# Statistical profile of O. Zabuzhko's idiostyle

Olena Levchenko†, Vitalii Karasov*,†

*Lviv Polytechnic National University, S.Bandera str., 12, Lviv, 79000, Ukraine*

**Abstract**

The study presents the results of research aimed at identifying markers of O. Zabuzhko's idiostyle through statistical analysis of her textual corpus, compared with writings by Y. Andrukhovych and M. Matios for reference. Utilizing the Python programming language in conjunction with the Natural Language Toolkit (NLTK) and the GRAK-17 corpus, we applied methodologies grounded in computational, quantitative, and corpus linguistics to examine the statistical attributes inherent to O. Zabuzhko's idiostyle. The analysis encompassed an examination of the following characteristics: TF-IDF indicators, text readability indices, emotional tone analysis, lexical diversity metrics, ratios between authorial and dialogic speech components, and the usage of specific groups of onymy, particularly toponyms. The statistical significance of the findings was assessed using Student's t-test. It was determined that the greatest occurrence of toponyms was observed in the texts authored by Y. Andrukhovych, whereas O. Zabuzhko's texts exhibited notably higher complexity, as evidenced by longer sentence and word lengths. Furthermore, textual structure analysis revealed a more pronounced ratio between authorial and dialogic speech, particularly evident in the works of O. Zabuzhko and M. Matios.

**Keywords**

Idiostyle, idiostyle markers, toponym, statistical analysis, Python, NLTK, Corpus

## 1. Introduction

The advancement of computer methods for processing natural language data has led linguists to increasingly employ statistical techniques in examining the idiostyle of specific authors, particularly for authorship attribution. Stylometry, a burgeoning field, is focused on discerning and analyzing the statistical attributes of distinct functional styles of language or speech associated with particular individuals [1]. Stylometric inquiries aim not only to ascertain typology but also to address attribution challenges (including authorial, temporal, and stylistic attribution, with potential applications in forensic and criminal linguistics), conduct diagnostics, and undertake text and segment reconstructions.

Here are several commonly utilized indicators in stylometric analysis:

- Word Frequency Analysis;

*CLW-2024: Computational Linguistics Workshop at 8th International Conference on Computational Linguistics and Intelligent Systems (CoLInS-2024), April 12–13, 2024, Lviv, Ukraine*

* Corresponding author.

† These authors contributed equally.

olena.p.levchenko@lpnu.ua (O. Levchenko); vitalii.v.karasov@lpnu.ua (V. Karasov)

0000-0002-7395-3772 (O. Levchenko); 0000-0002-8039-2811 (V. Karasov)

- N-gram Analysis;
- Function Word Analysis;
- Sentence and Paragraph Lengths;
- Lexical Diversity;
- Syntactic Analysis;
- Authorship Attribution;
- Authorial Signature;
- Machine Learning Approaches.

Thus, apart from traditional methods of authorship attribution rooted in philological approaches, there exist non-traditional research methodologies that leverage computer-based and statistical analysis. Within the domain of quantitative linguistics, these unconventional approaches rely on a series of linguistic and statistical hypotheses. Specific attributes within the approaches rely on a series of linguistic and statistical hypotheses. Specific attributes within a written text can act as distinguishing markers, disclosing the author's identity irrespective of social or historical contexts. For instance, M. Eder in study Style-Markers in Authorship Attribution A Cross-Language Study of the Authorial Fingerprint, posits that a sample text, such as a novel, authored by a renowned writer, is considered representative of their overall "language" usage. Importantly, Eder underscores that statistical analysis operates under the assumption that investigating common phenomena, like function words, provides more dependable conclusions compared to analyzing rare occurrences, such as hapax legomena [2].

To characterize idiostyle, researchers utilize various indicators including sentence and word length, punctuation mark frequency, parts of speech frequency, measures of lexical diversity (linguistic richness), metrics of rhythmicity, emotional tone of the text, and other related factors [3-7].

The paper proposes a study examining the statistical characteristics of O. Zabuzhko's idiostyle, including TF-IDF indicators, text readability indices, emotional tone analysis, lexical diversity metrics, ratios between authorial and dialogic speech volumes, and the utilization of specific groups of onymy, notably toponyms. All indicators were compared with those of the author's contemporaries, namely Y. Andrukhovych and M. Matios. Additionally, the statistical significance of these indicators was assessed.

To conduct the study, subcorpora consisting of texts authored by O. Zabuzhko, Y. Andrukhovych, and M. Matios were created. The investigation leveraged Python, an interpretable, object-oriented high-level programming language. Furthermore, the study made use of various libraries for natural language processing, such as NLTK, along with the GRAK-17 corpus, to process the data.

## 2. Related works

In both global and Ukrainian linguistics, the base of research on quantitative textual parameters and authorship attribution has experienced significant growth in recent decades [8-11]. Researchers have focused on developing statistical profiles for authors, while others have concentrated on identifying markers for idiostyles or language varieties [12-14]. Additionally, scientists working with stylometry have made considerable contribution to the field [15-17].

In the paper, The Statistical Parameters of Ivan Franko's Authorial Style Determined by the Chi-square Test, the authors employ a methodology that integrates the analysis of both absolute and relative frequencies of Ukrainian letters alongside the chi-square test. This approach aims to identify Ivan Franko's authorial style. Their findings indicate that when the differences between literary works are not statistically significant, it suggests that these works were authored by the same writer [18].

In her work, Statistical Research of the Color Component ЧОРНИЙ (BLACK) in Roman Ivanychuk's Text Corpus, N. Lototska undertakes a statistical examination of phrases containing the color designation ЧОРНИЙ (BLACK) within Roman Ivanychuk's fictional works. N. Lototska reveals that BLACK emerges as the most prevalent color designation in Ivanychuk's textual corpus. Employing a corpus-based approach, Lototska employs metrics such as absolute and relative frequency, as well as MI-score and t-score, to describe and analyze the word combinations [19].

In their work, Quantitative Parameters of Lucy Montgomery's Literary Style, N. Hrytsiv, T. Shestakevych, and Y. Shyika focus on fundamental aspects of quantitative comparative analysis and key elements of statistical linguistics. Specifically, they emphasize metrics such as the coefficient of diversity, the average frequency of lexeme repetitions within the text, the coefficient of uniqueness, and the index of vocabulary saturation [20]. These parameters enable an exploration of Lucy Montgomery's literary style through quantitative analysis.

In the paper, The Linguometric Approach for Co-authoring Author's Style Definition, authors employ linguometry technologies to discern the authorial style in publications. Through statistical linguistic analysis, the text undergoes content monitoring facilitated by Stemming Algorithms like the Stemmer by Martin F. Porter and NLP methods to identify a set of stop words. These stop words play a pivotal role in linguometric methods, aiding in the computation of text diversity coefficients to ascertain the degree of attribution to a particular author. Furthermore, a formal approach is proposed for defining the author's style in Ukrainian-language texts. Experimental results of this method, assessing the degree of authorship attribution of analyzed text to a specific author, are obtained by providing another author's text fragment as a reference. The experimental research is conducted using technical Ukrainian-language texts [21].

## 3. Methods

The statistical profile of O. Zabuzhko's idiostyle is enriched by several parameters. Previous research indicates that a distinctive trait of O. Zabuzhko's idiostyle is the length of sentences, which averages 58.5% longer than those of her contemporaries, along with the use of punctuation marks. The longest sentence observed comprises 862 words, while the longest word spans 67 letters. Additionally, the author demonstrates a tendency for creating new lexical units through the combination of two or more bases, known as compounds [3]. Furthermore, statistical analysis highlights the significance of the noun "anger" in the author's texts [4].

The TF-IDF indicator was employed to examine the statistical attributes of O. Zabuzhko's idiostyle alongside her contemporaries. TF-IDF (Term Frequency-Inverse Document Frequency) calculates the significance of each word within a document relative to its

frequency not only in that document but also across the entire collection of texts [22]. This method facilitates the identification of keywords and determines which words are more important within the context of the entire corpus. Leveraging the data derived from TF-IDF analysis, various tasks such as keyword extraction, document similarity measurement, and text classification can be effectively executed.

To obtain the results, we utilized libraries such as TfidfVectorizer for TF-IDF vectorization and NLTK for text preprocessing tasks. The text samples were tokenized, converted to lowercase, and punctuation removed. Additionally, stop words specific to the Ukrainian language were excluded, and lemmatization was applied to normalize the tokens. Subsequently, a TF-IDF vectorizer was created, and the preprocessed text was inputted to generate a TF-IDF matrix representation. From this matrix, TF-IDF scores were computed for each word within the text, indicating its significance within the document relative to the entire corpus.

The usage of toponyms in the authors' works was conducted using Python. SpaCy, a natural language processing library, was employed to extract place names (geographical names) from the Ukrainian text. By loading a model of the Ukrainian language and defining a function to extract place names based on the "LOC" label assigned to geographic entities, the script adeptly identifies and calculates place names within the text.

The Flesch-Kincaid Grade Level text readability index was calculated using the following formula:

$$FKGL = 0.39\left(\frac{Average\ words\ per\ sentence}{Average\ syllables\ per\ word}\right) + 11.8\left(\frac{Average\ syllables\ per\ word}{Average\ word\ per\ sentence}\right) - 15.59 \quad (1)$$

Utilizing the Python programming language, the emotional tone of each text was assessed using the TextBlob library. Classification into positive, neutral, or negative categories was based on polarity scores ranging from -1.0 to 1.0. Scores closer to -1.0 signify a predominantly negative tone, while those nearing 1.0 denote a largely positive tone; a score of 0 indicates neutral tone.

The lexical diversity of texts was evaluated using Python, employing the NLTK library for text processing. Specifically, we utilized functions from NLTK such as word_tokenize and stopword corpora. Additionally, Matplotlib, a widely used graphing library, facilitated the visualization of lexical diversity through a bar chart. The algorithm was structured to calculate lexical diversity and plot the outcomes. It began by processing sample texts from various Ukrainian authors, tokenizing the texts with NLTK tokenizer, eliminating stop words, and subsequently generating a histogram to depict the lexical diversity displayed by each author.

## 4. Results and discussion

In examining the statistical properties of the idiostyle, the TF-IDF metric was employed to assess the significance of words within the authors' text corpus. Consequently, keyword lists were derived for the texts of O. Zabuzhko, Y. Andrukhovych, and M. Matios (refer to Table 1 and Table 2). Notably, onymy were excluded from the analysis.

**Table 1**
The significance of words within a corpus of texts

| Zabuzhko | TF-IDF | Andrukhovych | TF-IDF |
|---|---|---|---|
| життя (life) [zhyttia] | 0,216 | життя (life) [zhyttia] | 0,103 (3) |
| пам'ятати (remember) [pamiataty] | 0,171 | пам'ятати (remember) [pamiataty] | 0,110 (2) |
| мати (have) [maty] | 0,117 | мати (have) [maty] | 0,083 (5) |
| нема (no) [nema] | 0,116 | – | – |
| знати (know) [znaty] | 0,111 | – | – |
| взагалі (generally) [vzahali] | 0,109 | – | – |
| разом (alongside) [razom] | 0,105 | разом (alongside) [razom] | 0,082 (6) |
| очі (eyes) [ochi] | 0,100 | – | – |
| мить (moment) [myt] | 0,092 | – | – |
| він (he) [vin] | 0,090 | – | – |

**Table 2**
The significance of words within a corpus of texts

| Zabuzhko | TF-IDF | Matios | TF-IDF |
|---|---|---|---|
| життя (life) [zhyttia] | 0,216 | життя (life) [zhyttia] | 0,150 (3) |
| пам'ятати (remember) [pamiataty] | 0,171 | – | – |
| мати (have) [maty] | 0,117 | мати (have) [maty] | 0,109 (8) |
| нема (no) [nema] | 0,116 | нема (no) [nema] | 0,096 (9) |
| знати (know) [znaty] | 0,111 | – | – |
| взагалі (generally) [vzahali] | 0,109 | – | – |
| разом (alongside) [razom] | 0,105 | – | – |
| очі (eyes) [ochi] | 0,100 | очі (eyes) [ochi] | 0,140 (4) |
| мить (moment) [myt] | 0,092 | – | – |
| він (he) [vin] | 0,090 | – | – |

The data reveals that the lexeme *життя* (life) [zhyttia] appears in the list of both the author and her contemporaries. Additionally, the verb *пам'ятати* (remember) [pamiataty] holds the second position in the lists of O. Zabuzhko and Y. Andrukhovych. Furthermore, the verb *мати* (have) [maty] is common among all three authors, albeit with varying frequencies within their respective corpora.

The hypothesis regarding the statistical significance of the lexeme *життя* (life) [zhyttia] in O. Zabuzhko's texts was tested utilizing a modified Student's t-test (refer to Table 3).

**Table 3**
Frequency of lexeme *життя* (life) [zhyttia]

| Corpora/Student's t-test | Relative Frequency of lexeme *життя* (life) [zhyttia]/ Student's t-test results |
|---|---|
| Zabuzhko | 0,115 |
| Subcorpus of literary texts | 0,096 |
| **MST**[2] | 6,04 |
| Andrukhovych | 0,086 |
| **MST** | 9,67 |
| Matios | 0,176 |
| **MST** | 13,83 |

The findings indicate statistical significance, supporting the hypothesis that the lexeme *життя* (life) [zhyttia] in the author's works the compared texts is statistically significant and can serve as a marker of idiostyle.

During the analysis of O. Zabuzhko's texts, it was identified the most common collocations following the "adjective + *життя* (life)" [zhyttia] model and juxtaposed them with those found in the texts of her contemporaries (refer to Table 4).

**Table 4**
The most frequent collocations of the adjective + *життя* (life) [zhyttia] model

| Zabuzhko | RF[3] | Andrukhovych | RF | Matios | RF |
|---|---|---|---|---|---|
| людський (human) [liudskyi] | 0,00555 | людський (human) [liudskyi] | 0,01051 | людський (human) [liudskyi] | 0,00591 |
| новий (new) [novyi] | 0,00196 | новий (new) [novyi] | 0,00370 | новий (new) [novyi] | 0,00414 |
| український (Ukrainian) [ukrainskyi] | 0,00131 | український (Ukrainian) [ukrainskyi] | 0,00061 | український (Ukrainian) [ukrainskyi] | 0,00059 |
| свідомий (conscious) [svidomyi] | 0,00098 | Свідомий (conscious) [svidomyi] | 0,00061 | свідомий (conscious) [svidomyi] | 0,00059 |
| попередній (previous) [poperednii] | 0,00065 | попередній (previous) [poperednii] | 0,00061 | попередній (previous) [poperednii] | 0,00059 |
| справжній (real) [spravzhnii] | 0,00032 | справжній (real) [spravzhnii] | 0,00061 | справжній (real) [spravzhnii] | 0,00059 |

The compounds most frequently occurring in O. Zabuzhko's texts include those composed of the elements *людський* (human) [liudskyi], *чужий* (foreign) [chuzhyi],

[2] Modified Student's Test
[3] Relative frequency

*минулий* (past) [mynulyi], *громадський* (public) [hromadskyi], and *український* (Ukrainian) [ukrainskyi] (refer to Figure 1).

A modified Student's t-test was employed to assess the statistical significance of the "adjective + *життя* (life)" [zhyttia] model. The obtained results are presented in Table 5.

**Table 5**

Frequency of adjective + *життя* (life) [zhyttia] model

| Corpora/Student's t-test | Relative Frequency of Adj + N model/ Student's t-test results |
|---|---|
| Zabuzhko | 0,0213 |
| Andrukhovych | 0,0150 |
| **MST** | 4,9317 |
| Matios | 0,0335 |
| **MST** | 6, 3495 |

The model was determined to be statistically significant, thereby confirming our hypothesis.

The hypothesis regarding the significance of certain groups of onymy, particularly toponyms, for the author's idiostyle was examined. The results are summarized in Table 6, which indicates that the greatest number of toponyms appears in the texts of Y. Andrukhovych. Employing a modified Student's t-test, we assessed the statistical significance of these findings. It was determined that the frequencies of toponyms in O. Zabuzhko's texts do not exhibit statistical significance when compared to those in the texts of her contemporaries. Consequently, we infer that the frequency of place names in O. Zabuzhko's texts cannot serve as a marker of idiostyle.

**Table 6**

Frequency of toponyms

| Corpora/Student's t-test | Relative Frequency of toponyms/ Student's t-test results |
|---|---|
| Zabuzhko | 0,0021678 |
| Andrukhovych | 0,00496516 |
| **MST** | 1,625 |
| Matios | 0,00316079 |
| **MST** | 0,980 |

In M. Matios's novels, the most frequent place name is *Вижниця* (Vyzhnytsia) with a relative frequency of 0,0002: Сусідня **Вижниця** святкує малий Пурім щороку на честь місцевого бургомістра, який не дав жидів на смерть черкесам (Every year, the neighboring town of Vyzhnytsia celebrates Small Purim in honor of their local mayor, who prevented Jews from being harmed by Circassians) [Susidnia **Vyzhnytsia** sviatkuie malyi Purim shchoroku na chest mistsevoho burhomistra, yakyi ne dav zhydiv na smert cherkesam] (M. Matios, Nation, 2001).

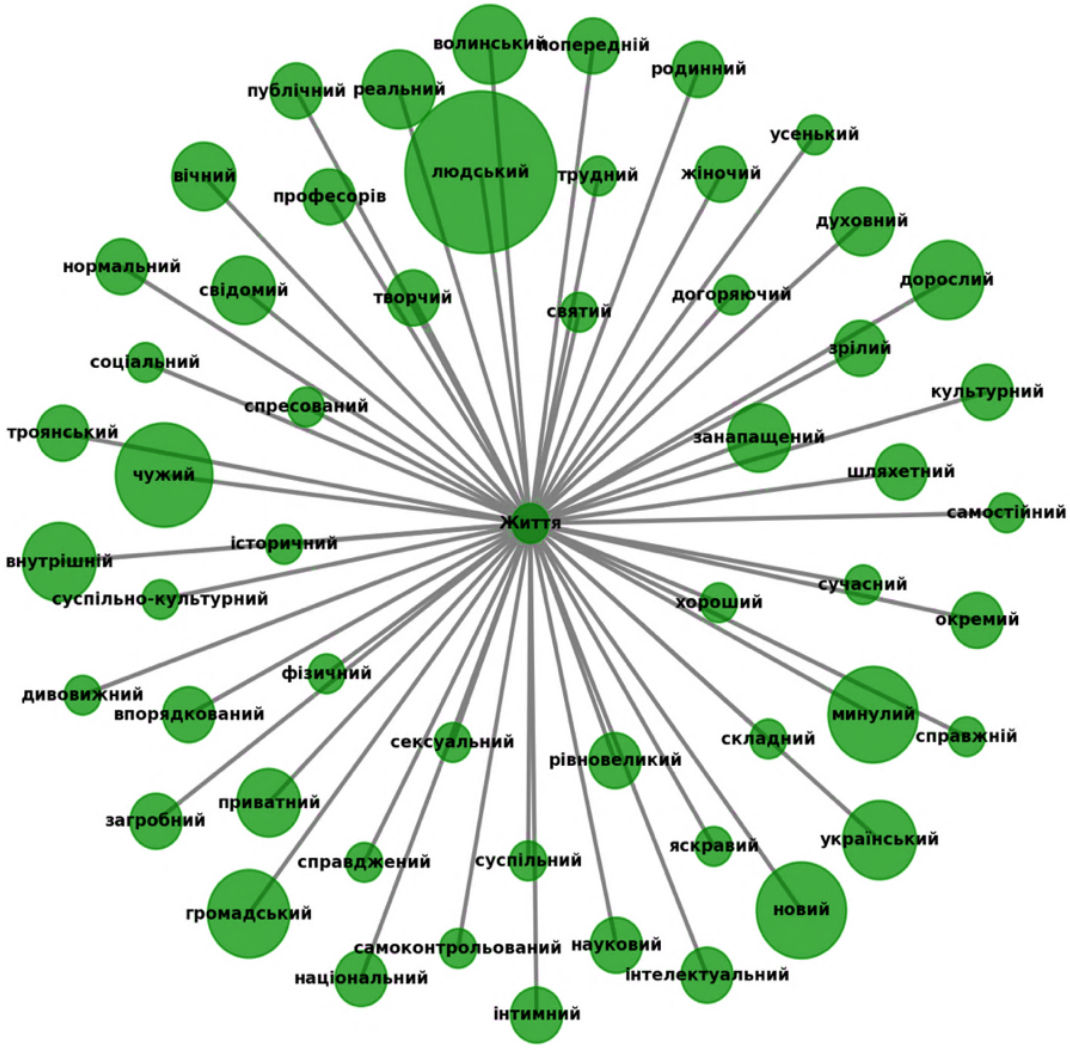


**Figure 1**: The most frequent compounds for token життя (life) [zhyttia]

It has been determined that in O. Zabuzhko's texts the most frequent toponym is *Київ* (Kyiv) with a relative frequency of 0,006: (...Ігор, замордований большевиками в дрогобицькій тюрмі так, що тільки сорочку мати впізнала на трупі, і Нестор, який пропав десь ув Аушвіці, арештований у вересні сорок першого, і Лодзьо, Лодзьо Дарецький, найздібніший із нашої кляси, що поїхав до Києва торік улітку, коли підпілля вже впало, і коли-небудь, Христом-Богом свідчу, я ще перестріну те стерво, котре вислало Лодзя до Києва просто в лапи ґестапо, щоб його там застрілили як собаку першого ж дня по приїзді…) (Ihor, murdered by the Bolsheviks in a Drohobych prison, leaving only his mother able to recognize his shirt on the corpse; Nestor, who vanished somewhere in Auschwitz after being arrested in September 1941; and Lodz, Lodz

Darecki, the most talented guy in our class, who headed to Kyiv last summer when the underground was already crumbling. And mark my words, one day, I'll come face to face with the witch who shipped Lodz off to Kyiv, straight into the clutches of the Gestapo, to be gunned down like a dog on his very first day there.) [...Ihor, zamordovanyi bolshevykamy v drohobytskii tiurmi tak, shcho tilky sorochku maty vpiznala na trupi, i Nestor, yakyi propav des uv Aushvitsi, areshtovanyi u veresni sorok pershoho, i Lodzo, Lodzo Daretskyi, naizdibnishyi iz nashoi kliasy, shcho poikhav do **Kyieva** torik ulitku, koly pidpillia vzhe vpalo, i koly-nebud, Khrystom-Bohom svidchu, ya shche perestrinu te stervo, kotre vyslalo Lodzia do **Kyieva** prosto v lapy gestapo, shchob yoho tam zastrilyly yak sobaku pershoho zh dnia po pryizdi…] (O. Zabuzhko, The Museum of Abandoned Secrets, 2009).

In Y. Andrukhovych's work, the most frequent place name is *Чортопіль* (Chortopil) with a relative frequency of 0, 040: Але цілком не виключено, що свою першу українську подорож Карл Йозеф здійснив під впливом родинного міту про прадіда, фанатично діяльного надлісничого з Ворохти, згодом службово переведеного до **Чортополя** (It's entirely plausible that Karl Josef embarked on his maiden voyage to Ukraine spurred by a family legend about his great-grandfather—an avid, fiercely dedicated forester from Vorokhta, rumored to have been officially reassigned to **Chortopil**) [Ale tsilkom ne vykliucheno, shcho svoiu pershu ukrainsku podorozh Karl Yozef zdiisnyv pid vplyvom rodynnoho mitu pro pradida, fanatychno diialnoho nadlisnychoho z Vorokhty, zghodom sluzhbovo perevedenoho do **Chortopolia**] (Y. Andrukhovych, Twelve Circles, 2003).

During the research of the characteristics of O. Zabuzhko's idiostyle and her contemporaries, another indicator examined was the Flesch-Kincaid Grade Level text readability index. The findings indicate that O. Zabuzhko's texts exhibit a significantly higher level of complexity compared to those of her contemporaries, as evidenced by metrics such as the average number of words per sentence and the average number of syllables per word (refer to Figure 2).

Therefore, it can be concluded that the data regarding the average sentence length of O. Zabuzhko's texts constitutes a significant characteristic of her idiostyle.

Lexical diversity of a writer's texts serves as one of the widely used markers of idiostyle. The following data were obtained (see Figure 3).

Analysis of the results conducted using Python and NLTK indicates that the lexical diversity of texts, when comparing data from O. Zabuzhko, Y. Andrukhovych, and M. Matios, does not yield statistically significant differences. However, it is important to note that this does not diminish the significance of the indicator itself.

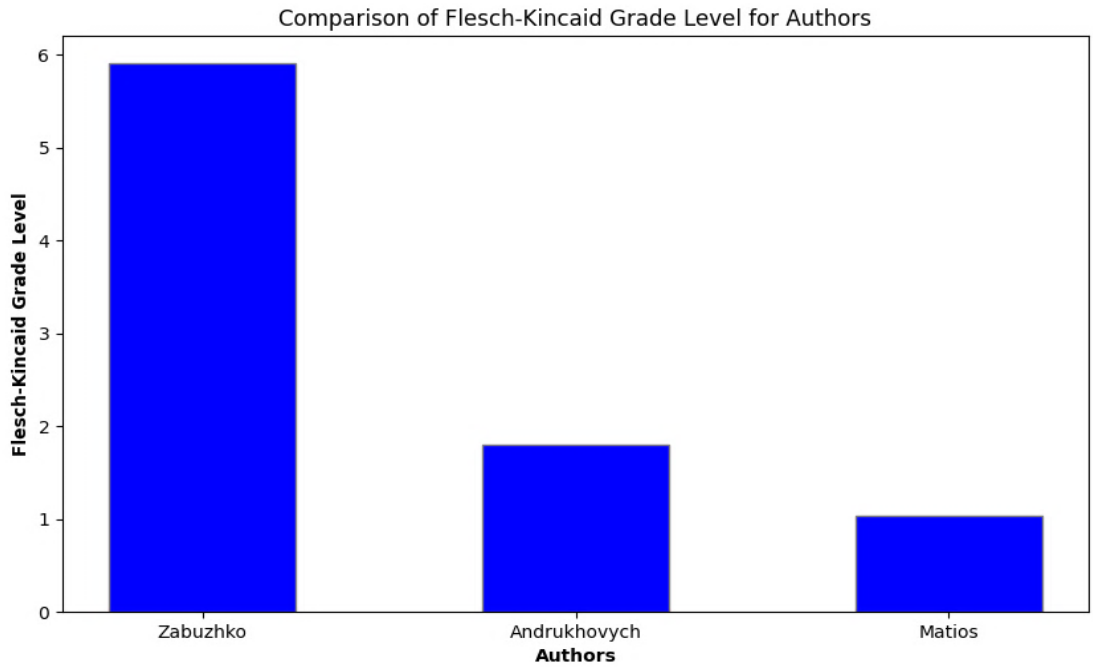


**Figure 2**: Comparison of Flesch-Kincaid Grade Level for Authors

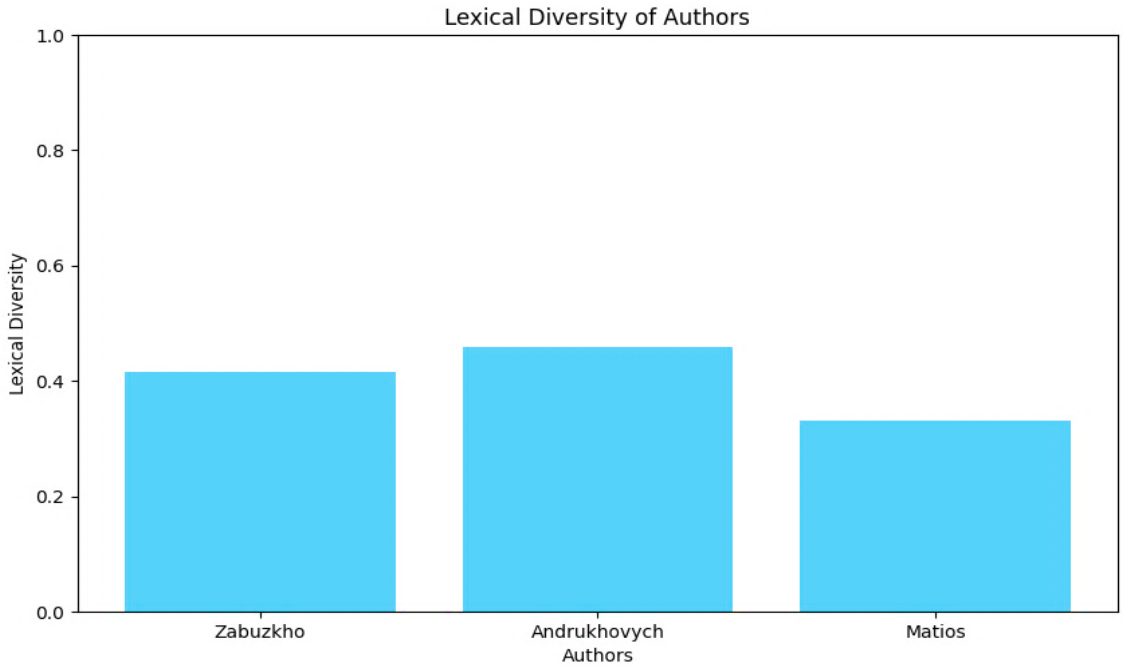


**Figure 3**: Lexical Diversity of Authors

The study scrutinized the text structure of O. Zabuzhko, Y. Andrukhovych, and M. Matios, focusing on the proportion of authorial and dialogic speech (refer to Table 7).

**Table 7**
Amount of dialogues in the corpus of texts

| Corpora/Student's t-test | Frequency of Dialogues |
|---|---|
| Zabuzhko | 1229 |
| Andrukhovych | 1108 |
| **MST** | 0,649 |
| Matios | 402 |
| **MST** | 4,020 |

The statistical significance of the results was assessed utilizing the modified Student's t-test. The data obtained demonstrate the significance of the mentioned marker of idiostyle. Specifically, the indicators concerning the amount of dialogic speech in the texts of O. Zabuzhko and M. Matios were found to be statistically significant.

An essential marker of idiostyle is the data regarding the emotional tone of texts. The following indicators were obtained during the study: Zabuzhko: 0.24, Andrukhovych: 0.22, Matios: 0.0 (refer to Figure 4).

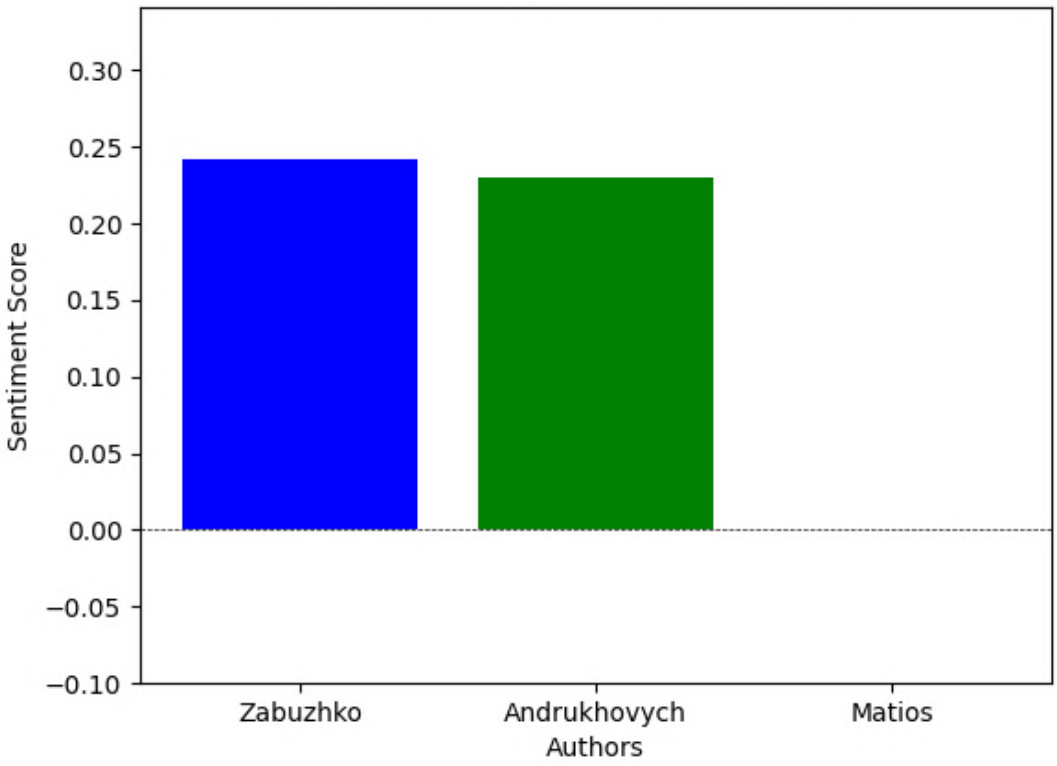


**Figure 4**: Sentiment Analysis of Ukrainian Authors

The results indicate that the emotional tone of O. Zabuzhko's and Y. Andrukhovych's texts leans towards positivity, whereas that of M. Matios' is neutral. It is important to emphasize the necessity for substantial refinement of the algorithm, particularly when applied to a broader corpus of texts. Special consideration should be given to capturing the emotional and expressive nuances embedded within metaphors found in literary texts.

## 5. Conclusions

Hence, the statistical profile of an author's texts should encompass the following metrics: TF-IDF values, text readability indices, emotional tone analysis, lexical diversity scores, ratio of authorial to dialogic speech, utilization of specific categories of onymy (particularly toponyms), sentence and word lengths, punctuation frequency, and parts of speech usage frequency.

The TF-IDF analysis conducted on texts authored by O. Zabuzhko, Y. Andrukhovych, and M. Matios revealed distinct sets of keywords. Notably, terms such as *життя* (life) [zhyttia] and *мати* (have) [maty] appear across all three corpora albeit with differing weights. Besides quantitative assessments, qualitative analysis, particularly considering the peculiarities of concept verbalization, proves crucial for characterizing the idiostyle. The core attributes of the concept of *ЖИТТЯ* (LIFE) [ZHUTTIA] are *людський* (human) [liudskyi], *чужий* (alien) [chuzhyi], *минулий* (past) [mynulyi], *громадський* (public) [hromadskyi], *український* (Ukrainian) [ukrainskyi], but the specificity of the idiostyle creates a number of low-frequency attributes that are inherent in the texts of a particular author.

It has been observed that the texts of Y. Andrukhovych contain the highest frequency of toponyms. However, the disparity in the frequency of place names between the texts of O. Zabuzhko and her contemporaries is not statistically significant. Conversely, O. Zabuzhko's texts exhibit higher complexity, notably reflected in the length of sentences, which are on average 59% longer. Additionally, an analysis of the text structure revealed a more pronounced ratio of authorial to dialogic speech, particularly evident in the works of O. Zabuzhko and M. Matios. Regarding the analysis of emotional tone in the studied texts, it is our assessment that objective results were not obtained, as the analysis algorithm did not consider the emotional and expressive aspects of metaphorical language.

The suggested research methodology ought to undergo refinement with a larger corpus of linguistic data. Such an approach, centered on the construction of statistical profiles of idiostyles will provide the valuable insights for the automatic processing of natural language information, notably in the domain of automated authorship detection.

## 6. References


[1] C. Klaussner, J. Nerbonne, Ç. Çöltekin, Finding Characteristic Features in Stylometric Analysis, in: Digital Scholarship in the Humanities, Vol. 30, 2015, pp. 114–129.

[2] M. Eder, Style-markers in authorship attribution: a cross-language study of the authorial fingerprint, in: Studies in Polish Linguistics, Vol. 6 (1), 2011.

[3] S. Buk, Quantitative analysis of the novel Ne Spytavńy Brodu by Ivan Franko in the Light of Statistical and Quantitative Linguistics, in: Speech and context, International Journal of Linguistics, Semiotics, and Literary Science, Vol. 1(6), 2014, pp. 100–112.
[4] S. Buk, Y. Krynytskyi, A. Rovenchak, Properties of autosemantic word networks in ukrainian texts, Advances in Complex Systems, Vol. 22(6), 2019, pp.1–22.
[5] V. Karasov, O. Levchenko, Statistical characteristics of O. Zabuzhko's idiolect, in: IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 2022, pp. 138–141.
[6] V. Karasov, O. Levchenko, ANGER Conceptual Metaphors in Literary Texts by Oksana Zabuzhko and Halyna Pahutyak, in: Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2023), Vol. 2, Kharkiv, Ukraine, 2023, pp. 323–333.
[7] I. Kulchytskyi, Statistical Analysis of the Short Stories by Roman Ivanychuk, in: Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2019), Vol. 1, Kharkiv, Ukraine, 2019, pp. 312–321.
[8] M. Łaziński, Słowa klucze polszczyzny w statystyce iw świadomości społecznej [Key words of the Polish language in statistics and in the public consciousness], in: Polszczyzna w dobie cyfryzacji [Polish in the age of digitization], Yearbook, 2020, pp. 267–276.
[9] M. Eder, Does size matter? Authorship attribution, small samples, big problem, in: Digital Scholarship in the Humanities, Volume 30(2), 2015, pp. 167–182.
[10] R. D. Peng, N. W. Hengartner, Quantitative Analysis of Literary Styles in: The American Statistician, Vol. 56(3), pp. 175–185.
[11] N. Lototska, Statistical Characteristics of Roman Ivanychuk's Idiolect (Based on Writer's Text Corpus), in: CEUR Workshop Proceedings, Vol. 3171, 2022, pp. 487–500.
[12] K. Geben and I. Fedorowicz, The Idiolect of Wojciech Piotrowicz: A Vocabulary of Autobiographical Prose, in: *Slavistica Vilnensis*, Vol. 65(1), 2020, pp. 87-102.
[13] I. Kulchytskyi, L. Tsiokh, M. Malaniuk, Quantitative equivalence level in poetry translation, in: IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), IEEE, 2018, pp. 51–54.
[14] I. Khomytska, V. Teslyuk, I. Bazylevych, I. Karamysheva, Automated Identification of Authorial Styles, in: Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2023), Vol. II, Kharkiv, Ukraine, 2023, pp. 323–333.
[15] A. Rovenchak, S. Buk, Part-of-speech sequences in literary text, in: Evidence from Ukrainian, Journal of Quantitative Linguistics, Vol. 25(1), 2018, pp. 1–21.
[16] D.T. Robbert, F.V. de Velde, Beyond mere text frequency: assessing subtle grammaticalization by different quantitative measures. A case study on the Dutch soort construction, in: Languages, Vol.5(4), 2020. https://doi.org/10.3390/languages5040055.
[17] R. Hou, C.-R. Huang, Robust stylometric analysis and author attribution based on tones and rimes, in: Natural Language Engineering, 2019, pp. 1–23.
[18] I. Khomytska, I. Bazylevych, V. Teslyuk, The Statistical Parameters of Ivan Franko's Authorial Style Determined by the Chi-square Test, in: Proceedings of the 17th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT (2022), Lviv, 2022, pp. 73–76.
[19] N. Lototska, Statistical Research of the Colour Component ЧОРНИЙ (BLACK) in R. Ivanychuk's Text Corpus, in: Proceedings of the 5th International Conference on

Computational Linguistics and Intelligent Systems (COLINS 2021),Vol. I, Lviv, Ukraine, 2021, pp. 486–497.
[20] N. Hrytsiv, T. Shestakevych, J. Shyyka, Quantitative Parameters of Lucy Montgomery's Literary Style, in: CEUR Workshop Proceedings, Vol. 2870, 2021, pp. 670–684.
[21] V. Lytvyn, V. Vysotska, Y. Burov, I. Bobyk and O. Ohirko, The Linguometric Approach for Co-authoring Author's Style Definition, in: *IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS)*, Lviv, Ukraine, 2018, pp. 29-34.
[22] V. Sundaram, S. Ahmed, S. A. Muqtadeer and R. Ravinder Reddy, Emotion Analysis in Text using TF-IDF, in: 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021, pp. 292-297, doi: 10.1109/Confluence51648.2021.9377159.