military discourse texts. It made it easier to determine which lexical items were most frequently used, which formed the text's lexical core. Furthermore, contextual equivalents of scientific and technical vocabulary were found by using text correlation, which streamlined the translation process and guaranteed the accuracy of the Ukrainian equivalents.

The next step in the study was to compile a glossary with an analysis of the transformations used, and thus the following results were obtained: the most common transformation was modulation, which was used in 10 cases. The second most frequent was transcoding with 13 cases, and the third was morphological substitution with 7 cases. Descriptive translation was used 4 times, and addition – 3 times. Specification (5 cases), antonymic translation, generalization and addition (2 cases each) were also used. The least used transformation was antonymic translation.

CONCLUSION

The study of language by the examination of large collections of real texts is known as corpus linguistics. This field of study divides corpora into groups according to characteristics like language (spoken, written, or multilingual), annotation, and text format. Due to its reliance on objective data analysis and ability to provide insights into actual language use, this methodology has advantages over previous methods.

Technical and scientific texts need to be translated with both subject matter and target language expertise. Translators face difficulties include correctly expressing technical jargon and staying up to date with the constantly changing nature of scientific language.

An extensive collection of parallel corpora encompassing a wide range of languages and domains is available through the OPUS corpus toolkit. These datasets are useful for several natural language processing (NLP) applications, such as sentiment analysis and cross-lingual information retrieval.

OPUS makes NLP research easier in a variety of languages and fields by providing processing tools for tasks like tokenization, data filtering, and sentence alignment in addition to data access. These characteristics quicken research procedures and improve data quality.

Collaboration within the NLP community is promoted via the OPUS project through resources such as tutorials, forums, and documentation. In the quickly changing field of natural language processing, this community-driven method guarantees the toolkit's ongoing development and applicability.

The OPUS corpus and the AntConc Corpus Manager AI tool were quite helpful in our study, especially when it came to detecting phrases from military discourse texts. The primary vocabulary of the texts was formed by these techniques, which made it easier to identify frequently used lexical terms. Furthermore, contextual equivalents of scientific and technical phrases were found by using text correlation, which streamlined the translation process and guaranteed the accuracy of Ukrainian counterparts.

Creating a glossary and looking at the transformations employed were additional analysis tasks. Thus, the results of this study reveal that lexical transformations accounted for 26% of the total, lexical semantic transformations for 24%, and lexical and grammatical transformations for 50%. Transcoding emerged as the most commonly used transformation, employed 13 times, or 26% of the total. This transformation involves replacing characters or codes of one alphabet with those of another, ensuring clarity in technical terms and abbreviations, crucial in military translation.

Following closely, Transposition was the second most frequent transformation, occurring 10 times, constituting approximately 20% of the total. This type of transposition involves replacing one word or phrase with another while maintaining the same grammatical category, aiding in faster and more efficient comprehension.

Other transformations used, in descending order of frequency, include Morphological replacement (7 times, 14%), Modulation (5 times, 10%), Concretization (5 times, 10%), Addition (3 times, 6%), Descriptive translation (4 times, 8%), and Generalization (2 times, 4%). Antonymic translation was the least utilized, appearing only once, accounting for 2% of the total.

In conclusion, the findings indicate a varied use of transformation techniques in military translation, with Transposition and Transcoding being the most prevalent. These transformations play a vital role in maintaining accuracy and clarity in translated texts, ensuring effective communication within the military domain.

BIBLIOGRAPHY

1. Baker M. (1993). *Corpus linguistics and translation studies: Implications and applications*. Amsterdam: Text and technology.
2. Baker P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
3. Barlow, M. (2000). *Parallel texts in language teaching. Multilingual corpora in teaching and research*. Amsterdam: Rodopi.
4. Biber D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*. Vol. 8, no 4, 243–257.
5. Chesterman, A. (2004). *Hypotheses about translation universals*. Amsterdam: John Benjamins Publishing Company.
6. Halverson S. (1998). *Translation Studies and Representative Corpora: Establishing Links between Translation Corpora, Theoretical. Descriptive Categories and a Conception of the Object*. Bergen: University of Bergen.
7. Johansson S. (1998) *Corpora and Cross-Linguistic Research*. Amsterdam: Rodopi.
8. Kennedy G. (1998). *Introduction to Corpus Linguistics*. London-New-York: Longman.
9. Kenny D. (2001). *Lexis and Creativity in Translation: A corpus-based study*. Manchester: St. Jerome.
10. Neumann, S., Freiwald, J., & Heilmann, A. (2022). On the Use of Multiple Methods in Empirical Translation Studies: A Combined Corpus and Experimental Analysis of Subject Identifiability in English and German. In S. Granger & M. Lefer (Authors), *Extending the Scope of Corpus-Based Translation Studies* (pp. 98–129). London: Bloomsbury Academic.
11. Sinclair J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
12. Tiedemann, J. (2009). News from OPUS – a Collection of Multilingual Parallel Corpora with Tools and Interfaces. In N. Nicolov, K. Bontcheva, G.

Angelova, & R. Mitkov. *Recent Advances in Natural Language Processing*, V, 237–248. John Benjamins, Amsterdam/Philadelphia, Borovets, Bulgaria.

13. Tiedemann, J. (2012). *Parallel Data, Tools and Interfaces in OPUS*. In LREC Conferences, 2214–2218.
14. Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work. Studies in Corpus Linguistics*. Amsterdam: John Benjamins
15. Абабілова Н. М., Білокамінська В. Л. (2015). Особливості перекладу термінів українською мовою. *Молодий вчений*. № 2(17), 126–128.
16. Бідненко Н. П. (2004). *Науково-технічний переклад з англійської мови*. Дніпропетровськ: Дніпропетровський університет Альфреда Нобеля.
17. Демська-Кульчицька О. М. (2003). Базові поняття корпусної лінгвістики. *Українська мова*, Київ. №1. 42–47.
18. Жуковська В. В. (2013). *Вступ до корпусної лінгвістики: навчальний посібник*. Житомир: Вид-во ЖДУ ім. І. Франка.
19. Жулавська О.О., Дорогокупля А.О. (2013). *Дослідження безособових форм дієслів програмними засобами корпусної лінгвістики*.
20. Зімомр І. (2010). *Про використання корпусних досліджень у вивченні української мови*. *Українська мова*, Київ. № 3(37). 26–33.
21. Кальниченко О.А., Черноватий Л.М. (2020). *Енциклопедія перекладознавства. Handbook of Translation Studies*. Т. 1. Вінниця : Нова Книга.
22. Карпенко А. В. (2006). *Корпусне дослідження української мови в інтернеті*. *Мовознавство*, Київ. № 1. 29–33.
23. Коваленко А. Я. (2002). *Загальний курс науково-технічного перекладу: посібник для студентів перекладацьких факультетів*. Київ: Інкос.
24. Коптілов В. В. (2003). *Теорія і практика перекладу*. Київ: Юніверс.
25. Корунець І. В. (2007). *Теорія і практика перекладу (аспектний переклад)* : підручник 5-те вид., виправ. і допов. Вінниця : Нова Книга.
26. Коянджан І. А., Репан Т. В. (2012). *Роль корпусної лінгвістики у вивченні історії мови*. *Мовознавство*, Київ. № 3. 48–53.

27. Латишев Л.К. (2008). *Технологія перекладу*. Київ:Академія.
28. Максимов С. Є. (2006). *Практичний курс перекладу (англійська та українська мови). Теорія та практика перекладацького аналізу тексту для студентів факультету перекладачів та факультету заочного та вечірнього навчання: навч. посіб.* Київ: Ленвіт.
29. Сухенко К. М. (2000). *Лексичні проблеми перекладу*. Київ: Вид-во Київ.
30. Шпак В. К. (2007). *Основи перекладу: граматичні та лексичні аспекти: навч. посібник*. Київ: Знання.

LIST OF REFERENCE SOURCES

31. OPUS - Corpora. OPUS - Corpora. URL: <https://opus.nlpl.eu/>
32. Cambridge Dictionary | English Dictionary, Translations & Thesaurus. Cambridge Dictionary | English Dictionary, Translations & Thesaurus. URL: <https://dictionary.cambridge.org/>
33. Merriam-Webster: America's Most Trusted Dictionary. Merriam-Webster: America's Most Trusted Dictionary. URL: <https://www.merriam-webster.com/>

LIST OF DATA SOURCES

1. Lindsay, J. M. (2024, April 3). The President's Inbox Recap: The war in Gaza. Council on Foreign Relations. <https://www.cfr.org/blog/presidents-inbox-recap-war-gaza>

ANNEX

№	The English term in OPUS	The Ukrainian terminological equivalent	Transformations
1	The military base is equipped with advanced surface-to-air missile defence systems that are...	Військова база обладнана передовими системами протиповітряної ракетної оборони, які...	transposition
2	...the implementation of the rotational Air Defence Model requires careful planning...	...впровадження Ротаційної моделі повітряної оборони потребує ретельного планування...	transposition
3	...soldiers were briefed on their air defence tasks and deployed to...	...військовослужбовці були проінформовані про свої завдання з повітряної оборони та розташовані на...	transposition
4	The country invested heavily in upgrading its air defence systems to counter evolving aerial threats.	Країна вклала значні кошти у модернізацію своїх систем повітряної оборони , щоб протистояти зростаючим повітряним загрозам.	transposition

5	...team identified the target reference point for the upcoming artillery strike...	...група визначила орієнтир для майбутнього артилерійського удару...	modulation
6	...the injured soldier was evacuated by medevac to receive urgent medical care...	...поранений військовослужбовець був евакуйований за допомогою медевакуації , щоб отримати невідкладну медичну допомогу...	transcoding
7	The rapid advancement in military electronics has revolutionized modern warfare tactics.	Швидкий прогрес у військовій електроніці революціонував тактику сучасної війни.	transcoding
8	The fighter jet comprises various subsystems that...	Винищувач складається з різних підсистем , які...	transcoding
9	The analysis provided valuable insights based on GlobalData intelligence reports, guiding strategic decision-making.	Аналіз надав корисні висновки на основі звітів розвідки GlobalData , що керує стратегічним прийняттям рішень.	transposition
10	...a prominent think tank to formulate policies	...видатний аналітичний центр ,	modulation

	on national security.	щоб сформулювати політику у сфері національної безпеки.	
11	The development of a new weapon system requires extensive testing...	Розробка нової збройної системи потребує широких випробувань...	transliteration
12	Air Policing missions aim to maintain airspace integrity...	Місії повітряного патрулювання спрямовані на збереження цілісності...	concretization
13	Nuclear arsenals serve as a deterrent against potential aggressors...	Ядерні арсенали служать як стримувальний ефект проти потенційних агресорів...	Morphological replacement + addition
14	The military aircraft employed advanced techniques for radar jamming during the mission.	Військові літаки використовували передові техніки для перешкоджання радіолокаційному спостереженню під час місії.	Morphological replacement
15	The success of the mission relied heavily on the tactical maneuver executed by the special	Успіх місії в значній мірі залежав від тактичного маневру , який виконав	transcoding

	forces team.	командний склад спецпідрозділу.	
16	The Air defence system was on high alert following reports of approaching enemy aircraft.	Система повітряної оборони була на високому ступені готовності після звітів про наближення ворожих літаків.	transposition
17	The military employed advanced Missile defence systems to intercept incoming projectiles.	Військові використовували передові системи протиракетної оборони , щоб перехопити вхідні снаряди.	antonymic translation
18	Airborne early warning (AEW) aircraft played a crucial role in detecting and tracking hostile aircraft.	Літаки дальнього радіолокаційного стеження (ДРЛС) відіграли важливу роль у виявленні та відстеженні ворожих літаків.	modulation
19	Electronic warfare (EW) capabilities disrupted enemy communications and radar systems.	Можливості радіоелектронної боротьби (РЕБ) перешкоджали ворожим зв'язкам та радіолокаційним системам.	concretization

20	The deployment of Ballistic missile defence (BMD) shields key assets against long-range missile threats.	Розгортання системи балістичної ракетної оборони (БРО) захищає ключові об'єкти від загроз дальнього радіусу дії ракет.	transcoding
21	Anti-aircraft artillery (AAA) units provided ground-based air defence against aerial threats.	Підрозділи протилітакової артилерії (ПЛА) забезпечували повітряну оборону на землі проти повітряних загроз.	concretization
22	The Radar-guided missile system accurately tracked and engaged hostile aircraft.	Система керованих ракет з радіолокаційним наведенням точно виявляла та атакувала ворожі літаки.	Descriptive translation
23	Tactical nuclear weapons provide a credible deterrence against potential adversaries.	Тактична ядерна зброя забезпечує вірогідний стримувальний ефект проти потенційних противників.	Transcoding
24	The military conducted extensive training to counter the threat posed	Військові провели широкомасштабне навчання для	Morphological replacement

	by Anti-tank guided missiles (ATGMs) .	протистояння загрозі від протитанкових керованих ракет (ПТКР) .	
25	The launch of an Intercontinental ballistic missile (ICBM) poses a significant threat to global security.	Запуск міжконтинентальної балістичної ракети (МБР) становить значну загрозу глобальній безпеці.	Transcoding
26	The naval fleet is equipped with Cruise missiles for precision strikes against hostile naval assets.	Морський флот обладнаний крейсерськими ракетами для точних ударів по ворожим морським об'єктам.	concretization
27	Multi-role fighter aircraft excel in performing a wide range of combat missions with versatility.	Багатоцільові винищувачі відзначаються виконанням широкого спектру бойових місій з універсальністю.	modulation
28	The deployment of Armored fighting vehicles (AFVs) provides ground forces with enhanced protection and mobility.	Розгортання броньованих машина піхоти (БМП) забезпечує сухопутним військам підвищену	modulation

		захищеність та мобільність.	
29	Attack helicopters conducted aerial reconnaissance missions to gather intelligence on enemy movements.	Бойові вертольоти виконували повітряні розвідувальні місії для збору інформації про рух ворожих військ.	generalizaion
30	The naval fleet is equipped with advanced Naval guns for naval warfare operations.	Морський флот обладнаний передовими морськими гарматами для морських бойових операцій.	generalisaion
31	The launch of a Submarine-launched ballistic missile (SLBM) poses a serious threat to maritime security.	Запуск балістичної ракети, запущеної з підводного човна (БРПЧ) , становить серйозну загрозу морській безпеці.	Descriptive translation
32	Anti-submarine warfare (ASW) operations are critical for protecting naval assets from underwater threats.	Операції антипідводної війни (АПВ) є критичними для захисту морських об'єктів від підводних загроз.	transcoding
33	Close air support (CAS) missions provide air	Місії ближньої повітряної	Morphological replacement

	assistance to ground forces engaged in combat operations.	підтримки (БПП) забезпечують повітряну допомогу сухопутним військам, що знаходяться в бойових операціях.	
34	Unmanned aerial vehicles (UAVs) play a pivotal role in conducting reconnaissance and surveillance missions.	Безпілотні літальні апарати (БЛА) відіграють ключову роль у проведенні розвідувальних та спостережних місій.	concretization
35	The military employed Surface-to-surface missiles for precise strikes against enemy targets.	Військові використовували ракету типу земля-земля для точних ударів по цілях ворога.	addition
36	Advanced Guided bombs ensure accuracy and effectiveness in targeting enemy installations.	Передові керовані бомби забезпечують точність та ефективність у нанесенні ударів по об'єктах ворога.	transcoding
37	The use of Chemical weapons in warfare is strictly prohibited by international conventions.	Використання зброї масового ураження у війні строго заборонено міжнародними	addition

		конвенціями.	
38	The deployment of Biological weapons poses a grave threat to civilian populations and the environment.	Розгортання біологічної зброї масового ураження становить серйозну загрозу для цивільного населення та навколишнього середовища.	addition
39	Amphibious assault vehicles (AAVs) are essential for amphibious landing operations during military campaigns.	Амфібійні бойові машини (АБМ) є необхідними для амфібійних десантних операцій під час військових кампаній.	transcoding
40	The deployment of Mine-resistant ambush-protected vehicles (MRAPs) ensures the safety of military personnel in hostile environments.	Розгортання броньованих машин зі захистом від пасток і засідок забезпечує безпеку військового персоналу в умовах ворожого середовища.	Descriptive translation
41	Tactical ballistic missiles (TBMs) are capable of delivering high-precision strikes against enemy targets.	Тактичні балістичні ракети (ТБР) здатні наносити високоточні удари по цілях ворога.	transcoding

42	The military utilized Rocket artillery to provide indirect fire support to ground troops.	Військові використовували ракетну артилерію , щоб надавати непряму вогневу підтримку сухопутним військам.	transcoding
43	Anti-ship missiles were deployed to target enemy naval vessels and disrupt maritime operations.	Проти-корабельні ракети були розгорнуті для ураження ворожих морських суден та руйнування морських операцій.	Morphological replacement
44	The military deployed an advanced Countermeasure system to jam enemy radar signals and protect against missile attacks.	Військові розгорнули передову систему протидії для перешкодження сигналам ворожої радіолокації та захисту від ракетних атак.	transposition
45	The sniper rifle was equipped with a recoil brake to minimize the impact of recoil on accuracy	Снайперська гвинтівка була оснащена гальмом відкату для мінімізації впливу відкату на точність.	transposition
46	Frontline soldiers are the	Військовослужбовці	Descriptive

	backbone of any military operation, facing the brunt of combat head-on.	передових підрозділів - це головний складник будь-якої військової операції, які стикаються з головним ударом бойових дій безпосередньо.	translation
47	The leaders are airborne and ranger qualified.	Командири всіх рівнів проходять парашютно-десантну підготовку і курси рейнджерів.	concretization
48	The missile followed a ballistic trajectory , accurately hitting its target hundreds of miles away.	Ракета слідувала балістичною траєкторією і точно вразила свою ціль на відстані сотень миль.	transcoding
49	The stealth aircraft was designed to minimize its radar signature, making it difficult for enemy detection.	Літак-невидимка був спроектований для мінімізації свого радіозору, що робить його важко виявити ворогам.	Morphological replacement
50	Cyber warfare has become increasingly prevalent in modern	Кібервійна стала все більш поширеною у сучасних конфліктах,	transcoding

	conflicts, with hackers targeting military infrastructure and communication systems.	коли хакери атакують військову інфраструктуру та комунікаційні системи.	
--	--	---	--

SUMMARY

Курсова робота присвячена дослідженню використанню корпусу текстів OPUS для пошуку термінологічних відповідників під час перекладу текстів з англійської на українську мову. Під час проведення дослідження розглянуто основні етапи використання корпусу текстів, укладання глосарію на основі AntConc та проаналізовано функціонування різних видів перекладацьких трансформацій під час процесу перекладу науково-технічної термінології військового. Детально проаналізовано відібрані термінологічні одиниці та здійснено перекладацький аналіз фактичного матеріалу дослідження, який включає 50 науково-технічних термінів. Крім того, у курсовій роботі розроблено діаграми, що містять результати проведеного дослідження, способи перекладу термінології у військовому контексті.

Ключові слова: переклад, перекладацький аналіз, термінологія, корпус текстів, військовий дискурс, науково-технічні терміни, OPUS.