

# Key Colour Terms in the Ukrainian Prose Fiction of the 21<sup>st</sup> Century

Olena Levchenko and Marianna Dilai

*Lviv Polytechnic National University, Bandera Str., 12, Lviv, 79013, Ukraine*

## Abstract

This paper presents a corpus-based study of key colour terms in the Ukrainian prose fiction of the 21<sup>st</sup> century. The aim of this study is to identify statistical characteristics of colour terms in the fiction texts, and to determine whether colour terms can be markers of idiolect / genderlect, in particular in terms of their frequency, derivation potential and collocability. The subcorpora of women's and men's prose fiction of the periods of 2000-2020 and 1980-1999 have been created in the GRAC (GRAC-11). A comprehensive research methodology includes quantitative methods supplemented by qualitative ones. A colour profile in the corpus and idiolect includes such parameters as the frequency of key colour terms, determining the colours which are used and which are not used by a language personality, structural grouping, calculating the median and the mode, identifying a collocate type (concrete / abstract) and thematic grouping of collocates, determining stabilized and individual author collocability, modelling dominant directions of metaphorization in the idiolect.

## Keywords

Colour terms, Ukrainian prose fiction, corpus-based approach, statistical profile, GRAC

## 1. Introduction

Today, colour terms are studied from different perspectives: conceptualization and verbalization in different languages in diachrony and synchrony, in particular the contrastive aspect, phrase-forming potential, individual worldview and discourse, translation, psycholinguistic features [1, 2, 3, 4, 5]. Much research on colour terms has been done in Ukrainian linguistics [6, 7, 8, 9, 10, 11]. The **relevance** of this paper lies in using corpus-based approach and analysing frequency of colour terms in the modern Ukrainian prose fiction on the basis of a large volume of corpus data.

The **purpose** of this paper is to identify statistical characteristics of colour terms (statistical profile of colour terms) in the Ukrainian fiction texts of the 21<sup>st</sup> century, and to determine whether colour terms can serve as markers of idiolect / genderlect, in particular in terms of their frequency, derivation potential and collocability. For the purpose of this research, we use the data and features of the GRAC [12]. The research **objectives** are as follows:

- to create subcorpora of the Ukrainian women's and men's prose fiction of two time periods, 1980-1999 and 2000-2020, in the GRAC (GRAC-11);
- to extract relative frequency of key colour terms and determine 'colour formulas' for the subcorpora, i.e., the order of decreasing / increasing frequency of colours;
- to perform structural grouping and determine the specificity of the use of colour terms calculating the median and the mode;
- to analyse collocations, determine thematic groups of colour terms collocates and identify the topics of distribution groups that prevail in the analysed texts;
- to view stabilized and individual author collocability and specify the main directions of metaphorization in the idiolect.

---

COLINS-2022: 6th International Conference on Computational Linguistics and Intelligent Systems, May 12–13, 2022, Gliwice, Poland.

EMAIL: levchenko.olena@gmail.com (O. Levchenko); mariannadilai@gmail.com (M. Dilai);

ORCID: 0000-0002-7395-3772 (O. Levchenko); 0000-0001-5182-9220 (M. Dilai);



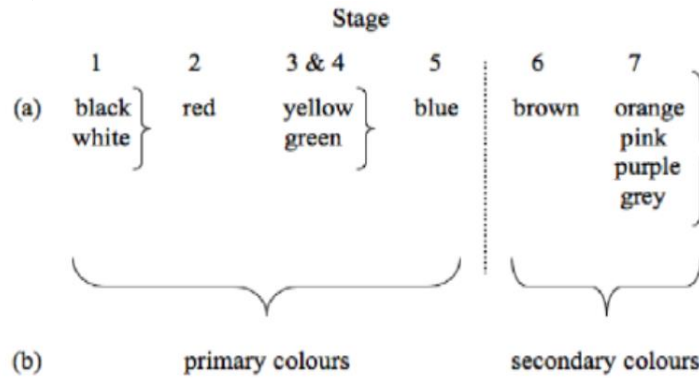
© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

## 2. Related Works

The seminal study by B. Berlin and P. Kay *Basic Colour Terms: Their Universality and Evolution* [13] had a strong resonance with linguists presenting valuable conclusions on the basic colour terms in a culture (see Figure 1).



**Figure 1:** Fixed development of colour term acquisition proposed by B. Berlin and P. Kay (1969)

The statistical study of colour terms was presented in the proceedings of the conference *Progress in Colour Studies in 2006* in the paper of A. Pawłowski *Quantitative Linguistics in the Study of Colour Terminology* [14]. The research describes a comprehensive quantitative analysis of the structure of the lexical field of colour in a multilingual corpus. Analysing several genetically and culturally related languages, the author comes to the conclusion on the moderate universality in the colour conceptualisation. The rank distribution of colour terms in different languages, as presented by A. Pawłowski [14], is shown in Figure 2.

	Czech	English	French (FDFW)	French (VRFDF)	Italian	Polish	Rumanian	Russian	Slovak	Spanish	Ukrainian
white	1	2	1	1	1	2	1	2	2	1	2
black	2	1	2	2	3	1	2	1	1	2	1
red	3	3	3	3	2	3	4	3	3	5	3
green	5	4	5	5	4	4	3	5	4	4	5
yellow	9	8	8	7	7	8	5	7	6	7	7
blue	4	5	4	4	5	5	6	4	5	3	4
brown	8	7	7	8	8	7	7	8	9	8	9
purple	12	11	9*	11*	11*	12*	8*	10	12	11*	11
pink	6	9	10*	12*	9	9	9*	9	7	12*	8
orange	11	10	11*	10	12*	10	10*	12	10	10	12*
grey	7	6	6	6	6	6	11*	6	8	6	6
violet	10	12*	12*	9	10	11	12*	11	11	9	10

**Figure 2:** Rank distribution of the colour categories in the corpora (languages) by A. Pawłowski (2006)

More recent studies proposed a mathematical algorithm for determining basic colour terms from color-naming data [15], designed quantitative methods to study colour categorization [16] and measured the differences in color categorization systems manifested by men and women [17].

Ukrainian scholar V. Starko used corpus data in the reconstruction of referential models underlying the Ukrainian colour terms *блакитний*, *голубий*, *синій* 'blue' [9]. The statistical research on the colour terms *білий* (*white*) and *чорний* (*black*) as markers of idiolect was carried out by O. Levchenko and N. Lototska [6, 7].

It is worth noting that the conceptual domain of colour is a rich source domain of metaphors. E. Leach claims that since all objects of the visible external world have a colour attribute, colour

differences always make a good basis for classification. However, one colour class usually includes an infinite variety of things, so social metaphors of colour are always potentially ambiguous [18].

### 3. Methods and Materials

To achieve the research objectives, the subcorpora of the Ukrainian women's and men's prose fiction of the periods 2000-2020 and 1980-1999 have been created in the GRAC (GRAC-11) [12]. The research subcorpora comprise the works of the following Ukrainian writers: Yu. Andrukhovych, Yu. Vynnychuk, M. Hrymych, L. Denysenko, L. Deresh, O. Zabuzhko, R. Ivanychuk, Br. Kapranov, I. Karpa, V. Kozhelyanko, A. Kokotyukha, Y. Kononenko, M. Matios, M. Mednikova, H. Pahutyak, S. Pyrkalo, S. Povalyaeva, T. Prokhasko, I. Rozdobudko, N. Snyadanko, M. Sokolyan, G. Tarasyuk, A. Chekh, V. Shklyar.

A comprehensive research **methodology** was used in this study. Quantitative methods of analysis are supplemented by qualitative ones. A colour profile in corpus and idiolect is built including such parameters as frequency ('colour formula', i.e., the order of decreasing / increasing frequency of colours) of key colour terms. In addition, the colours which are used and which are not used by a language personality are determined, i.e., the colour terms diversity. Using the method of structural grouping, the units are divided into groups that characterize their structure on the basis of frequency, the median and the mode are determined, which allows us to determine the specificity of the use of colour terms. The distribution of units is analysed not only by means of quantitative analysis, but also types of collocates (concrete / abstract) are determined. Thematic grouping of collocates is performed to obtain data on the topics of distribution groups that prevail in certain texts. Furthermore, stabilized and individual author collocability is analysed and employing the method of modelling the main directions of metaphorization in the idiolect are identified.

### 4. Experiment

In the course of research, the relative frequency of a range of colour terms in the GRAC was calculated. The list of colour terms selected for analysis includes colours in the colour spectrum. It was supplemented by semantic associates of the word *червоний* 'red'. The total list of the analysed units includes 26 colour terms (*багровий* 'crimson', *бежевий* 'beige', *білий* 'white', *бірюзовий* 'turquoise', *блакитний* 'blue', *бордовий* 'burgundy', *брунатний* 'brown', *бузковий* 'purple', *вохристий* 'ocher', *голубий* 'blue', *димчастий* 'smoky', *жовтавий* 'yellowish', *жовтий* 'yellow', *жовтуватий* 'yellowish', *зелений* 'green', *зеленуватий* 'greenish', *землистий* 'earthy', *золотавий* 'golden', *ліловий* 'purple', *малиновий* 'crimson', *однотонний* 'self-coloured', *оранжевий* 'orange', *синій* 'blue', *сірий* 'gray', *фіолетовий* 'violet', *червоний* 'red', *чорний* 'black'). This approach is used to carry out a pilot study due to the inability to cover an exhaustive list of colours and their shades. The semantic associates of colour terms were obtained with the help of vector analysis [19].

The analysis of the GRAC data shows that in terms of frequency colour terms are arranged in the order given in Figure 3.

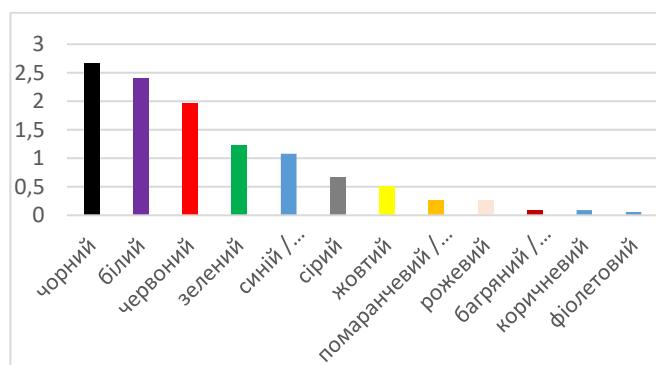


Figure 3: Frequency of colour terms in GRAC-13

Regarding the frequency of colour terms in the women's fiction prose of the 21<sup>st</sup> century, *сірий* 'gray' takes fourth place (in GRAC-13 – sixth), *зелений* 'green' is in fifth place (in GRAC-13 – fourth), *синій / блакитний / голубий* 'blue' take sixth place (in GRAC-13 – fifth), *рожевий* 'pink' overtakes *помаранчевий / оранжевий* 'orange' and comes in eighth (in GRAC-13 – ninth). *Фіолетовий* 'violet' is more frequent in the women's texts compared to the corpus data (in the female subcorpus it is tenth, in GRAC-13 – twelfth).

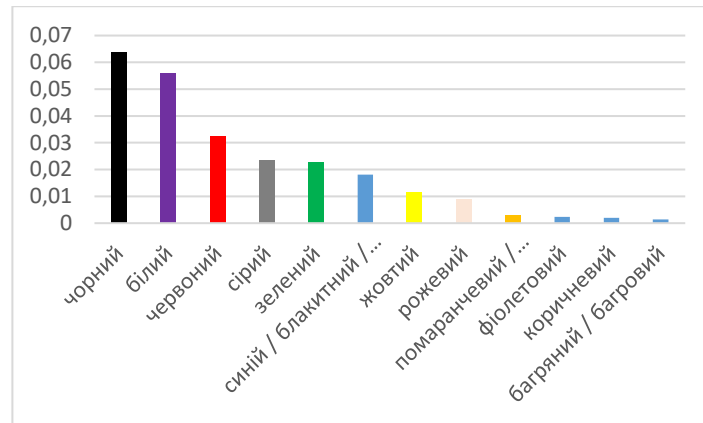


Figure 4: Frequency of colour terms in the women's texts

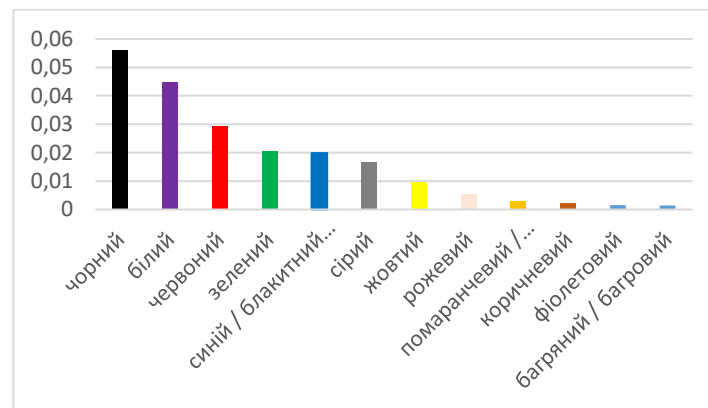


Figure 5: Frequency of colour terms in the men's texts

The difference in decreasing frequency order applies to the colours *рожевий* 'pink' (it is higher in the male subcorpus and comes in eighth; in GRAC-13 – ninth) and *помаранчевий / оранжевий* 'orange'.

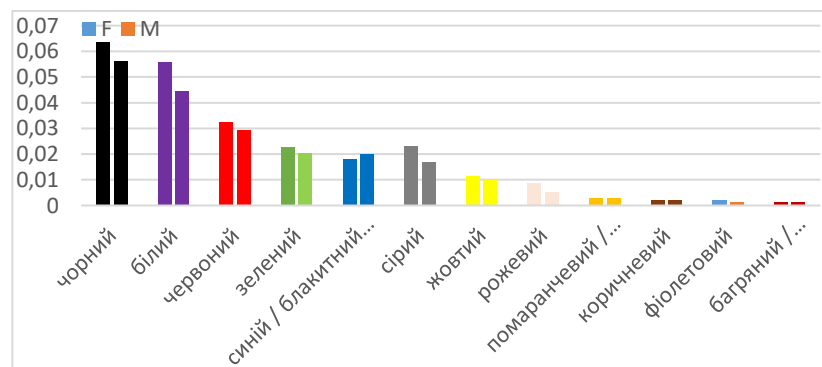


Figure 6: Comparison of colour terms frequency in the women's and the men's texts

In general, we can observe a higher frequency of the colour terms in the female subcorpus, except for the colour terms *синій / блакитний / голубий* 'blue', *коричневий* 'brown' and *багряний / багровий* 'crimson'.

The differences are revealed between the results obtained by A. Pawłowski and the results of Ukrainian fiction analysis (without taking into account the author's gender), which can be explained by the fact that calculations were carried out on the basis of the other texts, as well as by the fact that texts belong to another literary style. The order of decreasing frequency is the same for *чорний* 'black', *білий* 'white', *червоний* 'red', *жовтий* 'yellow', *рожевий* 'pink' (1, 2, 3 and 7, 8, respectively). However, in our corpus *зелений* 'green' takes fourth place (in A. Pawłowski's study *синій* 'blue' is fourth), *сірий* 'gray' comes in fifth (in A. Pawłowski's study it is *зелений* 'green'), *синій / блакитний / голубий* 'blue' are in sixth place (in A. Pawłowski's study it is *сірий* 'gray'). The end of the list also differs significantly. Similar differences are also observed when comparing data from GRAC-13 and A. Pawłowski's findings (see Table 1).

**Table 1**  
Colour terms in GRAC-13, WC, MC and A. Pawłowski's study

Colour terms	GRAC-13 (relative frequency)	WC	MC	A. Pawłowski			
чорний 'black'	2,667373253	1	0,063700521	1	1		
білий 'white'	2,404043977	2	0,055757130	2	0,044717257	2	
червоний 'red'	1,965700144	3	0,032413834	3	0,029500493	3	
зелений 'green'	1,237710902	4	0,022656336	5	0,020512371	4	синій 'blue' – 4
синій / блакитний / голубий 'blue'	1,082217051	5	0,018094389	6	0,019856725	5	зелений 'green' – 5
сірий 'gray'	0,674362208	6	0,023309949	4	0,016852278	6	6
жовтий 'yellow'	0,508055101	7	0,011324833	7	0,009834699	7	7
помаранчевий / оранжевий 'orange'	0,270266814	8	0,002894569	9	0,002856746	9	рожевий 'pink' – 8
рожевий 'pink'	0,267217915	9	0,008750401	8	0,0052812	8	коричневий 'brown' – 9
багряний / багровий 'crimson'	0,097241793	10	0,001347242	12	0,001361727	12	фіолетовий 'violet' – 10
коричневий 'brown'	0,091143995	11	0,002027532	11	0,002233521	10	багряний 'purple' – 11
фіолетовий 'violet'	0,059376017	12	0,002334330	10	0,001505826	11	оранжевий 'orange' – 12

## 5. Results

Comparing the dynamics of frequency in the Ukrainian fiction prose (subcorpora of 1980-1999 and 2000-2020), it has been revealed that the order of colour terms frequency is absolutely identical in the female and the male subcorpora of 1980-1999, although there are slight differences in frequency values. Furthermore, frequency of individual colour terms varies in different time periods, although frequency values, for example, of *зелений* 'green', do not change significantly. The change in the frequency of *синій / блакитний / голубий* 'blue' and *оранжевий / помаранчевий* 'orange' can be explained by extralinguistic factors, including political ones.

**Table 2**

Colour terms in WC (1980-1999), MC (1980-1999), WC (2000-2020) and MC (2000-2020)

Colour terms	WC 1980-1999	MC 1980-1999	WC 2000-2020	MC 2000-2020
Чорний 'black'	0,05656	1	0,06823	1
Білий 'white'	0,0559	2	0,06084	2
Червоний 'red'	0,03394	3	0,03651	3
синій/ блакитний/ голубий 'blue'	0,03186	4	0,0312	4
Зелений 'green'	0,02694	5	0,0282	5
Сірий 'gray'	0,01678	6	0,01888	6
Жовтий 'yellow'	0,01208	7	0,01353	7
рожевий 'pink'	0,009399313	8	0,006574623	8
помаранчевий / оранжевий 'orange'	0,00334	10	0,00185	10
Фіолетовий 'violet'	0,00186	11	0,00179	11
багряний / багровий 'crimson'	0,00005053	12	0,00177497	12

Based on the frequency of colour terms in the works of the studied writers, 'colour formulas' typical of the writers are built. In addition, the data for the women's (WC) and men's (MC) prose corpus of the period 2000-2020 are obtained (see Table 3).

**Table 3**

Colour formulas

Colour formula	Colour terms in subcorpora
<b>ЧорБіЧерСіЗЖоБлаСи</b>  black, white, red, gray, green, yellow, sky blue, blue	<b>WC</b> – чорний 'black', білий 'white', червоний 'red', сірий 'gray', зелений 'green', жовтий 'yellow', блакитний 'sky blue', синій 'blue', голубий 'blue', фіолетовий 'violet', помаранчевий 'orange', золотавий 'gold', бузковий 'purple', брунатний 'brown', малиновий 'crimson', блакить 'azure', жовтуватий 'yellowish', оранжевий 'orange', бордовий 'burgundy', бежевий 'beige', ліловий 'purple', зеленуватий 'greenish', жовтавий 'yellow', бірюзовий 'turquoise', димчастий 'smoky', землистий 'earthy', металік 'metallic', однотонний 'self-coloured', багровий 'murrey', павичевий 'peacock', вохристіий 'ocher', охристіий 'ocher'; <b>not found</b> – опаловий 'opal'.
<b>ЧорБіЧерЗСіСиЖоБла</b>  black, white, red, green, gray, blue, yellow, sky blue	<b>MC</b> – чорний 'black', білий 'white', червоний 'red', зелений 'green', сірий 'gray', синій 'blue', жовтий 'yellow', блакитний 'sky blue', голубий 'blue', помаранчевий 'orange', фіолетовий 'violet', малиновий 'crimson', золотавий 'gold', брунатний 'brown', жовтуватий 'yellowish', оранжевий 'orange', бузковий 'purple', блакить 'azure', жовтавий 'yellowish', зеленуватий 'greenish', багровий 'murrey', ліловий 'purple', бордовий 'burgundy', бежевий 'beige', землистий 'earthy', однотонний 'self-coloured', бірюзовий 'turquoise', димчастий 'smoky', павичевий 'peacock', металік 'metallic', вохристіий 'ocher', опаловий 'opal', охристіий 'ocher'.

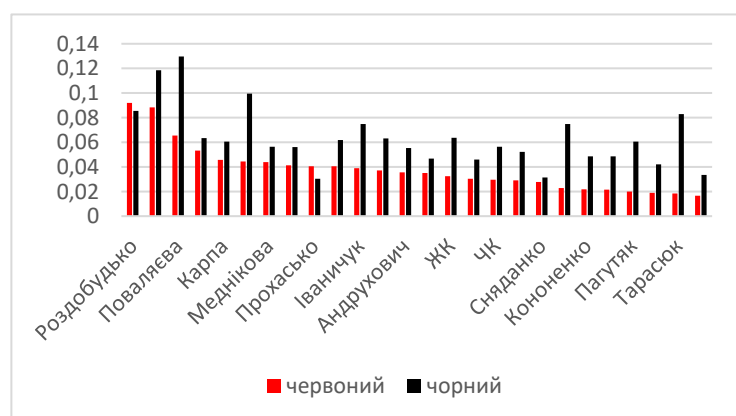
## 6. Discussion

The research results show that order of decreasing frequency is the same in WC and MC for *чорний* 'black', *білий* 'white' and *червоний* 'red'. We can see some differences concerning such colours as *сірий* 'gray', *зелений* 'green', *жовтий* 'yellow', *блакитний* 'blue', *синій* 'blue', although this list of colour terms is the most similar.

However, it should be noted that the frequencies of the colour terms used by the writers differ. In addition, the models include the names of colours and shades which are not used in these texts. For example, T. Prohasko's model is **БіЧерЗЧорСиЖоСіБла**, including *білий* 'white'; *червоний* 'red'; *зелений* 'green'; *чорний* 'black'; *синій* 'blue'; *жовтий* 'yellow'; *сірий* 'gray'; *блакитний* 'blue'; *помаранчевий* 'orange'; *фіолетовий* 'violet'; *малиновий* 'crimson'; *бордовий* 'burgundy'; *однотонний* 'self-coloured'. The writer does not use *голубий* 'blue'; *бузковий* 'lilac'; *бірюзовий* 'turquoise'; *жовтуватий* 'yellowish'; *жовтавий* 'yellowish'; *брунатний* 'brown'; *золотавий* 'golden'; *зеленуватий* 'greenish'; *землистий* 'earthy'; *оранжевий* 'orange'; *ліловий* 'purple'; *блакить* 'azure'; *димчастий* 'smoky'; *багровий* 'purple'; *бежевий* 'beige'; *вохристий* 'ocher'. I. Rozdobudko's colour terms model is **БіЧерЧорСиЗСіЖоБла**, including *білий* 'white', *червоний* 'red', *чорний* 'black', *синій* 'blue', *зелений* 'green', *сірий* 'gray', *жовтий* 'yellow', *блакитний* 'blue', *бузковий* 'purple', *золотавий* 'golden'; the writer does not use *багровий* 'crimson', *землистий* 'earthy', *оранжевий* 'orange', *бежевий* 'beige', *вохристий* 'ocher', *однотонний* 'self-coloured'.

The findings show that the colour term *багровий* 'crimson' is used only by A. Kokotyukha, V. Kozhelyanko, M. Matios, Y. Andrukhovych, I. Karpa; *димчастий* 'smoky' appears only in the texts of L. Deresh, M. Sokolyan, L. Denysenko, I. Rozdobudko; *вохристий* 'ocher' is used by H. Tarasyuk and L. Deresh. The quantitative data on the colour range used by the studied authors prove that such data can be used as idiolect markers. Here are some findings concerning the use of colour terms which are on the given list and the use of other colour terms not included into the list presented respectively for the writers: A. Kokotyukha – 24/5, L. Denysenko – 24/5, I. Karpa – 24/5, I. Rozdobudko – 23/6, R. Ivanychuk – 22/7, H. Pahutyak – 22/7, L. Deresh 21/8, S. Pyrkalo – 21/8, Y. Andrukhovych – 20/9, V. Shklyar – 20/9, H. Tarasyuk – 20/9, Yu. Vynnychuk – 19/10, V. Kozhelyanko 19/10, A. Chekh – 19/10, S. Povalyaeva – 19/10, M. Hrymych – 17/12, M. Mednikova – 17/12, N. Snyadanko – 17/12, M. Sokolyan – 17/12, O. Sabuzhko – 15/14, M. Matios – 15/14, Kapranov Brothers – 14/15, Ye. Kononenko – 14/15, T. Prokhasko – 13/16. It should be noted that these findings do not reflect the whole picture, because the whole corpus has not been semantically annotated yet. Therefore, at this stage of the study we can only talk about preliminary results obtained on a limited number of colour terms.

The diagram (see Figure 7) shows a comparison of the frequencies of the colour terms *червоний* 'red' and *чорний* 'black' in the studied texts.



**Figure 7:** The relative frequency of *червоний* 'red' and *чорний* 'black'

The data are arranged in descending order of the relative frequency of the colour term *червоний* 'red'. The analysis shows that the relative frequency in these texts differs significantly. It is noteworthy

that only in I. Rozdobudko's and T. Prokhasko's works the frequency of *червоний* 'red' is higher than the frequency of *чорний* 'black'. This frequency ratio can be viewed as the idiolect marker. On the other hand, the average data taken from the corpora of women's and men's texts cannot be used as a gender marker, as the variability of the indicators of an individual author is quite significant.

In addition, structural grouping is used for the homogeneous division into groups that characterize the structure on a certain basis. The number of groups is roughly calculated by the Sturges formula (1).

$$n = 1 + 3,322 \log n = 1 + 3,322 \log (26) = 6 \quad (1)$$

The width of the interval is:

$$h = \frac{x_{max} - x_{min}}{dx} = \frac{0.1297 - 0.0304}{6} = 0.0166, \quad (2)$$

where  $x_{max}$  is the maximum value of the grouping feature,  $x_{min}$  is the minimum value of the grouping feature.

Then, the boundaries of the groups are defined (see Table 4).

For each value of the series, it is calculated how many times it is within a particular interval.

The mode is the value of the feature in the data set which is the most common (3)

$$M_0 = x_0 + h \frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)}, \quad (3)$$

where  $x_0$  is the beginning of the mode interval;  $h$  is the value of the interval;  $f_2$  is the frequency corresponding to the modal interval;  $f_1$  is the pre-mode frequency;  $f_3$  is the post-mode frequency.

**Table 4**

Boundaries of colour groups

Group number	Lower limit	Upper limit
1	0.0304	0.047
2	0.047	0.0636
3	0.0636	0.0802
4	0.0802	0.0968
5	0.0968	0.1134
6	0.1134	0.13

The beginning of the interval is 0.047, because this is the interval common for most cases.

$$M_0 = 0.047 + 0.0166 \frac{12-6}{(12-6)+(12-3)} = 0.0536 \quad (4)$$

The most frequent value of the series is 0.0536.

As we can see (Table 5), the typical frequency of the colour term *чорний* 'black' is the relative frequency in the range 0.047–0.0636 (group II); the mode is 0.0536. We do not observe gender formal markers in the use of this colour term (like in case of most of the other lexemes analysed). The high frequency of this colour term is characteristic of the idiolect of V. Shklyar and S. Povalyaeva. We also revealed a higher frequency of shades of black in women's texts, in particular, this is typical of the texts of S. Povalyaeva, I. Karpa (Table 6).



**Table 5**The colour term *чорний* 'black' in the corpus

Чорний 'black' / groups	Writers	Frequency $f_i$
0.0304 – 0.047	T. Prokhasko, N. Snyadanko, A. Kokotyukha, A. Chekh, S. Denysenko, L. Deresh	6
0.047 – 0.0636	E. Kononenko, O. Zabuzhko, Br. Kapranov, Yu. Andrukhovych, Yu. Vynnychuk, MC, M. Mednikova, I. Karpa, H. Pahutyak, M. Hrymych, M. Sokolyan, S. Pyrkalo	12
0.0636 – 0.0802	WC, R. Ivanychuk, M. Matios	3
0.0802 – 0.0968	H. Tarasyuk, I. Rozdobudko	2
0.0968 – 0.113	V. Kozhelyanko	1
0.113 – 0.13	V. Shklyar, S. Povalyaeva	2

**Table 6**The colour term *чорний*- 'black-' in the corpus

Чорно- 'black-' / groups	Writers	Frequency $f_i$
0 – 0.000485	Yu. Andrukhovych, Br. Kapranov, Ye. Kononenko, A. Chekh, A. Kokotyukha, V. Shklyar, M. Matios, H. Pahutyak, N. Snyadanko	10
0.000485 – 0.00097	M. Hrymych, Yu. Vynnychuk, MC, R. Ivanychuk, M. Sokolyan, H. Tarasyuk, M. Mednikova, WC, O. Zabuzhko, I. Rosdobudko	10
0.00097 – 0.00146	L. Deresh, V. Kozhelyanko	2
0.00146 – 0.00194	T. Prokhasko, L. Denysenko, S. Pyrkalo	3
0.00194 – 0.00243		
0.00243 – 0.00291	S. Povalyaeva, I. Karpa	2

The findings show that one of the qualitative markers of the idiolect is the specificity of metaphorization, in particular the components of certain thematic groups in structural and semantic models of metaphors. We can talk about the colour attribution to abstract and specific concepts (determining the relationships and thematic groups). The semantically annotated corpus enables automated metaphor search by search request ([lemma="чорний"] [tag=".\*noun.\*" & semtag="1:abst.\*"]).

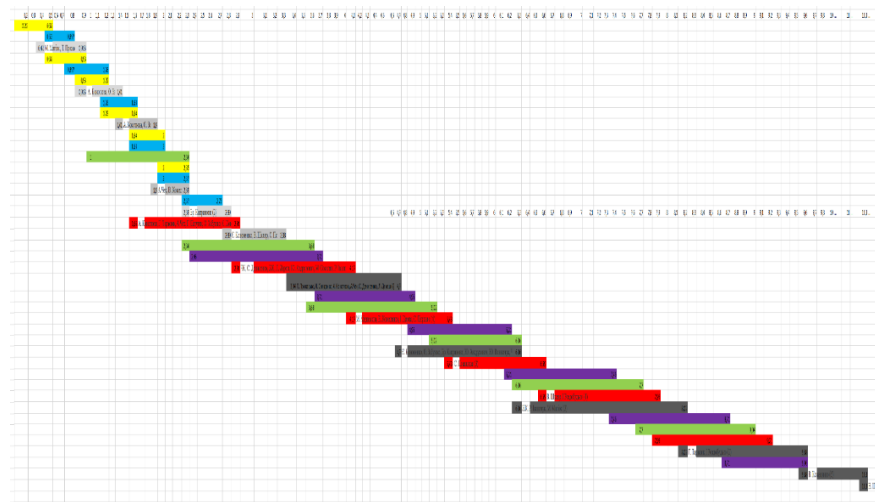
For instance, V. Shklyar most often uses the colour attribute *чорний*- 'black-' with such specific concepts as **clothes** – хустина 'kerchief', жупан 'zhupan (dressing gown)', бекеша 'bekesha (coat)', берет 'beret', бурка 'burka', черкеска 'cherkeska (male coat)', бушлат 'bushlat', сукня 'dress', шапчина 'hat', панчохи 'stockings', плац 'coat', костюм 'suit', краватка 'tie', фрак 'tailcoat', хламида 'chlamys', шкірянка 'leather jacket' etc.; **parts of body** – борода 'beard', борідка 'beard', обличчя 'face', рука 'hand', морда 'snout', кучері 'curls', око 'eye' etc.; **vehicles** – авто 'car', автомобіль 'car', джип 'jeep', "Пежо" 'Peugeot', "Форд" 'Ford', машина 'car' etc.; **landscape elements** – небо 'sky', ріка 'river', безодня 'abyss', кам'яні брили 'stone blocks', провалля 'chasm', твань 'tid' etc.

Furthermore, we revealed a significant ratio of stabilized (recorded in dictionaries) and individual author word combinations in the analysed texts. V. Shklyar uses a significant number of stabilized word combinations with the component *чорний* 'black', for example: *Тоді я ще не знав, що настане та чорна година 'black hour', коли я залишуся в лісі тільки з оцим китайцем, і ми з'їмо з ним периу сиру ворону без солі* (V. Shklyar, *Чорний ворон 'Black Crow'*); *Ще добре, що мала обачність, децю припасла, заощадила в ті часи, коли Нестор не рахував грошей, смітив ними на всі боки, не думаючи про чорний день 'black day'* (V. Shklyar, *Кров кажана 'The blood of a bat'*); *Негоже козакові скаржитися, але наприкінці листопада для нас настали чорні часи... 'black time'* (V. Shklyar, *Чорний ворон 'Black Crow'*); *Масток лежить на березі Азовського моря в селищі Урзуф, яке раніше належало до курортної зони, а тепер опинилося в зоні АТО, як називають четверту*

російсько-українську війну брехливі політики та шанувальники **чорного гумору** *'black humor'* (V. Shklyar, *Чорне сонце* *'Black Sun'*); Загнаний у глухий кут, він нікого не чув, **чорні думки** *'black thoughts'* одна по одній холодили мозок (V. Shklyar, *Маруся* *'Mariusya'*); Але це викликало в Петлюри **чорну заздрість** *'black envy'*, ревності ... (V. Shklyar, *Чорний ворон* *'Black Crow'*); Він силкувався усе те згадати, та заважали інші думки, що знагла накочувалися на нього — про **чорну зраду** *'black betrayal'* Тимоша Корча, про Матейя Мазура, хоч Василь ніколи не довідається, що саме Матей застрелив Дмитра крізь вікно... (V. Shklyar, *Маруся* *'Mariusya'*); Я, мабуть, забіг би дуже далеко, бо не чув ні втоми, ні болю у стертих до крові ногах, нічого не відчував, окрім **чорного розпачу** *'black despair'*, що виповнив усеньке моє єство (V. Shklyar, *Троця* *'Destruction'*); — Що, коли ми будемо чекати прихильності від Денікіна, а білі москалі робитимуть свою **чорну справу** *'black business'*? (V. Shklyar, *Маруся* *'Mariusya'*). The following examples show individual-authorial metaphorical expressions: Думаю, що, якби не **чорна безвихідь** *'black impasse'*, мало хто з повстанців заломився б, пішов на амністію, зрадив ліс (V. Shklyar, *Чорний ворон* *'Black Crow'*); **Голос** *'voice'* у нього теж був **чорний** *'black'* (В.Шкляр, *Чорний ворон* *'Black Crow'*); Я була впевнена, що докола мене снується якась **чорна змова** *'black conspiracy'* (V. Shklyar, *Кров кажана* *'The blood of a bat'*).

Another writer Yu. Vynnychuk most often attributes black colour to specific concepts (the most frequent are verbalizers of the conceptual domain **clothes** – білизна *'underwear'*, вбрання *'clothes'*, гарнітур *'suit'*, камізька *'vest'*, капелюх *'hat'*, костюм *'suit'*, мейти *'shoes'*, панчохи *'stockings'* etc. These stabilized expressions are used in the following examples: Руки тонюські, ноги такі, що здається, ось-ось підломляться, дупця як кулачок, **біда чорна** *'black misery'* (Yu. Vynnychuk, *Мальва Ланда* *'Malva Landa'*); Він захоплює карколомним сюжетом, присмаченим еротикою, **чорним гумором** *'black humor'*, нестримним потоком фантазії і, звичайно ж, розкішною мовою (Yu. Vynnychuk, *Мальва Ланда* *'Malva Landa'*); Настав **чорний день** *'black day'* для Львова (Yu. Vynnychuk, *Танго смерті* *'Tango of Death'*); Більшість із них сиділа роками, то були терті калачі з **чорним піднебінням** *'with black palate'* (Yu. Vynnychuk, *Цензор снів* *'Sensor of Dreams'*) (contamination of phraseological units). The following examples illustrate individual author metaphorical expressions: Але боялася Рута усілякої чортівні, яка у лісі водилася, а було її чимало, і чигала вона на кожному кроці, особливо під кінець дня у темних затінених місцинах, звідки уже починав витікати **чорний мед ночі** *'black honey of night'*, скрадаючись, наче звір... (Yu. Vynnychuk, *Аптекарь* *'Pharmacist'*); Най мене нагла **чорна кава залле!** *'Let the impudent black coffee flood me!'* (Yu. Vynnychuk, *Цензор снів* *'Sensor of Dreams'*) (intertext transformation). The range of units that are combined with the colour term **чорний** *'black'* differ significantly in the prose fiction by V. Shklyar and Y. Vynnychuk.

The diagram in Figure 8 shows the frequency of the colour terms **чорний** *'black'*, **білий** *'white'*, **червоний** *'red'*, **зелений** *'green'*, **сірий** *'gray'*, **синій** *'blue'*, **жовтий** *'yellow'*.



**Figure 8:** The range of the relative frequency of colour terms **чорний** *'black'*, **білий** *'white'*, **червоний** *'red'*, **зелений** *'green'*, **сірий** *'gray'*, **синій** *'blue'*, **жовтий** *'yellow'* (0,0022 - 0.13)

The relative frequency of this thematic group is within 0,0022 - 0.13.

## 7. Conclusions

Thus, the Ukrainian prose fiction is characterized by the frequency of the colour terms *чорний* 'black', *білий* 'white', *червоний* 'red', *зелений* 'green', *синій/блакитний/голубий* 'blue', *сірий* 'gray', *жовтий* 'yellow', *помаранчевий* 'orange', *рожевий* 'pink' etc. We revealed some differences between the results obtained by A. Pawłowski and the results obtained by analysing Ukrainian prose fiction using the GRAC (without taking into account the gender of the author). The descending order of frequency is the same for *чорний* 'black', *білий* 'white', *червоний* 'red' та *жовтий* 'yellow', *рожевий* 'pink' (1, 2, 3 and 7, 8, respectively). In our research corpus, *зелений* 'green' takes fourth place (in A. Pawłowski's study – *синій* 'blue'); *сірий* 'gray' comes in fifth (in A. Pawłowski's study – *зелений* 'green'), *синій/блакитний/голубий* 'blue' is in sixth place (in A. Pawłowski's study – *сірий* 'gray'). The end of the list also differs significantly. It should be noted that corpus findings showed a higher frequency of the colour terms in the women's subcorpus, except for the colour terms *синій/блакитний/голубий* 'blue', *коричневий* 'brown' та *багряний/багровий* 'crimson'.

It has been determined that in the studied subcorpus the typical frequency of the colour terms is within the ranges as follows: 0.047 - 0.0636 – *чорний* 'black'; 0.0621 - 0.0746 – *білий* 'white'; 0.01 - 0.0234 – *зелений* 'green'; 0.0053 - 0.0126 – *синій* 'blue'; 0.0141 - 0.019 – *сірий* 'gray'; 0.0093 - 0.0129 – *жовтий* 'yellow'.

Building the statistical profile of Ukrainian prose fiction is important in terms of its comparison with the characteristics of the statistical profile of the idiolect. We have proposed a model of the statistical profile of the writer's idiolect based on colour terms. The first characteristic is the colour 'formula', i.e., the list of colour terms used by writers in descending order of frequency, for example ЧорЧерБіЗСіСиЖоБла – *чорний* 'black'; *червоний* 'red'; *білий* 'white'; *зелений* 'green'; *сірий* 'gray'; *синій* 'blue'; *жовтий* 'yellow'; *блакитний* 'blue'. The formula does not include – *помаранчевий* 'orange'; *бірюзовий* 'turquoise'; *зеленуватий* 'greenish'; *оранжевий* 'orange'; *блакить* 'blue'; *димчастий* 'smoky'; *багровий* 'crimson'; *вохристий* 'ocher'; *однотонний* 'self-coloured'). Thus, V. Shklyar uses 20 out of 29 studied colour terms.

Another feature is the structural grouping of data, which characterizes the set based on the frequency of colour terms. Thus, the following frequency distribution is characteristic of V. Shklyar's works: *чорний* 'black' (0,113 – 0,13), *червоний* 'red' (0,0795 – 0,0921), *білий* 'white' (0,0621 – 0,0746), *сірий* 'gray' (0,0289 – 0,0338), *зелений* 'green' (0,0234 – 0,0368), *жовтий* 'yellow' (0,0129 – 0,0164), *синій* 'blue' (0,0126 – 0,0163).

In addition to formal quantitative characteristics, information on the collocability of the studied units is important. For example, V. Shklyar often uses colour attributes with specific concepts belonging to such thematic groups as: clothes, parts of the human body, vehicles, landscape elements. We have considered both the list of stabilized metaphors and the list of individual author's metaphors.

The semantic annotation of the corpus is important for this kind of corpus-based studies. After the completion of the semantic annotation of the GRAC, it is planned to conduct the research based on more extensive list of colour terms.

## 8. References

- [1] H. Kauppinen-Räsänen, M. N. Jauffret, Using colour semiotics to explore colour meanings, *Qualitative Market Research: An International Journal*, volume 21, 1 (2018) 101–111. doi:10.1108/QMR-03-2016-0033.
- [2] C. Mohr, D. Jonauskaitė, E. S. Dan-Glauser, M. Uusküla, N. Dael, Unifying research on colour and emotion: Time for a crosscultural survey on emotion associations with colour terms, in: L. W. MacDonald, C. P. Biggam, G. V Paramei (Eds.), *Progress in Colour Studies: Cognition, language, and beyond*, John Benjamin's Publishing Company, Amsterdam, 2018, pp. 209–222. doi: 10.1075/z.217.11moh.
- [3] D. T. Lindsey, A. M. Brown, Lexical Color Categories, *Annual Review of Vision Science*, volume 7, 1 (2021) 605-631. doi:10.1146/annurev-vision-093019-112420.

- [4] N. Zaslavsky, C. Kemp, T. Regier, N. Tishby, Efficient compression in color naming and its evolution, in: Proc. Natl. Acad. Sci. U.S.A., volume 115 (31) 2018, pp. 7937–7942. doi.org/10.1073/pnas.1800521115.
- [5] E. Gibson, R. Futrell, J. Jara-Ettinger, K. Mahowald, L. Bergen, S. Ratnasingam, M. Gibson, S. T. Piantadosi, B. R. Conway, Color naming across languages reflects color use, in: Proc. Natl. Acad. Sci. U.S.A., volume 114 (40), 2017, pp. 10785–10790.
- [6] O. Levchenko, N. Lototska, Kolyoronazva bilyy yak marker idiolektu (na materialy tekstiv R. Ivanychuka), *Linhvistychni studii* 41 (2021) 131-142. [O. Levchenko, N. Lototska, Colour nomination білий (white) as an idiolect marker (based on R. Ivanychuk's texts), *Linguistic studies* 41 (2021) 131-142].
- [7] N. Lototska, Statistical Research of the Colour Component ЧОРНИЙ (BLACK) in Roman Ivanychuk's Text Corpus, in: Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2021), Lviv, Ukraine, volume I, 2021, pp. 486–497.
- [8] K. O. Filatova, A Cognitive Linguistic Analysis of Ukrainian Compounds with Color-Terms, *Transcarpathian Philological Studies*, volume 12, Helvetica, 2019, pp. 180-186.
- [9] V. Starko, Korpusni dani v doslidzhenni ukrainskykh koloratyviv, *Ukrainska mova* 1 (2014) 51-60. [V. Starko, Corpus data in the Ukrainian colour terms research, *The Ukrainian language* 1 (2014) 51-60].
- [10] I. Inshakova, A. Inshakov, Funktsionuvannya kolyoronomeniv u frazeolohizmakh, in: *Naukovi pratsi Kamyanets-Podilskoho natsionalnoho universytetu imeni Ivana Ohienka, Filolohichni nauky*, 48, 2019, pp. 140-144. [A. Inshakov, Functioning of colour terms in phraseological units, in: *Research papers of Ivan Ohienko National University of Kamyanets-Podilskyy, Philological sciences*, volume 48, 2019, pp. 140-144].
- [11] N. Vasylyeva, Henderna verbalizatsiya pryrody: kolir, zvuk, zapakh (na materialy ukrainskoi, rosiyskoi, anhliyskoi, frantsuzkoi khudozhnyoi literatury kintsya XIX – pershoi polovyny XX stolittya), *Dysertatsiya kandydata filolohichnykh nauk, Pivdenoukrainskyi natsionalnyi pedahohichniy universytet imeni K. Ushynskoho*, Odesa, 2019. [N. Vasylyeva, Gender Interpretation of Nature: Colour, Sound, and Smell (based on Ukrainian, Russian, English, and French Belles-Lettres Prose of the Late XIX – the First Half of XX Century), PhD thesis in philology, K. Ushynskyy Southern Ukrainian Pedagogical University, Odesa, 2019].
- [12] M. Shvedova, R. von Waldenfels, S. Yarygin, A. Rysin, V. Starko, M. Wozniak, M. Kruk, General regionally annotated corpus of the Ukrainian language (GRAC), Kyiv, Lviv, Yena, 2017-2021.
- [13] B. Berlin, P. Kay, Basic color terms: their universality and evolution, The University of California Press, Berkeley and Los Angeles, 1969.
- [14] A. Pawłowski, Quantitative linguistics in the study of colour terminology, in: C. P. Biggam, C. Kay (Eds.), *Progress in color studies: language and culture*, volume 1, 2006, pp. 37-55. doi: 10.1075/z.pics1.07paw
- [15] N. Fider, L. Narens, K. Jameson, N. Komarova, Quantitative approach for defining basic color terms and color category best exemplars, *J. Opt. Soc. Am. A* 34 (2017) 1285-1300.
- [16] N. Fider, N. Komarova, Quantitative study of color category boundaries, *J. Opt. Soc. Am. A* 35 (2018) B165-B183.
- [17] N. Fider, N. Komarova, Differences in color categorization manifested by males and females: a quantitative World Color Survey study, *Palgrave Commun* 5, 142 (2019). <https://doi.org/10.1057/s41599-019-0341-7>.
- [18] E. Leach, *Culture and Communication: The Logic by which Symbols Are Connected. An Introduction to the Use of Structuralist Analysis in Social Anthropology (Themes in the Social Sciences)*, Cambridge, Cambridge University Press, 1976. doi:10.1017/CBO9780511607684.
- [19] Automated workplace of a researcher, Complex information system of scientific research, 2020. URL: <http://icybcluster.org.ua:34145/word2vec/similar/>