Quantitative Parameters of Some Novellas by Roman Ivanychuk

Ihor Kulchytskyy
Lviv Polytechnic National University, 12 Bandera street,
Lviv, Ukraine, 79013
bis.kim@gmail.com

Abstract. Nowadays there are many approaches and methods in the field of modern linguistics, although there has been an increasing tendency towards using quantitative methods for research. It is believed that on the verge of the two branches, namely linguistics and statistics, the modern scholars can obtain the most accurate and up to date results. This paper deals with the statistical analysis of the novellas written by the renowned Ukrainian writer Roman Ivanychuk. The analysis of the linguistic text by the means of statistics provide an in-depth perspective on the specific style of writing of the author.

Keywords: statistical analysis, quantitative parameters, novellas, idiolect, corpus linguistics

1 Introduction

At the current stage of the development of linguistics, the use of the electronic corpus of texts has become an integral part for many researches devoted to the individual style of author. Corpus linguistics is a methodology of linguistics that consists of computer-based empirical analysis (both quantitative and qualitative) of actual models of language usage, using large-scale collections of naturally occurring spoken and written texts available in electronic form, called corpora. An electronic corpus of texts if a useful tool for language learning, texts attribution and historical research of some linguistic phenomenon. The focus of this paper is on the individual style of writing of Roman Ivanychuk researched by the means of statistics in order to find some distinctive features of the Ivanychuk's writing as it is believed that he possessed an indeed special manner of writing and he has a passion to use extremely long sentences in his writing comparing to other Ukrainian authors. The results of the research will be useful for text attribution, language learning and historical research of the Ukrainian language.

2 The Interrelation of Corpus Linguistics, Statistics and Idiolect

From the historical standpoint, the use of quantitative criterion in the linguistic studies has long been among the most relevant applied methods of linguistic research. Looking

back to the XX century, it was Ferdinand de Saussure who one of many laid the foundations of such research methods [3, p. 123]. Later on, the evolvement of machine translation significantly spread up the use of mathematical methods in linguistics.

In the course of word processing for their input into the machine, various quantitative estimates of some particular features of language were obtained, which proved to be useful not only for the creation of mathematical language models, but also for linguistic theory. Since language is a probabilistic rather than a well-defined system, quantitative methods are needed to identify it, related to the study of probabilistic, gradual, frequency, and other illogical features.

When the texts were properly processed for further work in the computers, different quantitative indicators of the separate linguistic features were obtained. They turned out to be useful not only for creation of certain mathematical models, but for the linguistic theory in general. Since language is a probabilistic rather than a well-defined system, quantitative methods are getting more important aiming at proper identification of its specific features [11, p. 139].

Statistics is a mathematical science which purpose is to collect, analyze, explain, demonstrate and interpret data. Statistical methods also broadly used in the corpus linguistics as well. They have become one of the most efficient and time-saving tools of processing different sets of texts.

Since corpus linguistics is based on conducting linguistic analyzes, it can be used to explore many types of language issues, and it has the potential to generate interesting, fundamental, and often unexpected new perspectives on language. That is why corpus linguistics has become one of the most widely used methods of linguistic research in recent years.

Text corpus can be defined as a systematic set of natural texts (both written and spoken). The term systematicity means that the structure and content of the corpus comply with certain extra-linguistic principles (e.g. sampling principles on the basis of which the included texts were selected).

3 Material: Collection, Organization and Methods of Research

The material for the research is the following novellas of Roman Ivanychuk: “I zemlia, I zelo, I pisnia” (“And earth, and green, and song”) (further in the text this novella will be referred to as RI1) [4], “Lisova povist” (“Forest story”) (RI2) [5], “Nespokutne” (“No Atonement”) (RI3) [6], “Solo na fleiti” (“Flute Solo”) (RI4) [7]. To stick with the general requirement for the publication, the novellas titles are also presented in the author’s translation into English.

First of all, the texts of the given novellas were converted in electronic form with the help of the ABBYY Fine Reader software and saved in .docx format. The next step was the normalization of the texts in the MS Word editor. The normalization meant bringing the text in full compliance with the original, arranging the spelling and punctuation of

the text in accordance with the spelling standards [15], marking all foreign words with the relevant languages, etc.

The received normalized texts were formalized with the help of R2U software, access granted by Vasily Starko [14].

The results of the automatic lemmatization have been converted to the required format using native Python applications and have been validated and corrected with MS Access.

The next step was to structure the text using XML-style tags [10]. The following structural elements were distinguished:

- paragraph - … ;
- sentences - <s>… </s>;
- character language - <q>… </q>;
- epigraph - <motto>… </motto>;
- the text of the epigraph - <mottotext>… </mottotext>;
- source of the epigraph - <mottospring>… <mottospring>;
- the beginning of the original page with the number - <bp n = x />;
- place and date of writing - <place>… </place>.

The normalization, text recognition and verification of the automatic lemmatization were done within the master's thesis by the graduate student of the Department of Applied Linguistics of the Lviv National Polytechnic University Victoria Ogorodnik [12].

The received texts and the results of the lemmatization were subjected to statistical analysis. Statistics are calculated using standard methods and formulas adopted for mathematical statistics [Beginning Statistics]. The necessary software for analysis is written in Python language.

For the general statistical research of the abovementioned novellas, the following coefficients were calculated [2; 8; 13]:

Vocabulary richness. It is also called the diversity factor/coefficient. The greater the value of this indicator is, the more different words in a particular text can be found. It is calculated as the ratio of the number of words in the text to the number of words usage.

Average word repetition in text. It shows how many times each word is used in the text. It is calculated as the ratio of word usage to word count.

Exclusivity ratio. This indicator characterizes the variability of vocabulary. It is calculated separately for the text (the ratio of the number of word forms that are encountered in the text once to the total number of word forms) and for the vocabulary (the ratio of the number of words that are encountered in the text once to the total number of words).

Vocabulary concentration coefficient. This indicator is opposite to the exclusivity ratio. If for text, it is calculated as the ratio of the number of word forms that encountered in the text 10 or more times. Accordingly, for a text vocabulary, it is calculated as the ratio of the number of words that have appeared in the text 10 times or more to the total number of words. The relatively small number of high-frequency vocabulary (low

concentration ratio) and the relatively large number of words with frequency 1 (high exclusivity ratio) tend to indicate a considerable variety of vocabulary.

Automatic readability index (ARI) is a degree of readability of texts, the ratio of characters in the word and the number of sentences is calculated according to the formula: ARI = 4,71 * C / W + 0,5 * W / (S * 3) - 21,43, where C stands for characters, W for words and S for sentences.

Coefficient of lexical density is calculated as the ratio of the number of word forms of independent parts of speech in the text to the total number of word forms.

Adjectives to nouns ratio is also called the coefficient of epithelization. It is calculated as the ratio of the number of uses in the text of adjectives to the number of uses of nouns.

Adverb to verb ratio is the ratio of the number of uses of adverbs to the number of uses of verbs.

Nouns to verbs ratio is computed as the ratio of the number of uses of nouns to the number of uses of verbs.

Verbs to total number of words ratio is also known as aggressiveness ratio and is counted as the ratio of the use of verbs to the total number of all words in the text.

Coefficient of logical connectivity (conjunctions and prepositions to total number of sentences ratio) is basically calculated as the ratio of the number of uses of conjunctions and prepositions to the total number of sentences in the text.

Coefficient of speech "embolism" (clogging) (or exclamations & particles to total number of words ratio) is calculated as the ratio of the number of uses of exclamations and particles to the total number of words used.

Adjectives to nouns ratio, adverb to verb ratio, nouns to verbs ratio, and verbs to total number of words ratio generally define and partially describe the style of the novella. If the nouns to verbs ratio is bigger than 1, one can assume that the text is narration (or is written in nominal style).

Adjectives to nouns ratio (the number of adjectives to one noun) in the nominal style indicate the degree of a fiction style (as far as the text can be considered a fiction). This is due to the fact that adjectives are the main mean of the figures of speech expressions namely such as epithets and comparisons because of their relations with nouns. Verbs to total number of words ratio (also known as aggressiveness ratio) determines the ratio of the number of verbs and verb forms (adjectives and adverbs) to the total number of all words. High aggressiveness indicates high emotional intensity of the text, dynamics of events, intense emotional state of the author when writing the text. A logic ratio of magnitudes within 1 provides a sufficiently harmonious link between auxiliary parts of speech and syntax constructions. With a nominative ratio of less than 1 and a high verb ratio, we state the verbal idiostyle of the work, and the verb ratio (the number of adverbs per verb) indicates the level and number of speech figures used.

4 The Discussion of the Results of the Statistical Analysis of Novellas by Roman Ivanychuk

The general statistical indicators of the researched novellas: the researched novellas have the following general statistical indicators (table 1):

Table 1. Statistical indicators used in the research

Statistical Indicators	Novellas			
	RI1	**RI2**	**RI3**	**RI4**
Number of word usage	8775	7523	5098	4376
Number of word forms	3938	3472	2520	2178
Number of words	2614	2444	1825	1648
Hapax legomenon for word forms	2915	2570	1940	1660
Number of word forms used 10 times or more	101	76	50	46
Hapax legomenon for words	1636	1542	1213	1127
Number of words used 10 times or more	127	109	74	63
Number of letters in the text	43873	40222	25917	22819
Number of sentences in the text	398	165	168	105

The words distribution and the number of words according to parts of speech is presented as below. The results of the carried-out research have shown that the **novella "And earth, and green, and song"** contains the following parts of speech:

Words: noun — 974 (37,26%); verb — 759 (29,04%); adjective — 458 (17,52%); adverb — 173 (6,62%); pronoun — 70 (2,68%); gerund — 50 (1,91%); preposition — 45 (1,72%); conjunction — 39 (1,49%); particle — 26 (0,99%); numeral — 14 (0,54%); exclamation — 5 (0,19%); present participle — 1 (0,04%).

Words usage: noun — 2697 (30,74%); verb — 1478 (16,84%); adjective — 833 (9,49%); adverb — 362 (4,13%); pronoun — 976 (11,12%); gerund — 56 (0,64%); preposition — 956 (10,89%); conjunction — 937 (10,68%); particle — 435 (4,96%); numeral — 30 (0,34%); exclamation — 14 (0,16%); present participle — 1 (0,01%).

"Forest story" novella:

Words: noun — 747 (30,56%); verb — 696 (28,48%); adjective — 462 (18,90%); adverb — 240 (9,82%); gerund — 109 (4,46%); pronoun — 79 (3,23%); preposition — 42 (1,72%); conjunction — 34 (1,39%); particle — 30 (1,23%); numeral — 4 (0,16%); present participle — 1 (0,04%).

Words usage: noun — 2173 (28,88%); verb — 1199 (15,94%); adjective — 855 (11,37%); adverb — 464 (6,17%); gerund — 126 (1,67%); pronoun — 804 (10,69%); preposition — 852 (11,33%); conjunction — 751 (9,98%); particle — 281 (3,74%); numeral — 17 (0,23%); present participle — 1 (0,01%).

"No Atonement" novella:

Words: noun — 620 (33,97%); verb — 531 (29,10%); adjective — 299 (16,38%); adverb — 138 (7,56%); pronoun — 74 (4,05%); gerund — 58 (3,18%); preposition — 37 (2,03%); conjunction— 31 (1,70%); particle — 31 (1,70%); numeral— 4 (0,22%); exclamation — 2 (0,11%).

Words usage: noun — 1329 (26,07%); verb — 852 (16,71%); adjective — 456 (8,94%); adverb — 226 (4,43%); pronoun — 763 (14,97%); gerund — 64 (1,26%); preposition — 637 (12,50%); conjunction — 538 (10,55%); particle — 217 (4,26%); numeral — 14 (0,27%); exclamation — 2 (0,04%).

"Flute Solo" novella:

Words: noun — 620 (37,62%); verb — 407 (24,70%); adjective — 289 (17,54%); adverb — 134 (8,13%); pronoun — 68 (4,13%); preposition — 38 (2,31%); conjunction — 31 (1,88%); present participle — 30 (1,82%); particle — 23 (1,40%); numeral — 7 (0,42%); exclamation — 1 (0,06%).

Words usage: noun — 1188 (27,15%); verb — 695 (15,88%); adjective — 432 (9,87%); adverb — 217 (4,96%); pronoun — 665 (15,20%); preposition — 515 (11,77%); conjunction — 453 (10,35%); gerund — 31 (0,71%); particle — 163 (3,72%); numeral — 16 (0,37%); exclamation — 1 (0,02%).

The meanings of the statistical coefficients that characterize the researched novellas presented in the Table 1 below

Table 2. Total coefficients of words

Coefficient	**Novellas**			
	RI1	**RI2**	**RI3**	**RI4**
Vocabulary richness	0,30	0,32	0,36	0,38
Average word repetition in text	3,36	3,08	2,79	2,66
Exclusivity ratio for word forms	0,33	0,34	0,38	0,38
Exclusivity ratio for words	0,63	0,63	0,66	0,68
Vocabulary concentration coefficient for word forms	0,01	0,01	0,01	0,01
Vocabulary concentration coefficient for words	0,05	0,04	0,04	0,04
Automatic readability index	13,14	26,55	17,69	23,97

Table 3. General text coefficients

Coefficient	Novellas			
	RI1	**RI2**	**RI3**	**RI4**
Coefficient of lexical density	0,22	0,21	0,23	0,22
Adjectives to nouns ratio	3,24	2,54	2,91	2,75
Adverb to verb ratio	0,24	0,35	0,25	0,30
Nouns to verbs ratio	1,76	1,64	1,45	1,64
Verbs to total number of words ratio (aggressiveness)	0,17	0,18	0,18	0,17
Coefficient of logical connectivity	4,76	9,72	6,99	9,22
Coefficient of speech "embolism	0,05	0,04	0,04	0,04

It is important to mention that the percentage of parts of speech in different words usages and words slightly differs. The results are represented on the picture 1 below:

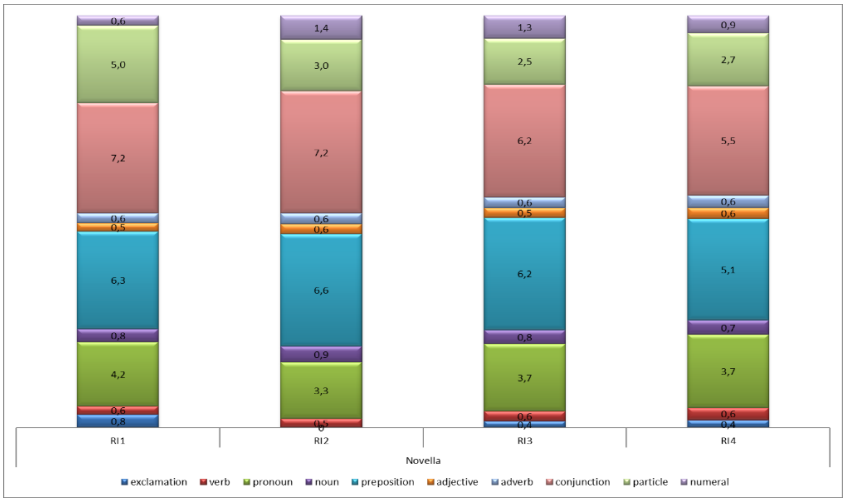

Fig. 1. The percentage difference of parts of speech in word usages and words in the text.

It should be noted that taking into account the fact that modern grammatical theories consider gerund and present participle as verbs classes, these two parts of speech were merged as verbs [1].

As it can be seen, for parts of the speech such as verb, noun, adjective and adverb, the percentage words decreased (on average: verb – in 0.6, noun – in 0.8, adjective – in 0.6, adverb – in 0.6). But it increased significantly for pronouns (3.7), prepositions (6.0), conjunctions (6.5), particles (3.3). The percentage number of the numerals did

not change at all (1.0) while the percentage of pronouns decreased (0.4). The reason is probably to be found in the method of constructing the statements. For further parts of speech analysis of texts, prepositions, conjunctions, and particles were grouped into "auxiliary parts of speech group" while the exclamations and numerals were grouped into the "miscellaneous" group, since in terms of quantity their selection is not big enough to carry out a general statistical analysis described in the paper.

The results were compared to the quantitative parts of speech distribution of the Dictionary of the Ukrainian language consisting of 11 volumes:

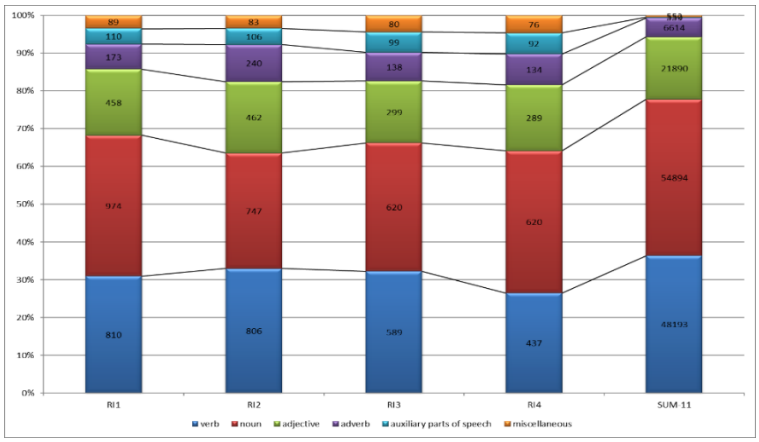

Fig. 2. The parts of speech distribution of Roman Ivanychuk's novellas comparing to the 11 volume the Dictionary of the Ukrainian language

The figure 3 below represents the parts of speech distribution for words encountered in the researched novellas. The figure 4 below represents the parts of speech distribution for word usage encountered in the researched novellas.

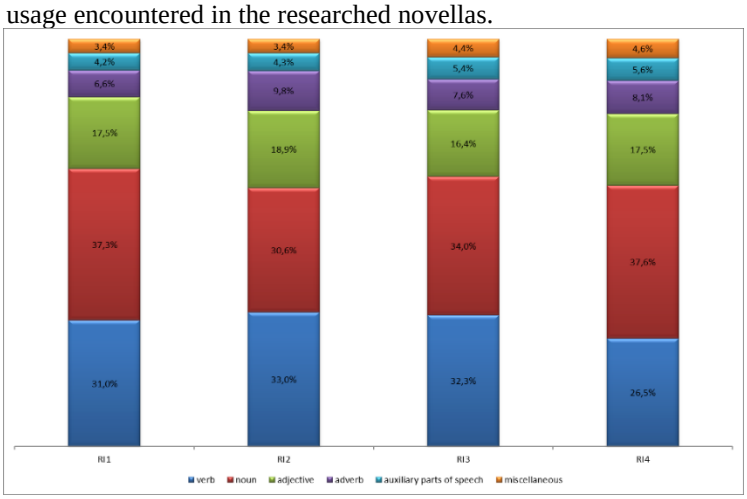

Fig. 3. Parts of speech distribution for words

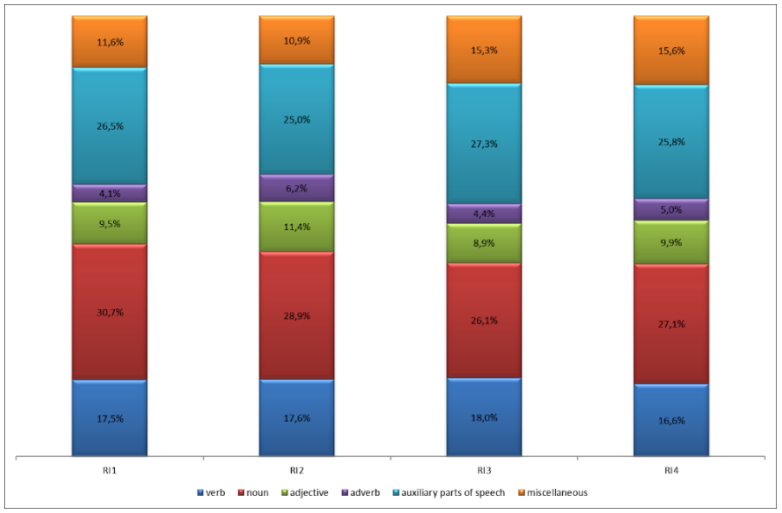

Fig. 4. Parts of speech distribution for word usages

The distribution of rank frequencies is shown on the figure 5. It mainly focuses on word forms, although it is important to mentioned that the distribution of rank frequencies for wards is identical as for wordforms.

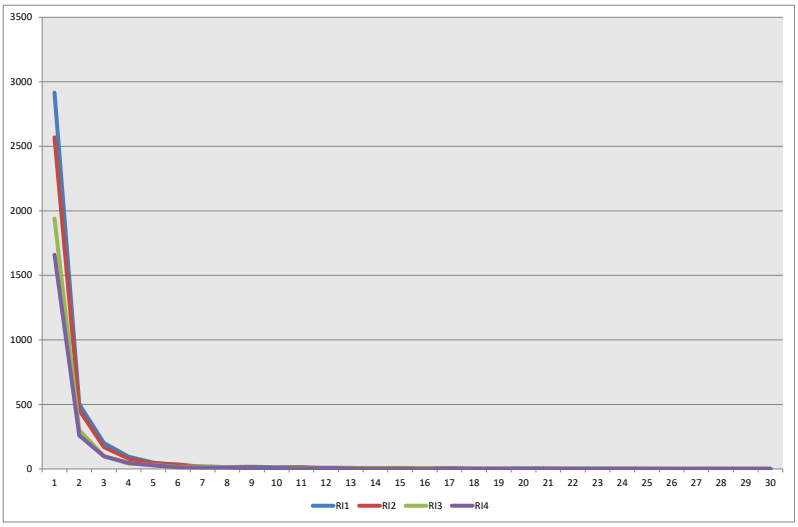

Fig. 5. The distribution of rank frequencies for word forms in the novellas by R. Ivanychuk

The frequencies distributions for each of novellas are as follows:

- **novella “And earth, and green, and song”:**

Words: 1 — 1636 (62,59%); 2 — 402 (15,38%); 3 — 186 (7,12%); 4 — 103 (3,94%); 5 — 67 (2,56%); 6 — 33 (1,26%); 9 — 22 (0,84%); 7 — 19 (0,73%); 8 — 19 (0,73%); 10 — 16 (0,61%); 12 — 10 (0,38%); 11 — 9 (0,34%); 13 — 7 (0,27%); 16 — 6 (0,23%); 20 — 5 (0,19%); 14 — 4 (0,15%); 15 — 4 (0,15%); 18 — 4 (0,15%); 21 — 4 (0,15%); 24 — 4 (0,15%); 28 — 4 (0,15%); 17 — 3 (0,11%); 19 — 2 (0,08%); 26 — 2 (0,08%); 27 — 2 (0,08%); 31 — 2 (0,08%); 33 — 2 (0,08%); 34 — 2 (0,08%); 38 — 2 (0,08%); 39 — 2 (0,08%); 44 — 2 (0,08%); 67 — 2 (0,08%); 22 — 1 (0,04%); 29 — 1 (0,04%); 36 — 1 (0,04%); 37 — 1 (0,04%); 42 — 1 (0,04%); 45 — 1 (0,04%); 48 — 1 (0,04%); 52 — 1 (0,04%); 55 — 1 (0,04%); 58 — 1 (0,04%); 68 — 1 (0,04%); 69 — 1 (0,04%); 73 — 1 (0,04%); 78 — 1 (0,04%); 80 — 1 (0,04%); 85 — 1 (0,04%); 86 — 1 (0,04%); 92 — 1 (0,04%); 100 — 1 (0,04%); 121 — 1 (0,04%); 123 — 1 (0,04%); 126 — 1 (0,04%); 139 — 1 (0,04%); 142 — 1 (0,04%); 159 — 1 (0,04%); 217 — 1 (0,04%); 254 — 1 (0,04%).

Word forms: 1 — 2915 (74,02%); 2 — 501 (12,72%); 3 — 201 (5,10%); 4 — 95 (2,41%); 5 — 49 (1,24%); 6 — 27 (0,69%); 7 — 17 (0,43%); 9 — 17 (0,43%); 8 — 15 (0,38%); 11 — 15 (0,38%); 10 — 13 (0,33%); 12 — 7 (0,18%); 13 — 7 (0,18%); 18 — 6 (0,15%); 14 — 5 (0,13%); 21 — 5 (0,13%); 15 — 4 (0,10%); 26 — 3 (0,08%); 27 — 3 (0,08%); 19 — 2 (0,05%); 20 — 2 (0,05%); 23 — 2 (0,05%); 17 — 1 (0,03%); 28 — 1 (0,03%); 30 — 1 (0,03%); 31 — 1 (0,03%); 32 — 1 (0,03%); 33 — 1 (0,03%); 34 — 1 (0,03%); 37 — 1 (0,03%); 40 — 1 (0,03%); 42 — 1 (0,03%); 44 — 1 (0,03%); 45 — 1 (0,03%); 50 — 1 (0,03%); 51 — 1 (0,03%); 53 — 1 (0,03%); 56 — 1 (0,03%); 59 — 1 (0,03%); 77 — 1 (0,03%); 82 — 1 (0,03%); 83 — 1 (0,03%); 86 — 1 (0,03%); 120 — 1 (0,03%); 126 — 1 (0,03%); 142 — 1 (0,03%); 156 — 1 (0,03%); 191 — 1 (0,03%); 235 — 1 (0,03%).

- “**Forest story” novella**

Words: 1 — 1542 (63,09%); 2 — 383 (15,67%); 3 — 169 (6,91%); 4 — 99 (4,05%); 5 — 50 (2,05%); 6 — 36 (1,47%); 7 — 28 (1,15%); 8 — 18 (0,74%); 10 — 12 (0,49%); 11 — 11 (0,45%); 9 — 10 (0,41%); 12 — 10 (0,41%); 14 — 10 (0,41%); 17 — 10 (0,41%); 13 — 9 (0,37%); 15 — 3 (0,12%); 19 — 2 (0,08%); 21 — 2 (0,08%); 24 — 2 (0,08%); 26 — 2 (0,08%); 39 — 2 (0,08%); 41 — 2 (0,08%); 50 — 2 (0,08%); 16 — 1 (0,04%); 18 — 1 (0,04%); 22 — 1 (0,04%); 23 — 1 (0,04%); 25 — 1 (0,04%); 27 — 1 (0,04%); 28 — 1 (0,04%); 29 — 1 (0,04%); 30 — 1 (0,04%); 31 — 1 (0,04%); 32 — 1 (0,04%); 33 — 1 (0,04%); 36 — 1 (0,04%); 37 — 1 (0,04%); 47 — 1 (0,04%); 48 — 1 (0,04%); 52 — 1 (0,04%); 72 — 1 (0,04%); 78 — 1 (0,04%); 86 — 1 (0,04%); 93 — 1 (0,04%); 109 — 1 (0,04%); 119 — 1 (0,04%); 123 — 1 (0,04%); 124 — 1 (0,04%); 125 — 1 (0,04%); 142 — 1 (0,04%); 159 — 1 (0,04%); 172 — 1 (0,04%); 207 — 1 (0,04%).

Word forms: 1 — 2570 (74,02%); 2 — 456 (13,13%); 3 — 169 (4,87%); 4 — 80 (2,30%); 5 — 45 (1,30%); 6 — 35 (1,01%); 7 — 16 (0,46%); 8 — 15 (0,43%); 9 — 10 (0,29%); 12 — 9 (0,26%); 13 — 9 (0,26%); 10 — 6 (0,17%); 11 — 6 (0,17%); 14 — 4 (0,12%); 15 — 4 (0,12%); 17 — 4 (0,12%); 16 — 3 (0,09%); 18 — 2 (0,06%); 19 — 2 (0,06%); 23 — 2 (0,06%); 27 — 2 (0,06%); 39 — 2 (0,06%); 21 — 1 (0,03%); 24 — 1 (0,03%); 25 — 1 (0,03%); 28 — 1 (0,03%); 29 — 1 (0,03%); 30 — 1 (0,03%); 32 — 1 (0,03%); 34 — 1 (0,03%); 62 — 1 (0,03%); 63 — 1 (0,03%); 64 — 1 (0,03%); 76 — 1 (0,03%); 81 — 1 (0,03%); 93 — 1 (0,03%); 105 — 1 (0,03%); 117 — 1 (0,03%); 121 — 1 (0,03%); 124 — 1 (0,03%); 134 — 1 (0,03%); 158 — 1 (0,03%); 201 — 1 (0,03%).

- **“No Atonement” novella**

Words: 1 — 1213 (66,47%); 2 — 280 (15,34%); 3 — 107 (5,86%); 4 — 53 (2,90%); 5 — 33 (1,81%); 6 — 23 (1,26%); 7 — 17 (0,93%); 8 — 14 (0,77%); 9 — 11 (0,60%); 13 — 9 (0,49%); 10 — 8 (0,44%); 11 — 7 (0,38%); 12 — 5 (0,27%); 15 — 5 (0,27%); 14 — 4 (0,22%); 22 — 4 (0,22%); 16 — 2 (0,11%); 17 — 2 (0,11%); 18 — 2 (0,11%); 20 — 2 (0,11%); 31 — 2 (0,11%); 43 — 2 (0,11%); 54 — 2 (0,11%); 71 — 2 (0,11%); 19 — 1 (0,05%); 21 — 1 (0,05%); 23 — 1 (0,05%); 30 — 1 (0,05%); 36 — 1 (0,05%); 39 — 1 (0,05%); 45 — 1 (0,05%); 46 — 1 (0,05%); 74 — 1 (0,05%); 77 — 1 (0,05%); 83 — 1 (0,05%); 93 — 1 (0,05%); 95 — 1 (0,05%); 99 — 1 (0,05%); 137 — 1 (0,05%); 149 — 1 (0,05%).

Word forms: 1 — 1940 (76,98%); 2 — 301 (11,94%); 3 — 100 (3,97%); 4 — 42 (1,67%); 5 — 31 (1,23%); 7 — 21 (0,83%); 6 — 15 (0,60%); 8 — 14 (0,56%); 13 — 8 (0,32%); 9 — 6 (0,24%); 11 — 5 (0,20%); 12 — 4 (0,16%); 15 — 4 (0,16%); 10 — 3 (0,12%); 14 — 3 (0,12%); 70 — 2 (0,08%); 16 — 1 (0,04%); 17 — 1 (0,04%); 18 — 1 (0,04%); 19 — 1 (0,04%); 21 — 1 (0,04%); 23 — 1 (0,04%); 24 — 1 (0,04%); 29 — 1 (0,04%); 30 — 1 (0,04%); 32 — 1 (0,04%); 38 — 1 (0,04%); 40 — 1 (0,04%); 42 — 1 (0,04%); 47 — 1 (0,04%); 53 — 1 (0,04%); 71 — 1 (0,04%); 73 — 1 (0,04%); 92 — 1 (0,04%); 95 — 1 (0,04%); 135 — 1 (0,04%); 136 — 1 (0,04%).

- **“Flute Solo” novella**

Words: 1 — 1127 (68,39%); 2 — 230 (13,96%); 3 — 97 (5,89%); 4 — 46 (2,79%); 6 — 31 (1,88%); 5 — 24 (1,46%); 7 — 17 (1,03%); 9 — 7 (0,42%); 12 — 7 (0,42%); 8 — 6 (0,36%); 10 — 6 (0,36%); 14 — 6 (0,36%); 11 — 4 (0,24%); 13 — 3 (0,18%); 15 — 3 (0,18%); 16 — 3 (0,18%); 17 — 3 (0,18%); 18 — 2 (0,12%); 21 — 2 (0,12%); 22 — 2 (0,12%); 25 — 2 (0,12%); 44 — 2 (0,12%); 49 — 2 (0,12%); 88 — 2 (0,12%); 19 — 1 (0,06%); 20 — 1 (0,06%); 23 — 1 (0,06%); 26 — 1 (0,06%); 28 — 1 (0,06%); 43 — 1 (0,06%); 50 — 1 (0,06%); 53 — 1 (0,06%); 58 — 1 (0,06%); 62 — 1 (0,06%); 64 — 1 (0,06%); 89 — 1 (0,06%); 116 — 1 (0,06%); 138 — 1 (0,06%).

Word forms: 1 — 1660 (76,22%); 2 — 259 (11,89%); 3 — 98 (4,50%); 4 — 47 (2,16%); 5 — 28 (1,29%); 6 — 13 (0,60%); 8 — 11 (0,51%); 7 — 10 (0,46%); 10 — 8 (0,37%); 12 — 8 (0,37%); 9 — 6 (0,28%); 13 — 4 (0,18%); 11 — 3 (0,14%); 16 — 3 (0,14%); 28 — 2 (0,09%); 49 — 2 (0,09%); 51 — 2 (0,09%); 85 — 2 (0,09%); 14 — 1 (0,05%); 21 — 1 (0,05%); 22 — 1 (0,05%); 23 — 1 (0,05%); 25 — 1 (0,05%); 26 —

1 (0,05%); 36 — 1 (0,05%); 48 — 1 (0,05%); 62 — 1 (0,05%); 70 — 1 (0,05%); 88 — 1 (0,05%); 116 — 1 (0,05%).

As it can be seen, words with frequency equal to 1 have been found in 65%-68% of the whole text (figure 6). Regarding the word forms, words with frequency equal to 1 are a bit higher in terms of quantity, and are equal to 73%-76% and 95–96% (figure 7).

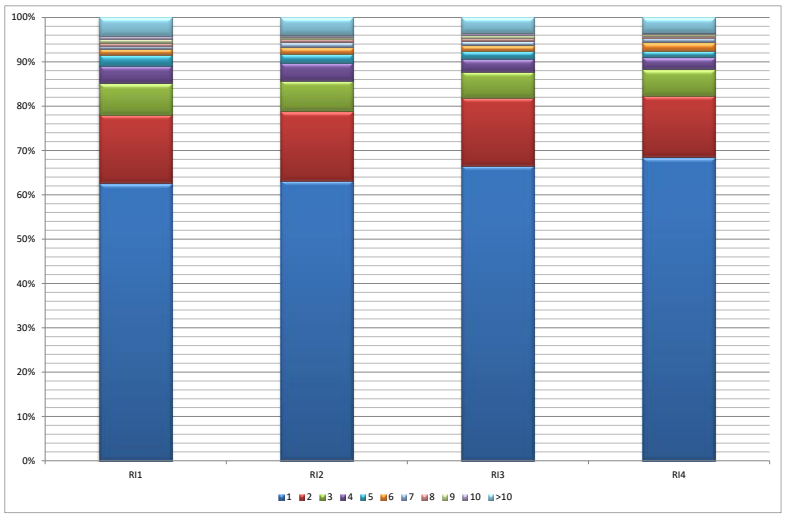

Fig. 6. Ranks (frequencies) of words for each novella

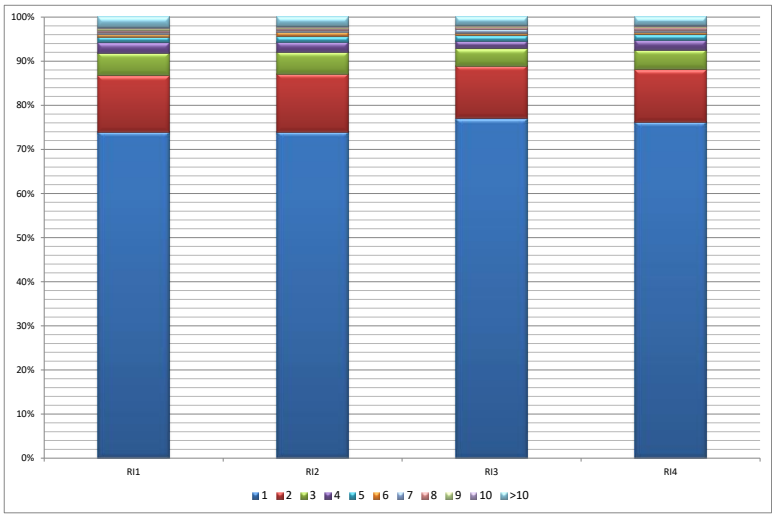

Fig. 7. Ranks (frequencies) of word forms for each novellas

The results shown above can help us to assume that the Ukrainian writer Roman Ivanychuk possessed an incredibly rich vocabulary that was indeed reflected in his manner of writing. At the same time the received results allowed to come up with the following statistical coefficients below:

Table 4. Words coefficient

Coefficient	**Novella**			
	RI1	**RI2**	**RI3**	**RI4**
Vocabulary richness	0,30	0,32	0,36	0,38
Average word repetition in text	3,36	3,08	2,79	2,66
Exclusivity ratio for word forms	0,33	0,34	0,38	0,38
Exclusivity ratio for words	0,63	0,63	0,66	0,68
Vocabulary concentration coefficient for word forms	0,01	0,01	0,01	0,01
Vocabulary concentration coefficient for words	0,05	0,04	0,04	0,04
Automated readability index	13,14	26,55	17,69	23,97

Table 5. Text coefficient

Coefficient	**Novella**			
	RI1	**RI2**	**RI3**	**RI4**
Coefficient of lexical density	0,22	0,21	0,23	0,22
Adjectives to nouns ratio	0,31	0,39	0,34	0,36
Adverb to verb ratio	0,24	0,35	0,25	0,30
Nouns to verbs ratio	1,76	1,64	1,45	1,64
Verbs to total number of words ratio (aggressiveness)	0,17	0,18	0,18	0,17
Coefficient of logical connectivity	1,59	3,24	1,34	1,89
Coefficient of speech “embolism”	0,05	0,04	0,04	0,04

The calculation made in this research show that the analyzed texts by R. Ivanychuk contain the equal number of nouns and verbs as the *nouns to verbs ratio* is big enough to conclude that all his novellas have a specific idiostyle that is characterized by robust, accurate, and informative account of Ivanychuk’s thoughts on the paper. In terms of

linguistics, the noun phrases and substantive groups significantly prevail in his writing. This prove that his writing has "nominative" style which also includes a wide and frequent usage of adjectives that specify and describe everything called by nouns.

The *adjectives to nouns ratio* (the number of adjectives per 1 noun) in the texts of the nominal idiostyle also characterizes the highly fiction level of the writing, as adjectives in general are main mean of metaphoric expressions of tropes (namely epithets and comparisons). The coefficient of the *adjective to nouns ratio* of the researched texts is pretty high (0,31-0,39) which means that Roman Ivanychuk used a lot of epithets in his writing. The nominative style of his writing also supports the fact that there is a pretty low *verbs to total number of words ratio (aggressiveness).* It indicates that the writing style focuses more on how to describe things rather than reflect some actions. It also shows that the writing is emotionally neutral. The presence of high *coefficient of logical connectivity* (within 1), harmonic connection between auxiliary parts of speech and syntactic constructions demonstrates that the sentences produced by the author tend to be complex and compound that is also a distinctive feature of the nominative idiostyle in general.

The length of words and sentences in the researched novellas of Roman Ivanychuk is presented in the table below:

Table 6. The statistical indicators of the distribution of words length in the novellas

	Max value	**Min value**	**Mean value**	**Mean square deviation**	**Medium frequency fluctuation**
RI1	22	1	5	2,8	0,0299
RI2	15	1	5,34	2,93	0,0338
RI3	17	1	5,08	2,93	0,0409
RI4	21	1	5,22	3,01	0,0454

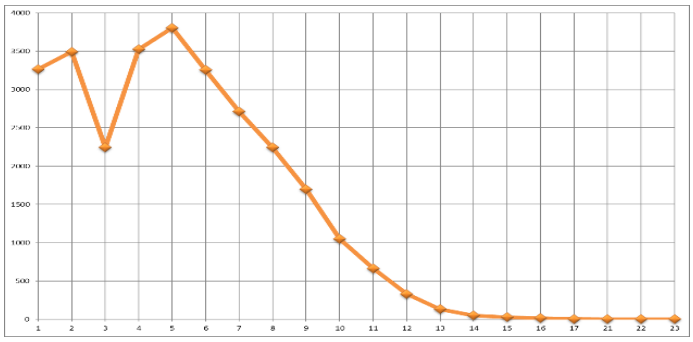

Fig. 8. Average number of the statistical indicators of the distribution of words length in the novellas

The table below represents the statistical indicators of length of words by R. Ivanychuk comparing to the same statistical indicators of other Ukrainian writers.

Table 7. the statistical indicators of length of words by R. Ivanychuk comparing to the same statistical indicators of other Ukrainian writers

Other Ukrainian writers	Mean value	Mean square deviation	Relative error
А. Головко (A. Holovko)	4,74	0,1	0,03
О. Гончар (O. Honchar)	5,41	0,07	0,02
О. Довженко (O. Dovzhenko)	4,73	0,08	0,03
П. Панч (P. Panch)	5,28	0,29	0,09
М. Стельмах (M. Stelmakh)	5,3	0,16	0,05
Ю. Яновський (Iu. Ianovskui)	5,06	0,13	0,04
Повісті Р. Іваничука	5,15	2,91	0,01

The analysis of the given indicators shows that according to the mean length of words, the novellas of R. Ivanychuk are close to the texts of Iu. Ianovskui and P. Panch. However, this also can reflect the specificity of this statistical indicator.

The table below represents the statistical indicators of the distribution of the sentence length in the novellas of R. Ivanychuk.

Table 8. The statistical indicators of the distribution of the sentence length in the novellas of R. Ivanychuk

Statistical indicator	Value received
Quantity of different lengths	926
Mean value	30,8
Mean square deviation	31,12
Medium frequency fluctuation	1,0105
Standard error	1,0228
Relative error	0,0651

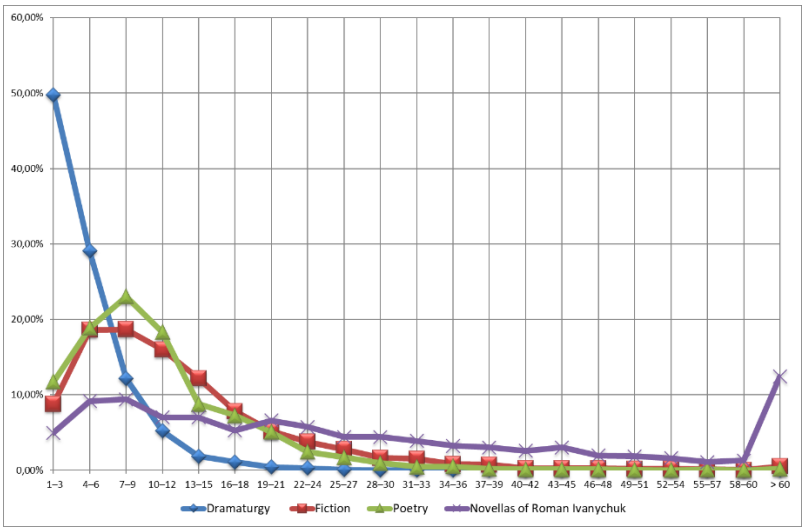

Fig. 9. The distribution of lengths of sentences of R. Ivanychuk's novellas in comparison with other genres

5 Conclusions

The carried-out research allows to concluded that the Ukrainian author Roman Ivanychuk possessed a special, perhaps unique and definitely interesting and eye-catching matter of writing. Not only his texts and plots are gripping, but the form itself is also very outstanding and out of ordinary for that period of time. First of all, his manner of writing has a nominative style (that is definitely a distinctive feature for his style) where nouns and adjectives significantly prevail over the other parts of speech. This proves that his intention of writing was to describe things, to reflect on the paper how he saw the world around. At the same time his writing was emotionally reserved. Moreover, Roman Ivanychuk tended to use large sentences to describe his ideas and thoughts. The length of sentences in his writings if probably the larger one (or among the largest ones) in the Ukrainian prose.

Additionally, it has to be mentioned that the level of statistical researches of the Ukrainian fiction is general is still evolving. The methods of research used this far are obsolete and need to be updated, the size of the selections for researches are generally small and need to be enlarged (which will provide wider and more accurate results).

Nowadays it is common to use symbols to measure the length of words and words – to measure the length of sentences. However, it is also possible to measure the length of sentences, passages, and even whole texts in symbols and words can be widely used for measuring the length of passages, chapters, etc.

In my research I decided to use the above described approach, although did not include all of the results in the paper as without presentation in comparison with other Ukrainian writers, these results are rare and does not provide much value this far. So,

this is the intention to continue the research in this direction, research other writers and compare Ivanychuk's manner of writing with theirs. Work definitely must go on and it will.

6 References

1. Aleksienko, L., Zuban, O., Kozlemkozh, I.: Suchasna ukraiinska mova. Znannia, Kyiv, 534 p (2013).
2. Buk, S.: Kilkisne zistavlennia tekstiv (na materiali redaktsii 1884 ta 1907 rokiv povisti Ivana Franka "BOA CONSTRICTOR"). Ukrainske literatyroznavstvo, 76, pp. 179-192 (2012).
3. Ferdinand de Saussure.: Kurs obshchei lingvistiky. Trudy po iazykozhanyiu, Moskwa, 269 p (1977).
4. Ivanychuk, R.: I zemlia, I zelo, I pisnia (eng. And earth, and green, and song). pp. 6-35 Sribne slovo. Lviv (2006).
5. Ivanychuk, R.: Lisova povist (eng. Forest story). Sribne slovo. Lviv, pp. 116-139 (2006).
6. Ivanychuk, R.: Nespokutne (eng. No Atonement). Sribne slovo. Lviv, pp. 106-115 (2006).
7. Ivanychuk, R.: Solo na fleiti (eng. Flute Solo). Sribne slovo. Lviv, pp. 86-104 (2006).
8. Kamińska-Szmaj, I.: Części mowy w słowniku i tekście pięciu stylów funkcjonalnych polszczyzny pisanej (na materiale słownika frekwencyjnego). Biuletyn Polskiego Towarzystwa Językoznawczeg, XLI, pp. 127–136 (1988).
9. Kulchytskyi, I.: Technolohichni apekty ukladannia korpusiv tekstiv. Monographia spilno z V., Shevchenko I., Zahnitko A. ta in. za redaktsiieiu Levchenko O. Vydavnytstvo Lvivska Politechnika, pp. 29-45 (2015).
10. Lawson, B., Sharp, R.: Introducing HTML5. Second Edition New Riders, CA, pp. 295 (2012).
11. Levytshkyi, V.: Kvantytatyvnoe metody v lynhvystyke. Ruta, Chernivtsi, p.190 (2004).
12. Ohorodnyk, V.: Kilkisnyi rozpodil rechen I slovoform u tvorakh Romana Ivanychuka. XIII Vseukrainska naukovo-metodychna konferentsiia molodykh naukovtsiv, Mykolaiv, 84 p (2018).
13. Ruszkowski, M.: Wskaźnik epitetyzacji w badaniach stylistycznych. Respectus Philologicus, № 5(10), pp. 48–53 (2004).
14. Starko, V.: Ukrainska: dykh is bukva v tsyphri, https://zbruc.eu/node/87161, last accessed 2019/12/26.
15. Ukrainskyi pravopus, https://mon.gov.ua/ua/osvita/zagalna-serednya-osvita/navchalni-programi/ukrayinskij-pravopis-2019, last accessed 2019/12/26.