Statistical Research of the Colour Component ЧОРНИЙ (BLACK) in Roman Ivanychuk's Text Corpus

Nataliia Lototska[1]
Lviv State University of Life Safety, Kleparivska str. 35, Lviv, 79013, Ukraine

Abstract

The article considers the statistical analysis of word combinations with colour component ЧОРНИЙ (BLACK) in Roman Ivanychuk's fiction. The research is based on Roman Ivanychuk's and Ukrainian prose fiction text corpora to compare statistical parameters and qualitative indicators and to detect the specific characteristics of the author's idiolect.

Colour nominations are important elements for modeling the world by a linguistic personality. The colour ЧОРНИЙ (BLACK) forms the core of colours in linguistic studies and is the most frequent colour nomination in Roman Ivanychuk's text corpus. Corpus-based approach, absolute / relative frequency, statistical association measures MI-score and t-score are used to describe and analyze the author's word combinations with colour nomination ЧОРНИЙ (BLACK) as a marker of his idiolect.

Structural and semantic models of collocations and collocations with colour component ЧОРНИЙ (BLACK) are found out; thematic groups of typical collocates for colour ЧОРНИЙ (BLACK) as an attribute in the model Adj. + N. are described; high-frequency collocates of the node ЧОРНИЙ (BLACK) are presented; statistical association measure MI-score allowed to extract author-individual collocations in Roman Ivanychuk's text corpus.

Keywords

Idiolect, collocation, colour nomination, text corpus, association measures, statistical analysis, word combination

1. Introduction

Language of literature reflects the linguistic competence of the author, predominance of using particular language constructs and words as well as features of the national language [10, p. 10]. Statistical analysis of the historical prose fiction of Roman Ivanychuk, a Ukrainian writer of the XX-XXI centuries, enables to demonstrate individual and unique manner of author's writing. The topicality of the research lies in the lack of thorough idiolect research of Roman Ivanychuk's historical prose, a need for an integrated study of writer's lexical system based on the text corpus and by means of modern methods of analysis.

[1] COLINS-2021: 5th International Conference on Computational Linguistics and Intelligent Systems, April 22–23, 2021, Kharkiv, Ukraine
EMAIL: nata07lototska@gmail.com (N. Lototska)
ORCID: 0000-0001-6692-196X (N. Lototska)

2. Corpus based and statistical approach in linguistic research

Corpus based approach consists in presenting, verifying linguistic facts, studying the frequency of language units and the compatibility of these units with other elements. Corpus-based research reveals unpredictable patterns of variation [24, p. 84]. In corpus based studies “description should be comprehensive in terms of data set”, and “language categories” follow “from constantly repeating patterns and frequency distributions that go out of context” [24, p. 84, 87].

Text corpus is a reliable material for statistical analysis of any language or speech units, it possesses useful statistical information such as number of word types, frequency, co-occurrences [1]. Statistical approaches become one of the most efficient and time-saving tools of processing different sets of texts” [11].

Text corpus is able to provide a detailed study of writer's lexicon and to open prospects for further researches [2]. Statistical methods applied to different writers’ texts may reveal statistical characteristics which differ them one from the others and therefore present particular individual creative manner of a writer [14].

Text quantitative characteristics allow to objectively determine the qualitative characteristics of the writer's idiolect. It is generally acknowledged there is an internal interdependence between the qualitative and quantitative features of language structure, which determines the subordination of frequency of language units in speech to certain statistical patterns [12, p. 5].

Statistical studies allow to obtain new data or to test obtained knowledge about a linguistic unit, when the researcher is sure of the probabilistic nature of the linguistic object and aims to describe it in quantitative characteristics [15].

In a text, particularly in a sentence, the choice of words is determined not only by their denotative and significative meanings, but rather dependent on the surrounding words which they are grammatically and semantically related to.

The text corpus and tools of corpus linguistics allow to study the connectivity of lexical units, enables to identify and expand the lexical fund of set phrases of various types and peculiarities of their use [27]. Elena Tognini-Bonelli developed the "corpus model of meaning" based on corpus data [24, p. 214] which supposes that the meaning does not focus solely on one lexical unit, but extends to a word sequence.

3. Statistical study of word combinations in a text corpus

The importance of statistical methods in studying the word co-occurrence in a text corpus is evident. Statistical methods provide reliable quantitative data on the compatibility of lexical units based on corpus texts, one can study lexical units in context and obtain the data about frequency of lexical forms, lexemes, grammatical categories to further investigate the compatibility of lexical units and their peculiarities etc. “The combinatory ability of language units, collocability, is one of the linguistic laws” [27, p. 333].

The word co-occurrence is presented by a term ‘collocation’. There are two different views on what counts as a collocation in general linguistics: 1) a highly frequent word combination (this includes many frequent, but mundane, phrases, e.g. red shirt); 2) a word combination in which the collocate is not chosen absolutely freely (e.g., hazel eyes). Stefan Evert suggests the following definition: “A collocation is a word combination whose semantic and syntactic properties can’t be fully predicted on the basis of information about its constituents and which therefore should be added to the dictionary (lexicon)” [3, p. 17].

Works by John Sinclair have developed a corpus approach of understanding of collocation. Within this approach collocation is considered as the word combination used in the text together more often than at random probability separately [21, p. 115–116], in other words collocation is understood as statistically stable set phrases.

One of the main approaches of working with corpus data is to study collocations is concordance – text lines in a corpus represent the word in context. Concordance lines are the source of information about patterns of usage of word (node) and the connection between other words (collocate).

Nowadays, there are several ways to calculate the degree of coherence of parts of a collocation. Collocations are studied by means of mathematical criteria – association measures, which are based on probability theory and mathematical statistics. Association measures are mathematical formulas determining the strength of association between two or more words based on their occurrences and co-occurrences in a text corpus, they serve to calculate the degree of syntagmatic closeness between the components of collocation. These measures play an important role in the automatic extraction of collocations. Viktor Zakharov and Mariia Khokhlova state that methods like MI-score, t-score and log-likelihood are predominantly used to detect collocations [26].

Studies of collocations in Ukrainian fiction texts aren't carried out at a sufficient level. Collocation research provides important information about the author's style peculiarities.

4. Colour nomination ЧОРНИЙ (BLACK) as a marker of the author's idiolect

Colour nominations are important elements for constructing a conceptual and language picture of the world of a linguistic and cultural community [18]. The use of certain colour nominations, which takes part in modeling the real world in writer's fiction, consists in a manifestation of author's writing and a presentation of color picture of his world. Colour units play a conceptual role in the text and help the writer (reader) shed more light on the human psychology and understand the world [9], reflect the mentality of the linguistic personality.

Colour nominations were examined by different Ukrainian linguists [5, 6, 9, 16, 18, 23]. Researches of colour nominations are dedicated to study writer's personality based on semantic, quantitative, conceptual analysis.

The semantic structure and functioning of individual author's colour adjectives in the poetic Ukrainian neoclassicists' vocabulary is analyzed by Nataliia Gavrilyuk, noting that author's innovations are characterized by a more complex semantic structure compared to traditional colour lexemes [5]. Sviatoslav Gordinsky's idiostyle is studied by Volodymyr Piven and the linguist emphasizes the peculiarities of individual world perception, arranges colour nominations into thematic groups [16].

While studying colour nominations in poetic language of Lina Kostenko Galyna Gubareva identifies seven microfields according to the dominant color nominations and determines the quantitative content of each microfield [6]. Liudmyla Suprun performs semantic, stylistic and quantitative analysis of color nominations to highlight the specifities of their use in the novels by Ukrainian writers (Oles Honchar, Mykhailo Stelmakh, Pavlo Zagrebelny) [23].

Ryszard Tokarski pays attention to a linguistic colour representation in Polish language, connects the main colour nominations with culturally determined colour prototype and presents the complete colour symbolics with corresponding semantic connotation of colour nominations [25, p. 9].

To accomplish similar study for Roman Ivanychuk the corpus of his prose fiction texts is created. This corpus comprises 16 historical novels and 1 historical trilogy written throughout 1962-2016 (total corpus size is 1,295 million words): *At The Edge Of The Paven Way (Krai bytoho shliakhu), Mallows (Mal'vy), Red Wine (Cherlene vyno), Manuscript From Ruska Street (Manuskrypt z vulytsi Rus'koyi), Water From The Stone (Voda z kameniu), The Fourth Dimension (Chetvertyi vymir), Scars On The Rock (Shramy na skali), Crane's Cry (Zhuravlynyi kryk), Because War Is War (Bo viyna viynoyu), Horde (Orda), The Gospel Of Thomas (Yevanheliye vid Tomy), Pillars Of Fire (Vohnenni stovpy), Saxaul In The Sands (Saksaul u piskakh), Across The Pass (Cherez pereval), Pilgrimage (Khresna proshcha), Voices From Above The Waters Of Kinneret (Holosy z-nad vod Henisareta), I Have Not Written About Donbass Yet (Ya shche ne pysav pro Donbas)*. First, the texts of the novels were converted into electronic

form, the next step was the normalization[2] of the texts in the MS Word editor [11]. "Text normalization process contains the following stages: normalization of coding, normalization of graphics, text proofreading, technical normalization of punctuation" [11, p. 58].

The next step was to upload these texts into GRAC v.8 [20] and thus creating the Roman Ivanychuk's subcorpus (RITC). The GRAC makes possible to search any linguistic phenomenon using NoSketchEngine interface that enables search by lemma, word form and grammatical tags, visualization of their frequencies as a concordance, customize text filters (texts of a given period, style, original language etc.) [19].

For an integrated research of the author's idiolect, the subcorpus of Ukrainian prose fiction (UPFTC) was created in the GRAC by applying filters like: style Fiction (DOC.STYLE – FIC), original language Ukrainian (DOC.ORIGINAL – UK), time span (DOC.DATE – 1960–2016). The size of this subcorpus is 73, 234 million words.

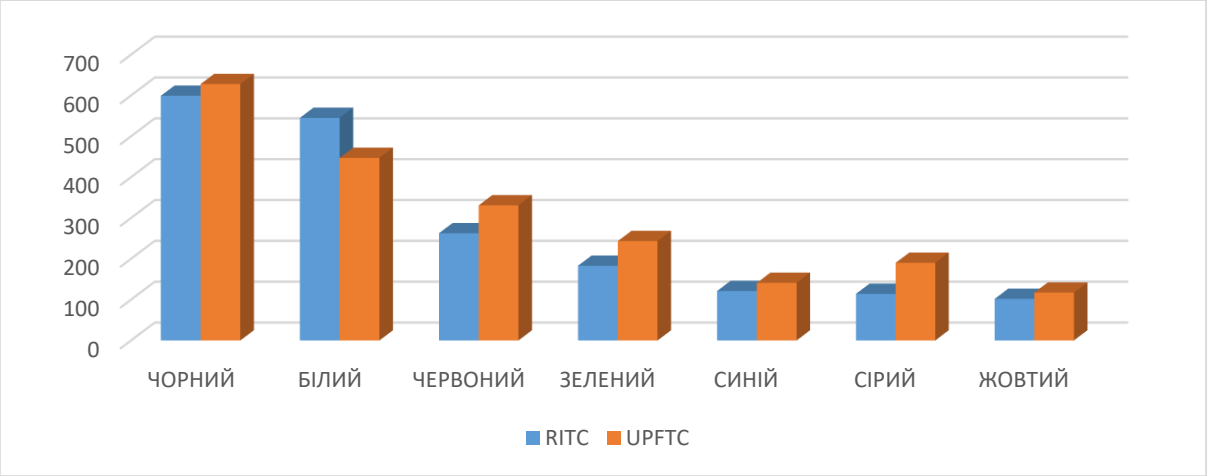
The obtained data from both corpora – RITC and UPFTC – allow to compare statistical parameters and qualitative indicators and reveal lexical markers of the author's idiolect. The texts and the results of the lemmatization were subjected to statistical analysis.

In our study the hypothesis is that colour nominations demonstrate the author's vision and his cultural experience, therefore collocations with colour nominations may be considered writer's idiolect markers. Student's t-test (t-value) is used to test the hypothesis and to reveal significant results.

The absolute and relative frequencies of colour nominations in Roman Ivanychuk's text corpus and Ukrainian prose fiction text corpus are manifested in the table 1 and in the figure 1.

Table 1. Absolute and relative frequencies of colour nominations use in RITC and UPFTC

Colour	Roman Ivanychuk		Ukrainian prose fiction	
	Absolute frequency	Relative frequency	Absolute frequency	Relative frequency
ЧОРНИЙ	776	599,2	46 800	627,9
БІЛИЙ	706	545,1	33 323	447,1
ЧЕРВОНИЙ	340	262,5	24 667	330,9
ЗЕЛЕНИЙ	237	183	18 143	243,4
СИНІЙ	157	121,2	10 521	141,1
СІРИЙ	148	114,2	14 170	190,1
ЖОВТИЙ	132	101,9	8 735	117,2



[2] Normalization means a set of information procedures that make the texts suitable for insertion into the corpus: bringing all texts to one code table, checking them for punctuation correctness (sense-identical entities should be marked with one character), eliminating unnecessary characters (for example, blank paragraphs, several gaps in a row, etc.), unification of formatting tools and methods, and more [11, p. 58].

Figure. 1. Absolute and relative frequencies of colour nominations use in RITC and UPFTC

The empirical t-value of frequency of colour nominations in RITC and UPFTC is 2.6 and this value is in the zone of indeterminacy and thus cannot be treated as idiolect markers.

In Roman Ivanychuk's text corpus the most frequent colour nomination is ЧОРНИЙ (BLACK). Figure 2 presents the frequency of ЧОРНИЙ (BLACK) in RITC and in Ukrainian writers' prose fiction (based on UPFTC data). The frequency of ЧОРНИЙ (BLACK) is the highest in texts by Pavlo Zagrebelny (13.3) and Oles Berdnyk (12.7), Roman Ivanychuk's texts (12.2) take the third place by the frequency.

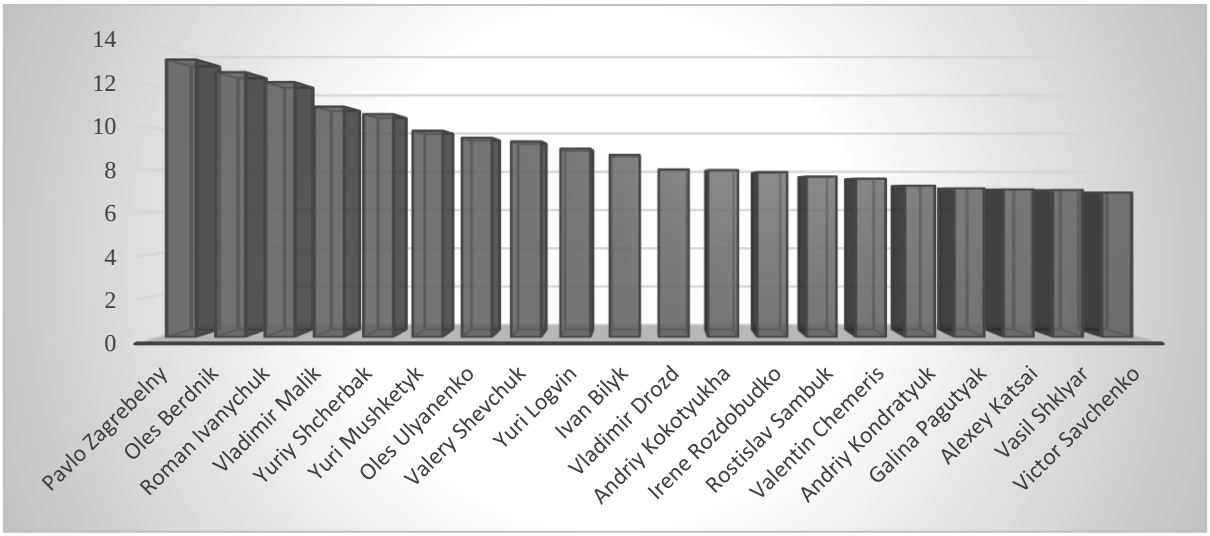


Figure. 2. The frequency of colour nomination ЧОРНИЙ (BLACK) in Ukrainian prose fiction text corpus

Words co-occurrences with colour component ЧОРНИЙ (BLACK) are described and extracted by means of the GRAC and its Collocation tool of the NoSketch Engine system. Absolute and relative frequency, association measures T-score and MI-score are applied to study the obtained collocations as markers of the author's idiolect.

In texts by Roman Ivanychuk and Ukrainian fiction writers the most frequent co-occurrences with BLACK are collocations with functional words (prepositions, conjunctions). Typical models are presented in the table 2.

Table 2. High-frequent collocations with the colour component ЧОРНИЙ (BLACK)

Model	**Roman Ivanychuk**		**Ukrainian fiction**	
	Word combination	**Absolute/Relative frequency**	**Word combination**	**Absolute/Relative frequency**
Prep. + **чорний** / **чорний** + Prep.	в |чорний|	79/61,00	в |чорний|	2436/33,7,
	з |чорний|	59/45,56	у |чорний|	2286/31,5
	у |чорний|	49/37,84	з |чорний|	1950/26,8
	на |чорний|	26 /20,08	на |чорний|	1677/22,9
	до |чорний|	20/15,44	до |чорний|	552/7,67
	за |чорний|	12/9,27		
	|чорний| з	8/6,18	|чорний| з	165/2,32
	|чорний| від	7/5,41	|чорний| від	261/3,6
	|чорний| по білому	5/3,86	|чорний| по білому	115/1,5
Conj. + **чорний** / **чорний** + Conj.	й |чорний|	20/15,44	і |чорний|	1313/17,9
	і |чорний|	20 /15,44	|чорний| і	520/7,07

Model	Roman Ivanychuk		Ukrainian fiction	
	Word combination	Absolute/Relative frequency	Word combination	Absolute/Relative frequency
	|чорний| й	8/6,18	й |чорний|	454/6,36
	|чорний| і	7/5,41	|чорний| й	301/4,15
			|чорний| та	179/2,4

*The relative frequency (RF) ≥ 2 is taken into consideration.

Comparative study of word combinations with colour component ЧОРНИЙ (BLACK) in RITC and UPFTC demonstrates that high-frequency word combinations are: *в |чорний|, у |чорний|, на |чорний|, і |чорний|*. In Roman Ivanychuk text corpus the frequency of collocations *в |чорний|, з |чорний|, у |чорний|, до |чорний|, й |чорний|* is much higher than in Ukrainian fiction, for example: *|чорний| й*. Moreover the collocation *за |чорний|* and the idiom *|чорний| по білому* are high-frequent in Roman Ivanychuk's texts. The obtained results in both text corpora are presented in the figure 3.

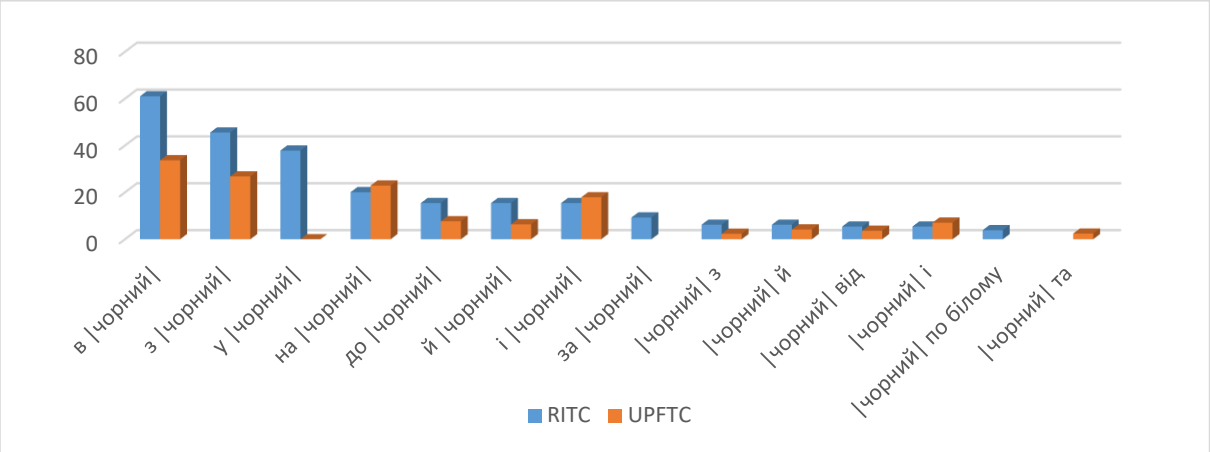


Figure. 3. High-frequent collocation with colour component ЧОРНИЙ (BLACK)

The empirical t-value of frequency of collocations with the colour component ЧОРНИЙ (BLACK) in RITC and UPFTC is 3,7; this value is in the zone of determinacy and is treated as idiolect markers.

In RITC among the high-frequent collocations with ЧОРНИЙ (BLACK) are word combinations with the meaning of cause (*чорний від праці 2/1,54, від безнадійної вістки 1/0,77, від кіптяви, від злоби, від поганої лайки*), co-occurrences with a conjunction to combine opposite colours (*чорний і білий 2/1,54, червоний і чорний 1/0,77*) and the common idiom (*чорний по білому 5/3,86*).

The colour nomination ЧОРНИЙ (BLACK) is mostly an attribute of nouns (collocates) in the model Adj + N. These attributes belong to such thematic groups[3] (absolute/relative frequency indicators are presented to the right from the example).

UNIVERSE. SPACE. TIME. EARTH. INANIMATE NATURE. **Space and its elements. Natural elements and phenomena**: *отвір 10/7,72, море 8/6,18; небо, хмара, нора 6/4,63; яма, рілля, шлях 5/3,86; земля, потік, провалля 4/3,09; діра, згарище 3/2,32; плесо, небозвід, озеро, пруг, рядок, смуга, темінь, темрява 2/1,54; дно, лід, рівнина, хмарка, обеліск, стовпець, буря, злива, тінь 1/0,77*. **Movement**: *хід 2/1,54, вихід 1/0,77.* **Time, time interval, time characteristics:** *ніч 3/2,32; година, майбуття, свят 1/0,77.* **Building, facility, accommodation and their parts:** *кам'яниця, каменица, стіна 4/3,09; бійниця, окуляри 3/2,32; клуб, руїна, підвал, стріха 2/1,54; будинок, фортеця, рама, димар, іконостас, стеля, баня, ніша 1/0,77.* **Material, substance, part of something:** *дим 9/6,95; вода 3/2,32; бляха, вуглинка, дерматин, клапоть, мідь, лава, накип, нитка, павутиння, папір, пил, пух, сажа 1/0,77.* **Object:** *книга 2/1,54; квадрат, ключ, обвідка, палітурка, подушка, предмет, знак, підківка, перука, свастика, дужка, пеньок, браслет, жердина, запона,*

[3] Mykhail Mukhin's and Natalia Snizhko's classifications have been modified [13; 22]

вервечка, хоругва, хрест, цятка 1/0,77. **Supernatural phenomena:** *магія 3/2,32; демон, мара, привид 1/0,77.*

EARTH: NATURAL WORLD. HUMAN AS A PART OF NATURAL WORLD. **Animals:** *ворон 6/4,63; дрізд, крук, метелик 3/2,32; гайвороння 2/1,54; буйвіл, галка, жук, дрохва, кіт, мул, орел, пташка 1/0,77.* **Plants:** *ліс 5/3,86, ягідка 2/1,54, дерево 1/0,77.* **Human and family relationships:** *люд, постать, натовп 2/1,54; євнух, чорнокнижник, караван, юрба, валка, орда, король, челядь, мужлан, чоловік 1/0,77.* **Human body, parts of the body:** *око 21/15,44, волосся 20/16,21, борода, борідка 7/5,41; брова, вуса, чуб 4/3,09; рука, кучері 3 /2,32; рот, крило, шкіра, очиці, паща, сльоза 2/1,54; надбрів'я, кістяки, душа, кров, рубець 1/0,77* . **Clothes:** *сурдут 9/6,95, сутана 8/6,18, ряса 7/5,41, капелюх 6 /4,63; сукман, хустка, пов'язка 5/3,86; кибитка, пелерина 4/3,09; кирея, шапка, мундир, стрічка, ярмулка, вбрання, намітка, туніка, каптур, сукня, фрак, хламида 3 /2,32; бинда, бурка, плащ, костюм, хустина 2/1,54; пройма, плаття, керсетка, піджак, шаль, краватка, камзол 1/0,77*. **Action**: *війна, процесія, робота 1/0,77*. **Emotion**: *ненависть 4/3,09, біда, горе, зрада 2/1,54; лють 1/0,77*. **Message**: *вість 2/1,54, звістка 1/0,77.* **Mentality, ability, intellect**: *сила 2/1,54, думка 1/0,77*. **Proper name**: *Чорна рада 4/3,09.* The obtained results are presented in the figure 4.

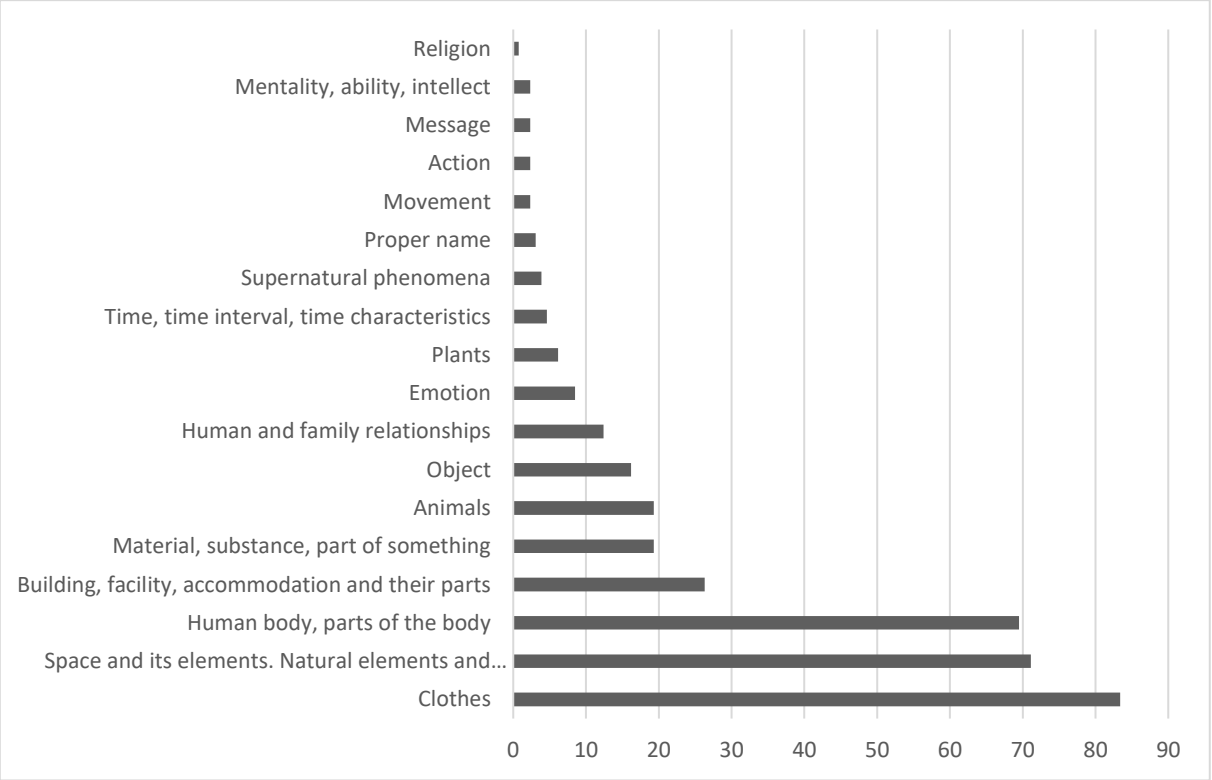


Figure. 4. Thematic groups of collocates with ЧОРНИЙ (BLACK) as an attribute to a noun in RITC

The analysis of these thematic groups demonstrates that this colour nomination is mostly combined with the notions of **Clothes, Space (Natural elements and phenomena)** and **Human body (Parts of the body**).

In UPFTC high-frequent collocates of colour ЧОРНИЙ (BLACK) as an attribute are: *око 13,95*[4], *волосся 8, море 7,72, хмара 6,88, діра 4,95, тінь 4,41, дим 4,36, ліс 3,88, брова 3,25, небо 3,21, костюм 3,03, дірка 3,01, хліб 2,7, пляма 2,68, плащ 2,66, колір 2,62, орда 2,58, отвір 2,54, хустка 2,51, день 2,42, вода 2,31, окуляри 2,25, пес 2,19* (20 most frequent collocates are taken into consideration). See the figure 5 which contains the results retrieved from RITC and UPFTC.

[4] Relative frequency

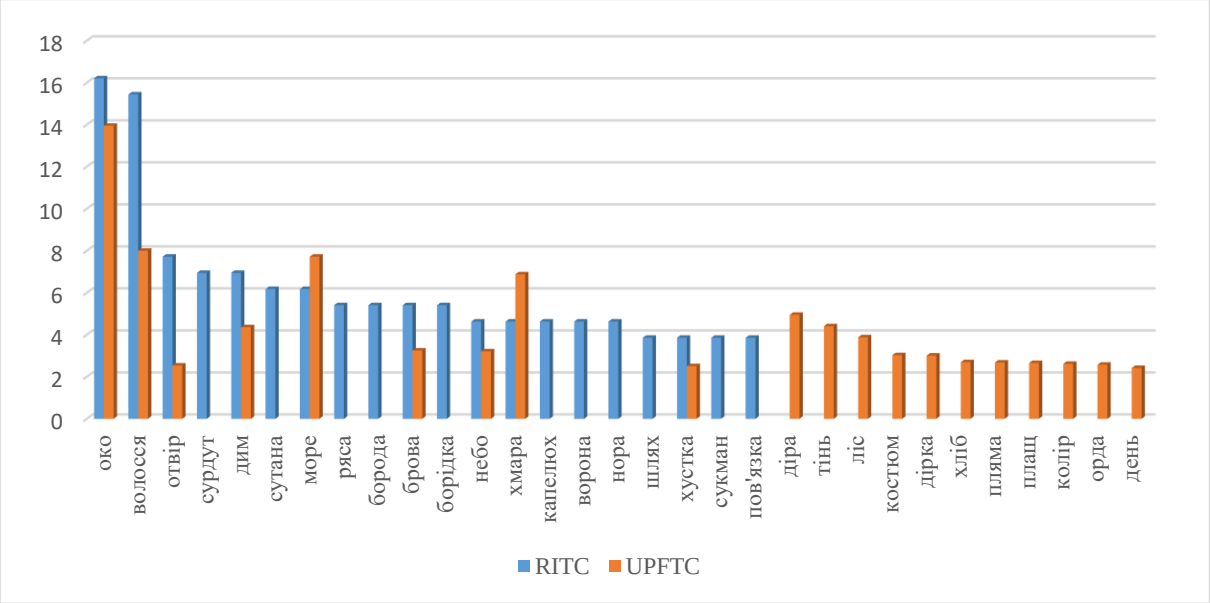


Figure. 5. High-frequent collocates of colour nominations ЧОРНИЙ (BLACK) as an attribute to a noun in RITC and UPFTC.

The empirical t-value of frequency of collocates of colour nominations ЧОРНИЙ (BLACK) in RITC and UPFTC is 6,2; this value is in the zone of determinacy and can be presented as idiolect markers.

The most frequent collocates of colour nomination ЧОРНИЙ (BLACK) in both corpora are: *око, волосся*, but their frequencies are different, especially for the collocate *волосся* (the frequency of collocation *чорне волосся* is twice much in RITC than in UPFTC). Collocates *отвір, дим, брова, небо, хустка* are much more frequent in Roman Ivanychuk's text corpus, collocates like *море, хмара* are much less frequent than in Ukrainian prose fiction. Meanwhile the words *сурдут, сутана, борідка, капелюх, ворона, нора, шлях, сукман, пов'язка* are high-frequent in Roman Ivanychuk's text corpus, but they were not retrieved to the list of 20 high-frequent collocates among colour nomination under study in UPFTC.

The use of rare words as collocates of colour nomination ЧОРНИЙ (BLACK) is a characteristic feature of Roman Ivanychuk's texts, e.g.: *бійниця, обвідка, сурдут, сукман, кибитка, кирея, ярмулка, намітка, бинда, пройма, керсетка, камзол, мужлан.*

The researchers claim that the simplest way to detect a collocation pair is based on the relative frequency, which gives the most common collocation associations, however, this method has a number of drawbacks. Considering this, it is obvious that one of the options could be MI-score [4]. Elena.Yagunova and Lidiya Pivovarova concluded that the lists of collocations obtained using MI-score and t-score differ fundamentally: MI-score is the best one for distinguishing object names, terms, complex nominations; t-score, on the contrary, works better when distinguishing between 'lexical bundles' (derivative functional words, discourse markers) and 'set expressions' [17].

In our research the statistical association measures MI-score and T-score are taken into consideration and are applied to word combinations with the node ЧОРНИЙ (BLACK). The obtained data are presented in Table 2 and Table 3.

Table 2. Co-occurrences with colour component ЧОРНИЙ (BLACK) extracted by T-score[5]

№	Word combination	Relative frequency	T-score
1.	в |чорний|	61	8.884
2.	з |чорний|	40	7.2073
3.	у |чорний|	37,8	6.996
4.	на |чорний|	20,08	5.091

[5] The 20 highest co-occurrences according to T-score

№	Word combination	Relative frequency	T-score
5.	чорне око	16,22	4.581
6.	чорне волосся	15,44	4.472
7.	до |чорний|	15,44	4.466
8.	й |чорний|	15,44	4.465
9.	і |чорний|	15,44	4.461
10.	за |чорний|	9,27	3.460
11.	чорний отвір	7,72	3.162
12.	чорний сурдут	6,95	2.999
13.	чорний дим	6,95	2.999
14.	чорна сутана	6,18	2.828
15.	чорне море	6,18	2.828
16.	з |чорний|	6,18	2.818
17.	чорна борідка	5,4	2.645
18.	чорна ряса	5,4	2.645
19.	чорні брова	5,4	2.645
20.	чорна борода	5,4	2.645

Table 3. Word combinations with colour component ЧОРНИЙ (BLACK) extracted by MI-score[6]

№	Word combination	Relative frequency	MI
1.	чорна дрохва	0,77	19.313
2.	чорний сукман	3,9	18.828
3.	чорний дрізд	2,32	18.576
4.	чорний потік	1,54	18.313
5.	чорна сажа	0,77	18.313
6.	чорний дерматин	0,77	18.313
7.	чорна галка	0,77	18.313
8.	чорний мужлан	0,77	18.313
9.	чорна хламида	2,32	18.091
10.	чорна палітурка	1,54	17.991
11.	чорна пов'язка	3,9	17.935
12.	чорна пелерина	3,09	17.854
13.	чорна камениця	3,09	17.854
14.	чорна сутана	6,18	17.854
15.	чорне надбрів'я	0,77	17.728
16.	чорна борідка	5,4	17.121
17.	чорний отвір	7,72	16.635
18.	чорний сурдут	6,95	16.625
19.	чорна ряса	5,4	16.566
20.	чорна хустка	3,9	16.050

As can be observed statistical measure t-score extracts the most frequent collocations, while the MI-score allows to reveal low-frequency co-occurrences. The MI-score measure is critical when rare multi-word terms are to be extracted.

[6] The 20 highest co-occurrences according to MI-score

It is worth noting that corpus based approach, values of relative frequency and statistical association measure allowed to reveal author-individual word combinations of Roman Ivanychuk's text corpus: *чорне богогульство, чорна війна, чорна дрохва, чорний іконостас, чорна каменица, чорна кибитка, чорне майбуття, чорний мужлан, чорне надбрів'я, чорна процесія, чорна челядь*.

5. Conclusion

Text corpus is a reliable material for statistical analysis of compatibility of lexical units. Statistical association measures allow us to identify co-occurrences in text corpus.

Colour collocations with the component ЧОРНИЙ (BLACK) are important markers used in the writer's idiolect study. The research of co-occurrences with colour ЧОРНИЙ (BLACK) presents their list and helps to analyze them. The statistical study of colour nominations in Roman Ivanychuk's text corpus and Ukrainian prose fiction text corpus allows to reveal author-individual characteristics of Roman Ivanychuk which are significant.

Colour nomination ЧОРНИЙ (BLACK) is the most frequent in Roman Ivanychuk's text corpus. High-frequent co-occurrences with this colour component are prepositional models like *Prep. + ЧОРНИЙ (BLACK)*, with the meaning of cause. The colour nomination ЧОРНИЙ (BLACK) acts mostly as an attribute of nouns (collocates) in the model *Adj + N*. Thematic groups of collocates of ЧОРНИЙ (BLACK) as an attribute are described and analyzed. The component ЧОРНИЙ (BLACK) is often associated with the nominations of Clothes, Space (Natural elements and phenomena) and Human body (Parts of the body). High-frequent and low-frequent collocates of colour ЧОРНИЙ (BLACK) are studied by means of absolute / relative frequency and statistical association measures MI-score and T-score. As a result MI association index is more suitable for determining author-individual constructions.

List of obtained high-frequency colour co-occurrences in Roman Ivanychuk's text corpus is different from the data obtained from Ukrainian literary prose text corpus, which, in its turn, determines the specificity of the author's idiolect. The practical results of the study can be applied for text attribution and further research of writer's individual written language. Obtained statistical data may be useful to compile frequency and collocation dictionaries of Roman Ivanychuk's and Ukrainian prose fiction texts.

6. References

[1] D. Biber and S. Conrad, Register, genre, and style, Cambridge University Press, 2009, 344 pp.

[2] S. Buk, Kvantytatyvna parametryzatsiia tekstiv Ivana Franka: proekt ta yoho realizatsiia, in: Visnyk Lvivskoho universytetu, Seriia filolohichna, Vyp. 58, 2013, s. 290–307. [S. Buk, Quantitative parameterization of Ivan Franko's texts: project and its realization, in: Bulletin of Lviv University, Philological Series, Vol. 58, 2013, pp. 290–307]

[3] S. Evert, The Statistics of Word Cooccurrences: Word Pairs and Collocations, Ph.D.thesis, University of Stuttgart, 2004 (Published in 2005).

[4] O. Gorina, Primenenie metodov korpusnoj lingvistiki dlya opredeleniya kontekstno-specificheskikh slov i kollokacij, in: Vestnik Leningradskogo gosudarstvennogo universiteta imeni A.S. Pushkina. Seriya: Ehkonomika, T. 3, 2011. [O. Gorina, Application of corpus linguistics methods to the definition of con-text-specific words and collocations, in: Bulletin of A.S. Pushkin LSU, vol. 3, 2011]

[5] N. V. Havryliuk, Semantychna struktura ta funktsionuvannia indyvidualno-avtorskykh prykmetnykiv-koloronazv u tvorchomu dorobku neoklasykiv, in: Kultura narodov Prychernomoria, № 32, 2002, s. 332–334. [N. V. Gavrilyuk, The semantic structure is the function of individual-author's color adjectives in the texts of neoclassicals, in: Culture of the peoples of the Black Sea region, no. 32, 2002, pp. 332–334]

[6] H. A. Hubareva, Semantyka ta stylistychni funktsii koloratyviv u poetychnii movi Liny Kostenko, avtoref. dys. … kand. filol. nauk, Kharkiv, 2002, 18 s. [G. A. Gubareva, Semantics and stylistic

functions of colour names in the poetic language of Lina Kostenko, Ph.D.thesis, philological sciences, Kharkiv, 2002, 18 p.]
[7] S. D. Kacnel'son, Soderzhanie slova, znachenie i oboznachenie, in: M.–L., 1965, s. 81–81. [S. D. Katznelson, The content of the word, meaning and notation, in: M.–L., 1965, pp. 81–81]
[8] M. Khokhlova, Yssledovanye leksyko-syntaksycheskoy sochetaemosti v russkom yazyike s pomoshchiyu statystycheskykh metodov (na baze korpusov tekstov): avtoref. dys. na soysk. uch. step. kand. fylol. nauk, Sankt-Peterburg, 2010, 218 s. [M. Khokhlova, The study of lexical and syntactic collocatibility in the Russian language using statistical methods (based on Text Corpus), Ph.D.thesis, philological sciences, Sankt-Peterburg, 2010, 218 p.]
[9] N. V. Kopteva, Funkcionirovanie kolorizmov v hudozhestvennom tekste kak rezul'tat vzaimodejstviya lingvokul'turnogo i kreativnogo faktorov: na materiale tvorchestva russkikh pisatelej XIX – XX, dis.... kand. filol. nauk, Rostov-na-Donu, 2005, 247 s. [N. V. Kopteva, Colour functioning in the artistic text as a result of interaction of linguocultural and creative factors: based of texts of Russian writers of XIX – XX centuries, Ph.D. thesis, philological sciences, Rostov-na-Donu, 2005, 247 p.]
[10] I. Kulchytskyi, Statistical analysis of the short stories by Roman Ivanychuk, in: Proceedings of the 3rd International conference on computational linguistics and intelligent systems, COLINS–2019, Kharkiv, Ukraine, 2019, pp.312–321.
[11] I. M. Kulchytskyi, Unormuvannia tekstu pid chas dokorpusnoho opratsiuvannia: dosvid zastosuvannia, in: Visnyk Natsionalnoho universytetu “Lvivska politekhnika”, Seriia: Informatsiini systemy ta merezhi. Vyp. 7, 2020, S. 51–58. [I.M. Kulchytskyy, Text normalization during pre-corpus preparation: experience of application, in: Bulletin of the National University "Lviv Polytechnic", Series: Information systems and networks, vol. 7, 2020, pp. 51–58
[12] V. V. Levitskiy, Kvantitativnyie metodyi v lingvistike, Nova Kniga, Vinnitsa, 2007, 264 s. [V. V. Levitsky, Quantitative methods in linguistics, Nova Kniga, Vinnitsa, 2007, 264 p.]
[13] M. Yu. Mukhin, Leksicheskaya statistika i idiostil' avtora: korpusnoe ideograficheskoe issledovanie (na materiale proizvedenij M. Bulgakova, V. Nabokova, A. Platonova i M. Sholokhova), avtoreferat dissertacii na soiskanie uchènoj stepeni doktora filologicheskikh nauk, Ekaterinburg, 2011, 43 s. [M. Yu. Mukhin, Lexical statistics and author’s idiostyle: corpus ideographic research (based on the works of M. Bulgakov, V. Nabokov, A. Platonov and M. Sholokhov), Ph.D. thesis, philological sciences, Ural State University, Ekaterinburg, 2011, 43 p.].
[14] O. O. Pavlychko, Shchodo statystychnykh parametriv avtorskoho styliu (na materiali tvoriv E.M. Remarka), in: Movni i kontseptualni kartyny svitu, Vyp. 29, VPTs «Kyivskyi un-t», Kyiv, 2010, s. 186–191. [O. O. Pavlichko, Regarding the statistical parameters of the author's style (based on the texts of E.M.Remark), in: Linguistic and conceptual worldview, Vol. 29, PPC Kyiv University, Kyiv, 2010, pp. 186–191]
[15] V. S. Perebyinis, M. P. Muravytska, N. P. Darchuk, Chastotni slovnyky ta yikh vykorystannia. Kyiv, 1985, 204 s. [V. S. Perebiynis, M. P. Muravitska, N. P. Darchuk, Frequency dictionaries and their use. Kyiv, 1985, 204 p.]
[16] V. F. Piven, Idiostyl poetychnykh tvoriv Sviatoslava Hordynskoho, avtoref. dys. ... kand. filol. nauk, Zaporizhzhia, 2007, 20 s. (2007). [V. F. Piven, Idiostyle of poetic texts s of Sviatoslav Hordynsky, Ph. D. thesis, philological sciences, Zaporizhzhia, 2007, 20 p.]
[17] L. M. Pivovarova, E. V. Yagunova, Izvlecheniye i klassifikatsiya terminologicheskikh kollokatsiy na materiale lingvisticheskikh nauchnykh tekstov (predvaritelnyye nablyudeniya), in: Materialy Simpoziuma «Terminologiya i znaniye», Moskva, 2010, s. 214–229. [L. M. Pivovarova, E. V. Yagunova, The extraction and classification of term collocations based on linguistic scientific texts (preliminary observation), in: Symposium materials “Terminology and knowledge”, Moscow, 2010, p. 214–229]
[18] T. F. Semashko, Osoblyvosti semantyky ta funktsionuvannia sliv-koloratyviv v ukrainskii frazeolohii, avtorepherat dysertatsii na zdobuttia naukovoho stupenia kand. filol. nauk, 2008, 27 s.

[T. F. Semashko, Semantics peculiarities and word-colour functioning in Ukrainian phraseology, Ph.D.thesis, 2008, Kyiv, 27 p.]

[19] M. Shvedova, The General Regionally Annotated Corpus of Ukrainian (GRAC, uacorpus.org): Architecture and Functionality, in: Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems, COLINS 2020, vol. I, Lviv, Ukraine, 2020, pp. 489–506.

[20] M. Shvedova, R. von Waldenfels, S. Yarygin, M. Kruk, A. Rysin, V. Starko, M. Woźniak: GRAC: General Regionally Annotated Corpus of Ukrainian, uacorpus.org.

[21] J. Sinclair, Corpus, concordance, collocation: Describing English language, Oxford University Press, 1991, 179 p.

[22] N. Snizhko, Ukrainska ideohrafiia: istoriia, suchasnyi stan ta perspektyvy, in: Ukrainska mova, № 3, 2016, Ss. 28–43. [N. Snizhko, Ukrainian ideography: history, current state and prospects, in: Ukrainian language, № 3, 2016, pp. 28–43]

[23] L. O. Suprun, Semantyka i prahmatyka nazv koloriv v ukrainskomu romannomu teksti seredyny – druhoi polovyny KhKh st. (na materiali tvoriv O. Honchara, P. Zahrebelnoho, M. Stelmakha), dys. … kand. filol. nauk: 10.02.01, Kharkiv, 2009, 235 s. [L. O. Suprun, Semantics and Pragmatics of coulor nominatios in the Ukrainian Novel Text of middle – second half of the twentieth century (based on texts of O. Gonchar, P. Zagrebelny, M. Stelmakh), Ph.D. thesis, philological sciences, Kharkiv, 2009, 235 p.]

[24] E. Tognini-Bonelli, Corpus Linguistics at Work. Amsterdam, John-Benjamins, 2001, 236 p.

[25] R. Tokarski, Semantyka barw we współćzesnej polszczyznie, Lublin, UMCS, 1995, 249 s.

[26] V. P. Zakharov, M. V. Khokhlova, Avtomaticheskoye vyyavleniye terminologicheskikh slovosochetaniy, in: Strukturnaya i prikladnaya lingvistika, vyp.10, Sankt-Peterburg, 2014, s. 182–200 [V.P. Zakharov, M.V. Khokhlova, Automatic detection of terminological word combinations, in: Structural and Applied Linguistics, issue 10, Sankt-Petersburg, 2014, pp. 182–200].

[27] V. P. Zakharov, Sochetaemost cherez pryzmu korpusov, in: Kompiuternaia lynhvystyka y yntellektualnыe tekhnolohyy, 14 (21), Moskva, 2015, s. 667–682. [V.P. Zakharov, Word collocatibility in Text Corpus, in: Computer linguistics and information technologies, 14 (21), Moscow, 2015, pp. 667–682]