УДК 81'271.12
DOI https://doi.org/10.32838/2663-6069/2020.2-1/22

Kalymon Yu. O.
Lviv State University of Physical Culture named after Ivan Boberskyj

NARRATIVE AND LANGUAGE TYPES OF SHORT STORIES BY VASYL STEFANYK: QUANTITATIVE APPROACH

The article is devoted to the quantitative analysis of Vasyl Stefanyk's short stories, namely its dialectal part as well as different types of narrative (author's and character's speech patterns). The research aims at expanding the methods of studying the vocabulary of a certain author by means of computer technologies and statistical tools, which will complement traditional methods and clarify the existing statements regarding the work of Vasyl Stefanyk. The corpus of his works (short stories) was created to conduct quantitative studies. By linguistic corpus, which is the object of corpus linguistics, we mean a large, unified, structured, philologically competent mass of linguistic data submitted electronically and intended for various linguistic studies. The main idea is that corpora is more an auxiliary tool and facilitate the process of triangulation of linguistic phenomena, that is their study through several approaches and techniques, which helps to refine or re-examine the object of study. In general, our research may contribute to an in-depth study of the Western variant of the Ukrainian language, which existed at the turn of the nineteenth and twentieth centuries. Such kind of research has complemented from a corpus of the short stories structurally annotated according to the generally acknowledged principles. The first part of this article is about discussing the features of dialectal language on the scale of all short stories' language. Combining this information with the data about different types of narrative (author's and character's speech) contribute to defining peculiar features of short stories by Vasyl Stefanyk and his idiolect in general. Further research may lie in analyzing similar features among other writers and their comparison with those of Vasyl Stefanyk.

Key words: *corpus linguistics, short stories, narrative, quantitative linguistics, dialectal language, national language, direct speech.*

Introduction. Nowadays researchers are focusing on quantitative and qualitative approaches for analyzing a writer's literary heritage. Quantitative and statistical methods, which are gaining popularity with the development of computer technologies, corpora, and software capable of processing large amounts of data and analyzing them according to the researcher's requests, are considered to be the latest methods of investigating author's style features. Such research is more representative and effective due to quantitative and concordance methods, especially in the field of stylometry, for language expression is determined not only by the author, but also by a number of other factors (text type, addressee and circumstances), so it can be very diverse [16; 18, p. 27–30; 19, p. 255–264; 22, p. 259–267; 24, p. 212–241].

The quantitative approach supposes the analysis of linguistic units at different levels (phonemes, morphological, lexical and syntactic elements), their frequency and distribution in the text, while qualitative approach establishes patterns of functioning of these elements of linguistic structure in context. Since the middle of the last century, there has been a need for a comprehensive approach to the analysis of a work of art that involves the merging of linguistic, mathematical and literary methods [23, p. 212–241].

The object of our research is to show the application of these methods in analysis of Vasyl Stefanyk's idiolect. **The subject** of our article is the author's short stories and their narrative types. The main **purpose** of this paper is to show how different types of narrative are represent within the author's short stories. The tasks are the following: to describe different approaches to studying types of narration; to use corpus linguistics methods and quantitative analysis to analyze short stories by Vasyl Stefanyk; to interpret the obtained results as features of short stories.

Methodology. According to David L. Hoover in [17, p. 17] there are two main modes of style variation analysis – macro- and microanalysis, the latter being the most suitable one for studying any author's idiolect or even within one literary work "as a way of better understanding authorship and style". Nowadays corpora of texts as well as statistical tools play vital roles in such kind of research. Among Slavic nations the greatest projects have been dedicated to the style of Karel Chapek in the Czech Republic [15] and Fiodor Dostoievskyi in Russia [13]. In Poland similar approach

has been adopted thoroughly by Jan Rybicki for Henryk Sienkiewicz's trilogy [21]. In Ukraine Solomiia Buk has applied such type of analysis to Ivan Franko's six novels [1]. As it can be seen from the above such researches are very few in Ukrainian studies but are highly needed. Our aim is to try out this method on Vasyl Stefanyk's short stories as one of the aspects of studying his idiolect.

Results and Discussion. The research is based on the corpus of the author's short stories published during his lifetime. Most of them were taken from the anniversary issue published in 1933 and three short stories taken from the last lifetime publications (Stefanyk). As a result, we have obtained a corpus of 57 short stories that were normalized according to general rules as described by Ihor Kulchytskyi [5]. According to the common definition of text corpus and its main features [2; 14; 20] ours has acquired the following characteristics:

1) creation and purpose: referential (information about the language in a certain time period); synchronous (representing the language of a certain time interval); static (fixes the status of the language at a certain time period); special (representing a particular language period, a fragment of a language aimed at solving partial, specific scientific research tasks);

2) by type of text material: full text (contains full texts of short stories by Vasyl Stefanyk); written (the constituent material of the corpus is only printed texts); monolingual;

3) by type of software processing: partially annotated (contains structural, morphological and extralinguistic annotation).

As for structural annotation it includes the division of text into chapters, sections, paragraphs, sentences, etc., with the special status of headings, notes, references, etc., depending on the genre-stylistic features of the text. In general, the structural composition of a text is considered an element of its architectonics and, above all, contains a division into paragraphs, sentences, words [12]. The text of the short stories itself is further annotated at the sentence level that we mean as "a syntactic unit constructed on a fixed pattern, expressing a relatively complete thought or a certain emotion and acting as the primary means of communication" [3]. Capital letter at the beginning of the sentence and one of these characters at the end of the sentence (/? /! /… / !!! / ??? /?!) we consider as the beginning and the ending of a sentence, respectively.

The language of Vasyl Stefanyk's stories has caused a great deal of debate and controversy whether it is oversaturated with dialectal linguistic units that impede the perception of text by a reader unfamiliar with such vocabulary [4; 6; 7]. Based on the case, the quantitative distribution (absolute and relative) of the writer's general and dialectal vocabulary units for each novel have been calculated. A generalized diagram of the correlation of these types of vocabulary across all texts is provided below (Diagram 1).

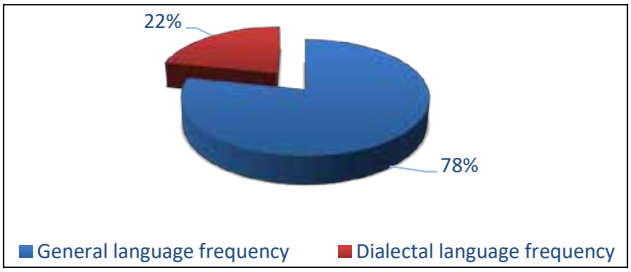

Diagram 1

As one can see, in general dialect vocabulary comprises less than a quarter of all word forms.

The structural annotation of the corpus also contains distinction between the author's speech and direct speech of the characters. These indicators for all the short stories are summarized on a diagram below (Diagram 2).

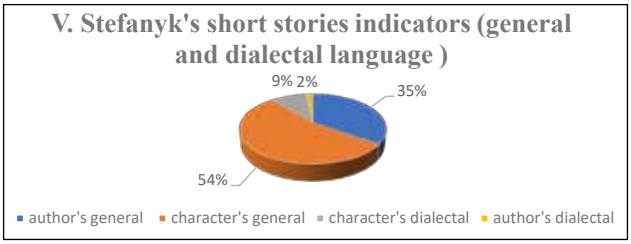

Diagram 2

The data obtained indicate that in the author's language the dialectal vocabulary is 2% and the direct language of the characters – 9%. The general vocabulary ratio in these types of narratives is 35% and 54%, respectively. Thus, the language of characters in Vasyl Stefanyk's short stories contains a large proportion of dialectal elements, which confirms the fact that it was in the direct speech of the characters that the author used the language that was inherent to them.

This data also allows us to analyze the ratio of direct and indirect speech to total language (Diagram 3).

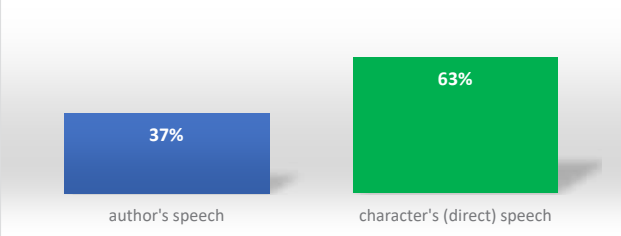

Diagram 3. The ratio of the character's and the author's language in Vasyl Stefanyk's short stories

As one may see the former indicator contains 63% and the latter 37% of the whole text respectively, which indicates a high degree of direct speech that might be explained by the genre features the author used throughout his life.

Conclusions. Although such calculations provide useful information about Vasyl Stefanyk's idiolect, it would be even informative to compare them with the same data obtained from other authors, including his contemporaries. As a result, it may help to underline peculiarities of the Western variant of the Ukrainian national language that was prevailing on the territory of Western Ukraine at the end of the XIX – beginning of the XX centuries.

References:

1. Buk C. Пряма й авторська мова великої прози Івана Франка: лінгвостатистичне дослідження у контексті корпусної лінгвістики. *Вісник Львівського університету. (Серія філологічна, вип. 52 : Мовознавство)*. 2011. С. 199–209.
2. Демська О. М. Текстовий корпус: ідея іншої форми. Київ : ВПЦ НаУКМА. 2011. 282 с.
3. Загнітко А. П. Український синтаксис: навч.-практ. комплекс: [в 2 ч.]. Донецьк-Слов'янськ : ДонНУ. 2010. 651 с.
4. Кобилянський Б. В. Діалект і літературна мова. К., 1960. 276 с.
5. Кульчицький І. М. Технічні аспекти опрацювання комп'ютером природномовних текстів. *Вісник Національного університету Львівська політехніка. Інформаційні системи та мережі*. Видавництво Львівської політехніки. Львів. 2014. С. 344-353.
6. Матвіяс І. Взаємодія діалектної лексики з лексикою української літературної мови. *Мовознавство (№ 2)*. 1985. С. 37-43.
7. Матвіяс І. Г. Відображення особливостей говорів у мові української художньої літератури. *Мовознавство (№ 6)*. 2008. С. 3–13.
8. Стефаник В. Межа. *Літературно-науковий вісник, т. 92, кн. 2*. Львів. 1927. С. 97–98.
9. Стефаник В. Твори (передмова В. Коряка. До друку виготував Ів. Лизанівський. 3-є вид.). ДВУ. 1929. С. 94–95.
10. Стефаник В. Твори (з дереворитами В. Касіяна і М. Бутовича). Львів : З друкарні Видавничої Спілки «Діло». 1933. 222 с.
11. Стефаник В. Шкільник. *«Рідна школа» №1*. Львів, 1932. С. 2–4.
12. Тишківська Н. Архітектоніка новел В. Стефаника. *Василь Стефаник і українська культура (тези). Ч. ІІ. Секція ІІІ. Мова і стиль творів В. Стефаника*. Івано-Франківськ, 1991. С. 71–73.
13. Шайкевич А. Я., Андрющенко В. М., Ребецкая Н. А. Статистический словарь Ф. М. Достоевского. Москва : 2002. URL: http://cfrl.ruslang.ru/dost_cd0/dostoevski.htm (дата звернення: 25.02.2019).
14. Широков В. Корпусна лінгвістика. НАН України, УМІФ. Київ : Довіра. 2005. 472 с.
15. Čermák František (ed.): Slovník Karla Čapka. Praha : Nakladatelství Lidové noviny. Ústav Českého národního korpusu. 2007. 714 p.
16. Hoover D. Quantitative Analysis and Literary Studies. In: A Companion to Digital Literary Studies. URL: http://digitalhumanities.org/companion/view?docId=blackwell/9781405148641/9781405148641.xml&chunk.id=ss1-6-9&toc.id=0&brand=9781405148641_brand (дата звернення: 24.03.2019).
17. Hoover D. The microanalysis of style variation. Digital Scholarship in the Humanities, *Volume 32 (2).* 2017. P. 17–30.
18. Jones S. When Computers Read: Literary Analysis And Digital Technology. *Bulletin of the American Society for Information Science and Technology. Vol. 38.* 2012. P. 27–30.
19. Kešelj V., Peng F., Cercone N., Thomas C. N-gram based author profiles for authorship attribution. *Proceedings of the conference Pacific Association for Computational Linguistics, PACLING. Vol. 3.* 2008. p. 255–264.
20. MacEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. Cambridge University Press. 2012. 294 p.
21. Rybicki J. Policzmy Trylogię Sienkiewicza. *Annales Universitatis Paedagogicae Cracoviensis.* 2010. P. 85–109.
22. Siemens R. A New Computer-assisted Literary Criticism? *Computers and the Humanities*. 2002. p. 259–267.
23. Vaughan E., O'Keeffe A. Corpus Analysis. The International Encyclopedia of Language and Social Interaction. 2015. P. 1–17.
24. Wright D. Using word n-grams to identify authors and idiolects: A corpus approach to a forensic linguistic problem. *International Journal of Corpus Linguistics Vol. 22(2).* 2017. p. 212–241.

Калимон Ю. О. ТИПИ НАРАТИВУ ТА ЛЕКСИКИ У НОВЕЛАХ ВАСИЛЯ СТЕФАНИКА: КІЛЬКІСНИЙ ПІДХІД

Стаття присвячена кількісному аналізу новел Василя Стефаника, а саме його діалектному складнику, а також різним типам оповіді (авторське мовлення та мова персонажів). Такого типу дослідження поодинокі в україністиці. Метою дослідження є розширення методів вивчення ідіолекту певного автора за допомогою комп'ютерних технологій та статистичних засобів, що доповнить традиційні методи та уточнить наявні твердження щодо творчості Василя Стефаника. Корпус його творів (новел) створений для проведення кількісних досліджень. Під лінгвістичним корпусом, що є об'єктом корпусної лінгвістики, ми маємо на увазі велику, єдину, структуровану, філологічно компетентну масу мовних даних, що подаються в електронному вигляді та призначені для різних лінгвістичних досліджень. Основна ідея полягає в тому, що корпуси є допоміжним інструментом і полегшують процес тріангуляції мовних явищ, тобто їх вивчення за допомогою декількох підходів і прийомів, що допомагає уточнити або поглянути по-новому на предмет дослідження. Загалом наші дослідження можуть сприяти поглибленому вивченню західного варіанту української мови, що існував на межі XIX–XX ст. Такий тип дослідження доповнив наш корпус новел, структурно анотованих за загальновизнаними принципами. Перша частина цієї статті стосується обговорення особливостей діалектної мови в масштабі мови всіх новел. Поєднання цієї інформації з даними про різні типи оповіді (авторська та персонажів) сприяє визначенню особливостей новел Василя Стефаника та його ідіолекта загалом. Подальше дослідження полягає у переспективі аналізу таких показників серед інших письменників та їх порівнянні із Василем Стефаником.

Ключові слова: *корпусна лінгвістика, наратив, розповідь, кількісна лінгвістика, діалектна мова, національна мова, мова автора, мова персонажа.*