

A method of statistical idiolect/genderlect profiling for automated authorship attribution

Olena Levchenko[†], Nataliia Lototska^{*,†} and Olesya Nakhlik[†]

Lviv Polytechnic National University, Bandera Str., 12, Lviv, 79000, Ukraine

Abstract

The article presents the results of the statistical profiling of the idiolect based on the analysis of taste names in Ukrainian fiction 1991-2023. The study aims to identify the list of idiolect markers relying on corpus data. It was found that one of the markers of an idiolect is the peculiarities of metaphorization. Metaphorical collocations with the verbalisers of the CMAK (TASTE) [SMAK] concept were automatically identified in the General Regionally Annotated Corpus of Ukrainian, which, in turn, made it possible to determine stabilized/individual author's collocations and taste prototypes in the analysed texts. The created statistical profile of the idiolect/genderlect comprises data on the frequency of the main names of tastes and their verbalizers, structural grouping, calculation of the median and mode, determination of the type of collocates (concrete/abstract) with taste attributes, thematic grouping of the obtained collocates.

Keywords

Statistical genderlect/idiolect profiling, automatic metaphor identification, authorship attribution, idiolect marker, prototype

1. Introduction

The verbalisers of a concept, in particular, their formal indicator such as frequency, can serve as markers of an idiolect/genderlect, which in turn creates the basis for establishing authorship or creating a statistical profile of an idiolect. In addition, the specificity of an idiolect/genderlect is manifested in metaphors, as they reflect the individual author's linguistic picture of the world, worldview, and experience. In view of this, it is important to develop a methodology for the automatic identification of metaphorical collocations with a concept verbaliser component.

The relevance of this article lies in the need for automated/automatic authorship attribution, in particular, determination of the author's gender. To perform this task, it is necessary to identify verbal data that can serve as markers of an idiolect.

The novelty of this study stems from the fact that for the first time the objects of statistical analysis are the names of taste in fiction texts by male and female authors as markers of an idiolect.

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

* Corresponding author.

[†] These authors contributed equally.

✉ olena.p.levchenko@lpnu.ua (O. Levchenko), nataliia.y.lototska@lpnu.ua (N. Lototska), olesia.s.nakhlik@lpnu.ua (O. Nakhlik)

ORCID 0000-0002-7395-3772 (O. Levchenko); 0000-0001-6692-196X (N. Lototska); 0000-0003-0086-4469 (O. Nakhlik)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The purpose of this paper is to develop a methodology for automated corpus data analysis in order to identify metaphorical collocations with verbalisers of the CMAK (TASTE) [SMAK] concept. The verbalisers of the CMAK (TASTE) [SMAK] concept belong to the perceptual vocabulary, which plays an important role in reflecting the objective reality. The issues of studying conceptualisation and verbalisation are of particular interest, since it is the conceptual level data that most clearly reflect the linguistic community's perception of the world and contain information about the national specificity of the language. Within the framework of Ukrainian linguistics, the study of CMAK (TASTE) [SMAK] concept verbalisers has been carried out mainly within the framework of linguistics, without the use of modern information technologies, in particular, the developments of corpus and statistical linguistics. In previous studies, an attempt was made to analyse verbalisation in terms of the KOJIP (COLOUR) [KOLIR] concept [23], since the authorship attribution should be carried out taking into account a comprehensive analysis of statistical parameters, in particular, quantitative and qualitative analysis of metaphorical features.

The objectives of the study are as follows:

- creating a statistical profile of the CMAK (TASTE) [SMAK] concept verbalisers;
- confirmation of the research hypothesis about the possibility of identifying the author of a text with the help of the dominant type of metaphor used in his/her previous texts;
- analysis of the effectiveness of automatic identification of taste metaphors in the studied male and female authors' texts using the corpus features.

2. Related works

Taste names are of interest to researchers for their etymology and semantics, as they are characterised by fluctuating meanings depending on cultural, linguistic and historical contexts. The concept of CMAK (TASTE) [SMAK] has been in the focus of attention recently [2, 3, 4, 6, 10, 15, 16, 19, 28, 29, 33, 36, 37, 38, 39, 40, 41].

In a number of Ukrainian studies, authors have in one way or another touched upon the analysis of vocabulary denoting taste [5, 13, 22]. However, these studies dealt with a purely linguistic consideration of the problem and did not address the issues of automatic identification of metaphorisation. Instead, linguistic and gender studies increasingly use quantitative analysis and a linguistic and cognitive approach to interpret data. In particular, the results of studies based on frequency analysis are presented in [18, 21, 23, 24, 27, 35].

The research of writers' fiction texts using quantitative methods to detect individual peculiarities is traced in the studies [23, 24, 26]. In the Ukrainian language, statistical studies of gendered texts were conducted on the material of large corpora of texts to identify statistical and qualitative characteristics of fiction texts that can serve as a marker for determining the text author's gender [23, 24, 34, 35].

To detect collocations, in particular metaphorical ones, statistical criteria, referred to as association indicators, are used to calculate the strength of statistical dependence of two or more qualitative variables [9, 30]. The association indicators most commonly used for studying collocations include MI, t-score, log-likelihood, Dice, gmean, etc. Thus, MI-score is best suited for detecting terminological collocations, proper names, and complex nominations; t-score is suitable for common language stable collocations; MI-score is better for finding similarities, t-score is applicable for establishing differences; log-likelihood is suitable for different sizes of text corpora; z-score and t-score are not recommended for low-frequency candidates and small

samples [7, 9, 14]. Association indicators are primarily intended for automatic detection of collocations/constructions in a text (corpus of texts) based on the determination of randomness/nonrandomness of a certain sequence of words in a text (corpus of texts) [8, 9, 30, 32]. However, the analysis of collocation lists obtained by the above methods does not allow obtaining lists of metaphorical compounds. Text corpora and their functionality, in particular, the availability of semantic corpus annotation, make it possible to automatically identify metaphors [11, 17, 25], which is important for identifying markers of writers' idiolects.

3. Methods and materials

"The language sciences are undergoing a quantitative revolution. There is ever more data, an ever-growing toolkit of statistical methods, and a dedicated push towards incorporating more empirical research into linguistic theorising" [38, p. xii]. In quantitative research, linguistic features are classified and counted, and even more complex statistical models are constructed to explain these observed facts. In qualitative research, however, we use the data only for identifying and describing features of language use and for providing real occurrences/examples of phenomena [12, p. xi]. Statistics is an important tool for analysing linguistic data, as quantitative methods ensure the reliability of the results and allow us to identify the properties of linguistic units and text structure. Statistical methods are widely used in linguistics and have become one of the most efficient and time-saving tools for processing different sets of texts [20]. Statistical profiles of texts by Ukrainian writers based on word combinations were compiled to identify characteristic markers [23, 26].

A corpus of texts is a reliable material for statistical analysis of lexical collocations, as it helps to increase the accuracy of the data and the objectivity of the results. The quantitative collocation analysis aids in extracting concordance lines for the in-depth qualitative analysis at the lexical level studying adjectives and noun collocates, and at the lexico-grammatical level [1].

Methodologically, collocation analysis is viewed as essentially a probabilistic phenomenon consisting of identifying statistically significant collocations and excluding chance combinations. In collocation analysis, interest usually focuses on the extent to which the actual pattern of these occurrences differs from the pattern that would be expected, assuming a random distribution of forms. Any significant difference can be taken as, at least, preliminary evidence that the presence of one word in the text affects the occurrence of another in some way [12, p. 196].

To study the CMAK (TASTE) [SMAK] concept, two research subcorpora of contemporary fiction of 1991-2023 were created in GRAC-17 [31]: male authors' research subcorpus (male subcorpus/MS) comprising 3,589,285 words and female authors' research subcorpus (female subcorpus/FS) of 2,770,296 words. The filters used to create the research subcorpora included literary style (DOC.STYLE – FIC), creation in the relevant time period (DOC.DATE – 1991-2023), language of the original text (DOC.ORIGINAL – UK) and the surname and name of the author (DOC.AUTHOR – Y. Andrukhovych, Y. Vynnychuk, V. Lys, V. Shevchuk, V. Shkliar (for male authors' texts) and S. Andrukhovych, O. Zabuzhko, M. Matios, H. Pahutiak, I. Rozdobudko (for female authors' texts).

Taste adjectives are characterised by their own semantics, not by their relation to other objects, actions or circumstances. For example, *гіркий* (*bitter*) [*hirkyi*], *солодкий* (*sweet*) [*solodkyi*], *кислий* (*sour*) [*kyslyi*], *солоний* (*salty*) [*solonyi*]. These adjectives are the main

components of the concept of СМАК (TASTE) [SMAK] by physiological sensation. Both in Ukrainian and in other languages, the system of words for taste has a clearly defined core: *солодкий* (sweet) [solodkyi], *гіркий* (bitter) [hirkyi], *солоний* (salty) [solonyi], *кислий* (sour) [kyslyi]. The taste names denoting the main types of elementary taste are joined by the lexemes *прісний* (bland) [prisnyi], *квасний* (tangy) [kvasnyi], *нудотний* (cloying) [nudotnyi], *приторний* (sugary) [prytorny], *терпкий* (tart) [terpkyi]. This group also includes lexemes which denote complex taste sensations (*гіркувато-солодкий* (bittersweet) [hirkuvato-solodkyi]), characterising the taste with a core name with an additional connotation.

The study considers the main four types of tastes according to the physiological sensation of *солодкий* (sweet) [solodkyi], *солоний* (salty) [solonyi], *гіркий* (bitter) [hirkyi], *кислий* (sour) [kyslyi], as well as the words *прісний* (bland) [prisnyi], *квасний* (tangy) [kvasnyi], *нудотний* (cloying) [nudotnyi], *приторний* (sugary) [prytorny] and their common root verbalizers. This approach was used to conduct a pilot study of the СМАК (TASTE) [SMAK] concept due to the impossibility of covering an exhaustive list of taste attributes.

This study relies on a comprehensive research methodology. Quantitative methods of analysis complement the qualitative interpretation of the results. Thus, the study uses grouping of the population into groups homogeneous by a certain feature; determines the median and mode, which will allow us to find out the specifics of the use of linguistic units to denote taste; presents typical collocations, the components of which are adjectives denoting the main tastes. A thematic grouping of collocates was carried out to obtain data on the dominant groups of typical collocates of taste attributes. The stabilised and individual author's collocations were analysed, and the quantitative characteristics of collocations for taste attributes used in the literal and figurative sense in male and female texts are presented. The effectiveness of automatic identification of taste metaphors in the studied male and female authors' texts using the corpus functionality was analysed [see [Data Set](#)].

The statistical significance of the results was assessed using the chi-square test and p-value. In the context of the chi-square test, the p-value analysis helps to determine whether there is a statistically significant relationship between variables. If the p-value is < 0.05, the difference in frequencies is statistically significant.

4. Experiment and results

The collected and analysed factual material makes it possible to illustrate information about the main taste sensations selected by male and female writers in fiction. The frequency of such verbalisers of the СМАК (TASTE) [SMAK] concept as *солодкий* (sweet) [solodkyi], *солоний* (salty) [solonyi], *гіркий* (bitter) [hirkyi], *кислий* (sour) [kyslyi], *прісний* (bland) [prisnyi], *квасний* (tangy) [kvasnyi], *нудотний* (cloying) [nudotnyi], *приторний* (sugary) [prytorny] is analysed, and their absolute (AF) and relative (RF) frequencies are presented (for convenience, all data are RF 10^{-5}) in male and female authors' texts (see Table 1).

Table 1

Frequency of 'taste' adjectives in MS and FS

Taste adjective	MS		FS		Chi-square test statistic	P-value
	AF	RF,%.	AF	RF,%.		
<i>солодкий (sweet) [solodkyi]</i>	229	638.01	298	1075.7	35.621	2.4*10 ⁻⁹
<i>гіркий (bitter) [hirkyi]</i>	145	403.98	127	458.43	0.960	0.327
<i>солоний (salty) [solonyi]</i>	59	164.44	43	155.22	0.034	0.852
<i>кислий (sour) [kyslyi]</i>	38	105.87	19	68.58	2.027	0.154
<i>прісний (bland) [prisnyi]</i>	13	36.22	19	68.58	2.643	0.103
<i>квасний (tangy) [kvasnyi]</i>	3	8.36	11	39.7	5.628	0.017
<i>нудотний (cloying) [nudotnyi]</i>	7	19.5	2	7.22	0.0	1.0
<i>приторний (sugary) [prytorny]</i>	1	2.78	2	7.22	0.050	0.822

The data obtained show that the frequency of use of the adjectives *солодкий (sweet) [solodkyi]* is 1075.7, *прісний (bland) [prisnyi]* 68.58, *квасний (tangy) [kvasnyi]* 39.7, and *приторний (sugary) [prytorny]* 7.22 in the female authors' texts, respectively, which is higher than in the male authors' texts: *солодкий (sweet) [solodkyi]* 638.01, *прісний (bland) [prisnyi]* 36.22, *квасний (tangy) [kvasnyi]* 8.36, and *приторний (sugary) [prytorny]* 2.78. The frequency of the words *солодкий (sweet) [solodkyi]*, *прісний (bland) [prisnyi]*, and *квасний (tangy) [kvasnyi]* is 1.7, 1.9, and 2.6 times higher in the female authors' texts, respectively. In the male texts, the frequency of the words *кислий (sour) [kyslyi]* is 105.87, *нудотний (cloying) [nudotnyi]* 19.5, which is higher than in the female texts – *кислий (sour) [kyslyi]* 68.58, *нудотний (cloying) [nudotnyi]* 7.22, respectively. The frequency of the words *кислий (sour) [kyslyi]*, *нудотний (cloying) [nudotnyi]* is 1.5 and 2.7 times higher in the male texts, respectively. The frequencies of the words *гіркий (bitter) [hirkyi]* and *солоний (salty) [solonyi]* correlate in the two subcorpora: *гіркий (bitter) [hirkyi]* – male 403.98, female 458.43; *солоний (salty) [solonyi]* – 164.44 and 155.22, respectively. It should be noted that for *солодкий (sweet) [solodkyi]* the Chi-square test statistic is 35.62 and p-value is 2.4*10⁻⁹ and for *квасний (tangy) [kvasnyi]*, the Chi-square test statistic is 5.63 and p-value is 0.018. The words *солодкий (sweet) [solodkyi]* and *квасний (tangy) [kvasnyi]* show statistically significant differences in the frequency of use in male and female authors' texts. An extremely low p-value for the word *солодкий (sweet) [solodkyi]* indicates a significant relationship with gender, while a lower p-value for the word *квасний (tangy) [kvasnyi]* indicates a much weaker, but still noteworthy relationship.

The adjectives *солодкий (sweet) [solodkyi]*, *гіркий (bitter) [hirkyi]*, and *солоний (salty) [solonyi]* rank first, second, and third in terms of frequency of use in the male and female authors' texts, respectively. In male texts, *кислий (sour) [kyslyi]* ranks fourth, *прісний (bland) [prisnyi]* fifth, *нудотний (cloying) [nudotnyi]* sixth, *квасний (tangy) [kvasnyi]* seventh, and *приторний (sugary) [prytorny]* eighth, respectively. In the female authors' texts, the adjectives *кислий (sour) [kyslyi]* and *прісний (bland) [prisnyi]* rank fourth, *квасний (tangy) [kvasnyi]* fifth, *нудотний (cloying) [nudotnyi]* and *приторний (sugary) [prytorny]* sixth, respectively.

Thus, the most frequent taste adjective in both male and female texts is *солодкий* (*sweet*) [*solodkyi*], while the least frequent ones are *квасний* (*tangy*) [*kvasnyi*] and *приторний* (*sugary*) [*prytorny*] in male texts, *нудотний* (*cloying*) [*nudotnyi*] and *приторний* (*sugary*) [*prytorny*] in female texts.

For the purposes of authorship attribution, it is also important to trace the frequency of use of taste adjectives in the texts of a particular writer (see Figure 1).

The diagram shows that V. Lys and M. Matios most often use taste adjectives, 593.43 and 588.38, respectively. The second position among the male writers is occupied by Y. Vynnychuk (337.11), while among the female writers, I. Rozdobudko and H. Pahutiak rank second (494.53 and 480.09, respectively) (RF data correlate), which is significantly higher than Y. Vynnychuk's. The third position among the male authors belongs to V. Shkliar (256.32), and among the female writers to S. Andrukhovych (209.36). V. Shevchuk with 86.37 is at the bottom of the male authors' list and O. Zabuzhko with 108.29 is the last in the female authors' rating, but the frequency of using taste adjectives in her texts is 1.25 times higher than in V. Shkliar's. The obtained results make it possible to state that V. Lys tops the ranking in terms of the frequency of use of taste adjectives, while V. Shevchuk takes the last position.

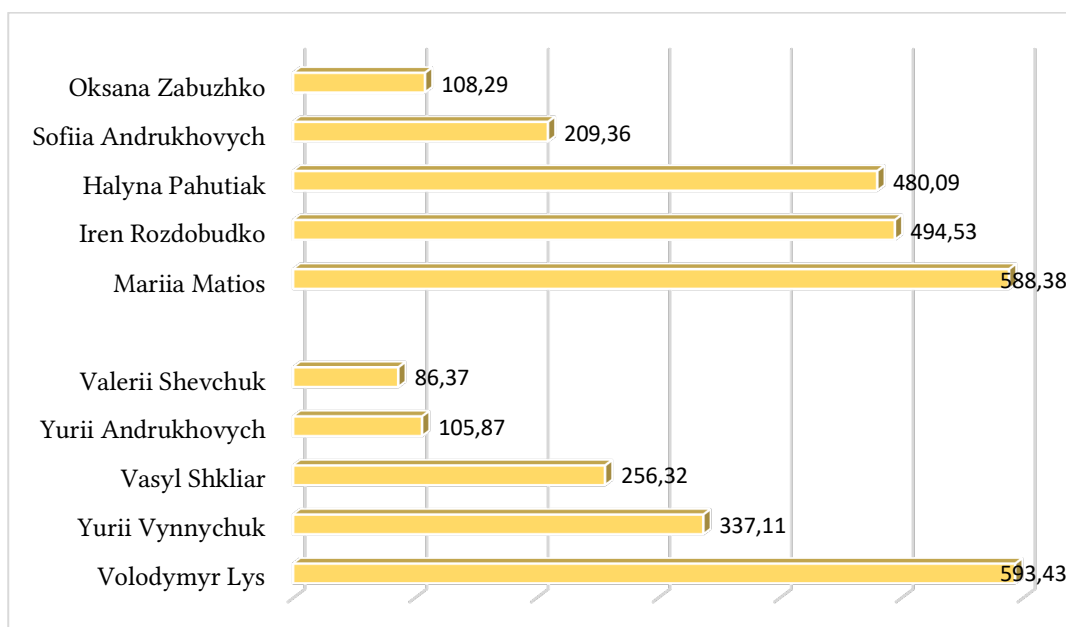


Figure 1: Generalised information about taste adjectives in the author's texts

However, it is important to trace the quantitative characteristics of the use of the taste adjectives *солодкий* (*sweet*) [*solodkyi*], *солоний* (*salty*) [*solonyi*], *гіркий* (*bitter*) [*hirkyi*], *кислий* (*sour*) [*kyslyi*], *прісний* (*bland*) [*prisnyi*], *квасний* (*tangy*) [*kvasnyi*], *нудотний* (*cloying*) [*nudotnyi*], *приторний* (*sugary*) [*prytorny*] in the texts of each of the studied writers (see Table 2 and Figure 2).

Table 2

The frequency of the studied taste adjectives in the author's texts

Taste adjective	MS, RF%.					FS, RF%.				
	V. Lys	Y. Vynnychuk	V. Shkliar	Y. Andrukhovych	V. Shevchuk	M. Matios	I. Rozdobudko	H. Palutiak	S. Andrukhovych	O. Zabuzhko
<i>солодкий</i> (sweet) [solodkyi]	231.24	192.23	117.0	55.72	41.79	476.48	241.85	176.87	126.34	54.14
<i>гіркий</i> (bitter) [hirkyi]	222.88	52.93	47.36	47.36	33.43	64.97	101.07	209.36	46.92	36.09
<i>солоний</i> (salty) [solonyi]	86.36	30.64	41.79	2.79	2.79	3.61	101.07	32.48	10.83	7.22
<i>кислий</i> (sour) [kyslyi]	39.00	41.79	22.28		2.79	7.22	18.05	43.31		
<i>прісний</i> (bland) [prisnyi]	8.36	8.36	13.93		5.57	32.49	18.05	3.61	3.61	10.83
<i>квасний</i> (tangy) [kvasnyi]		2.79	5.57			3.61		14.44	21.66	
<i>нудотний</i> (cloying) [nudotnyi]	2.79	8.36	8.36				7.22			
<i>пригорний</i> (sugary) [prytorny]	2.79						7.22			

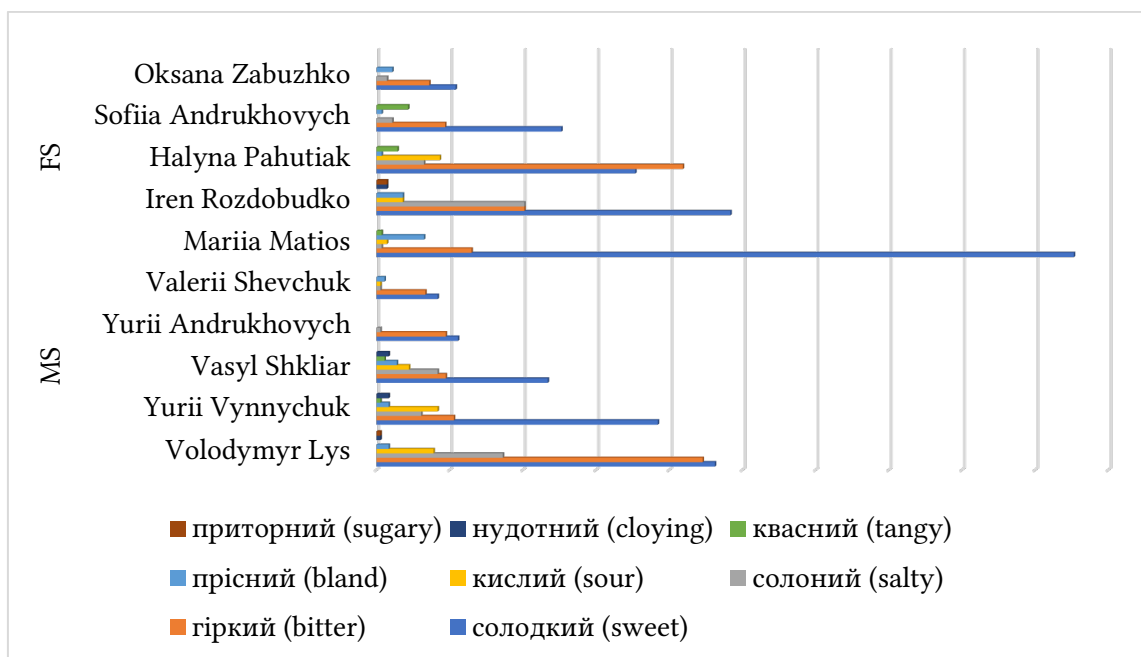


Figure 2: The studied taste adjectives in the author's texts

The structural grouping was carried out taking into account the frequency of use of the adjective *солодкий* (*sweet*) [*solodkyi*] in the texts of the studied authors. The number of groups was calculated using the Sturges formula, where n is the number of the writers studied (1).

$$k = 1 + 3,322 \log n = 1 + 3,322 \log (10) \approx 1 + 3.322 \approx 4.322 \quad (1)$$

Table 3

The results of grouping (adjective *солодкий* (*sweet*) [*solodkyi*]) in the studied texts of writers

Groups	Writers	Frequency, f_i
41.79 – 150.46	V. Shevchuk, O. Zabuzhko, Y. Andrukhovych, V. Shkliar,	5
	S. Andrukhovych	4
150.46 – 259.13	H. Pahutiak, Y. Vynnychuk, I. Rozdobudko, V. Lys	0
259.13 – 367.80		0
367.80 – 476.48	M. Matios	1

f_i is the number of values from the population that fall into a certain interval or group.

The obtained data demonstrate that M. Matios is the leader in the use of the adjective *солодкий* (*sweet*) [*solodkyi*] (476.48). At the same time, this adjective is also the most frequent in the male authors' texts, where V. Lys occupies the first position with a frequency of 231.24, which is 2 times lower than that of M. Matios. The high frequency of the word *солодкий* (*sweet*) [*solodkyi*] in the texts by the above-mentioned writer is due to the metaphor of the novel "*Солодка Даруся*" ("*Sweet Darusia*") [*Solodka Darusia*"].

Table 4

The results of grouping (adjective *гіркий* (*bitter*) [*hirkyi*]) in the studied texts of the writers

Groups	Writers	Frequency, f_i
33.43 – 80.795	V. Shevchuk, O. Zabuzhko, S. Andrukhovych, Y. Andrukhovych, V. Shkliar, Y. Vynnychuk, M. Matios	7
80.795 – 128.16	I. Rozdobudko	1
128.16 – 175.525		0
175.525 – 222.89	H. Pahutiak, V. Lys	2

The highest frequency of use of the adjective *гіркий* (*bitter*) [*hirkyi*] was observed for V. Lys and H. Pahutiak. The obtained frequency data in their texts correlate, 222.88 and 209.36, respectively.

Table 5

The results of grouping (adjective *солоний* (*salty*) [*solonyi*]) in the studied texts of the writers

Groups	Writers	Frequency, f_i
2.79 – 27.36	Y. Andrukhovych, V. Shevchuk, M. Matios, O. Zabuzhko, S. Andrukhovych	5
27.36 – 51.93	Y. Vynnychuk, H. Pahutiak, V. Shkliar	3
51.93 – 76.5		0
76.5 – 101.07	V. Lys, I. Rozdobudko	2

The highest frequency of the adjective *солоний* (*salty*) [*solonyi*], 101.07, was found in I. Rozdobudko's texts, while in V. Lys's texts the frequency of this word was 86.36.

Table 6

The results of grouping (adjective *кислий* (*sour*) [*kyslyi*]) in the studied texts of the writers

Groups	Writers	Frequency, f_i
0 – 10.83	S. Andrukhovych, Y. Andrukhovych, O. Zabuzhko, V. Shevchuk, M. Matios	5
10.83 – 21.66	I. Rozdobudko	1
21.66 – 32.49	V. Shkliar	1
32.49 – 43.32	V. Lys, Y. Vynnychuk, H. Pahutiak	3

The adjective *кислий* (*sour*) [*kyslyi*] is most often used by H. Pahutiak (43.31), Y. Vynnychuk (41.79), and V. Lys (39.00), and the data obtained from their texts correlate.

Table 7

The results of grouping (adjective *квасний (tangy) [kvasnyi]*) in the studied texts of the writers

Groups	Writers	Frequency, f_i
0 – 5.415	Y. Andrukhovych, O. Zabuzhko, V. Lys, I. Rozdobudko, V. Shevchuk, Y. Vynnychuk, M. Matios	7
5.415 – 10.83	V. Shkliar	1
10.83 – 16.245	H. Pahutiak	1
16.245 – 21.66	S. Andrukhovych	1

The adjective *квасний (tangy) [kvasnyi]* has a high frequency rate in the texts of S. Andrukhovych with 21.66 and H. Pahutiak with 14.44.

Table 8

The results of grouping (adjective *прісний (bland) [prisnyi]*) in the studied texts of the writers

Groups	Writers	Frequency, f_i
0 – 8.123	Y. Andrukhovych, S. Andrukhovych, H. Pahutiak, V. Shevchuk	4
8.123 – 16.246	Y. Vynnychuk, V. Lys, O. Zabuzhko, V. Shkliar	4
16.246 – 24.369	I. Rozdobudko	1
24.369 – 32.492	M. Matios	1

The adjective *прісний (bland) [prisnyi]* most often occurred in the texts of M. Matios (32.39), while in the male authors' texts, the highest rate of this adjective was found in V. Shkliar's (13.93), which is 2.3 times less.

Table 9

The results of grouping (adjective *нудотний (cloying) [nudotnyi]*) in the studied texts of the writers

Groups	Writers	Frequency, f_i
0 – 2.09	S. Andrukhovych, Y. Andrukhovych, O. Zabuzhko, V. Shevchuk, M. Matios, H. Pahutiak	6
2.09 – 4.18	V. Lys	1
4.18 – 6.27		0
6.27 – 8.36	I. Rozdobudko, Y. Vynnychuk, V. Shkliar	3

The adjective *нудотний* (*cloying*) [*nudotnyi*] has high frequency rates in the texts by I. Rozdobudko, 7.22, Y. Vynnychuk, 8.36, and V. Shkliar, 8.36.

Table 10

The results of grouping (adjective *приторний* (*sugary*) [*prytorny*]) in the studied texts of the writers

Groups	Writers	Frequency, f_i
0 – 1.805	S. Andrukhovych, Y. Andrukhovych, Y. Vynnychuk, O. Zabuzhko, M. Matios, H. Pahutiak, V. Shevchuk, V. Shkliar	8
1.805 – 3.61	V. Lys	1
3.61 – 5.415		0
5.415 – 7.22	I. Rozdobudko	1

The adjective *приторний* (*sugary*) [*prytorny*] is most often used only in texts by I. Rozdobudko (7.22) and V. Lys (2.79), which is 2.6 times less.

It is worth noting that male writers take the lead in using various taste adjectives. V. Lys, Y. Vynnychuk, and V. Shkliar used seven out of eight taste adjectives. Among the female writers, this is characteristic of I. Rozdobudko only.

This frequency ratio of taste attributes in the texts of different authors can be considered as a marker of an idiolect. On the other hand, the averaged data taken from the male and female authors' corpora cannot be used as a marker of a genderlect, since the variability of the indicators of an individual author is quite significant.

It is important to determine the frequency and repertoire of common root verbalisers of the main components of the СМАК (TASTE) [SMAK] concept for male and female texts. The number of different verbalisers is 34 for male texts and 27 for female texts.

MS (RF,%): *солодкавий* (*sweetish*) [*solodkavyi*] 25.07; *гірко-солодкий* (*bittersweet*) [*hirko-solodkyi*] 16.72; *гіркуватий* (*bitterish*) [*hirkuvatyi*] 13.93; *гіркавий* (*slightly bitter*) [*hirkavyi*], *згірклий* (*rancid*) [*zghirklyi*], *кисло-солодкий* (*sour-sweet*) [*kyslo-solodkyi*], *прогірклий* (*rank*) [*prohirklyi*] 11.14; *підсолоджений* (*sweetened*) [*pidsolodzhenyi*] 8.36, *кислуватий* (*sourish*) [*kysluvatyi*], *пересолоджений* (*over-sweetened*) [*peresolodzhenyi*], *солоденький* (*pleasantly sweet*) [*solodenkyi*], *солодко-гіркий* (*sweet-bitter*) [*solodko-hirkyi*] 5.57; *болісно-солодкий* (*painfully sweet*) [*bolisno-solodkyi*], *гіркаво-солодкий* (*bitter sweet*) [*hirkavo-solodkyi*], *дражливо-солодкий* (*teasingly sweet*) [*drazhlyvo-solodkyi*], *жахливо-солодкий* (*awfully sweet*) [*zhakhlyvo-solodkyi*], *мазохістично-солодкий* (*masochistically sweet*) [*mazokhistychno-solodkyi*], *млосно-солодкий* (*intoxicatingly sweet*) [*mosno-solodkyi*], *нудно-солодкий* (*sickeningly sweet*) [*nudno-solodkyi*], *солодкаво-нудотний* (*sickeningly sweetish*) [*solodkavo-nudotnyi*], *солодкий-пресолодкий* (*very sweet*) [*solodkyi-presolodkyi*], *солодко-небажаний* (*sweet but undesirable*) [*solodko-nebazhanyi*], *солодко-щільний* (*sweet and thick*) [*solodko-shchilnyi*], *услєливо-солодкий* (*flattering sweet*) [*uleslyvo-solodkyi*], *khrypkuvato-solodkyi*, *солонуватий* (*saltish*) [*solonuvatyi*], *кислувато-їдкий* (*sharp sourish*) [*kysluvato-yidkyi*], *кисло-застояний* (*stale and sour*) [*kyslo-zastoianyi*], *гіркий-гіркий* (*very bitter*) [*hirkyi-hirkyi*], *гірко-іронічний* (*bitterly ironic*) [*hirko-ironichnyi*], *гірко-терпкий*

(bitter and tart) [hirko-terpkyi], гіркувато-пряний (bitterish-spicy) [hirkuvato-prianyi], солодково-гіркий (sweetish-bitter) [solodkavo-hirkyi], пріснуватий (blandish) [prisnuvatyi] 2.78.

FS (RF, %): гіркуватий (bitterish) [hirkuvatyi], солодкавий (sweetish) [solodkavyyi] 50.54; згіркий (rancid) [zghirklyi] 10.83; гіркая (bitter) [hirkaia], нудотно-солодкий (sickeningly sweet) [nudotno-solodkyi], солодкуватий (sweetish) [nudotno-solodkyi], солонкуватий (saltish) [solonkuvatyi], солонуватий (saltish) [solonuvaty] 7.2; гіркий (bitterish) [hirkavyyi], гірко-пекучий (bitter hot) [hirko-pekuchyi], гірко-солоний (bitter salty) [hirko-solonyi], малосолоний (low-salt) [malosolonyi], м'яко-солодкий (intoxicatingly sweet) [mlosno-solodkyi], підсолоджений (sweetened) [pidsolodzhenyi], прило-солодково-болотяний (musty sweetish murshy) [prilo-solodkavo-bolotianyi], прогіркий (rank) [prohirklyi], кислуватий (sourish) [kysluvatyi], кислувато-гіркий (sourish bitter) [kysluvatohirkyi], кислувато-сирний (sourish cheesy) [kysluvatohyryni], свіжо-солодкий (fresh sweet) [svizho-solodkyi], солоджений (sweetened) [solodzhenyi], солодково-кислий (sweetish sour) [solodkavo-kyslyi], солодково-нудотний (sickeningly sweetish) [solodkavo-nudotnyi], солодко-незавершений (sweet unfinished) [solodko-nezavershenyi], солодко-терпко-теплий (sweet, tart and warm) [solodko-terpko-teplyi], солодовий (malty) [solodovyyi], цукрово-солодкий (sugary sweet) [tsukrovo-solodkyi] 3.61.

In the male authors' texts, the most frequent common root verbalisers are солодкавий (sweetish) [solodkavyyi], гірко-солодкий (bittersweet) [hirko-solodkyi], гіркуватий (bitterish) [hirkuvatyi], гіркий (slightly bitter) [hirkavyyi] згіркий (rank) [zghirklyi], кисло-солодкий (sour-sweet) [kyslo-solodkyi], прогіркий (rancid) [prohirklyi], підсолоджений (sweetened) [pidsolodzhenyi], and in the female authors' texts гіркуватий (bitterish) [hirkuvatyi], солодкавий (sweetish) [solodkavyyi], згіркий (rank) [zghirklyi]. The analysis shows that the most frequent verbalisers of taste are гіркуватий (bitterish) [hirkuvatyi], солодкавий (sweetish) [solodkavyyi] in both the male and female texts.

For example, 23 different forms of compound adjectives were traced in the male texts, and 10 forms in the female texts. The largest number of verbalisers was traced for солодкий (sweet) [solodkyi] – 17 forms in the male / 14 in the female texts, гіркий (bitter) [hirkyi] – MS 12 / FS 8, кислий (sour) [kyslyi] – MS 4 / FS 3, солоний (salty) [solonyi] – MS 1 / FS 4, прісний (bland) [prisnyi] – MS 1 / FS 0.

The study of common root verbalisers of taste revealed forms that occur in both male and female authors' texts (see Figure 3).

The analysis of the quantitative data shows that the words гіркуватий (bitterish) [hirkuvatyi] and солодкавий (sweetish) [solodkavyyi] have the highest frequency in the studied texts. Their frequency is 50.54, while in the male authors' texts the frequency of these words is 13.93 and 25.07, which is 3.6 and 2 times less than in the female texts, respectively. However, in the male texts, the frequency of the words гіркий (slightly bitter) [hirkavyyi] is 11.14, підсолоджений (sweetened) [pidsolodzhenyi] 8.36, and прогіркий (rank) [prohirklyi] 11.14, which is 3, 2.3, and 3 times higher than in the female authors' texts. It should be noted that in the studied text arrays, the Chi-square test statistic is 5.84 and p-value is 0.015 for the word гіркуватий (bitterish) [hirkuvatyi], which indicates the statistical significance of the difference in results.

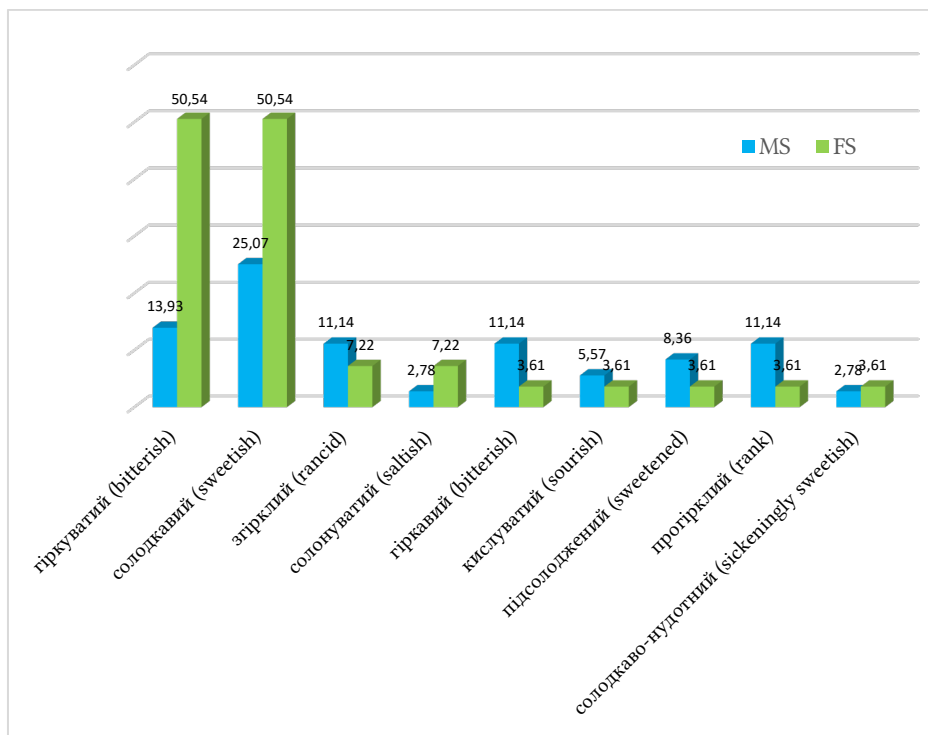


Figure 3: Frequency of the studied taste adjectives in the author's texts

It is worth noting that the most frequent taste adjective in the studied texts is *солодкий* (*sweet*) [*solodkyi*]. Thus, M. Matios takes the lead in using the adjective *солодкий* (*sweet*) [*solodkyi*], 476.48 (see Figure 4). However, it should be noted that almost half of the word combinations with *солодкий* (*sweet*) [*solodkyi*] in the writer's texts are represented by the collocations "*Солодка Даруся*" ("*Sweet Darusia*") [*Solodka Darusia*] (the title of the novel), 111.9 and *Солодка Даруся* (*Sweet Darusia*) [*Solodka Darusia*] (a character in the novel of the same name), 104.68.

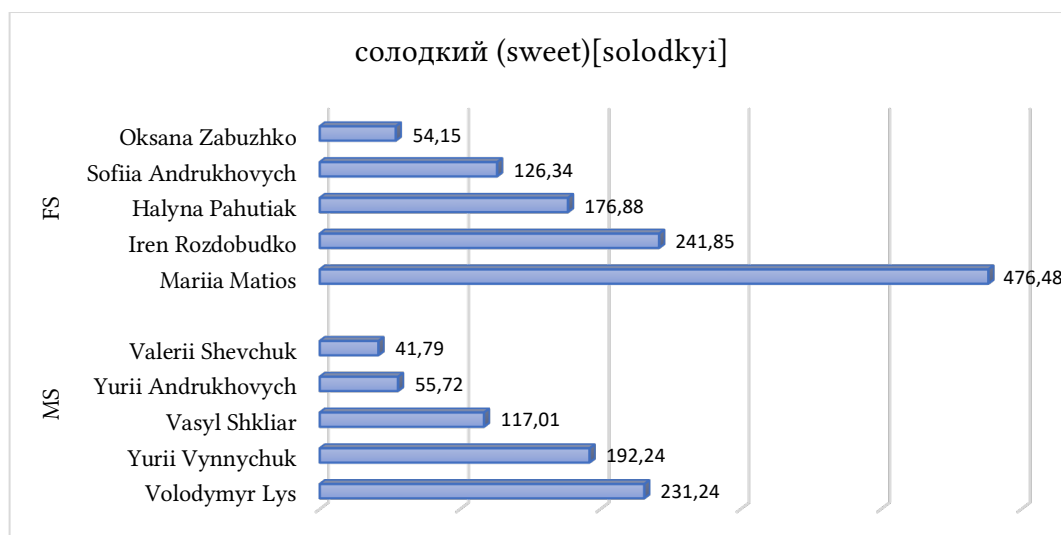


Figure 4: Frequency of *солодкий* (*sweet*) [*solodkyi*] attribute

5. Discussions

The taste feature is not independent, it is closely connected in the process of communication with the object. This feature is reflected in a special type of nomination – nominal attributive phrases [13, p.186], models of *taste adjective + noun* type. The lexical compatibility of adjectives denoting taste reflects their semantic structure, which, in turn, demonstrates the peculiarities of verbalisation of the CMAK (TASTE) [SMAK] concept.

The use of statistical association indicators such as MI-score and T-score to identify collocations does not provide a complete list of metaphors. Instead, the method of identifying metaphors based on semantic tags is effective and enables obtaining a list of metaphors.

The structural model *солодкий (sweet) [solodkyi] + noun* is the most frequent among the models of *taste adjective + noun* type. In GRAC-17, the query [lemma="*.солод.*"&tag="adj.*"] [tag="noun.*"&tag!="pron.*"] was used to automatically detect *солодкий (sweet) [solodkyi] + noun* collocations.

The taste attributes were applied to nouns belonging to a number of thematic groups (see Figure 5). It should be noted that collocations with common root verbalisers of the adjective *солодкий (sweet) [solodkyi]* were also taken into account. In the study, 138 different collocates with a total frequency of 426.26 in the male authors' texts and 109 different collocates with a total frequency of 736.38 in the female authors' texts were identified within this model.

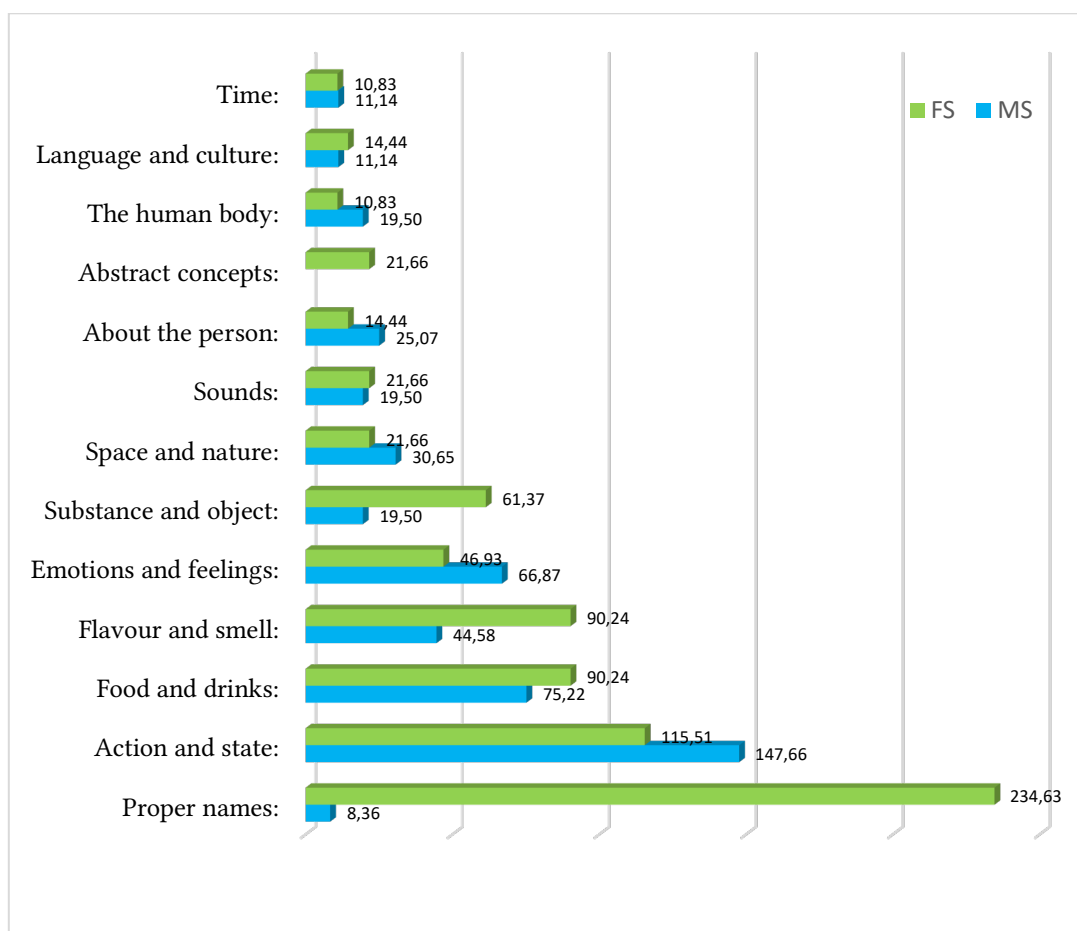


Figure 5: Thematic groups of collocates with *солодкий (sweet) [solodkyi]* component

In the female authors' texts, the taste adjective *солодкий* (*sweet*) [*solodkyi*] is most often an attribute of proper names, actions and states, food and drinks; in male authors' texts, it is an attribute of actions and states, food and drinks, emotions and feelings. The frequency of adjectives of the thematic group *Proper Names* is 28 times higher in the female texts than in male ones. In addition, the absence of taste attributes in the thematic group *Abstract Concepts* in the male texts is observed.

The most frequent collocates of the attribute *солодкий* (*sweet*) [*solodkyi*] and its verbalisers are:

MS (RF, %) – *дух* (*spirit*) [*dukh*] 13.93; *запах* (*smell*) [*zapakh*], *сон* (*sleep*) [*son*], *туга* (*longing*) [*tuha*] 11.14; *біль* (*pain*) [*bil*], *напій* (*drink*) [*napii*], *обійми* (*hug*) [*obiimy*], *пахощі* (*fragrance*) [*pakhoshchi*], *передчуття* (*anticipation*) [*peredchuttia*], *усмішка* (*smile*) [*usmishka*] 8.36; *виродження* (*degeneration*) [*vyrodzhennia*], *втома* (*fatigue*) [*vtoma*], *голос* (*voice*) [*holos*], *груша-скороспілка* (*early pear*) [*hrusha-skorospilka*], *дихання* (*breath*) [*dykhannia*], *дівчинка* (*girl*) [*divchynka*], *дурничка* (*trifle*) [*durnychka*], *знемога* (*exhaustion*) [*znetoha*], *кава* (*coffee*) [*kava*], *млість* (*languor*) [*mlist*], *отрута* (*poison*) [*otruta*], *парочка* (*couple*) [*parochka*], *перець* (*pepper*) [*perets*], *слово* (*word*) [*slovo*], *спогад* (*memory*) [*sprohad*], *хвилина* (*minute*) [*khvylyna*] 5.57;

FS (RF, %) – *Даруся* (*Darusia*) [*Darusia*] 216.58, *запах* (*smell*) [*zapakh*] 39.71; *сон* (*sleep*) [*son*], *тісто* (*dough*) [*tisto*] 18.05, *вино* (*wine*) [*vyno*], *голос* (*voice*) [*holos*], *нудота* (*nausea*) [*nudota*], *сік* (*juice*) [*sik*], *узвар* (*stewed fruit*) [*uzvar*] 10.83; *Буковина* (*Bukovina*) [*Bukovyna*], *втома* (*fatigue*) [*vtoma*], *дух* (*spirit*) [*dukh*], *мука* (*torture*) [*muca*], *отрута* (*poison*) [*otruta*], *парочка* (*голуб'ят*) (*a couple (of doves)*) [*parochka (holub'iat)*], *передчуття* (*anticipation*) [*peredchuttia*], *смак бажання* (*taste of desire*) [*smak bazhannia*], *тістечко* (*cookie*) [*tistechko*] 7.22.

One of the markers of the writers' idiolect is the specificity of the use of taste names, in particular in the direct and figurative senses. In other words, we can talk about the taste attribution of abstract and concrete concepts. The relative frequency of attributive collocations used in the direct sense with the adjective *солодкий* (*sweet*) [*solodkyi*] is 75.22% in the male authors' texts and 90.24% in female authors' texts. The number of different attributive collocations used in the figurative sense with the adjective *солодкий* (*sweet*) [*solodkyi*] is 103 with a total frequency of 351.04 for MS and 91 with a total frequency of 646.14 for FS. Moreover, the frequency of collocations used in the direct sense is 1.2 times higher for FS than for MS, while the frequency of metaphorical collocations is 1.8 times higher for FS.

At present, the semantically annotated corpus of GRAC-17 enables the automated identification of metaphorical collocations with the adjective *солодкий* (*sweet*) [*solodkyi*] using the following queries:

- [lemma=".*солод.*"&tag="adj.*"] [tag="noun.*"&tag!="pron.*"&tag!=".*prop.*"&semtag="1:conc.*"] – to search for collocations used in the direct sense, excluding collocations with proper names;
- [lemma=".*солод.*"&tag="adj.*"] [tag="noun.*"&tag!="pron.*"&semtag="1:abst.*"] – for collocations used in the figurative sense;
- [lemma=".*солод.*"&tag="adj.*"] [tag="noun.*"&tag!="pron.*"&tag=".*prop.*"&semtag="1:conc.*"] – for collocations including proper names.

It is worth noting that since the semantic annotation of the corpus is still being processed, the error rate should be taken into account when evaluating the results.

The following metaphorical collocations with the word *солодкий* (*sweet*) [*solodkyi*] were found using the above queries: MS (RF, %) – *солодкий сон* (*sweet sleep*) [*solodkyi son*] 11.14, *біль* (*pain*) [*bil*] 8.36, *хвилини* (*minutes*) [*khvylyny*] 5.57, *мить* (*moment*) [*myt*], *гра* (*game*) [*hra*], *страх* (*fear*) [*strakh*], *жаль* (*regret*) [*zhal*], *почуття* (*feeling*) [*pochuttia*], *музика* (*music*) [*muzyka*], *кисло-солодкий дух* (*sour-sweet spirit*) [*kyslo-solodkyi dukh*], *гірко-солодкий дух* (*bittersweet spirit*) [*hirko-solodkyi dukh*], *гірко-солодкий вітер* (*bittersweet wind*) [*irko-solodkyi viter*], *болісно-солодка тема* (*painfully sweet theme*) [*bolisno-solodka tema*], *солодкий Рома, Павлик* (*sweet Roma, Pavlyk*) [*solodkyi Roma, Pavlyk*] 2.78; FS (RF, %) – *Солодка Даруся* (*Sweet Darusia*) [*Solodka Darusia*] 216.58, *сон* (*sleep*) [*son*] 10.82, *солодко-незавершена думка* (*sweet unfinished thought*) [*solodko-nezavershena dumka*], *солодкавий дух* (*sweetish spirit*) [*solodkavyi dukh*], *солодкий дух* (*sweet spirit*) [*solodkyi dukh*], *біль* (*pain*) [*bil*], *хвилина* (*minute*) [*khvylyna*], *процес* (*process*) [*protses*], *ніч* (*night*) [*nich*], *година* (*hour*) [*hodyna*], *сон* (*sleep*) [*son*], *життя* (*life*) [*zhyttia*], *добро* (*goodness*) [*dobro*], *відчуття* (*feeling*) [*vidchuttia*], *смерть* (*death*) [*smert*], *мова* (*speech*) [*tova*], *музика* (*music*) [*muzyka*], *Карпати* (*Carpathians*) [*Karpaty*], *Буковина* (*Bukovina*) [*Bukovyna*] 3.61.

The query to find collocations with specific nouns revealed a number of collocations, including *солодка дівчинка* (*sweet girl*) [*solodka divchynka*], *солодка пташка* (*sweet bird*) [*solodka ptashka*], *солодка сучка* (*sweet bitch*) [*solodka suchka*], *солодка панна* (*sweet lady*) [*solodka panna*], *солодкий мурчик* (*sweet pussycat*) [*solodkyi murchyk*] (MS), which acquire a metaphorical meaning with the word *солодкий* (*sweet*) [*solodkyi*] in relation to a person. A similar situation is observed with the collocations *солодка назва* (*sweet name*) [*solodka nazva*] (MS, FS) and *солодке місце* (*sweet place*) [*solodke mistse*], *солодке повітря* (*sweet air*) [*solodke povitria*] (FS), since the words *назва* (*name*) [*nazva*], *місце* (*place*) [*mistse*], *повітря* (*air*) [*povitria*] are concrete concepts, but in combination with the word *солодкий* (*sweet*) [*solodkyi*] they acquire metaphorical meaning. The collocations to describe aroma and smell are used with the meaning 'causing pleasant sensations, giving pleasure', and therefore we interpret them as metaphorical.

Phrases with the word *солодкий* (*sweet*) [*solodkyi*] in the direct sense are used to describe food and drinks, substances: MS – *вино* (*wine*) [*vyno*], *водичка* (*water*) [*vodychka*], *кава* (*coffee*) [*kava*], *карамелька* (*caramel*) [*karamelka*], *лизак* (*lollipop*) [*lyzak*], *напій* (*drink*) [*napii*], *пиріг* (*pie*) [*pyrih*], *пляцок* (*cake*) [*pliatok*], *сік* (*juice*) [*sik*], *чай* (*tea*) [*chai*], *пахощі* (*fragrance*) [*prakhoshchi*], *СМАК* (*TASTE*) [*SMAK*], *маса* (*paste*) [*masa*], *отрута* (*poison*) [*otruta*], etc.; FS – *вино* (*wine*) [*vyno*], *горілка* (*vodka*) [*horilka*], *віскі* (*whiskey*) [*viski*], *желе* (*jelly*) [*zhele*], *малина* (*raspberry*) [*malyna*], *пляцок* (*cake*) [*pliatok*], *узвар* (*stewed fruit*) [*uzvar*], *фрукт* (*fruit*) [*frukt*], *чай* (*tea*) [*chai*], *СМАК* (*TASTE*) [*SMAK*], *тісто* (*dough*) [*tisto*], *вата* (*cotton wool*) [*vata*], *пагока* (*molasses*) [*patoka*], *рідина* (*liquid*) [*ridyna*], etc.

From the above examples, it can be noted that the collocation *солодке вино* (*sweet wine*) [*solodke vyno*] is used in both male and female texts, but the phrases *солодка горілка* (*sweet vodka*) [*solodka horilka*] and *солодке віскі* (*sweet whiskey*) [*solodke viski*] (strong alcoholic beverages) can be traced only in the female texts. The collocations describing food and drinks have the meanings 'sweet', 'made with sugar' and 'pleasant'. It should be noted that the collocation *солодке вино* (*sweet wine*) [*solodke vyno*] has the meaning 'made from grape varieties that contain a significant amount of sugar'.

Having analysed the attributive patterns with the word *солодкий* (*sweet*) [*solodkyi*] used in a figurative, including metaphorical, sense, we can say that the typical meaning is 'pleasant', 'calm' and 'happy': *біль* (*pain*) [*bil*], *втіха* (*comfort*) [*vtikha*], *сон* (*sleep*) [*son*], *втома* (*fatigue*)

[*vtoma*], *усмішка* (*smile*) [*usmishka*], *голос* (*voice*) [*holos*], *музика* (*music*) [*muzyka*], *долоня* (*palm*) [*dolonia*], *губи* (*lips*) [*huby*], *мить* (*moment*) [*myt*], *хвилинка* (*minute*) [*khvylyna*], *слово* (*word*) [*slovo*], *життя* (*life*) [*zhyttia*], *думка* (*thought*) [*dumka*], etc.

For example, *голос* (*voice*) [*holos*], *музика* (*music*) [*muzyka*] (MS, FS), *звук* (*sound*), *стогін* (*moan*), *шепіт* (*whisper*) [*shepit*] (MS), *спів* (*singing*) [*spiv*], *мурчання* (*purring*) [*murchannia*] (FS) has the meaning 'pleasing to the ear, heartfelt, sincere, pleasant'; *слово* (*word*) [*slovo*] (MS, FS), *мова* (*speech*) [*mova*] (FS) – 'excessively affectionate, flattering expression'; *благодать* (*grace*) [*blahodat*], *кохання* (*love*) [*kokhannia*], *надія* (*hope*) [*nadiia*] (MS), *насолода* (*pleasure*) [*nasoloda*], *добро* (*goodness*) [*dobro*], *слабкість* (*weakness*) [*slabist*] (FS) – 'giving moral satisfaction, comfort'; *втома* (*tiredness*) [*vtoma*], *поцілунок* (*kiss*) [*potsilunok*], *усмішка* (*smile*) [*usmishka*] (MS, FS), *брехня* (*lie*) [*brekhnia*], *звичка* (*habit*) [*zvychka*], *мрія* (*dream*) [*triiia*], *обійми* (*hug*) [*obiitmy*], *оргазм* (*orgasm*) [*orhazm*], *чекання* (*waiting*) [*chekannia*] (MS), *диво* (*miracle*) [*dyvo*], *очманіння* (*daze*) [*ochmaninnia*], *смерть* (*death*) [*smert*], *спокій* (*peace*) [*spokii*], *хвилювання* (*excitement*) [*khvyliuvannia*] (FS) – 'giving a feeling of pleasure or expressing satisfaction, joy, enjoyment (of a person's thoughts, feelings, state)'; *сон* (*sleep*) [*son*] (MS, FS) – 'a calm, deep, serene sleep'; *мить* (*moment*) [*myt*], *хвилинка* (*minute*) [*khvylyna*], *юність* (*youth*) [*yunist*] (MS), *життя* (*life*) [*zhyttia*] (FS) – 'full of abundance, happiness, joy, pleasure; happy'.

The use of such collocations with *солодкий* (*sweet*) [*solodkyi*] as a *парочка* (*couple*) [*parochka*] 'people in love' (MS, FS), *пташка* (*birdie*) [*ptashka*], *мурчик* (*pussycat*) [*murchyk*] (MS) to describe people in the sense of 'attractiveness, cuteness, charm', *сучка* (*bitch*) [*suchka*] (MS) is used in a negative context: 'a woman who uses her attractiveness to achieve her goals', *бицюра* (*goon*) [*bytsiura*] (MS) is used to describe physical strength; the names *Ізюм* (*Izium*) [*Izium*], *Павлик* (*Pavlyk*) [*Pavlyk*], *Рома* (*Roma*) [*Roma*], *дівчинка* (*girl*) [*divchynka*], *панна* (*panna*) [*panna*] (MS), *Даруся* (*Darusia*) [*Darusia*], *ROSHEN*, *Буковина* (*Bukovina*) [*Bukovyna*], *Карпати* (*Carpathians*) [*Karpaty*] (FS) describe places and people, attitudes towards them, relationships with them, with the meaning 'close and dear to the heart'. For example, *Можна обижатися хоч сотню разів, що вона, та солодка сучка* (*sweet bitch*) [*solodka suchka*], *продала любов ...* (V. Lys, "Століття Якова" ("Jacob's Century") ["Stolittia Yakova"]); *У голові моїй уже розтікалася солодка благодать* (*sweet grace*) [*solodka blahodat*]; *усе, чого я прагнув у цей вечір – напитися і звалити додому* (Y. Vynnychuk, "Весняні ігри в осінніх садах" ("Spring Games in Autumn Gardens") ["Vesniani ihry v osinnikh sadakh"]).

The analysis of comparisons with the component *солодкий* (*sweet*) [*solodkyi*] makes it possible to identify prototypes of *солодкий* (*sweet*) [*solodkyi*]: *вишнева наливка солодка* (*sweet*) [*solodka*], *як її пупчик* (Y. Vynnychuk, "Цензор снів" ("Sensor of Dreams") ["Tsenzor sniv"]); *вечір, солодкий* (*sweet*) [*solodkyi*], *як квіти рожевих лип* (Y. Vynnychuk, "Цензор снів" ("Sensor of Dreams") ["Tsenzor sniv"]); *думка про Євцю – солодка* (*sweet*) [*solodka*], *як груша на Спаса* (V. Lys, "Стара холера" ("Old Cholera") ["Stara kholera"]); *звістка була солодка* (*sweet*) [*solodka*], *як мед* (V. Lys, "Соло для Соломії" ("Solo for Solomiya") ["Solo dlia Solomii"]); *слива [...] солодка* (*sweet*) [*solodka*], *як мед* (V. Shkliar, "Чорне Сонце" ("Black Sun") ["Chorne Sontse"]); *любоффф солодка* (*sweet*) [*solodka*] *як повія* (Y. Andrukhovych, "Дванадцять обручів" ("Twelve Hoops") ["Dvanadtsiat obruchiv"]); *сестрички [...] солоденькі* (*sweet*) [*solodki*], *як мармулядки* (Y. Vynnychuk, "Мальва Ланда" ("Malva Landa") ["Malva Landa"]); *сливи [...] солодкі* (*sweet*) [*solodki*], *як мед* (I. Rozdobudko, "Гудзик" ("The Button") ["Gudzyk"]); *повітря [...] солодким* (*sweet*)

[*solodke*], як молочний коктейль (I. Rozdobudko, "Шості двері" ("The Sixth Door") ["Shosti dveri"]); життя її солодке (*sweet*) [*solodke*] наче мед (H. Pahutiak, "Брат мій Енкіду" ("My Brother Enkidu") ["Brat mii Enkidu"]). It is noteworthy that in the studied texts the traditional prototype of sweet for the Ukrainian linguistic community – honey – is actualised. Y. Vynnychuk and Y. Andrukhovych associate sweet taste with pleasure and enjoyment.

As already mentioned, among the authors studied, the highest frequency of the *солодкий* (*sweet*) [*solodkyi*] attribute is observed in the texts of V. Lys and M. Matios. M. Matios gives the attribute *солодкий* (*sweet*) [*solodkyi*] to nouns belonging to the following thematic groups: **proper names**: "Солодка Даруся" ("Sweet Darusia") ["Solodka Darusia"] (the title of the novel) 111.9 and Солодка Даруся (*Sweet Darusia*) [*Solodka Darusia*] 104.7 (the name of the character of the novel of the same name), Буковина (*Bukovina*) [*Bukovyna*] 7.22; **food and drinks**: малина (*raspberries*) [*malyna*], узвар (*stewed fruit*) [*uzvar*] 7.22; желе (*jelly*) [*zhele*], їжачки-піки (*hedgehogs-piques (a dish)*) [*jizhachky-piky*], перець (*pepper*) [*perets*], піраміда десерту (*dessert pyramid*) [*piramida desertu*] 3.61; **emotions and feelings**: добро (*goodness*) [*dobro*], радість (*joy*) [*radist*], сентиментальність (*sentimentality*) [*sentymentalnist*], смак бажання (*taste of desire*) [*smak bazhannia*] 3.61; **action and state**: дихання (*breathing*) [*dykhannia*], дрож-пропасниця (*shivering-fever*) [*drozh-propasnytsia*], розстання (*parting*) [*rozstannia*], сон (*sleep*) [*son*] 3.61; **substance and object**: отрута (*poison*) [*otruta*] 7.22; стіл (*table*) [*stil*], цемент (*cement (about cream)*) [*cement*] 3.61; **space and nature**: вода (*water*) [*voda*], житло (*dwelling*) [*zhytlo*], ружа (*rose*) [*ruzha*] 7.22; **about a person**: тітонька (*auntie*) [*titonka*], ягідка (*berry (about a person)*) [*yahidka*] 3.61; **human body**: губи (*lips*) [*huby*], долоня (*palm*) [*dolonia*] 3.61; **abstract concepts**: життя (*life*) [*zhyttia*] 3.61; **aroma and smell**: запах (*smell*) [*zapakh*] 7.22; **sounds**: голос (*voice*) [*holos*] 3.61; **language and culture**: слово (*word*) [*slovo*] 3.61; **time**: ніч (*night*) [*nich*] 3.61.

The fact of attributing the attribute *sweet* to emotions, states and actions is important for authorship attribution (... тонким наманікюремим пальчиком робить гостеві екскурсію не просто заставленим солодощами столом – а солодким смаком бажання (*sweet taste of desire*) [*solodkym smakom bazhannia*] (M. Matios, "Кулінарні фіґлі" ("Culinary Tricks") ["Kulinarni figli"]); Скажу Вам одне: кучерявим амуром – цим солодким дивом (*sweet miracle*) [*solodkym dyvom*] – Ви можете здивувати спадкового кронпринца, не те що кучерявого коханця (M. Matios, "Кулінарні фіґлі" ("Culinary Tricks") ["Kulinarni figli"]). The collocations *солодкий сон* (*sweet dream*) [*solodkyi son*], *солодке життя* (*sweet life*) [*solodke zhyttia*], *солодке слово* (*sweet word*) [*solodke slovo*] are stabilised metaphors: А друге спало, як у мами за пазухою, леда всміхаючись солодкому сниву (*sweet dream*) [*solodkomu snyvu*] кутиками уст та смішно ворущачи бровенятками (M. Matios, "Щоденник страченої" ("Diary of the Executed") ["Shchodennyk strachenoii"]); Повірте, моя порада почати спільне солодке життя (*sweet life*) [*solodke zhyttia*] дамськими пальчиками обернеться Вам сторницею (M. Matios, "Кулінарні фіґлі" ("Culinary Tricks") ["Kulinarni figli"]); Жодне вино, жодне солодке слово (*sweet word*) [*solodke slovo*] і найдурманливіші обійми в земному житті не дали мені такої веселості (M. Matios, "Чотири пори життя" ("Four Seasons of Life") ["Chotyry pory zhyttia"]). The writer uses diminutives with the adjective *солодкий* (*sweet*) [*solodkyi*] to express a favourable and affectionate attitude: І тоді мій син – моє любе дитятко, моя солодка ягідка (*sweet berry*) [*solodka yahidka*] – стрепенувся під чоловіковою рукою, потягнувся до пташки і заплакав (M. Matios, "Чотири пори життя" ("Four Seasons of Life") ["Chotyry pory zhyttia"]). The collocation *солодка вода* (*sweet water*) [*solodka voda*] is used in the literal sense of

‘without a salty taste’: *Твій орел дістає мене і тут, у піднебессі [...] він клює дзьобиком квіти на полонинах і пляях [...] і гейзери солодких вод (sweet waters) [solodkykh vod] салютують із-під укритого мохом каміння* (M. Matios, “Чотири пори життя” (“Four Seasons of Life”) [“Chotyry pory zhyttia”]).

In the texts of V. Lys, the adjective *солодкий (sweet) [solodkyi]* is an attribute of nouns in such thematic groups as **emotions and feelings** *трем (chills) [trem]* 5.57; *біль (pain) [bil]*, *воля (will) [volia]*, *задоволення (pleasure) [zadovolennia]*, *гріх кохання (sin of love) [hrikh kokhannia]*, *надія (hope) [nadiia]*, *хвиля радості (wave of joy) [khvyliia radosti]* 2.78; **action and state**: *передчуття (anticipation) [peredchuttia]* 8.36; *брехня (lie) [brekhnia]*, *дихання (breath) [dykhannia]*, *звичка (habit) [zvychka]*, *знемога (exhaustion) [znemoha]*, *ілюзія (illusion) [iliuziia]*, *марення (delusion) [marennia]*, *обійми (hug) [obiimy]*, *усмішка (smile) [usmishka]*, *сон (sleep) [son]*, *спогад (memory) [spohad]*, *сподівання (hope) [spodivannia]*, *чекання (expectation) [chekannia]*, *шум спогаду (noise of memory) [shum spohadu]* 2.78; **food and drinks**: *груша-скороспілка (early pear) [hrusha-skorospilka]* 5.57; *півник (cockerel) [pivnyk]*, *чай (tea) [chai]* 2.78; **space and nature**: *глибина (depth) [hlybuna]* 2.78; **sounds**: *голос (voice) [holos]*, *скрикування (screaming) [skrykuvannia]* 2.78; **human body**: *долоня (palm) [dolonia]*, *палець (finger) [palets]* 2.78; **substance and object**: *цукор (sugar) [tsukor]*, *крупинка (grain) [krupynka]*, *покривало (bedspread) [pokryvalo]* 2.78; **proper names**: *Ізіум (Izium) [Izium]*, *Павлик (Pavlyk) [Pavlyk]* 2.78; **about a person**: *бицюра (goon) [bytsiura]*, *сучка (bitch) [suchka]* 2.78.

V. Lys mostly uses the adjective *солодкий (sweet) [solodkyi]* to describe emotions, feelings, physiological states that bring joy, comfort, pleasure, or calmness, for example: *Я став провалюватися кудись униз [...], відчуваючи солодке задоволення (sweet pleasure) [solodke zadovolennia] од самого цього польоту* (V. Lys, “Камінь посеред саду” (“Stone in the Middle of the Garden”) [“Kamin posered sadu”]); *У тілі жив солодкий трем (sweet chills) [solodkyi trem] сьогоднішнього кохання* (V. Lys, “Діва Млинища” (“The Maiden of Mlynushche”) [“Diva Mlynushcha”]); *Але хай не заважає чекати тата, не перешкоджає її такому солодкому чеканню (sweet expectation) [solodkomu chekanniu]* (V. Lys, “Країна гіркої ніжності” (“The Country of Bitter Tenderness”) [“Kraina hirkoii nizhnosti”]); the use of the stabilised metaphor *солодкий сон (sweet dream) [solodkyi son]* is also traced: *Тітка, висока худа жінка з нервовими жилавими руками [...], подивилася на мене так, наче прокидалася з солодкого сну (sweet sleep) [solodkoho snu]* (V. Lys, “І прибуде суддя” (“And the Judge Will Arrive”) [“I prybude suddia”]). To describe a person, the writer uses coarse forms to express a positive, admiring attitude, which are individual authorial collocations: – *Мати мене хочете, солодкий бицюро (sweet goon) [solodkyi bytsiuoro]? Не дивуйтеся, вас у селі бицюрою прозивають* (V. Lys, “Соло для Соломії” (“Solo for Solomiya”) [“Solo dlia Solomii”]); *Можна обижатися хоч сотню разів, що вона, та солодка сучка (sweet bitch) [solodka suchka], продала любов* (V. Lys, “Століття Якова” (“Jacob’s Century”) [“Stolittia Yakova”]).

6. Conclusion

The obtained factual material makes it possible to present information on the frequency of the verbalisers of the СМАК (TASTE) [SMAK] concept of different types used by male and female writers in prose fiction. The statistical analysis of the studied texts shows that the adjectives *солодкий (sweet) [solodkyi]*, *гіркий (bitter) [hirkyi]*, *солоний (salty) [solonyi]* are the first,

second, and third most frequently used in the male and female authors' texts, respectively. In the male texts, *кислий (sour) [kyslyi]* ranks fourth, *прісний (bland) [prisnyi]* fifth, *нудотний (cloying) [nudotnyi]* sixth, *квасний (tangy) [kvasnyi]* seventh, and *приторний (sugary) [prytorny]* eighth, respectively. In the female authors' texts, the adjectives *кислий (sour) [kyslyi]* and *прісний (bland) [prisnyi]* take the fourth position, *квасний (tangy) [kvasnyi]* is ranked fifth, *нудотний (cloying) [nudotnyi]* and *приторний (sugary) [prytorny]* are in the sixth position. The words *солодкий (sweet) [solodkyi]* and *квасний (tangy) [kvasnyi]* show statistically significant differences in the frequency of use in the male and female texts, indicating a significant correlation with gender.

For the taste adjectives *солодкий (sweet) [solodkyi]*, *солоний (salty) [solonyi]*, *гіркий (bitter) [hirkyi]*, *кислий (sour) [kyslyi]*, *нудотний (cloying) [nudotnyi]*, *квасний (tangy) [kvasnyi]*, *приторний (sugary) [prytorny]* in the texts of each of the studied writers, the quantitative characteristics of occurrence are presented and a structural grouping is made taking into account the frequency of use of the adjective *солодкий (sweet) [solodkyi]* in the texts of the studied authors. Thus, in the texts under study, the typical frequency of taste adjectives is within the following ranges: 41.79 – 150.46 – *солодкий (sweet) [solodkyi]*; 33.43 – 80.795 – *гіркий (bitter) [hirkyi]*; 2.79 – 27.36 – *солоний (salty) [solonyi]*; 0 – 10.83 – *кислий (sour) [kyslyi]*; 0 – 5.415 – *квасний (tangy) [kvasnyi]*; 0 – 16.246 – *прісний (bland) [prisnyi]*; 0 – 2.09 – *нудотний (cloying) [nudotnyi]*; 0 – 1.805 – *приторний (sugary) [prytorny]*.

For the purposes of authorship attribution and creation of a statistical profile of the idiolect, the frequency of use of verbalisers of the СМАК (TASTE) [SMAK] concept in the texts of a particular writer was identified. Thus, V. Lys, 593.43 and M. Matios, 588.38 most often use taste adjectives; V. Shevchuk, 86.37 holds the last position in terms of frequency of use among the male authors, and O. Zabuzhko, with 108.29, is rated last among the female writers. The frequency ratio of taste attributes in the texts of different authors can be considered as a marker of the idiolect, but the averaged data taken from the male and female authors' corpora cannot be used as a gender marker.

The compatibility of adjectives denoting taste reflects their semantics, which, in turn, shows the peculiarities of the СМАК (TASTE) [SMAK] concept verbalisation. The thematic grouping of the collocates of attributive collocations demonstrates the dominant groups of typical collocates of taste attributes, which is shown by the example of the verbaliser *солодкий (sweet) [solodkyi]*. In the female authors' texts, the adjective *солодкий (sweet) [solodkyi]* is most often an attribute of proper names, action and state, food and drinks; in the male texts, it is an attribute of action and state, food and drinks, emotions and feelings.

One of the markers of the writers' idiolect is the specificity of the use of names of taste, in particular in direct and figurative senses. To identify collocations with the studied components, the article proposes a methodology for automated corpus data analysis based on semantic corpus annotation. The analysis of the obtained data shows that the frequency of collocations with the adjective *солодкий (sweet) [solodkyi]*, used in the direct sense, is 75.22 in the male authors' texts and 90.24 in the female authors' texts. The frequency of such collocations used in a figurative sense is 351.04 for MS and 646.14 for FS. Phrases with the word *солодкий (sweet) [solodkyi]* in the literal sense are used to describe food and drinks, substances. Typical meanings of attributive models with the word *солодкий (sweet) [solodkyi]* used in a figurative, including metaphorical, sense are 'pleasant', 'calm', 'happy'.

Among the authors studied, the highest frequency of the *солодкий (sweet) [solodkyi]* attribute was observed in the texts of V. Lys and M. Matios. For instance, they use the attribute

солодкий (*sweet*) [*solodkyi*] to describe emotions, feelings, physiological states that bring joy, pleasure or calmness. In V. Lys's texts, we can trace the use of coarse forms to express a positive admiring attitude, which are individually authorial collocations (*солодкий бицюра* (*sweet goon*) [*solodkyi bytsiura*], *солодка сучка* (*sweet bitch*) [*solodka suchka*]).

The developed methodology of automated corpus data analysis for identifying metaphorical collocations with the CMAK (TASTE) [SMAK] concept enables automated/automatic authorship attribution based on statistical data, in particular, determination of the author's gender and identification of the markers of the idiolect of the studied authors. For this kind of corpus research, semantic annotation of the corpus is important. After the GRAC semantic annotation is completed, it is planned to conduct a study based on a larger list of taste names.

References

- [1] O. A. M. S. Al-Nakeeb, B. A. H. Mufleh, Collocations and Collocational Networks of Characters: A Corpus-based Feminist Stylistic Analysis, *Language in India*, 18(9), 2018.
- [2] J. A. Avery, A. G. Liu, M. Carrington, A. Martin, Taste metaphors ground emotion concepts through the shared attribute of valence, *Frontiers in Psychology*, 2022, 13, 938663.
- [3] M. Bagli, *Tastes We Live By: The Linguistic Conceptualisation of Taste in English*, Vol. 50, Walter de Gruyter GmbH & Co KG, 2021
- [4] A.-S. Barwich, *Smellosophy: What the Nose Tells the Mind*, Harvard University Press, 2020.
- [5] A. D. Belova, Gustatory, Olfactory, Tactile Modalities in Connoisseurial Food Reviews, *Studia Linguistica*, Vol. 22, 2023, pp. 9-22.
- [6] J.-A. Brillat-Savarin, *The Physiology of Taste: or Meditations on Transcendental Gastronomy*, Dover Publications, Inc., 2019.
- [7] N. W. Brown, *An Architecture of Participation for Computational Social Biology*, University of California, Los Angeles, 2013.
- [8] K. Church, P. Hanks, Word association norms, mutual information, and lexicography, *Computational Linguistics*, Vol. 16 (1), 1990, pp. 22–29.
- [9] S. Evert, *The Statistics of Word Cooccurrences: Word Pairs and Collocations*, PhD thesis, University of Stuttgart, 2004.
- [10] P. Freedman, *Food: the history of taste*, Vol. 21. Univ of California Press, 2007.
- [11] L. Gandy, N. Allan, M. Atallah, O. Frieder, N. Howard, S. Kanareykin,... & S. Argamon, Automatic identification of conceptual metaphors with limited knowledge, In: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 27(1), 2013, pp. 328-334.
- [12] P. C. Gómez, *Statistical methods in language and linguistic research*, London, UK, Equinox, 2013.
- [13] V. Haidaienko, *Nazvy na poznachennia smaku: etymolohiia, semantyka, funktsionuvannia*, Dysertatsiia kandydata filolohichnykh nauk, Khersonskyi derzhavnyi pedahohichnyi universytet, Kherson, 2002. [I. V. Gaidaienko, *Taste names: etymology, semantics, functioning*, PhD thesis in philology, Kherson State Pedagogical University, Kherson, 2002].
- [14] A. Hazaa, N. Omar, F. M. Ba-Alwi, & M. Albared, Automatic Extraction of Malay Compound Nouns Using A Hybrid of Statistical and Machine Learning Methods. *International Journal of Electrical and Computer Engineering*, Vol. 6, No. 3, 2016, pp. 925–935. doi: 10.11591/ijece.v6i3.9663
- [15] V. Henshaw, K. McLean, D. Medway, C. Perkins, G. Warnaby, *Designing with smell: Practices, techniques and challenges*, Routledge, 2017.

- [16] J. Hoegg, J. W. Alba, Linguistic framing of sensory experience: there is some accounting for taste, *Psycholinguistic phenomena in marketing communications*, 2020, pp. 3-22.
- [17] T. Igamberdiev, H. Shin, Metaphor identification with paragraph and word vectorization: An attention-based neural approach, in: *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*, 2018, pp. 212–221.
- [18] M. K. Jamieson, G. H. Govaart, M. Pownall, Reflexivity in quantitative research: A rationale and beginner's guide, *Social and Personality Psychology Compass*, 2023, 17(4), e12735.
- [19] E. Keshet, *A matter of taste, Evidentials and modals*, Brill, 2020, pp. 69-81.
- [20] M. Kulchytskyi, Unormuvannia tekstu pid chas dokorpusnoho opratsiuvannia: dosvid zastosuvannia. *Visnyk Natsionalnoho universytetu "Lvivska politekhnika"*, Seriya: Informatsiini systemy ta merezhi, Vyp. 7, 2020, ss. 51–58. [I. M. Kulchytskyi, Text normalization during pre-corpus preparation: experience of application, *Bulletin of the National University "Lviv Polytechnic"*, Series: Information systems and networks, Vol. 7, 2020, pp. 51–58].
- [21] A. Kumcu, Linguistic synesthesia in Turkish: A corpus-based study of crossmodal directionality, *Metaphor and Symbol*, 2021, 36(4), 241-255.
- [22] O. P. Levchenko, *Symvoly u frazeolohichnykh systemakh ukrainskoi ta rosiiskoi mov: linhvokulturolohichnyi aspekt*, Dysertatsiya doktora filolohichnykh nauk, Instytut movoznavstva im. O.O. Potebni NAN Ukrainy, Lviv, 2007. [O. P. Levchenko, Symbols in the phraseological systems of the Ukrainian and Russian languages: linguistic and cultural aspect, Doctor of Science thesis in philology, O.O. Potebnia Institute of Linguistics of the National Academy of Sciences of Ukraine, Lviv, 2007.]
- [23] O. Levchenko, M. Dilai, Key Colour Terms in the Ukrainian Prose Fiction of the 21st Century, in: *CEUR Workshop Proceedings*, Vol. 1, Gliwice, Poland, 2022, pp. 49-60.
- [24] O. Levchenko, V. Holtvian, M. Dilai, Statistical profiles of Ukrainian prose fiction: Gender aspect, in: *2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)*, IEEE, 2021, pp. 97-100.
- [25] O. Levchenko, N. Romanyshyn, Modern approaches to automated identification of metaphor, *Visnyk of the Lviv University. Series Philology*, Vol. 70, 2019, pp. 288–298.
- [26] N. Lototska, O. Saban, R. Ivanychuk's Idiolect: quantitative parameterization of the language used in the text, in: *2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT)*, IEEE, 2023, pp. 1-4.
- [27] H. K. Mohajan, Quantitative research: A successful investigation in natural and social sciences, *Journal of Economic Development, Environment and People*, 2020, 9(4), pp. 50-79.
- [28] R. Muchembled, *Smells: a cultural history of odours in early modern times*, John Wiley & Sons, 2020.
- [29] A. Pathak, G.A. Calvert, K. Motoki, J. Park, How early acquired phonemes present in words (or brand names) can evoke the expectations of sweet tastes, *Food Quality and Preference*, 2022, 96, 104392.
- [30] V. Seretan, *Syntax-based collocation extraction*, Series: Text Speech and Language Technology, Vol. 44, Dordrecht: Springer, 2011. [24+8]
- [31] M. Shvedova, R. von Waldenfels, S. Yarygin, A. Rysin, V. Starko, M. Woźniak, M. Kruk et al. GRAC: General Regionally Annotated Corpus of Ukrainian. URL: <http://uacorpus.org/>
- [32] M. Stubbs, Collocations and semantic profiles: On the cause of the trouble with quantitative studies, *Functions of Language*, Vol. 1, 1995, pp. 23–55.

- [33] B. Stuckey, *Taste: surprising stories and science about why food tastes good*, Simon and Schuster, 2012.
- [34] F. Tripodi, Ms. Categorized: Gender, notability, and inequality on Wikipedia, *New media & society*, 2023, 25(7), pp. 1687-1707.
- [35] 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, Pivdennoukrainskyi natsionalnyi pedahohichnyi 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].
- [36] L. Vercelloni, *The invention of taste: a cultural account of desire, delight and disgust in fashion, food and art*, Bloomsbury publishing, 2017.
- [37] J.-P. Willem, *Alzheimer's, Aromatherapy, and the Sense of Smell: Essential Oils to Prevent Cognitive Loss and Restore Memory*, Simon and Schuster, 2022.
- [38] B. Winter, *Statistics for linguists: An introduction using R*, Routledge, 2019.
- [39] Wyatt, J. Zakkou, D. Zeman, *Perspectives on Taste: Aesthetics, Language, Metaphysics, and Experimental Philosophy*, Routledge, 2022.
- [40] Y. Zhong, C. R. Huang, *Sweetness or Mouthfeel: A corpus-based study of the conceptualization of taste*, *Linguistic Research*, 2020, 37(3), 359-387.
- [41] Y. Zhou, C. S. Tse, *The taste of emotion: Metaphoric association between taste words and emotion/emotion-laden words*, *Frontiers in psychology*, 2020, 11, 526623.