

# Statistical Analysis of the Short Stories by Roman Ivanychuk

Ihor Kulchytskyi<sup>[0000-0001-9550-9739]</sup>  
Lviv Polytechnic National University, 12 Bandera street,  
Lviv, Ukraine, 79013  
bis.kim@gmail.com

**Abstract.** The notion of idiostyle has been of a high interest of the modern linguistics so far. The personal style of an author possesses a significant value when it comes to the historical heritage of any language. The reason behind this is the fact that an author is the person living in a certain area, society and historical period that influence the way they speak and therefore write. When dealing with the idiolect, the best tools for researching and measuring it are the linguistic and statistics methods, as they show the quantitative characteristics representing the specific features of the writer's language in terms of numbers. The scope of statistics, though a mathematical science, is extremely broad, ranging from mathematics and social sciences to business and politics. It is known that statistics is increasingly used for linguistic research, in particular to study the features of the author's language.

**Keywords:** statistical analysis, short stories, idiostyle, euphony, length of word, length of sentences.

## 1 Introduction

The paper presents the results of the statistical analysis of the short stories by the Ukrainian writer of the 20th-21st centuries Roman Ivanychuk. He has got a completely individual and unique manner of writing. This is an artist who modernized Ukrainian history in a very professional and profound manner.

A range of different researches, articles, and dissertations have been dedicated to the specific features of the literature legacy of Roman Ivanychuk. They in details analyzed the works of the writer and compared them with the works of the other Ukrainian writers [1, 2, 3, 4, 8, 9, 15]. However, the specific features of Ivanychuk's historical prose, the individual use of lexical means, his speech and the peculiarity of his style remained unstudied what caused the topicality of the research.

The object is represented by R. Ivanychuk's discourse that is studied for individual specificity. The subject is the statistical indicators of the short stories by R. Ivanychuk.

In this paper the notion of idiostyle is interpreted as the set of those features of the author's language, which distinguish it from other authors. More narrowly, it is realized by means of certain stylistic, grammatical and lexical means of the author's writing that distinguishes one person from others [16, 17].

The artistic language reflects not only the linguistic competence of the author, and the advantages of using one or another language constructs and words, but also features of the national language [10].

The author's idiosyncrasy should be investigated in the context of his creative personality, including extra-linguistic and cultural factors and the literary era when the particular writer lived and created. A general idea of the idiosyncrasy of the writer can only be obtained by studying a large number of his texts [11].

## **2 Statistics in the language research**

In his work, the Ukrainian scholar V. Perebyinis claims that all complex systems are subject to the laws of statistics. Such systems are living organisms, economics, as well as science and human language. The scientist points out that it is necessary to use statistical methods in linguistics, since language is a complex system, and therefore it is a subject to the laws of statistics [14].

V. Perebyinis argues that any language is a system of discrete units, and any discrete unit has quantitative characteristics [14, 15]. Clearly or implicitly, but linguists recognize the fact that language has quantitative characteristics. V. Perebyinis gives examples of words such as 'more', 'less', 'many' others. Such estimates have no accuracy [10].

Statistical methods are widely used in linguistics. For example, in the Ukrainian lexicography statistics is used to compile frequency dictionaries, which help to find the most commonly used vocabulary. Also, statistical methods are related to the creation of teaching methods of a foreign language and learning optimization [12]. Statistical methods play an important role in applied linguistics (for the automatic processing of natural language and speech, for the creation of systems of machine translation and electronic dictionaries), psycholinguistics, sociolinguistics, comparative-historical linguistics, pragmalinguistics, and others [12].

The quantitative indicators provide some information about the text itself, the fact that there are certain differences between the language styles and genres that are of a statistical nature and are the basis of the so-called statistical stylistics.

One of the regularities of linguistic units in functional styles, from the standpoint of statistics, is a strict quantitative relationship between the abstract textual content prevailing in it and the type of author's mindset and peculiarities of the speech, in other words between the frequency use of words and grammatical forms with their lexical-grammatical meaning.

## **3 The research corpus**

The research corpus consists of the forty-three short stories from the following collections: 'The Rod Brings Ice' [7] that includes twenty-five short stories in general, 'Autumn Patterns' [6] with eleven short stories and 'One Loaf for Two' [5] consisting of seven short stories. The material of the research is presented in the table 1 with the number of tokens for each collection. Overall, the total number of tokens taken as a

basis of the research, makes up 97375 units. It is also to be mentioned that to receive the clean material for further processing, the texts from each collection were converted into electronic form, proofread and checked if the encoding was made correctly.

#### 4 The most frequent tokens

The first stage of the research is to identify the most frequent tokens in general. The 10 most frequent tokens are listed in the following table.

**Table 1. The most frequent tokens: top 10**

<b>Rank</b>	<b>Word</b>	<b>Quantity</b>	<b>Frequency</b>
1	<i>He (not)</i>	1909	2,12%
2	<i>I (and)</i>	1905	2,12%
3	<i>Ha (on/at)</i>	1778	1,98%
4	<i>ĪĪ (and)</i>	1539	1,71%
5	<i>B (in)</i>	1292	1,44%
6	З (from/with)	1264	1,41%
7	Що (that)	1135	1,26%
8	До (to)	1034	1,15%
9	Я (I)	966	1,07%
10	А (but/and)	914	1,02%

Obviously, among the most frequent tokens are the ones that belong to the auxiliary parts of speech. However, the first thing that attracts attention is a large number of personal pronouns.

So, the first 20 such word forms contain personal pronouns, while there are 11 personal pronouns in the set of 50 the most frequent tokens. This is so, despite the fact that, the first 20 most frequent words is the 19% of all words of the text, and the first 50 words make up the 29% of the whole selection.

The list of pronouns is presented in the table below:

**Table 2. The list of the first 20 tokens: pronouns**

<b>Rank</b>	<b>Worform</b>	<b>Quantity</b>	<b>Frequency</b>
9	Я (I)	966	1,07%
14	Він (he)	482	0,54%
15	Мене (me)	480	0,53%
16	Його (him)	449	0,50%
19	Її (her)	385	0,43%
22	Я (I)	346	0,38%

25	Мені (me)	295	0,33%
26	Вона (she)	292	0,32%
37	Ти (you)	210	0,23%
40	Ми (we)	198	0,22%
50	Їх (them)	138	0,15%

The most frequent nouns were placed in the range of ranks from 93 to 111. It has to be noted that the notion of rank is the position of the token in the selection according to its frequency. The word with the highest rank is the least frequent, while the one with the rank = 1 is the most frequent. Thus, the frequency of nouns in the s stories of R. Ivanychuk is much smaller than the frequency of pronouns. It statistically proves that most of his short stores were written from the first person singular.

**Table 3. The 5 most frequent nouns**

Rank	Wordform	Quantity	Frequency
93	Очі (eyes)	84	0,09%
95	Життя (life)	83	0,09%
101	Руки (hands)	81	0,09%
106	Обличчя (face)	79	0,09%
111	Людей (people)	75	0,08%

The next table contains the most frequent nouns and pronouns (only lemmas are given).

**Table 4. The most frequent nouns and pronouns**

Rank	Word	Quantity	Frequency	
Words	Word-forms			
1	9	Я (I)	2214	2,46%
2	14	Він (he)	1417	1,58%
3	19	Вона (she)	995	1,11%
4	37	Ти (you)	482	0,54%
5	40	Ми (we)	444	0,49%
6	54	Вони (they)	420	0,47%
7	101	Рука (hand)	261	0,28%
8	111	Людина (human)	259	0,29%
9	93	Око (eye)	207	0,22%
10	106	Обличчя (face)	129	0,14%
11	95	Життя (life)	119	0,13%

## 5 Other statistical characteristics of the short stories by R. Ivanychuk

The next stage of the research of the legacy of R. Ivanychuk was to determine such statistical characteristics as the degree of euphony of his texts, the length of words and the length of sentence and compare these indicators with the similar indicators of other Ukrainian writers and Ukrainian prose in general.

The degree of euphony was counted as the percentage of vowels, sonorant and voiced consonant phonemes taken comparing to all phonemes in the text. For this purpose, for each short story, the frequency of each phoneme was calculated on the basis of which the degree of the euphony of each short story is determined. The tool for such analysis was the author's program that transformed the sequence of letters into the sequence of phonemes according to the rules of theoretical phonology of the Ukrainian language. The results of this processing are given in a table.

**Table 7. The degree of euphony of Ivanychuk's short stories**

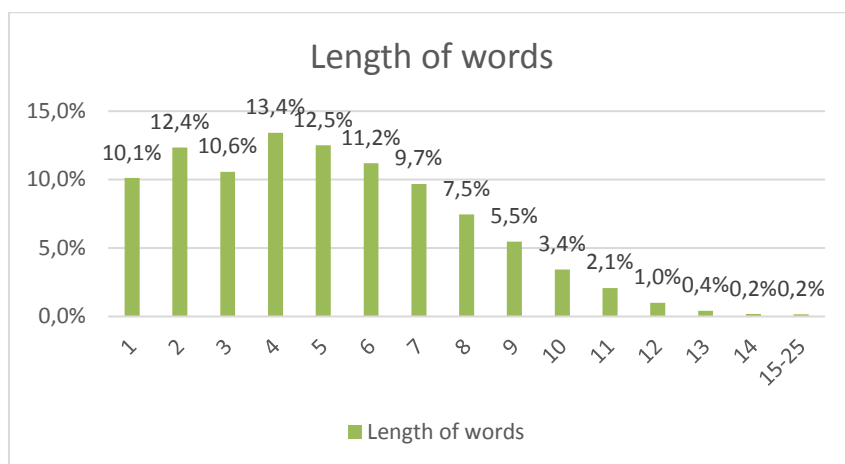
<b>Indicator</b>	<b>Value</b>
Maximum value	0,81
Minimal value	0,774
Mean value	0,7954
Mean square deviation	0,008
Medium frequency fluctuation	0,0012
Standard error	0,0012
Relative error	0,0031

The obtained results were compared to the degree of euphony of the Ukrainian fiction in general [14]. As one would expect, the value of the euphony index for the works of Roman Ivanychuk is the closest to the values of the mean euphony index of Ukrainian prose in general.

**Table 8. The mean indicators of euphony in the short stories of R. Ivanychuk and in the styles of the Ukrainian literature**

<b>Genre</b>	<b>Mean value</b>	<b>Mean square deviation</b>	<b>Relative error</b>
Drama	0,779	0,0018	0,0045
Prose	0,793	0,0024	0,0059
Poetry	0,789	0,0035	0,0088
Short Stories of R. Ivanychuk	0,795	0,008	0,0031

According to Perebyinis V. [13], the length of the words of the short stories of Roman Ivanuchyuk was measured with phonemes. The length of words distribution is shown on the figure:



**Fig. 1. The distribution of the length of words in the short stories of Roman Ivanychuk**

As it can be seen, the most frequent are words consisting of 4 phonemes – 12001 units (13.41%), 5 phonemes – 11199 units (12.51%) and 2 phonemes – 11058 units (12.35%). The middle group consists of words with the length of 6 phonemes – 10015 units (11.19%), 3 phonemes – 9449 units (10.56%), 1 phoneme – 9047 units (10.11%), 7 phonemes – 8667 units (9.68%), 8 phonemes – 6670 units (7.45%) and 9 phonemes – 4895 units (5.47%).

Words with the length of more than 10 phonemes make up only 3.8% of the whole selection: 10 phonemes – 3072 units (3.43%), 11 phonemes – 1861 units (2.08%), 12 phonemes – 906 units (1.01%), 13 phonemes – 369 units (0.41%), 14 phonemes – 155 units (0.17%), 15 phonemes – 71 (0.08%), 16 phonemes – 28 units (0.03%), 17 phonemes – 21 (0.02%), 18 phonemes – 12 units (0.01%), 19 phonemes – 5 units (0.01%), 20 phonemes – 3 units (0.0%), 22 phonemes – 2 units (0.0%), 25 phonemes – 1 units (0.0%) of all words. Words, the length of which is greater than 19 phonemes are compounds, with a hyphen, which were considered as one word. The result of the statistical processing of the distribution of the lengths of words is presented in the table below:

**Table 9. Statistical characteristics of words length in the short stories of Roman Ivanychuk**

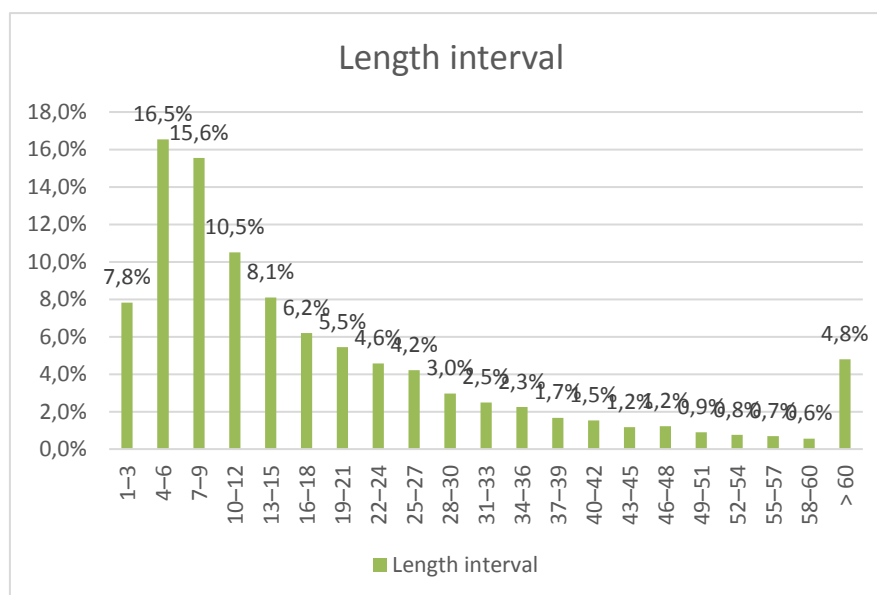
Indicator	Value
Number of different lengths	22
Mean value	5,06
Mean value	2,83
Medium frequency fluctuation	0,01
Standard error	0,01
Relative error	0

The comparison of the obtained results with the average results of the styles of the Ukrainian language shows that they are completely consistent. The average length of words in the novels by R. Ivanychuk is the closest to the works by another Ukrainian writer Yu Yanovsky.

**Table 10. Length of words in the works of some Ukrainian writers**

Other Ukrainian writers	Mean value	Mean square deviation	Relative error
А. Головка (A. Holovko)	4,74	0,1	0,034
О. Гончар (O. Honchar)	5,41	0,07	0,021
О. Довженко (O. Dovzhenko)	4,73	0,08	0,027
П. Панч (P. Panch)	5,28	0,29	0,091
М. Стельмах (M. Stelmakh)	5,3	0,16	0,048
Ю. Яновський (Yu. Ianovskui)	5,06	0,13	0,041
Новели Р. Іваничука (Short stories of R. Ivanychuk)	5,06	2,83	0

R. Ivanychuk is known for using long sentences in his works. This is a distinguishing feature of his short stores, so it was naturally to count and statistically prove this statement. The results are shown on the figure below:



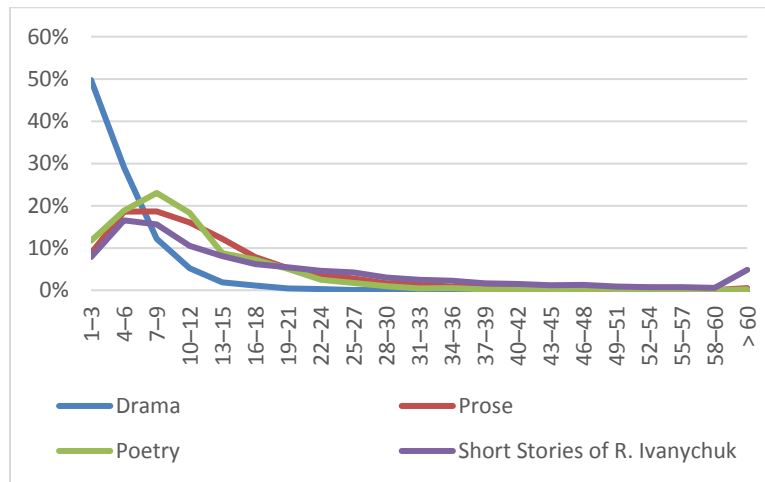
**Fig. 2. Sentences length in the short stories of Roman Ivanychuk**

From the obtained results, it is seen that R. Ivanychuk indeed used very large sentences that has more than 60 words. They make up almost 5% of all sentences he used. The highest percentage was given to the length of 4-5 words in the sentence. It is interesting to mention that sentences with the length from 16 words to 30 words are also pretty frequent.

**Table 21. The statistical indicators of the length of sentences in the short stories of R. Ivanychuk**

Indicator	Value
Quantity of different lengths	141
Mean value	19,57
Mean square deviation	22,46
Medium frequency fluctuation	0,32
Standard error	0,32
Relative error	0,03

The comparison of the obtained data to the similar data from works of different styles of the Ukrainian language (namely drama, prose, and poetry) is presented below:



**Fig. 3. The length of sentences in the different styles of the Ukrainian literature and in the short stories of R. Ivanychuk**

The only thing that can be noted after comparing the data, is that another Ukrainian writer Yu. Yanovsky also liked to use such long sentences as Ivanychuk. The results



obtained in the study correspond to the data for the Ukrainian prose. The only and significant difference is the large percentage of long sentences, which makes it possible to distinguish this feature, as distinctive characteristic of the novels of R. Ivanychuk.

## 6 Conclusions

Roman Ivanychuk is undoubtedly a talented Ukrainian writer who left a large legacy of interesting and worth reading novels, short stories and novelets. From the statistical standpoint, his works have not been researched enough to the date and the analysis presented in the paper already shows that Ivanychuk had a specific, perhaps unique, manner of writing.

The short stories taken as a basis of the research have some specific features that distinguish them from other works. The most frequent words used by the author are pronouns and nouns. That allows to conclude that in the most of his short stories the person of narrator is the main character.

Furthermore, the degree of euphony of R. Ivanychuk's works is similar to the Ukrainian prose in general. The most frequently used words are 4-, 5-, 6-, 2-phoneme words. This is also within the norm for the Ukrainian prose and other Ukrainian authors. What stands out is the length of sentences used by the author. He tended to use larger sentences comparing to other Ukrainian writers and this is a representative feature for text attribution. Like in no other Ukrainian writers' texts, the percentage of sentences longer than 60 words in the texts of R. Ivanychuk makes up almost 5%.

Other differences or specific features of short stories of R. Ivanychuk, are yet to be defined. The results of the research presented in the paper will be applied a further research in this direction. Also they already have a practical value for the text attribution of the Ukrainian literature.

## 7 References

1. Bezlichenko, M. A.: Poetyka istorychnoi prosy Romana Ivanychuka (osoblyvosti chasoprostorovykh kontseptsii). In: avtorepherat dysertatsii na zdobuttia naukovoho stupenia kand. philol. nauk: spetsialnist 10.01.01 "Ukrainska literatura", 20. (2007)
2. Burchenia, V. V.: Kategoria istoryzmu ta stylovi poshuky v prozi Romana Ivanychuka. In: avtorepherat dysertatsii na zdobuttia naukovoho stupenia kand. philol. nauk: spetsialnist 10.01.01 "Ukrainska literatura", 20. (2010)
3. Hadzii, O. A.: Publitsystyka Romana Ivanychuka: problematyka I poetyka. In: avtorepherat dysertatsii na zdobuttia naukovoho stupenia kand. philol. nauk: spetsialnist 10.01.01 "Ukrainska literatura", 20. (2011)
4. Horodylovska, H. P.: Elementy naukovoho ta publitsystychnoho slyliv v istorychnukh tvorakh Romana Ivanychuka. In: avtorepherat dysertatsii na zdobuttia naukovoho stupenia kand. philol. nauk: spetsialnist 10.01.01 "Ukrainska literatura", 18. (2008)
5. Ivanuchyuk, Roman.: Odnahlibyna na dvokh, Sribne Slovo, Lviv, 160 (2004)
6. Ivanuchyuk, Roman.: Osinni uzory: novely ta opovidannia, Pholio, Kharkiv, 299. (2015)
7. Ivanuchyuk, Roman.: Prut nese krygu, Kn.- zhurnl. vydavnytstvo, Lviv, 83 (1958)

8. Kapliuk, K. M.: Khudozhnii khronotop istorichnoi romanistyky R. Ivanychuka (na materialii romaniv "Malvy", "Zhuravlynyi kryk", "Orda"). In: avtorepherat dysertatsii na zdobuttia naukovoho stupenia kand. philol. nauk: spetsialnist 10.01.01 "Ukrainska literatura", 18 (2010)
9. Koliadych, Iu. V.: Osoblyvosti poetyky maloi prozy Romana Ivanychuka. In: avtorepherat dysertatsii na zdobuttia naukovoho stupenia kand. philol. nauk: spetsialnist 10.01.01 "Ukrainska literatura", 20. (2008)
10. Kulchystkyii, I. M.: Doslidzhennia dovzhyky rechennia ta slova u tvorakh Romana Ivanychuka. In: Inphormatsiini systemy ta merezhi, Visnyk Natsionalnoho universytetu "Lvivska Politehnika", Lviv, vol. 872, 139-149 (2017)
11. Kulchystkyii, I. M.: Idiolekt, idiosyl, indyvidualnyi syl. Totozne chy rizne?. In: Naukovyi visnyk Volynskoho natsionalnoho universytetu imeni L. Ukrainky, Lutsk, vol. 5 (282), 226-2291 (2014)
12. Levytskyi, V.: Kvantytatyvnoe metody v linhvistike. Nova Knyha, Vinnystia, 234 (2007)
13. Perebyinis, V. I.: Statystychni metody dlia linhvistiv. Nova Knyha, Vinnystia, 176 (2013)
14. Perebyinis, V. S.: Statystychni parametry styliv, Naukova dumka, Kyiv (1967)
15. Pidopryhora, S. V.: Romanni tryptykh "Vohnenni stovpy" u zhanrovii systemi istorichnoi prozy Romana Ivanychuka. In: avtorepherat dysertatsii na zdobuttia naukovoho stupenia kand. philol. nauk: spetsialnist 10.01.01 "Ukrainska literatura", 18 (2008)
16. Stavyska, L. O.: Pro termin idiolect. In: Ukrainska mova, vol. 4, 3-17 (2009)  
Uirina, U. M.: Idiosyl Oleny Telihi. In: dysertatsia kandydata filosofskykh nauk 10.02.01, Herson, 194 (2016)