

UDC 81'33: 81'25

DOI <https://doi.org/10.32782/tps2663-4880/2026.46.1.43>**UKRAINIAN TRANSLATIONS OF ERNEST HEMINGWAY'S NOVEL «FOR WHOM THE BELL TOLLS»: QUANTITATIVE MARKERS OF IDIOSTYLE****УКРАЇНСЬКІ ПЕРЕКЛАДИ РОМАНУ ЕРНЕСТА ГЕМІНГВЕЯ «FOR WHOM THE BELL TOLLS»: КВАНТИТАТИВНІ МАРКЕРИ ІДІОСТИЛЮ****Romanyshyn N.I.,***orcid.org/0000-0001-5918-5423**PhD in Philology, Associate Professor,**Associate Professor at the Department of Applied Linguistics**Lviv Polytechnic National University***Tsyokh L.Y.,***orcid.org/0000-0003-2695-4411**PhD in Philology,**Associate Professor at the Department of Applied Linguistics**Lviv Polytechnic National University***Shyika Yu.I.,***orcid.org/0000-0003-2474-0479**PhD in Education, Associate Professor,**Associate Professor at the Department of Applied Linguistics**Lviv Polytechnic National University*

The article aims to examine the challenges of preserving quantitative markers of idiosyle in Ernest Hemingway's novel "For Whom the Bell Tolls" based on a parallel corpus (the original text and its Ukrainian translations).

To conduct quantitative analysis using the InterText programme, a parallel corpus consisting of the novel "For Whom the Bell Tolls" and its two Ukrainian translations by M. Pinchevskyi and A. Savenets has been compiled for quantitative analysis using the InterText software. The corpus has been further analysed using the AntConc toolkit. The results have been processed with custom Python scripts, which enabled the calculation of key quantitative markers of idiosyle.

The topicality of the study lies in the proposed algorithm: through automated corpus processing and statistical analysis, a number of significant parameters have been identified that characterise the author's idiosyle. These include text length, number of word forms, number of lemmas, number of words in the lexicon with a frequency of 10 or higher, and the number of Hapax Legomena. Based on these data, several coefficients have been calculated to determine quantitative markers of idiosyle, including lexical diversity, average word repetition rate, text and vocabulary uniqueness coefficient, text concentration coefficient, lexical density, logical coherence coefficient, and the Automatic Readability Index. The application of the chi-square test ( $\chi^2$ ) for homogeneity confirms that the differences in the quantitative markers of idiosyle between the source text and its translations are statistically insignificant; therefore, the author's idiosyle is largely preserved in translation.

Quantitative correlations between the lexical-semantic features of the source text and the target text make it possible to objectively assess the degree of equivalence, independently of the researcher's or translator's subjective interpretation. Statistically insignificant differences in lexical-level coefficients may serve as one of the objective indicators of translation quality, although they still require further semantic and stylistic interpretation.

**Key words:** idiosyle, quantitative markers, parallel corpus, part-of-speech distribution, chi-square test.

Стаття має на меті висвітлити проблеми збереження кількісних маркерів ідіостилю роману Е. Гемінгвея «For Whom the Bell Tolls» на основі укладеного паралельного корпусу (оригінал та українські переклади).

Для проведення квантитативних досліджень за допомогою програми InterText було укладено паралельний корпус роману «For Whom the Bell Tolls» та двох його українських перекладів – М. Пінчевського та А. Савенця, який було проаналізовано в середовищі програми AntConc. Результати було опрацьовано власними програмами на мові Python, що дозволило обчислити основні квантитативні маркери ідіостилю.

Наукова новизна дослідження полягає у запропонованому алгоритмі: за допомогою автоматичного опрацювання корпусу та статистичних обрахунків було встановлено ряд важливих характеристик, які дозволяють з'ясувати риси ідіостилю письменника: обсяг тексту, кількість словоформ, кількість лем, кількість слів у словнику і з частотою 10 і більше, кількість Нарах Legomena. На основі встановлених даних обчислено ряд коефіцієнтів, що дозволяють визначити квантитативні маркери ідіостилю: текстову різноманітність, середню повторюваність слова в тексті, коефіцієнт винятковості тексту та словника, коефіцієнт концентрації тексту, коефіцієнт лексичної щільності, коефіцієнт логічної зв'язності, автоматичний індекс читабельності. Розрахунки критерію однорідності ( $\chi^2$ ) засвідчують, що розходження у квантитативних маркерах ідіостилю оригіналу і перекладу статистично допустимі, а отже в перекладі збережено ідіостиль оригіналу.

Кількісні співвідношення між лексико-семантичними засобами першотвору й перекладу дають можливість об'єктивно встановити міру еквівалентності, незалежно від особистого сприйняття дослідника чи перекладача.

Статистично допустиме розходження між коефіцієнтами лексичного рівня становить один з об'єктивних критеріїв якості перекладу, який, однак, потребує семантико-стилістичних уточнень.

**Ключові слова:** ідіостиль, квантитативні маркери, паралельний корпус, частиномовний розподіл тексту, критерій однорідності.

**Statement of the problem.** Modern translation studies demonstrate the tendency to apply complex methodologies of ST and TT comparative analysis. Linguistic statistical analysis constitutes an integral component of such approaches, as it reveals the quantitative characteristics of language phenomena [4]. Quantitative analysis is considered one of the most effective methods for processing probabilistic and statistical data, as well as for identifying the most frequent elements within a text. This approach plays a crucial role in determining stylistic dominance within a literary work, as well as in examining the unique linguistic features that reflect the particular style of the author [6; 7]. Furthermore, the application of quantitative analysis, particularly when combined with corpus technologies and computational data processing, provides new insights into how a writer's idiosyncrasy is preserved in translation [15].

The aim of this research is to perform a quantitative analysis of a parallel corpus specifically created for the purpose of conducted experiment. The corpus includes the original text of *For Whom the Bell Tolls* [9] and its two Ukrainian translations [1; 3].

The literary history of one of Ernest Hemingway's most famous novels, *For Whom the Bell Tolls* (1940) is complex and distinctive. The novel, dedicated to the events of the Spanish Civil War (1936–1939), in which the author himself participated, marked significant changes in the thematic and stylistic dominance of his writing. While American literary criticism has recognized this "Spanish" novel as one of Hemingway's finest works, its reception in Ukrainian translation has followed a different trajectory. The first Ukrainian translation by M. Pinchevskyi was published in 1969 under the title *For Whom the Bell Mourns*. This version was, in fact, heavily edited, featuring ideological, religious, and moral-ethical adaptations and paraphrases. It was not until 2018 that a new, more faithful version by A. Savenets was published by the Lviv-based Publishing House Vydavnytstvo Staroho Leva. Literary scholar N. Romanenko has conducted a comparative analysis of the content of both translations [2]. In this research, however, we adopt different comparative perspective, namely, linguo-statistical approach focusing on those parameters of the author's idiosyncrasy that can be subjected to statistical analysis.

**Methods and materials.** To conduct quantitative analysis using the InterText software, a parallel

corpus has been compiled consisting of the original text and its two Ukrainian translations. This software is designed to create an alignment file from two text files (the source and its translation), by pairing each paragraph of the original with its corresponding paragraph in the translation. The program automatically aligns the texts, displays the parallel content in both languages within the interface, and highlights mismatched (i.e. non 1:1) segments that require correction. Furthermore, it provides tools for merging, splitting, and repositioning segments to achieve precise alignment. By combining three alignment outputs – original to Pinchevskyi, original to Savenets, and Pinchevskyi to Savenets – we have constructed a parallel corpus with the following typological and applicative characteristics: 1) illustrative, as it has been created for the purpose of linguistic-statistical comparison between the original and its translations; 2) full-text, as it includes the complete texts of both the original and the translations; 3) static, as it does not anticipate the ongoing adding of new texts to the corpus; 4) bilingual, comprising texts in English and Ukrainian; 5) aligned, with English text fragments matched to their corresponding Ukrainian equivalents; 6) written, consisting exclusively of written texts; 8) annotated, with textual data tagged at the syntactic and lexical levels by the program.

The next stage, that involves generating the lists of word forms used in the novel and its translations, has been carried out using the AntConc software environment. The resulting data have been processed with custom Python scripts, which has enabled the calculation of overall corpus parameters, part-of-speech distribution statistics, and quantitative lexical coefficients. These indicators constitute the primary markers of the author's idiosyncrasy.

**Presentation of the main research material.** In this research, primary attention has been directed to the vocabulary of the source text and its translations. The automatic processing of the corpus, combined with statistical analysis, made it possible to identify a set of significant features that form the basis for distinguishing the author's idiosyncrasy, substantiating the aesthetic value of both the original and the translations, and establishing their degree of equivalence.

#### *General characteristics of the corpus.*

Table 1 presents a set of general characteristics identified on the basis of the electronic annotated corpus comprising the original text and its translations.

Table 1  
**General characteristics of the original  
 and translations**

Ratio	ST	TT1	TT2
Number of word usages	173150	140699	139200
Number of word forms	8586	21057	21349
Number of words	6001	11326	11483
Hapax legomena: text	3909	12557	12755
Hapax legomena: vocabulary	2454	5079	5159
Number of words with a frequency $\geq 10$	1078	1174	1191
Number of letters	129159	131558	131558
Number of sentences	2088	2255	2255

(ST – original, TT1 – Pinchevskyi's translation, TT2 – Savenets's translation)

The data indicate that the original text is the largest in volume, while the Ukrainian translations are nearly identical in size, each being 20% shorter than the original. It is important to note that this contradicts the theory of translational S-universals and T-universals proposed by E. Chesterman, according to which translated texts tend to be longer than their source texts [5]. In contrast, the comparison of the author's lexicon with the lexicons of the translations reveals the opposite tendency: Pinchevskyi's lexicon is 47% larger than that of the original, while Savenets' lexicon exceeds it by 44%.

The number of Hapax legomena in the Ukrainian texts is also higher in the Ukrainian texts: in Pinchevskyi's translation, Hapax legomena constitute 54% of the lexicon; in Savenets' translation – 56%; while in the original text they account for 49%. The amount of high-frequency vocabulary, despite the significant presence of articles in the English text, is also greater in the translations: 0.62% in the original, 0.83% in Pinchevskyi's version, and 0.86% in Savenets' version.

#### **Correlation of statistical characteristics of lexico-grammatical classes of words in the subcorpora**

Stylistic analysis demonstrates that the texts differ across all linguistic levels; however, the stylistic differentiation potential of linguistic units varies, being lowest at the phonetic level and highest at the lexical level [10]. Therefore, the present study prioritizes the analysis of lexical units (words).

It should be noted that words are classified into categories commonly referred to as parts of speech. The close interrelationship between lexical and grammatical meanings, which determine the structural and semantic properties of a word, provides a basis for identifying these categories as lexico-grammatical classes [11]. In this study, these classes are considered integral components of an

author's style. The analysis of their distribution and usage enables the identification of distinctive features of the author's idiostyle within the literary text. Figure 1 below illustrates the distribution of word usage across parts of speech in the source text (Fig. 1).

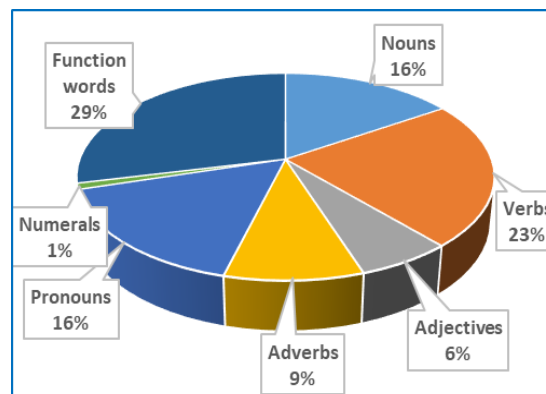


Fig. 1. Distribution of word usage in the ST

The highest stylistic differentiation potential in literary texts is associated with nouns, verbs, adjectives, and adverbs. In the original text, the most frequent parts of speech are verbs (23%), nouns (16%), and pronouns (16%), which slightly surpass adverbs (9%) and adjectives (6%).

The distribution of word usage across parts of speech is illustrated in Figures 2 and 3.

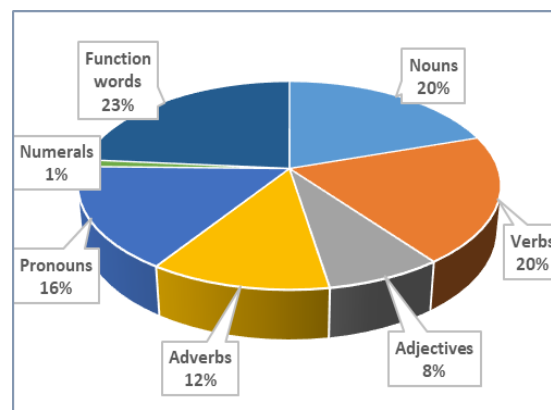


Fig. 2. Distribution of words usage across parts of speech in Pinchevskyi's translation (TT1)

As the diagrams demonstrate, verbs occur with statistically significant frequency in the original novel, highlighting the verbal nature of E. Hemingway's idiostyle. Verbs contribute to the dynamism and intensity of the narrative while preserving a concise writing style and an external objectivity.

However, this feature of the author's idiostyle is not fully preserved in the translations. In

Pinchevskiy's version, nouns and verbs account for an equal proportion (19.86% each), whereas in Savenets' translation, nouns constitute 22.80% and verbs 18.43%. Thus, in TT1, both verbal and nominal modes of expression are equally represented, whereas TT2 demonstrates a tendency toward nominalization, although no significant difference in the use of nouns and verbs is observed in the translated texts.

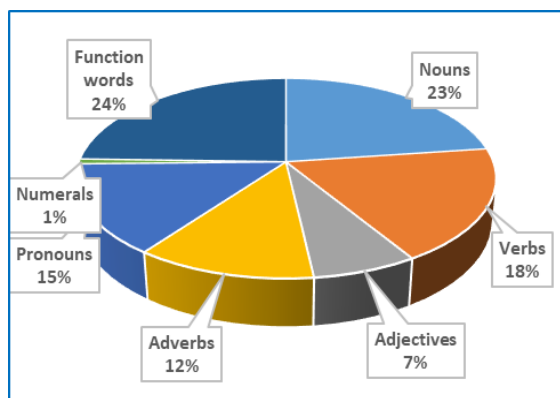


Fig. 3. Distribution of words usage across parts of speech in Savenets's translation (TT2)

In addition, quantitative relationships between parts of speech have been calculated, as they constitute a key component in the statistical characterization of an author's idiosyncrasy. These include the coefficient of nominal modifiers, the coefficient of verbal modifiers, the coefficient of nominality, and the coefficient of aggressiveness.

The coefficient of nominal modifiers is calculated as the ratio of the total number of adjectives to the total number of nouns. The coefficient of verbal modifiers is calculated as the ratio of the total number of adverbs to the total number of verbs. The coefficient of nominality is the ratio of the total number of nouns to the total number of verbs. Finally, the coefficient of aggressiveness (verbality) is determined as the ratio of the number of verbs and verb forms (including participles and gerunds) to the total number of words in the text.

The calculated values for E. Hemingway's novel *For Whom the Bell Tolls* and its Ukrainian translations are presented in Table 2.

Table 2

**Quantitative correlations of parts of speech**

Coefficient	ST	TT1	TT2
Noun modifiers coefficient	2,48	2,59	2,58
Verbal modifiers coefficient	0,40	0,60	0,51
Nominality coefficient	0,69	1,00	1,20
Aggressiveness coefficient	0,23	0,20	0,30

However, determining only the frequency of parts of speech is insufficient, as it does not

indicate whether the observed frequencies of lexicogrammatical classes in the original and translated texts differ significantly from theoretically expected value. This expected value would indicate the statistical significance or insignificance of the differences between the coefficient values. Therefore, in order to ensure a reliable quantitative analysis of the data presented in Table 2, the chi-square ( $\chi^2$ ) test is employed [13].

The value of  $\chi^2$  is calculated according to the following formula:

$$\chi^2 = S \times \left( \sum \frac{(knTn)^2}{\sum kn \sum Tn} - 1 \right) \quad (1)$$

After performing the necessary calculations for the dataset, the following results have been obtained:  $\chi^2 = 0.01$ .

To determine whether this value indicates a statistically significant difference, it must be compared with the critical  $\chi^2$  values [8]. A difference is considered significant if the calculated  $\chi^2$  value exceeds the critical value. In the present case, the obtained value (0.01) is substantially lower than the minimum critical threshold. This indicates that the difference in the quantitative parameters of part-of-speech distribution between the source text and its translations are not statistically significant.

**Comparing quantitative markers of the original idiosyncrasy and the translations**

The obtained quantitative data make it possible to calculate a set of coefficients that characterize the lexical level of the text and can serve as quantitative markers of idiosyncrasy [12; 14].

Lexical Diversity Coefficient is a highly effective tool in text analysis. Its significance extends beyond identifying lexical properties; it also plays a crucial role in delineating the features of an author's idiosyncrasy in literary works.

Lexical Diversity Coefficient ( $K_{ld}$ ) is defined as the ratio of the vocabulary size (i.e., the number of lemmas) to the overall text size (i.e., the number of word tokens). Higher values of this coefficient indicate greater lexical diversity.

Average Word Frequency Coefficient ( $K_{awf}$ ) represents the ratio of the total number of word tokens to the size of the lexicon (i.e., the number of unique lexemes). It reflects how frequently, on average, each lexical unit occurs in the text.

Uniqueness Coefficient is calculated separately for the lexicon and the text. It indicates the proportion of Hapax legomena (words occurring only once) within the lexicon or the text.

The opposite to the Uniqueness Coefficient is the Text Concentration Coefficient ( $K_{tc}$ ), which reflects

the proportion of the text occupied by the words with the frequency (i.e., those with a frequency of 10 or more).

Text Lexical Density Coefficient ( $K_{ld}$ ) displaying the correlation between notional parts of speech in the text and the general number of words. Thus, lexically dense text are those that contain less auxiliary parts of speech.

Logical Coherence Coefficient ( $K_{lc}$ ) is defined as the ratio of the total number of functional words (e.g., conjunctions and prepositions) to the total number of sentences. Values exceeding one are generally interpreted as indicating a sufficiently cohesive relationship between functional elements and syntactic structures.

Automated Readability Index (ARI) was developed by E. Smith and R. Senter at the University of Cincinnati. Initially, this index was used to assess the readability of textbooks and various documents [16]. The formula is as follows:

$$ARI = 4,41 \times \frac{Smb}{N} + 0,5 \times \frac{N}{S} \quad (2)$$

where Smb is the total number of characters in the text, N is the number of words in the text, S is the number of sentences in the text. Higher ARI values indicate greater textual complexity and stylistic density.

Table 3 presents the results of the calculations of these quantitative markers for the source text and its translations.

To determine the statistical significance of the differences between the coefficient values for the source text and its translations, the  $\chi^2$  test has been applied as described above. Comparisons have been conducted between the original and Pinchevskyi's translation, the original and Savenets' translation, as well as across all three texts in order to assess the overall homogeneity

of the idiostyle. The obtained  $\chi^2$  values are 1.88, 1.04, and 3.29, respectively. In all cases, the differences in the quantitative markers are statistically insignificant and fall within acceptable limits. This indicates that the idiostyle of the original novel has been largely preserved in the Ukrainian translations.

Table 3

Quantitative markers of original's and translations' idiostyle

Coefficient	ST	TT1	TT2
$K_{ld}$	0.035	0.080	0.060
$K_{awf}$	28.854	12.423	14.423
$K_{ve}$	0.409	0.448	0.441
$K_{fe}$	0.023	0.089	0.072
$K_{lc}$	0.006	0.008	0.008
$K_{ld}$	0.713	0.764	0.764
$K_{lc}$	3.644	2.360	2.960
ARI	7.154	8.964	8.115

**Conclusions.** The quantitative correlation between lexico-semantic units in the source text and its translations creates a basis for an objective assessment of their degree of equivalence, independent of the subjective judgments of the researcher or translator. Statistically acceptable deviations between lexical-level coefficients may therefore serve as one of the objective criteria for evaluating translation quality. At the same time, quantitative findings require further semantic and stylistic interpretation. The advantages of statistical methods lie in the objectivity of the conclusions they enable. The present study demonstrates the effectiveness of quantitative approaches not only for the objective and reliable analysis of linguistic data, but also as a powerful tool for revealing underlying mechanisms of language use that are not readily accessible through direct observation.

#### REFERENCES:

1. Гемінґвей Е. По кому подзвін : роман / пер. з англ. Андрія Савенця. Львів: Видавництво Старого Лева, 2018. 608 с.
2. Романенко Н. Рецепція роману «По кому подзвін» Е. Гемінґвея українською літературою ХХ–ХХІ століть (етапи, перешкоди, новітня версія). *Актуальні питання гуманітарних наук*. 2019. Вип 24. Т. 2. С. 108–114.
3. Хемінґвей Е. По кому подзвін: роман / пер. з англ. М. Пінчевського. Твори : в 4 т. Київ : Дніпро, 1981. Т. 3. С. 6–404.
4. Buk, S., Rovenchak, A. Rank-Frequency Analysis for Functional Style Corpora of Ukrainian. *Journal of Quantitative Linguistics*. 2004. No 11 (3). P. 161–171. doi: <https://doi.org/10.1080/0929617042000314912>.
5. Chesterman, A. What is a unique item? *Doubts and Directions in Translation Studies: Selected contributions from the EST Congress, Lisbon 2004*. Amsterdam/Philadelphia: Benjamins, 2007. P. 3–13
6. Contrastive corpus linguistics: patterns in lexicogrammar and discourse / Eds. A. Čermakova, H. Hasselgård, M. Malá, D. Šebestová. London; New York : Bloomsbury Academic, 2024. 312 p.
7. Coulthard M. Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics*, 2004. No. 25(4). P. 431–447. doi: <https://doi.org/10.1093/applin/25.4.431>
8. Critical Values of the Chi-Square Distribution. URL: <https://www.itl.nist.gov/div898/handbook/eda/section3/eda3674.htm> (дата звернення: 12.10.2025).

9. Hemingway E. *For Whom the Bell Tolls*. London: Arrow Books, 1994. 506 p.
10. Karasov V., Levchenko O. Statistical characteristics of O. Zabuzhko's idiolect. *IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT)*. 10-12 November 2022. Lviv, Ukraine. P. 138-141. doi: 10.1109/CSIT56902.2022.10000546.
11. Langacker R. W. Grammatical Classes. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press, 2008. P. 93–127.
12. Lototska N., Saban O. (2023). Ivanychuk's Idiolect: quantitative parameterization of the language used in the text. *IEEE 18th International Conference on Computer Science and Information Technologies (CSIT)*. October 19-21. Lviv, Ukraine. P. 1-4. doi: 10.1109/CSIT61576.2023.10324093.
13. Oakes M., Farrow, M. Use of the Chi-Squared Test to Examine Vocabulary Differences in English-Language Corpora Representing Seven Different Countries. *Literary and Linguistic Computing*. 2006. No. 22(1). P. 85-99. doi: <https://doi.org/10.1093/lc/fql044>.
14. Seifart F., Mundry R. Quantitative Comparative Linguistics based on Tiny Corpora: N-gram Language Identification of Wordlists of Known and Unknown Languages from Amazonia and Beyond. *Journal of Quantitative Linguistics*. 2015. No. 22. P. 202–214. doi: <https://doi.org/10.1080/09296174.2015.1037161>.
15. Shymko, V. Identifying Key Idiolect Markers in Sociolinguistic Profiling: A Scoping Review and Analytical Framework for Real-world Applications. *SAGE Open April-June 2025*. P. 1–32. doi: <https://doi.org/10.1177/21582440251334276>.
16. What Is the Automated Readability Index? URL: <https://www.webfx.com/tools/read-able/automated-readability-index/> (дата звернення: 19.12.2025).



Стаття поширюється на умовах ліцензії відкритого доступу CC BY 4.0

*Дата першого надходження статті до видання: 27.04.2026*  
*Дата прийняття статті до друку після рецензування: 22.05.2026*  
*Дата публікації (оприлюднення) статті: 29.05.2026*