

Palatovska O. V.

<https://orcid.org/0000-0002-6084-6505>

Kyiv National Linguistic University

Tomilin O. M.

<https://orcid.org/0000-0003-1535-9342>

National Technical University «Kharkiv Polytechnic Institute»

PARALLEL CORPORA IN THE STUDY OF THE TRANSLATION OF UKRAINIAN CULTURE-SPECIFIC VOCABULARY

This article explores the potential of using parallel language corpora in the translation of Ukrainian culture-specific vocabulary into English. The relevance of this research stems from the active development of corpus linguistics, the expansion of Ukrainian-English parallel corpora, and the need to refine methods for analyzing the translation of linguistic units that represent national and cultural particularities. Particular attention is paid to culture-specific vocabulary as a vehicle for historical, sociocultural, symbolic, and ideological meanings, the reproduction of which poses one of the most complex challenges in contemporary translation studies. The aim of this article is to develop and test a corpus-oriented approach to analyzing the translation of Ukrainian culture-specific vocabulary into English by using parallel language corpora to identify translation strategies, techniques, and patterns of variability in translation equivalents. The theoretical foundation of the study consists of works devoted to culture-specific items, translation strategies, and corpus-based methods in translation studies. The empirical data were obtained using OPUS, the largest open multilingual corpus. Using the translation of the lexeme “Maidan” as an example, the study analyzes the implementation of key translation strategies and techniques, and traces the influence of context and discursive factors on the choice of a translation decision. The results of the study showed that parallel language corpora make it possible to establish the actual variability of translation, determine the frequency of use of specific strategies and techniques, and identify how translation decisions depend on the genre, communicative, and cultural characteristics of the text. The paper demonstrates that the corpus-based approach is an effective tool for studying the translation of Ukrainian culture-specific vocabulary and has significant potential for the development of modern translation studies, the improvement of translators’ professional competence, and further linguistic research.

Keywords: culture-specific units, parallel language corpora, translation studies, translation strategies and techniques, domestication, foreignization, conservation, substitution.

Statement of the Problem. The rapid development of corpus linguistics opens up enormous opportunities for the use of language corpora in translation studies. Language corpora are large electronic databases of texts that reflect the actual functioning of language in both diachronic and synchronic aspects, provide online access to authentic language material, allow direct observation of the use of words and grammatical constructions in real-world language practice, and help clarify the semantics of specific lexical units, their grammatical forms, and stylistic characteristics. Parallel language corpora – that is, electronic collections of original texts and their translations aligned at the sentence or text fragment level – occupy a spe-

cial place in translation studies, as they allow one to clearly see how the same idea can be expressed in different language means [12, 13].

In this context, the importance of using parallel corpora in the translation of non-equivalent, culture-specific vocabulary cannot be overstated, since it is precisely this vocabulary that directly represents the connection between language and culture, historical memory, way of life, social practices, and value systems of a particular community and, as a rule, has no direct equivalents in the target language. The use of parallel corpora in the translation of such linguistic units has several significant advantages: 1) a corpus makes it possible to identify not just one standard

equivalent for a given word, but a range of actual translation options; 2) it allows determining the frequency of translation strategies and techniques used for a particular linguistic unit; 3) corpus analysis helps trace how translation decisions depend on discursive factors: genre, subject matter, target audience, stylistic register, the degree of cultural “foreignness” of a linguistic unit, and so on; 4) parallel corpora are an effective tool for translation training, as they develop the ability to work with authentic translation data, compare options, and justify one’s own choices.

This issue is particularly relevant for Ukrainian-English translation, as the Ukrainian language contains a significant body of culture-specific vocabulary related to history, everyday life, political events, national symbols, and social realities, etc. Such units include, in particular, the names of historical phenomena, ethnocultural objects, socio-political events, symbols, and concepts: *Cossack*, *hetman*, *vyshyvanka*, *varenyky*, *rushnyk*, *trident*, *Holodomor*, and others. Their translation into English often fluctuates between retaining the Ukrainian form, calquing, providing an explanation, or selecting a functionally similar equivalent. This is precisely why parallel corpora are a productive tool for identifying actual translation variability and describing the patterns of rendering Ukrainian culture-specific items in English.

Taking into account that currently there is a lack of research in Ukrainian academia examining the potential use of parallel corpora in translation, this article aims to fill this research gap.

Analysis of recent research and publications.

In linguistics, systematic research into linguistic units that convey cultural semantics gained momentum in the second half of the 20th century and has continued to develop at the intersection of ethnolinguistics, cognitive semantics, and translation studies. Linguistic units that reflect the connection between language and culture have been labeled with various terms by different researchers: “**cultural words**” (P. Newmark), “**culture-specific items**” (J. Aixelá), “**cultural key words**” (A. Wierzbicka), and “**cultural concepts**” (J. Bartmiński). These definitions can be considered synonymous; on the whole, their meanings are identical, differing only in the perspective from which a particular scholar studied these linguistic items [1, 3, 9, 16]. However, despite this terminological variability, researchers agree that the translation of such items cannot be reduced merely to finding a formal equivalent, as it requires consideration of cultural content, pragmatic function, and the translation recipient’s expectations.

In this paper, we use the term “**culture-specific items**” (CSIs), or culture-specific vocabulary, as we

consider it the most relevant to the topic of our study, and, following J. Aixelá, we define these linguistic units as follows: “Culture-specific items are those textually actualized items whose function and connotations in a source text involve a translation problem in their transfer to a target text” [1, p. 58].

The theoretical foundations for the research of CSIs in translation studies were laid in the works by J. Aixelá, P. Newmark, E. Nida, L. Venuti, and other scholars. In these works, the translation of such vocabulary is examined through the prism of the oppositions “conservation–substitution,” “foreignization–domestication,” and “adequacy–acceptability”. A significant number of reputable scholarly studies are devoted to the strategies and techniques of translating CSIs. Let us consider the works most relevant to our study, along with their key definitions.

One of the earliest works to outline “translation procedures” for “cultural words” is the previously mentioned work by R. Newmark [9, pp. 81–93]. The scholar offers a detailed list of techniques that are still used today in the translation of CSIs: *transference* (*pysanka*, *hopak*, *oblast*, *kobza*), *naturalization* (adaption of a word to the norms of the target language), *cultural equivalent* (*the Verkhovna Rada* → *Parliament*), *functional equivalent* (*hromada* → *local community*), *descriptive equivalent* (*kolyada* → *traditional Christmas caroling ritual*), *componential analysis* (breaking down meaning into components), *through-translation* (calque: *the Black Sea*, *Western Ukraine*), *synonym selection* (approximate translation: *hetman* → *leader*), *shifts / transpositions* (grammatical changes), *modulation* (change of perspective), *recognized translation* (established translation), *paraphrase / expansion* (expanded explanation), a combination of techniques (couplet), and glosses.

L. Venuti proposes one of the most influential models in translation studies, which is based on the opposition between two translation strategies: **domestication** vs. **foreignization**. Domestication is a translation strategy oriented toward the target culture (*Майдан Незалежності* → *Independence Square*, *вареники* → *dumplings*). The text becomes understandable to the reader, but its national cultural specificity is lost. Foreignization, on the other hand, is oriented toward the source culture, preserving the text’s “foreignness” and its cultural identity (*Майдан* → *Maidan*, *вишиванка* → *vyshyvanka*). Thus, L. Venuti emphasizes that translation is not a neutral process, but rather an ideological choice made by the translator. This dichotomy forms the basis for analyzing the translation of culture-specific vocabulary, as it allows determining the degree to which cultural elements are preserved or transformed [15].

J. F. Aixelá proposes a classification of translation techniques based on the contrast between strategies of **conservation** and **substitution**. In general, this approach corresponds to L. Venuti's concepts of domestication and foreignization. Conservation includes the following procedures: repetition (direct reproduction), orthographic adaptation, linguistic (non-cultural) translation (literal translation without cultural adaptation), extratextual gloss (explanation outside the text), and intratextual gloss (explanation within the text). Substitution consists of synonymy (replacement with a word having a similar meaning), limited universalization (generalization), absolute universalization (complete neutralization), naturalization (finding an equivalent in the target culture), deletion (omission), and autonomous creation (adding a new lexical unit). In this way, J. Aixelá views translation as a balance between preserving the cultural specificity of the source text and adapting it to the target culture [1].

It is worth noting that a **translation strategy** is a general approach by which a translator decides how to translate a text, taking into account the purpose, audience, and type of text. A **translation technique (procedure)** is a specific operation performed on an individual word or phrase that a translator applies within the framework of a chosen strategy.

Culture-specific vocabulary is increasingly analyzed not as an isolated lexical unit, but as an element of a text and discourse. This approach makes it possible to take into account not only the denotative meaning of a concept, but also its associative, evaluative, symbolic, and ideological potential. Contextual analysis is particularly important in this approach, since the same unit can perform different functions in different texts: it can create a sense of local coloring, mark a historical or social space, serve as a sign of national identity, or bring political or cultural meanings to the fore [1, 4, 7, 11].

Corpus-oriented studies of translation constitute a distinct branch of contemporary translation studies. The works by M. Baker, S. Fantinuoli, F. Zanettin, and other researchers [2, 6] have shown that corpus-based methods make it possible to move from isolated illustrative examples to a systematic analysis of a large body of translation decisions. Recent works devoted to Ukrainian-English corpora indicate a growing interest in the creation and use of such resources in translation studies [5, 8, 13]. In spite of this, Ukrainian science still lacks studies describing the potential uses of parallel Ukrainian corpora in the translation of culture-specific vocabulary.

Task statement. The purpose of this article is to develop and test a corpus-based approach to analyzing the translation of Ukrainian culture-specific vocabulary into English by using parallel language corpora to identify translation strategies, techniques, and patterns of variability in translation equivalents.

Outline of the main material of the study.

Approaches to CSIs in translation studies are based on understanding them as units closely linked to national culture and a specific sociocultural context. Within this methodology, particular attention is paid to the selection of translation strategies and techniques that ensure the preservation of the semantics and national-cultural flavor of these units in the target language. We will examine the main strategies and techniques for translating the lexeme **"Maidan"** using the data from OPUS, the world's largest open multilingual corpus.

OPUS (Open Parallel Corpus) is a collection of texts available in two or more languages. A distinctive feature of OPUS is sentence alignment, which allows users to quickly find the translation of a specific unit; analyze different translation variants; examine the contexts in which linguistic units are used; and compare translation strategies and techniques. In other words, each sentence in the source language corresponds to its translation into another language, including Ukrainian-English parallels. OPUS consists of many thematic subcorpora (OpenSubtitles, Europarl, TED Talks, Wikipedia, WikiMatrix, etc.). The most significant advantages of this corpus are the following ones: its very large size, the availability of open-access and free resources, and the inclusion of texts from various genres (ranging from spoken language and news websites to technical and fiction texts with parallel translations). In this way, OPUS provides an empirical basis for analyzing the functioning of culture-specific vocabulary in translation.

Let us consider the results of the translation options analysis of the lexeme **"Maidan"** in journalistic texts using the OPUS corpus (Table 1). Today, the lexeme **"Maidan"** is one of the most widely recognized words associated with Ukraine worldwide, making it important to examine its translation into English. This word became internationally known after the Orange Revolution (2004) and has since acquired an extended meaning: it no longer refers simply to a "square" but also functions as a symbol embodying fundamental value of the Ukrainian people, such as freedom, resilience, dignity, and patriotism. The analysis was conducted using the OPUS concordance interface with the Ukrainian-English (uk-en) language pair, drawing on the CCAIined, CCMatrix, MultiParaCrawl, and Wikipedia corpora. This approach made it possible to obtain aligned parallel examples of the use of this lexeme in various journalistic contexts.

The given examples of the lexeme **"Maidan"** translations demonstrate the application of several translation techniques described in Aixelá's classification, which are used to reproduce its cultural specificity. The most used translation procedure is repetition, which is implemented through the transliteration of the lexeme as **"Maidan"** (examples 1, 3, 5, 8, 9). This approach allows preserving the cultural marking and awareness of the realia.

Table 1

Translation of the lexeme “Maidan” (OPUS. (n.d.))

	Ukrainian text	English translation	Translation technique
1.	Протести на Майдані змінили хід історії України	The protests on the Maidan changed the course of Ukraine’s history.	repetition
2.	Люди зібралися на Майдані Незалежності .	People gathered at Independence Square .	absolute universalization
3.	Майдан став символом свободи.	The Maidan became a symbol of freedom.	repetition + conceptualization
4.	Учасники Майдану вимагали змін.	The participants of the Maidan movement demanded change.	repetition + intratextual gloss
5.	Після Майдану країна змінилася.	After the Maidan , the country changed.	repetition
6.	Студенти вийшли на Майдан.	Students came out to the square .	linguistic universalization
7.	Майдан Незалежності — центральна площа Києва.	Maidan Nezalezhnosti (Independence Square) is the central square of Kyiv.	combined translation: repetition + extratextual gloss
8.	Настрої Майдану були революційними.	The spirit of the Maidan was revolutionary.	repetition + conceptualization
9.	Події на Майдані привернули увагу світу.	Events on the Maidan drew global attention.	repetition
10.	Євромайдан розпочався в листопаді 2013 року.	Euromaidan began in November 2013.	repetition as conventionalized form

When the lexeme occurs in its literal meaning in neutral contexts, the translators applied absolute universalization (example 2), where the nationally specific name is replaced by a generally understandable equivalent (“*Independence Square*”), and linguistic universalization (generalization), where “*Maidan*” is generalized as “*the square*” without considering cultural specificity (example 6). And in example 7, repetition is used together with extratextual gloss: the transliterated name “*Maidan Nezalezhnosti*” is accompanied by a translation in brackets (“*Independence Square*”).

In examples 4 and 8, the lexeme “*Maidan*” undergoes conceptualization and is used in the meaning of ‘revolution, protests, people’, repetition is used for translation, which is combined (example 4) with intratextual gloss: the addition of “*the movement*” component clarifies this meaning for a foreign-speaking recipient.

Example 10 illustrates the conventionalized form “*Euromaidan*”, which became established as an international term in 2014 following the widespread use of the lexeme “*Maidan*” since 2004.

Thus, the analysis of the OPUS parallel corpora proves that the lexeme “*Maidan*”, according to the foreignization strategy, is mainly reproduced in the form of transliteration, which reflects its status as a culturally and politically marked concept. In some cases (in neutral contexts) it is adapted to the target culture (“*Independence Square*”). The variability in the lexeme “*Maidan*” translation confirms the thesis that CSIs should not be analyzed in isolation, but as an element of a text and discourse, since their meanings are conveyed through context, pragmatics, symbolism, and ideological connotations. This is fully

consistent with contemporary trends in translation studies and discourse analysis.

Conclusions. The analysis of translation variants of the lexeme “*Maidan*” based on parallel journalistic texts has clearly demonstrated that CSIs convey cultural and ideological connotations. They not only name the objects and phenomena of the source text but also represent another culture and shape culture’s image in the target language, which requires special translational efforts to preserve meaning, national flavor, and communicative functionality. Therefore, equivalence in translation is established precisely within the context, rather than at the level of individual lexical equivalents. Consequently, due to the semantic multidimensionality of CSI and the difficulties in translating them, any promotion (educational, translational, media-related, or artistic) of a borrowed concept contributes to its faster recognition and identification in a target-language environment.

Currently, the research on the translation of Ukrainian culture-specific vocabulary using parallel corpora is still in its early stages and is not yet comprehensive. A promising approach appears to be combining qualitative translation studies analysis with quantitative corpus-based methods: concordance searches, frequency analysis of equivalents, and their classification from the perspective of translation theory. Furthermore, the translation of culture-specific vocabulary requires an interdisciplinary approach that combines linguistic cultural, functional, pragmatic, discursive, and corpus-based aspects. In this context, parallel language corpora serve not only as a source of empirical data but also as a methodological foundation for studying translation strategies, identifying translation variability, and improving professional translation competence.

Bibliography:

1. Aixela J. Culture-specific items in translation // *Translation, power, subversion* / ed. by R. Alvarez, M. Carmen-Africa Vidal. Clevedon : Multilingual Matters, 1996. P. 52–78. URL: <https://www.scribd.com/document/669559460/03-Aixela-Ch4-Culture-Specific-Items> (дата звернення: 14.07.2026).
2. Baker M. Corpora in translation studies: An overview and some suggestions for future research // *Target*. 1995. Vol. 7, № 2. P. 223–243. DOI: <https://doi.org/10.1075/target.7.2.03bak>
3. Bartmiński J. *Aspects of cognitive ethnolinguistics*. London : Equinox, 2009.
4. Bassnett S. *Translation Studies*. 4th ed. London ; New York : Routledge, 2014.
5. Burda-Lassen O. Ukrainian-To-English Folktale Corpus: Parallel Corpus Creation and Augmentation for Machine Translation in Low-Resource Languages // *Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas (Workshop 2: Corpus Generation and Corpus Augmentation for Machine Translation)*. Association for Machine Translation in the Americas, 2022. P. 28–31. URL: <https://aclanthology.org/2022.amta-coco4mt.4/> (дата звернення: 14.07.2026).
6. Fantinuoli C., Zanettin F. Creating and using multilingual corpora in translation studies // *New directions in corpus-based translation studies* / ed. by C. Fantinuoli, F. Zanettin. Berlin : Language Science Press, 2015. P. 1–11.
7. Luo W., Huang D. Translation Strategies of Tourism and Culture Promotion Signs in the Context of Communication Translation // *Open Journal of Modern Linguistics*. 2025. Vol. 15. P. 461–472. DOI: <https://doi.org/10.4236/ojml.2025.153026>
8. Matvieieva S. A., Lemish N. Ye., Zernetska A. A., Babych V. O., Torgovets M. A. English-Ukrainian parallel corpus: Prerequisites for building and practical use in Translation Studies // *Studies about Languages (Kalby studijos)*. 2022. Vol. 40. P. 61–74. DOI: <https://doi.org/10.5755/j01.sal.40.1.30735>
9. Newmark P. *A Textbook of Translation*. Hertfordshire : Prentice Hall, 1988. URL: <https://files.znu.edu.ua/files/Bibliobooks/Zapolskikh/0041038.pdf> (дата звернення: 14.07.2026).
10. Nord C. *Translating as a Purposeful Activity: Functionalist Approaches Explained*. Manchester : St. Jerome Publishing, 1997. DOI: <https://doi.org/10.4324/9781315760506>
11. Pedersen J. *Subtitling Norms for Television: An Exploration Focusing on Extralinguistic Cultural References*. Amsterdam ; Philadelphia : John Benjamins Publishing Company, 2011.
12. Rizun N., Palatovska O., Sova Y., Tomilin O., Shvedova M. Corpus study of lexicon dynamics in Ukrainian-Polish economic and political relations // *CEUR Workshop Proceedings*. 2024. Vol. 3722. P. 131–146. DOI: <https://doi.org/10.31110/COLINS/2024-4/009>
13. Shvedova M., Lukashevskyi A. Creating Parallel Corpora for Ukrainian: A German-Ukrainian Parallel Corpus (ParaRook|DE-UK) // *Proceedings of the Third Ukrainian Natural Language Processing Workshop (UNLP) @ LREC-COLING 2024*. Torino : ELRA and ICCL, 2024. P. 14–22. URL: <https://aclanthology.org/2024.unlp-1.3/> (дата звернення: 14.07.2026).
14. Tiedemann J. OPUS: Open Parallel Corpus. URL: <https://opus.nlpl.eu/> (дата звернення: 14.07.2026).
15. Venuti L. *The Translator's Invisibility: A History of Translation*. 2nd ed. London; New York: Routledge, 2008. DOI: <https://doi.org/10.4324/9780203553190>
16. Wierzbicka A. *Understanding Cultures through Their Keywords: English, Russian, Polish, German, and Japanese*. New York : Oxford University Press, 1997. DOI: <https://doi.org/10.1093/oso/9780195088359.001.0001>

Палатовська О. В., Томілін О. М. ПАРАЛЕЛЬНІ КОРПУСИ У ДОСЛІДЖЕННІ ПЕРЕКЛАДУ УКРАЇНСЬКОЇ КУЛЬТУРНО-СПЕЦИФІЧНОЇ ЛЕКСИКИ

Статтю присвячено дослідженню можливостей використання паралельних мовних корпусів у перекладі української культурно-специфічної лексики англійською мовою. Актуальність дослідження зумовлена активним розвитком корпусної лінгвістики, розширенням українсько-англійських паралельних корпусів та необхідністю вдосконалення методів аналізу перекладу мовних одиниць, що репрезентують національно-культурну специфіку. Особливу увагу приділено культурно-специфічній лексиці як носію історичних, соціокультурних, символічних та ідеологічних смислів, відтворення яких становить одну з найскладніших проблем сучасного перекладознавства. Метою статті є розроблення та апробація корпусно орієнтованого підходу до аналізу перекладу української культурно-специфічної лексики англійською мовою шляхом використання паралельних мовних корпусів для виявлення перекладацьких стратегій, прийомів і закономірностей варіювання перекладних відповідей. Теоретичним підґрунтям дослідження є праці, присвячені культурно-специфічним одиницям, перекладацьким стратегіям і корпусним методам у перекладознавстві. Емпіричний матеріал отримано за допомогою найбільшого відкритого багатомовного корпусу OPUS. На прикладі перекладу лексеми “Майдан” проаналізовано реалізацію основних перекладацьких стратегій і прийомів, а також простежено вплив контексту і дискурсивних чинників на вибір перекладацького рішення. Результати дослідження засвідчили, що паралельні мовні корпуси дають змогу встановити реальну варіативність перекладу, визначити частотність використання окремих стратегій і прийомів, а також виявити залежність перекладацьких рішень від жанрових, комунікативних і культурних характеристик тексту. У роботі обґрунтовано, що корпусний підхід є ефективним інструментом дослідження перекладу української культурно-специфічної лексики та має значний потенціал для розвитку сучасного перекладознавства, вдосконалення професійної компетентності перекладачів і подальших лінгвістичних досліджень.

Ключові слова: культурно-специфічні одиниці, паралельні мовні корпуси, перекладознавство, перекладацькі стратегії і прийоми, доместикація, форенізація, збереження, заміна.

Дата першого надходження статті до видання: 18.07.2026
 Дата прийняття статті до друку після рецензування: 10.08.2026
 Дата публікації (оприлюднення) статті: 09.10.2026