Vitalii KARASOV,
orcid.org/0000-0002-8039-2811
Postgraduate student of the Department of Applied Linguistics
Lviv Polytechnic National University
(Lviv, Ukraine) vitalii.v.karasov@lpnu.ua

Olena LEVCHENKO,
orcid.org/0000-0002-7395-3772
Doctor of Philology, Professor,
Head of the Department of Applied Linguistics
Lviv Polytechnic National University
(Lviv, Ukraine) olena.p.levchenko@lpnu.ua

QUANTITATIVE MARKERS OF IDIOLECT: STYLOMETRIC ANALYSIS OF OKSANA ZABUZHKO'S LITERARY TEXTS

The article investigates the specificity of Oksana Zabuzhko's idiostyle through the prism of quantitative stylistic indicators, applying the methodology originally developed by Suitbert Ertel and later adapted to Slavic language by M. and T. Obrębski. The research focuses on four central parameters that serve as statistical markers of idiolect: the Impersonality Quotient, the Plurality Quotient, the Nominalization Quotient. In addition to these, the study also incorporates the Epithetization Index. By conducting a comparative analysis of Zabuzhko's works and the texts of her contemporaries, both male and female, the study identifies distinctive features of the author's style.

The Impersonality Quotient demonstrates that Zabuzhko's texts occupy an intermediate position between the high impersonality characteristic of her female contemporaries and the lower impersonality observed in male authors. The Plurality Quotient in Zabuzhko's prose is lower compared to both male and female contemporaries. The Nominalization Quotient shows higher values in Zabuzhko's texts, reflecting her tendency toward cognitive condensation of information, abstraction of processes, and intellectualization of discourse. Finally, the Epithetization Index reveals an increased frequency of adjectival modifications.

The findings confirm that Zabuzhko's idiostyle is defined by the integration of subjective voice and objectified narration, by a persistent orientation toward abstraction and conceptualization, and by intensified stylistic expressiveness. Thus, the research highlights the relevance of statistical markers for capturing the functional, cognitive, and expressive specificity of literary language and substantiates their role as reliable tools for idiolectal studies, authorship attribution, and corpus-based literary stylistics.

Key words: *idiolect, stylometric analysis, statistical indicators of style, nominalization, plurality, impersonality, epithetization, linguistic identity.*

Віталій КАРАСЬОВ,
orcid.org/0000-0002-8039-2811
аспірант кафедри прикладної лінгвістики
Національного університету «Львівська політехніка»
(Львів, Україна) vitalii.v.karasov@lpnu.ua

Олена ЛЕВЧЕНКО,
orcid.org/0000-0002-7395-3772
доктор філологічних наук, професор,
завідувач кафедри прикладної лінгвістики
Національного університету «Львівська політехніка»
(Львів, Україна) olena.p.levchenko@lpnu.ua

КІЛЬКІСНІ МАРКЕРИ ІДІОСТИЛЮ: СТИЛЕМЕТРИЧНИЙ АНАЛІЗ ХУДОЖНІХ ТЕКСТІВ ОКСАНИ ЗАБУЖКО

У статті досліджено специфіку ідіостилю Оксани Забужко через призму кількісних стилістичних показників, застосовуючи методологію, спочатку розроблену С. Ертелем, а згодом адаптовану до слов'янської мови М. та Т. Обребськими. Дослідження зосереджено на чотирьох основних параметрах, які слугують

статистичними маркерами ідіолекту: коефіцієнт безособовості, вказівника множинності, вказівника номіналізації. Крім того, у дослідженні також використовується індекс епітетизації. Через порівняльний аналіз творів Забужко та текстів її сучасників та сучасниць, дослідження виявляє особливості стилю авторки. Коефіцієнт безособовості демонструє, що тексти Забужко займають проміжну позицію між високим рівнем безособовості, характерним для її сучасниць, та нижчим рівнем безособовості, що спостерігається у творах сучасників. Вказівник множинності в прозі Забужко нижчий порівняно з творами як сучасників, так і сучасниць. Вказівник номіналізації має вищі значення в текстах Забужко, що відображає її схильність до когнітивної конденсації інформації, абстрагування процесів та інтелектуалізації дискурсу. Нарешті, індекс епітетизації виявляє підвищену частоту прикметникових модифікацій.

Результати підтверджують, що ідіостиль Забужко визначається інтеграцією суб'єктивного голосу та об'єктивованої розповіді, постійною орієнтацією на абстракцію та концептуалізацію, а також посиленою стилістичною виразністю. Таким чином, дослідження підкреслює актуальність статистичних маркерів для фіксації функціональної, когнітивної та виразної специфіки літературної мови та обґрунтовує їхню роль як надійних інструментів для ідіолектних досліджень, атрибуції авторства та корпусної літературної стилістики.

Ключові слова: *ідіостиль, стилометричний аналіз, показники стилю мовлення, номіналізація, множинність, безособовість мовлення, епітетизація, мовна особистість.*

Problem statement. When studying idiolect and stylistic features of speech, several quantitative indicators are employed (Darchuk et al., 2024: 224-247; Pavlychenko, 2013: 218-226; Plecháč & Flaišman, 2017: 743-769). In addition, M. Obrębska and T. Obrębski proposed a modified methodology based on the research of S. Ertel, since the specifics of Polish grammar do not allow for the direct transfer of counting methods developed for German-language texts (Obrębska, Obrębski, 2011: 161-170). This methodology is based, in particular, on the analysis of the impersonality quotient, plurality, and nominalization (Obrębska, Obrębski, 2011: 162-164, 166-168). First, when studying the impersonality quotient, not only pronouns are taken into account, but also personal forms of verbs, since in Polish the subject expressed by a pronoun is often zero, and information about the person is expressed by verb inflections (Obrębska, Obrębski, 2011: 166-167). Secondly, the plurality quotient was not calculated by directly counting all words in the plural (as in the classical version by S. Ertel), but only those grammatical categories where the number form is morphologically marked, i.e., nouns and verbs, were taken into account. This was due to the small volume of the analyzed texts and the need to ensure the correctness of the results (Obrębska, Obrębski, 2011, pp. 167-168). Thirdly, two separate formulas were used to study the nominalization quotient: (a) the ratio of nouns and verbs in the text as a whole, (b) the ratio of nominalized verb forms to their non-nominalized counterparts. This approach differs from S. Ertel's original model, which was based on pre-prepared lists of lexemes (Obrębska, Obrębski, 2011: 168).

Analysis of research. In both international and Ukrainian linguistics, the study of statistical markers of idiolect is increasingly carried out through corpus-based and computational methods. For instance, in his work Statistical Analysis of the Short Stories by Roman Ivanychuk, I. Kulchytskiy demonstrates that pronouns and nouns are the most frequently used parts of speech in Ivanychuk's prose. This observation suggests that the narrator often assumes the role of the central character in his stories. Moreover, the degree of euphony in Ivanychuk's texts closely corresponds to the general tendencies of Ukrainian prose. With regard to sentence length, Ivanychuk tends to employ longer constructions in comparison with other Ukrainian writers, a tendency that serves as a distinctive marker for attributing authorship to his texts (Kulchytskiy, 2019: 312).

As demonstrated in the study by L. Shytyk and T. Shcherbyna (2024), quantitative parameters such as text length, vocabulary size, diversity index, number of hapax legomena, as well as indices of exclusivity and readability, serve as informative criteria for constructing a statistical profile of an author's writing. An analysis of Lyuko Dashvar's literary works revealed that, despite noticeable differences among the novels in terms of length and the number of syntactic units, general stylistic stability indices (particularly lexical diversity and the proportion of high-frequency vocabulary) remain relatively consistent (Shytyk, Shcherbyna, 2024: 102).

Statistical markers of idiolect have also been widely applied in sociolinguistic research. According to V. Shymko, the specific features of an individual's idiolect can be identified through quantitative indicators such as lexeme frequency distribution, syntactic constructions, lexical preferences, and patterns of speech activity – all of which are amenable to statistical processing. The use of quantitative linguistic tools, including cluster analysis, correlation metrics, and machine-learning algorithms, has enabled the systematic detection of idiolectal traits across large corpora of textual and audiovisual data from digital communication on social media to spontaneous spoken discourse (Shymko, 2025: 10).

Thus, statistical markers of a writer's idiolect constitute essential tools for analyzing and defining the distinctive features of authorial style. They provide objective data that remain independent of the researcher's subjective interpretation. Applying these markers in text analysis makes it possible to uncover the unique traits of each author's style, including lexical preferences and characteristic linguistic constructions.

Methods such as word frequency counts, lexical diversity measures, and indices of exclusivity and concentration help to identify authorship, detect plagiarism, and trace the stylistic evolution of writers. Since every author tends to rely on certain "preferred" words and constructions, these recurrent patterns contribute to the uniqueness of an individual style.

With the advent of modern computational technologies and corpus-linguistic methods, the study of statistical markers of idiolect has become increasingly precise and efficient. These approaches not only allow for the automatic attribution of texts but also facilitate the examination of other stylistic characteristics – such as sentence and word length, frequency of punctuation use, and the identification of distinctive lexical units.

Topicality. Contemporary linguistics increasingly employs quantitative methodologies to investigate the individual style of writers. The application of statistical indicators, particularly the speech style indices developed by S. Ertel, provides opportunities for an objective characterization of idiolects and for identifying underlying cognitive and stylistic strategies of authors. The analysis of Oksana Zabuzhko's works through the lens of these indicators is relevant, as it enables the detection of distinctive features of her writing in the context of the contemporary literary process. This methodology provides a foundation for automated authorship detection.

Object of the Research. The object of the research is the idiolect of Oksana Zabuzhko in comparison with the texts of her contemporaries.

Subject of the Research. The subject of the research is the set of quantitative stylometric indicators – namely, the Impersonality Quotient, the Plurality Quotient, the Nominalization Quotient, the Epithetization Index, and their manifestations in the texts of Oksana Zabuzhko compared to the corpus of texts by her contemporaries.

Purpose of the Research. The purpose of the study is to analyze Oksana Zabuzhko's idiolect by means of selected quantitative stylometric indicators and to determine its specificity in a comparative perspective.

Overview of the Research. The study analyzes texts by O. Zabuzhko and her contemporaries, in particular, calculating the impersonality quotient (according to Obrębska, Obrębski) (Obrębska, Obrębski, 2011: 166-167). This index makes it possible to assess the extent to which the author's speech is oriented toward their own person or interlocutor (through first- and second-person pronouns and corresponding verb forms) or, conversely, focused on third persons, events, or objects in the external world. The impersonality quotient was calculated as the ratio of the number of third-person pronouns to the total number of personal pronouns (first, second, and third person), taking into account both pronouns and morphological features of verbs (in particular, the category of person). The analysis yielded the following values: O. Zabuzhko (0.556), male contemporaries (0.538), female contemporaries (0.624) (see Diagram 1).
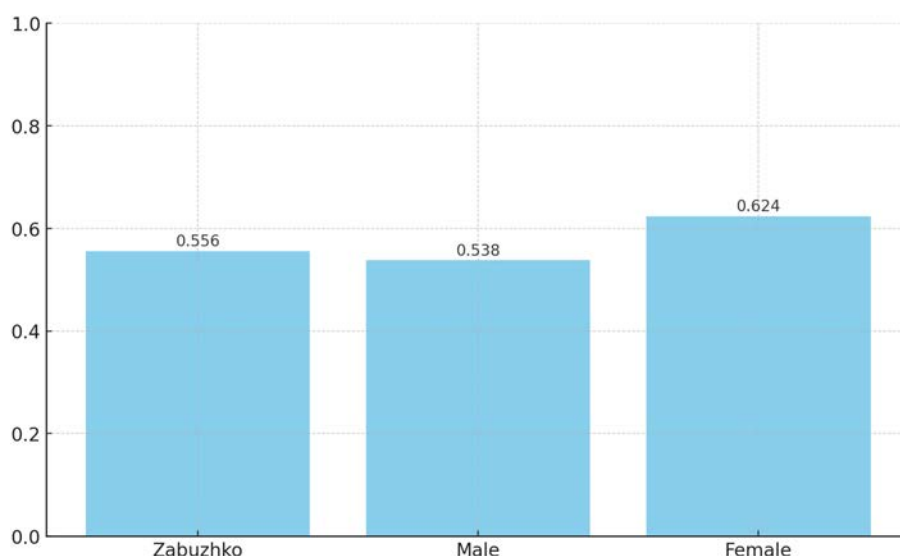

Diagram 1. The Impersonality Quotient in the Texts of O. Zabuzhko and Her Contemporaries

The highest indicator of impersonality quotient was recorded in the subcorpus of texts by Zabuzhko's contemporaries (0.624), which may indicate the dominance of a distanced, objectified speech strategy focused on describing external reality rather than subjective self-interpretation. This may be a sign of a general stylistic tendency among female authors in the contemporary literary field to formalize narrative or avoid direct emotional self-expression (Ryzhkova, 2008: 2-23). In contrast, the subcorpus of texts by Zabuzhko's contemporaries shows the lowest level of impersonal speech (0.538). Oksana Zabuzhko's texts demonstrate an average level of impersonality of speech (0.556). This balance between the presence of the pronoun "I" and objectified narration may indicate an intellectual style of writing that combines autobiographical reflections with philosophical and critical analysis. Her authorial strategy is distinguished by a dynamic combination of personal voice and cultural-social discourse, which corresponds to her identity as a public intellectual (Galich, 2022: 192-213), (Tomusiak, Raichuk, 2023: 104-108). In addition, the plurality quotient has been calculated (see Diagram 2).

Instead of counting all words in the plural, the index was defined as the proportion of words in the plural form among the total number of nouns and verbs in the text. In other words, an automated morphological analysis of each text was performed to identify only those parts of speech that grammatically express number. Next, the ratio of forms with the feature quantity=plural to the total number of forms marked as nouns or verbs was determined (Obrębska, Obrębski, 2011: 167-168).

According to the results obtained, the plurality quotient in O. Zabuzhko's texts is 0.1859, which is lower than that of her male contemporaries (0.2217) and female contemporaries (0.2087) (see Diagram 2). Higher index values in the subcorpora of her male and female contemporaries indicate a higher frequency of plural forms, which, according to S. Ertel's interpretation, indicates a tendency toward generalization, social representativeness, or impersonal presentation.

In contrast, O. Zabuzhko displays a somewhat narrower cognitive perspective, a tendency toward singular experiences and individualized expressions, which indicates a heightened subjectivity in speech.

One of the key parameters for analyzing idiolect is the nominalization quotient – the ratio between the use of noun and verb forms in the text. This coefficient reflects a tendency toward abstraction, generalization, and substantivization of actions (Obrębska, Obrębski, 2011, p. 168), which indicates the cognitive complexity and intellectualization of written discourse. The calculation of the nominalization quotient, defined as the ratio of the number of nouns to the sum of nouns and verbs in the text, allows us to quantitatively assess the degree of substantivation of speech as a manifestation of the author's abstract and conceptual thinking. The study analyzed texts by Oksana Zabuzhko (0.5814), a corpus of contemporary Ukrainian prose (0.5507), and a corpus of texts by the author's contemporaries (0.5477) (see Diagram 3).

The results show a higher frequency of noun forms in O. Zabuzhko's texts, which indicates her tendency toward cognitive condensation of information and abstraction of phenomena and actions. Comparatively lower indicators in the corpora of her contemporaries may indicate a more narrative-descriptive or expressive-emotional nature of writing, with an emphasis on the dynamics of action rather than on the nominative definition of concepts. Thus, the coefficient attests to the functional and stylistic specificity of O. Zabu-
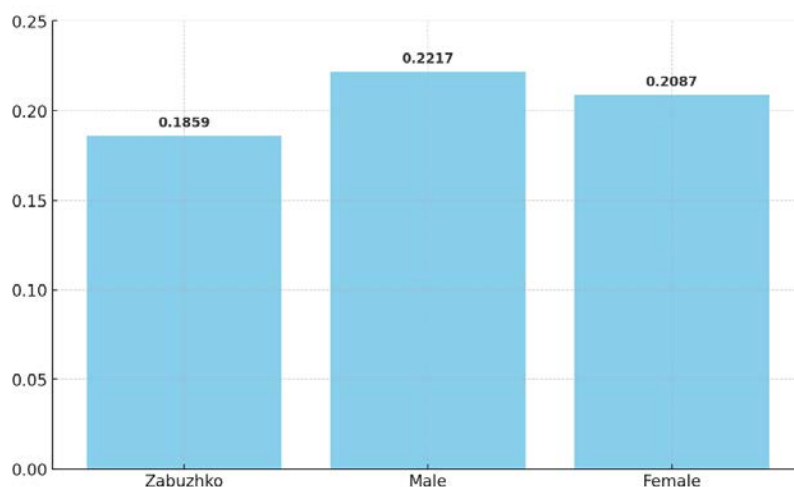

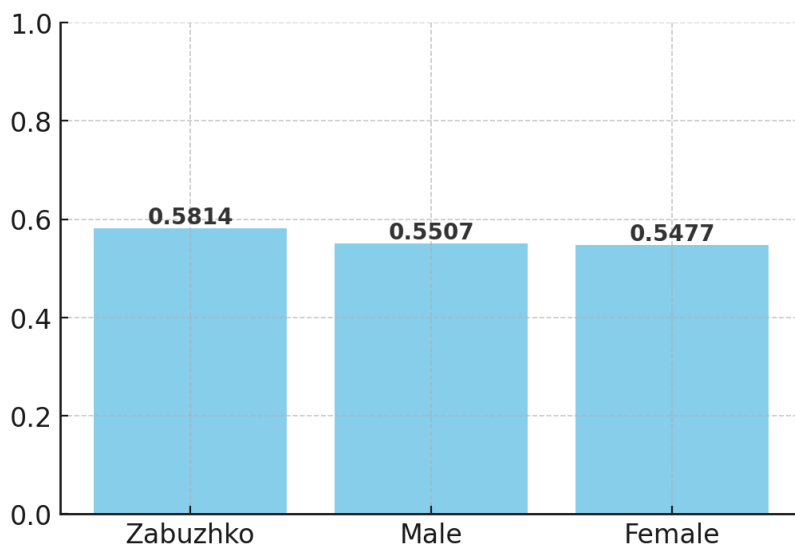**Diagram 2. The Plurality Quotient in the Texts of O. Zabuzhko and Her Contemporaries**

Diagram 3. The Nominalization Quotient in the Texts of O. Zabuzhko and Her Contemporaries

zhko's idiolect, which tends toward journalistic and philosophical stylistics with a higher degree of generalization and abstraction.

The epithetization index is an important indicator of the stylistic expressiveness of a text and allows for a quantitative characterization of the frequency of use of attributes. This index is calculated as the ratio of the number of nouns to the number of adjectives in the text: the lower the value, the more epithetized the speech is. The epithetization index as a statistical marker is widely used in the study of writers' idiolects, the analysis of stylistic norms of a particular era or genre, as well as in the comparison of language practices within the framework of comparative corpus analysis (Buk, 2012: 73-85).

According to the results of calculating the epithetization index, this indicator is 4.28% (0.0428) in Oksana Zabuzhko's texts, while in the corpus of her male contemporaries it is 3.95% (0.0395), and in the corpus of her female contemporaries it is 3.59% (0.0359). This indicates a higher frequency of adjectival descriptions in Oksana Zabuzhko's speech compared to her literary environment.

The data obtained confirm an increased level of epithetization in O. Zabuzhko's texts, which indicates the emotional and semantic saturation of the text. Thus, the writer's idiolect is stylistically expressive.

Conclusions. The results confirm the feasibility of using quantitative stylometric indicators developed by S. Ertel and modified by Obrembski to analyze the individual style of writers. Oksana Zabuzhko's texts have a number of features that testify to the specificity of her idiolect: a combination of personal and objective modes of linguistic representation, a tendency toward individualized experience while maintaining a high level of cognitive abstraction, and increased stylistic expressiveness.

Compared to the texts of her contemporaries, O. Zabuzhko's idiolect is characterized by a lower level of plurality, which indicates an orientation towards individual experience, and a higher level of nominalization and epithetization, which indicate the intellectualization and metaphorization of speech. Thus, the identified quantitative parameters make it possible to trace the author's individual linguistic and stylistic features.

BIBLIOGRAPHY

1. Darchuk N., et al. The System for Automatic Stylometric Analysis of Ukrainian Media Texts TextAttributor 1.0 (Techniques, Means, Functionality). Acta Linguistica Lithuanica. 2024. № 91. C. 224–247.

2. Kulchytskyi I. Statistical Analysis of the Short Stories by Roman Ivanychuk. In: COLINS. 2019. C. 312–321.

3. Obrębska M., Obrębski T. Style mówienia w schizofrenii. Analiza porównawcza z wykorzystaniem Wskaźników Stylów Mówienia Suitberta Ertela. Czasopismo Psychologiczne. 2011. Vol. 17, No. 2. P. 161–170.

4. Pavlyshenko B. Classification analysis of authorship fiction texts in the space of semantic fields. Journal of Quantitative Linguistics. 2013. Vol. 20, No. 3. P. 218–226.

5. Plecháč P., Flaišman J. Problém Barák-Neruda z pohledu současné stylometrie. Česká literatura. 2017. P. 743–769.

6. Shymko V. Identifying Key Idiolect Markers in Sociolinguistic Profiling: A Scoping Review and Analytical Framework for Real-world Applications. SAGE Open. 2025. T. 15, № 2. C. 21582440251334276.

7. Шитик Л., Щербина Т. Методи квантитативної лінгвістики в дослідженні ідіостилю письменника (на матеріалі творів Люко Дашвар). Linguistic Bulletin. 2024. Вип. 37. С. 102–111.

REFERENCES

1. Darchuk, N., et al. (2024). The System for Automatic Stylometric Analysis of Ukrainian Media Texts TextAttributor 1.0 (Techniques, Means, Functionality). Acta Linguistica Lithuanica, 91, 224–247.

2. Kulchytskyi, I. (2019). Statistical Analysis of the Short Stories by Roman Ivanychuk. In COLINS (pp. 312–321).

3. Obrębska, M., & Obrębski, T. (2011). Style mówienia w schizofrenii. Analiza porównawcza z wykorzystaniem Wskaźników Stylów Mówienia Suitberta Ertela [Styles of speech in schizophrenia. A comparative analysis using Suitbert Ertel's Speech Styles Quotients]. Czasopismo Psychologiczne, 17(2), 161–170. [in Polish].

4. Pavlyshenko, B. (2013). Classification analysis of authorship fiction texts in the space of semantic fields. Journal of Quantitative Linguistics, 20(3), 218–226.

5. Plecháč, P., & Flaišman, J. (2017). Problém Barák-Neruda z pohledu současné stylometrie [The Barák–Neruda problem from the perspective of contemporary stylometry]. Česká literatura, 743–769. [in Czech].

6. Shymko, V. (2025). Identifying Key Idiolect Markers in Sociolinguistic Profiling: A Scoping Review and Analytical Framework for Real-world Applications. SAGE Open, 15(2), 21582440251334276.

7. Shytyk, L., & Shcherbyna, T. (2024). Metody kvantytatyvnoi linhvistyky v doslidzhenni idiostyliu pysmennyka (na materiali tvoriv Liuko Dashvar) [Methods of Quantitative Linguistics in the Study of Writer's Idiostyle (Based on the Works of Liuko Dashvar)]. Linguistic Bulletin, 37, 102–111. [in Ukrainian].