Oksana MYKYTYUK
PhD in Philology, Associate Professor,
Lviv Polytechnic National University (Ukraine)
https://orcid.org/0000-0003-3922-4610

Liliia KHARCHUK
PhD in Philology, Associate Professor,
Lviv Polytechnic National University (Ukraine)
https://orcid.org/0000-0003-0063-1956

DMYTRO DONTSOV'S TEXTS FROM THE ASPECT OF CORPUS LINGUISTICS AND LEXICOGRAPHY

Незважаючи на велику кількість лінгвістичних напрацювань, що стосуються творчости Дмитра Донцова, все ж таки є низка мовознавчих питань, які на сьогодні ще не були вивчені. Серед них – створення лінгвістичного корпусу мови та конкордансу цього видатного мислителя, громадського діяча, публіциста у світлі новітніх методик – обчисленні та фіксації лексем з метою виявлення їхньої атрибутивної сполучуваности. З'ясовано, що будь-які наявні засоби програмування, які полегшують опрацювання великого обширу мовного матеріалу, є безсумнівно продуктивним способом здійснення власне концептуального аналізу і слугують додатковим інструментом дослідження. *Мета* дослідження – визначити особливості функціювання сентенцій Дмитра Донцова в аспекті корпусної лінгвістики і лексикографії на основі корпусного простору, що формує лінгвістичне бачення мовної картини світу та є джерелом для творення лексикографічних праць (конкордансів, мови творів, мови письменника тощо). Наголошуємо, що створення конкордації творів Донцова вперше стала предметом розгляду в корпусній лінгвістиці. Об'єктом вивчення є твори Дмитра Донцова, їхня лексикографічна параметризація, що подає усі можливі слова з описом (фонетичним, словотвірним, граматичним), з кількісними показниками, тобто це результат вивчення багатьох мовознавчих дисциплін, об'єднаних словником. Методологією дослідження є поєднання загальнотеоретичних *методів* (аналізу, узагальнення, пояснення) з прикладними методами лінгвістики. Аналіз досліджуваного матеріалу ґрунтується на основі створеного текстового корпусу «Дух нашої давнини» зі застосуванням програми Sketch Engine. На підставі проведеного аналізу встановлено, що електронний корпус дає можливість прискорити дослідження мови та багатократно підвищити їхню ефективність, вірогідність та перевірюваність. У статті розкрито евристичний потенціал, практичну ефективність корпусу і застосування конкордантних технологій у концептуальних дослідженнях. З'ясовано, що побудова повного конкордансу творів Дмитра Донцова дасть змогу показати картину світу на основі вивчення лексичного багатства автора та відтворити його розуміння політичної ситуації. Створений конкорданс є етапом для формування лексикографічних праць про Донцова, та дає розуміння етапів, методики, принципів та специфіки укладання словника мови творів Дмитра Донцова.

Ключові слова: корпусна лінгвістика, Sketch Engine, конкорданс, текст, мовна картина світу, Дмитро Донцов, лексикографічна практика

Introduction

In the context of an ever-increasing flow of information in everyday life, the challenge of automated processing of natural-sounding speech information becomes increasingly important. Increasingly, computer technologies in linguistics offer their advanced tools to process, store and select language data, which has triggered the fast development of the actual branch of linguistic studies – corpus linguistics. Through the use of large-scale empirical data and advanced computer technologies to reach objective insights into language function, linguistic corpora have quickly become invaluable resources. Data obtained through corpus analysis facilitate the drawing of qualitatively new conclusions about language and highlight research directions that previously received little attention.

The concept of "text corpus" is determinative for corpus linguistics, since it becomes the main aim corpus linguistics sets for itself, consequently, it becomes the object to study for this linguistic branch. Uniqueness of the linguistic text corpus is based on the fact that it is a voluminous collection of electronic texts, which can be used in a wide range of studies, since quite a large (representative) volume of the corpus guarantees that data are typical and secures completeness of representation of the whole spectrum of linguistic phenomena; data of different types remain in the corpus in their natural contextual form which provides the opportunity for their comprehensive and objective study; compiled and prepared array of data can be used repeatedly, by many researchers in different fields of knowledge and for different purposes.

Supplementation of the semantic word structure is changeable in a historical layer, so, nowadays, the necessity of producing dictionaries of writers' languages (or politicians' ones), the frequency lists, or any others to show the way of language supplementation with new words and new shades of words is extremely actual. As a result, online-concordances, which show lexical items as they appear in authors' texts, highlight the accents in the offset of a lexical norm and provide us with understanding the lexical-grammatical system of the Ukrainian language, are important.

Contemporary corpus linguistics encompasses various corpora of different types and sizes. This diversity reflects the wide range of research and practical applications, as well as the peculiarities of the linguistic material they represent.

The analysis of the latest studies and publications where the solving of the problem was initiated

The study based on data instead of mechanical text analysis is well-established in contemporary linguistics. At present, quite a large number of corpora are available – both unprocessed and linguistically annotated – supplying data that are current and essential for scientific investigations. Researchers mention that, in most cases, corpora are large, ranged from a few million to a few billion words, which is obviously unsuitable for study by the "word-by-word" method of careful reading. In general, there are two ways to obtain data from the text corpora: through the query interface or immediately by means of automated text processing. The studies [Jakubíček et al., 2020; Kilgarriff et al., 2014; Kunilovskaya, 2017; Martín, Trekker, 2021; Sass, 2022] present in detail the principles of how to reliably and effectively collect linguistic data out of corpora by means of queries, i.e., without the need for programming knowledge to manipulate data directly; they discuss what is worth consideration, which tools to use, what to do by default and how to solve problematic cases. As a result, the methodology for obtaining accurate and comprehensive data from corpora is crucial for conducting the linguistic studies.

To study authors' idiolects in writers' and statistical lexicography, linguists employ quantitative descriptions of writers' works, which provide precise information about the peculiarities of vocabulary functioning in their writings. A clear tendency towards using text corpora to compile authors' language dictionaries and frequency lists is traced [Buk, 2011].

The development of software for concordances has repeatedly been the subject of examination by both Ukrainian and foreign scientists, as it is an effective way to automate the work of researchers [Bouso, San Segundo, 2021; O'Keeffe, McCarthy, 2010; Simpson, 2011; Üveges, 2022; Villayandre Llamazares, 2008]. Currently, a range of corpora for the Ukrainian and Slavic languages has been produced (Czech National Corpus [Křen, 2024]; KTUM Corpus of Ukrainian Texts [Darchuk, 2024]; Laboratory of Ukrainian [Moskalevskyi, 2024]; Brown Corpus of Ukrainian [Kučera, Nelson Francis, 2024]; NER annotation corpus [Чаплинський, Дьомкін, 2024]; Corpus of the Chtyvo library [Chtyvo, 2024]; UA-GEC [Nahorna, 2024]; General Regionally Annotated Corpus of Ukrainian [Shvedova et al.,

2017–2022]; National Corpus of Polish [Wójtowicz, 2024]; Slovak National Corpus [Levická, 2024]; Bulgarian National Corpus [Koeva, 2024]) along with authors' concordances (Onlain-konkordans romanu Ivana Franka "Perekhresni stezhky" [Бук, Ровенчак, 2010]; Online Concordance to the Complete Works of Hryhorii Skovoroda [Pylypiuk et al., 2008]) of varying levels of detail, composition, volume and access.

Like most scientists, we believe that forming concordances is extremely important, as serves as a specific type dictionary, namely: the list of all the examples of an entry (a word or a phrase) found in a certain corpus is supplied with information about a source (metadata) [Shvedova et al., 2017–2022].

The necessity to compile the concordance of the texts of Dmytro Dontsov, study actuality argumentation

Despite substantial linguistic work on Dmytro Dontsov's writings, there still exist a number of questions not studied so far. Among these are compiling a linguistic language corpus as well as a concordance of this prominent thinker, public figure, publicist utilizing modern methods to calculate and fix lexemes in order to identify their attributive collocation. Currently, linguistics shows an increased interest in studying the interconnection between the world, the man (the nation) and the language, which predetermines a certain aspect of a fiction speech analysis. Given current realities, it is essential to study Dmytro Dontsov's writing, taking into account the peculiarities of national-language and personal author's pictures of the world, representing the category of the lexical image.

As mentioned, electronic text corpora now allow researchers to view texts or text fragments within the context of a particular author's language. The structure of any text is hierarchical and reflects those lexical items which reflect the author's understanding of reality. Thus, we have an object (texts) to model and a number of platforms allow researchers to design text corpora that facilitate high-quality, efficient studies. The compilation of all available texts in a corpus and capabilities of the modern applied linguistics have to unite the intellectual, political, spiritual, cultural, social life of a man and to meet the values an author proclaims.

The purpose of the study is to determine the peculiarities of the unique features of Dmytro Dontsov's maxims in the aspect of corpus linguistics and lexicography on the basis of the corpus space that forms the linguistic vision of the world and is a source of creating lexicographical pieces of work (concordances, a writing language, a writer language, etc.). We emphasize that the creation of Dmytro Dontsov's writing concordance has become the subject of corpus linguistics study for the first time.

The subject of study is the writings of Dmytro Dontsov, who is known as a thinker, politician, ideologist of Ukrainian nationalism, critic, media figure and magazine editor. It is important that the atmosphere the future thinker used to live in was intellectual, whilst Dontsov's language environment was not homogeneous – a mix of different languages. Dontsov's multilingual environment extended across different stages of his life; he studied in Petersburg (the Russian language), then – lifetime in Lviv (the Ukrainian language), from 1909 until 1911 a free student of Vienna University (thus the language was German), another lifetime in Lviv where he was issuing magazines (Ukrainian environment). After onset of World War II, Dontsov resided in multiple countries, including Romania, Germany, Riga, Paris, Great Britain, the USA, but from 1947 and up to his death he was living in Canada.

The political life of Dmytro Dontsov is important since he actively entered the party life (in student years he was a member of the Ukrainian Social-Democratic Party, was two times arrested for his activism. Later, in 1914, he became a member of the Ukrainian Liberation Union (ULU), sided with UPR), as well as his literature life (for two last years he had been lecturing on literature in Montreal University). Moreover, he wrote in different languages (Ukrainian, Polish, German, Franch, Italian, Russian), so the concordance of his writings can be compiled on the basis of different languages.

Studying Dontsov's language and conducting a quantitative analysis of his word-usage are active areas of contemporary linguistics, related to the anthropo-centrism of modern science as well as to interdisciplinary studies, carefully examining language objects.

Thus, the subject of our study is – lexicographical parametrization which provides all possible words with their description (phonetic, word-creating, grammatical), along with quantitative indices, i.e. it is the result of learning many linguistic disciplines, unified by the dictionary. Therefore, the purpose of our study is to define the lexicographical parameters of Dmytro Dontsov's texts, using the newly created corpus of "The Spirit of Our Antiquity" ("Dukh nashoi davnyny").

We define a concordance as the combination of the word-index, writer's language vocabulary, orthographic, word-created, frequency list, unified under a single name (in our case, D. Dontsov). In our view, the application of new technologies will provide the opportunity to find needed keywords, lemmas, quantify the number of their meanings, to realize the amount of word-usage, to study all the phonetic and word-creating forms of a word (with a quantitative characteristic), to analyze the grammatical forms used in the writings of D. Dontsov. However, the frequency list prioritizes the most commonly used words, whilst the concordance usually contains a neutral entry (in order to facilitate the search).

Scholars also suggest dictionaries of integrated type, "whose macrostructure includes an interpretation, a concordance, a frequency list, as well as certain elements like epithets,comparisons, metaphors etc., the use of demonstrative means such as drawings, tables, charts, graphs, diagrams" [Buk, 2013, p. 292].

Thus, compiling a concordance of Dontsov's writings forms a complex hierarchical and meaningly branching system, which has a number of vectors to study.

Concordances enable understanding of the language as building material, where words perform the function similar to erythrocytes, providing the body with oxygen. Compiling an author's concordance not only demonstrates contextual word functioning, but also reveals thought-creating, nation-centric, aesthetic function of a language, i.e. the whole author's laboratory becomes visible on the surface. In the context as a background we can observe the semantic and structural (phonetic-word-creating-grammatical) aspects of a language.

Adaptation of Sketch Engine for the compilation of the concordance of D. Dontsov's writings

At present, there are a great number of programs that allow for text processing, such as Intelligent Miner for text (IBM), TextAnalyst, WebAnalyst, Text Miner, SemioMap, InterMedia Text, Oracle Text, Autonomy Knowledge Server, Galaktika-ZOOM, InfoStream etc.. All of these programs provide reliability and trustworthiness in collecting and processing information.

We propose adapting information to study the semantic-word-creating system of of D. Dontsov writings using Sketch Engine software, which provides a number of search functions.
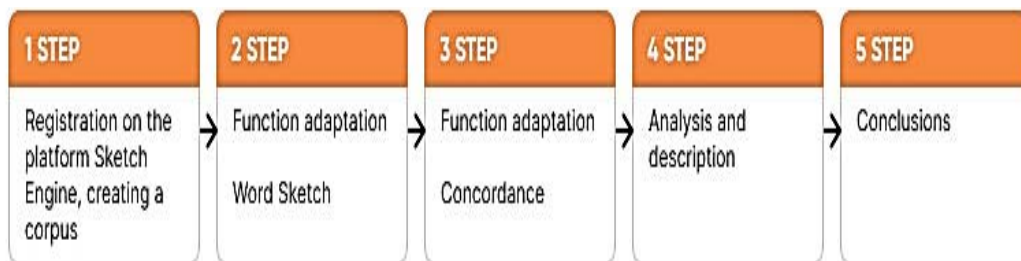

1 STEP	2 STEP	3 STEP	4 STEP	5 STEP
Registration on the platform Sketch Engine, creating a corpus	Function adaptation Word Sketch	Function adaptation Concordance	Analysis and description	Conclusions

Fig. 1. Structural and logical scheme covering the main stages (steps) of the study

The process of text corpus compilation became the first stage of the study. Note that the Sketch Engine program has several method for doing it. We utilized the downloading of a digitized text from a web source. The work of D. Dontsov, "The Spirit of our Antiquity" ("Dukh nashoi davnyny") [Dontsov, 1951], served as the material for the corpus.

It is important to note that the information can be found in both journalistic and literary notexts, as we do not have an opportunity to listen to D. Dontsov's speech. The same applies to the search for concepts. Later, while compiling additional subcorpora, we can group all the texts according to the themes to locate specific concepts. Words that are exceptions, i.e., contextually unclear, dialectal words, jargon words, are presented only in literature. Contiguous fields can be identified in philosophical or political writings. Data visualization is only feasible in politics, as it involves the formation of a certain party with its program, ideas, and principles. Answers to questions and classification of a text are possible across all areas. In the case of epistolary, revealing and searching for the information related to answering a question is considered appropriate. The information about D. Dontsov in the media sphere and in the educational process is also significant. In terms of the media sphere, answers to questions and data visualization are needed; for example, compiling a list

of informational quotes that would meet societal needs. The collection of the aphorisms "Happiness is to be strong. Aphorisms and maxims of Dmytro Dontsov" ("Shchastia – buty sylnym. Aforyzmy i sententsii Dmytra Dontsova") [Микитюк, 2021] could serve as an analogue to such a quote list. The informational component is integrated into the educational process, as the conceptual field allows for matching the concept-sphere of different chronological phases or for determining if the representatives of political parties share similar concepts.
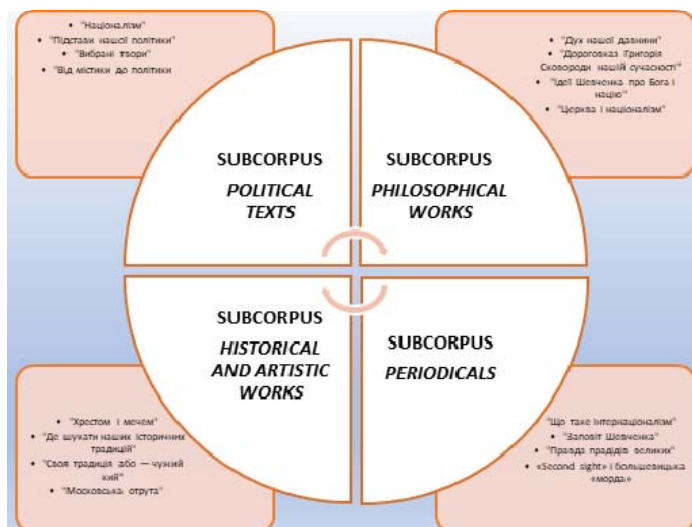

Fig. 2. The project of the compilation of the corpus of the writings of D. Dontsov according to the thematic fillin

It is noteworthy that the Sketch Engine program provides the opportunities to process texts of billions of words in seconds, to find the variants of a word, phrase, or sentence, and to present search results in the form of word sketches, concordances, or word lists. It also facilitates idiomatically accurate combinations of words and enables users to use words as native speakers do. Furthermore, it allows for text analysis, extraction of terms, production of frequency lists, and benefits from morphological analysis and part-of-speech tagging, as well as identifying neologisms and popular words in diachronic corpora, etc.
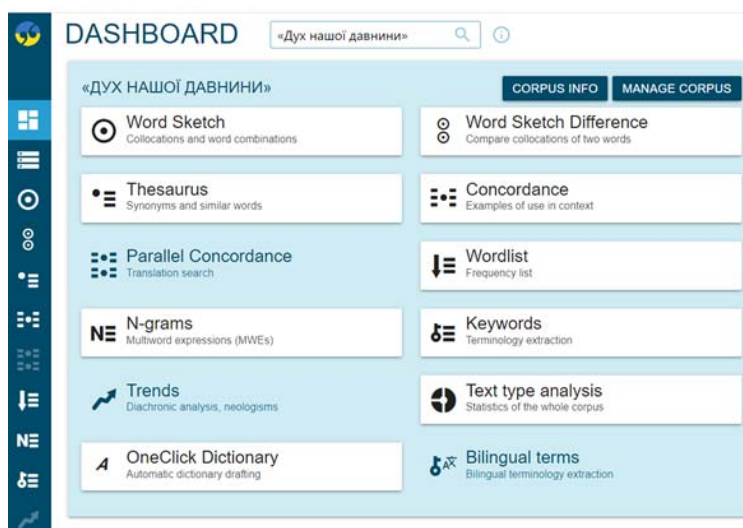

Fig. 3. The tool panel of Sketch Engine

The methodology and methods of study

The methodology of study combines general theoretical methods (analysis, generalization, explanation) with applied methods of linguistics. The scientific study presupposes a range of the following general scientific and linguistic methods:

1) Due to rapid development of computer technologies, new methods of text study, particularly of authors' language, have emerged, including statistical and corpus methods, which have drawn the special attention of linguists [Buk, Rovenchak, 2019; Lin et al., 2019; Pollak, Gantar, Holdt, 2019]. It is recognized that contemporary linguistics widely uses statistical methods, since they comprise a totality of modes and principles, according to which collection, systematization, processing, and interpretation of statistical data are arranged to reach scientific and practical conclusions. Thus, these methods, as purpose-oriented, are considered to be the most effective and valid methods of study in applied linguistics [Hadzalo, 2020].

2) The content-analysis method is a qualitative-quantitative approach to text processing, which includes the quantitative characteristization of the content and interpretation of results. This method involves preparing material and selecting texts (literary, philosophical, political). The material may chosen based on the date or place of writing. Typically, the content-analysis method can be sensibly combined with the distributive analysis of speech phenomena, which allows for the division of linguistic items and understanding of their properties based on their prominence in the speech flow, i.e., taking the context into account. Distributive modeling in lexicography clarifies the functioning of a word and its collocation. This aspect of study is evident among applied linguistics experts [Rovenchak, Buk, 2018].

Approbation of the language corpus of "The spirit of our Antiquity" ("Dukh nashoi davnyny") through the Sketch Engine functions

The Sketch Engine function provided us with the opportunity to find assigned lexemes in a word combination. For example, the lexeme "love" in the corpus of "The Spirit of our Antiquity" ("Dukh nashoi davnyny") is represented in the following results:

The modifiers of *'любов'* (*'love'*): *літепло* (літепла *любов*) / *warmth* (*love's* warmth), *вікінізький* (вікінізьку *любов*) / *vikinian* (vikinian *love*), *звідки* (*звідки* його *любов*) / *from where* (*from where* his *love*), *вона* (її свою *любов*) / *she* (her own *love*), *воно* (його *любов*) / *it* (its own *love*), *свій* (свою *любов*) / *own* (own *love*).

The verbs with *'любов'* (*'love'*) as an object: *ставляти* (ставляли *любов*) / *to set* (set *love*), *заступити* (заступить літепла *любов*) / *to substitute* (substitute *love's* warmth).

The modifiers of the genitive case of *'любов'* (*'love'*): *свобода* (*любов* свободи) / *freedom* (*love* for freedom), *відчизна* (*любов* отчизни) / *homeland* (*love* for homeland).

The prepositional word combinations *'любов'* (*'love'*) to...: *посад* (*любов* до державних посад) / *position* (*love* for state positions), *море* (*любов* до моря) / *sea* (*love* for the sea), *людина* (*любов* до людини) / *human* (*love* for human), *порядок* (*любов* до порядку) / *order* (*love* for order), *річ* (*любов* до речей) / *thing* (*love* for things).
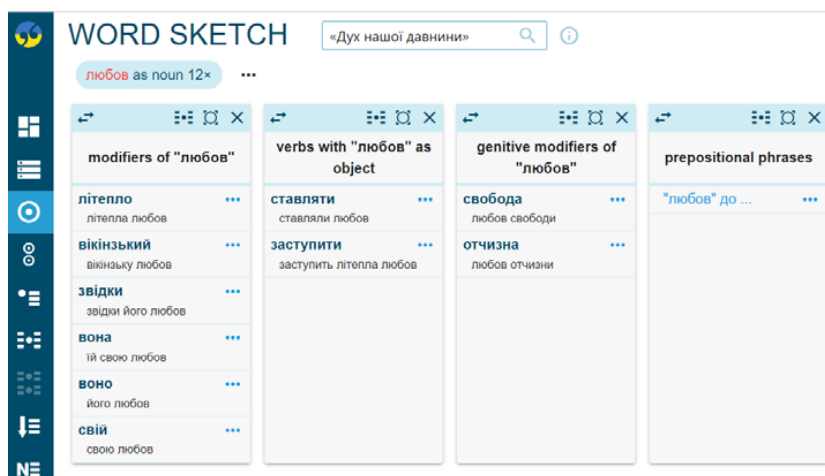

Fig. 4. The results of the Word Sketch function use for the lexeme 'любов / love' as an example

The statistical results of finding the lexeme "*love*" in the corpus text provide the opportunity to make a more detailed description of the lexeme, taking into consideration the meaning filling according to different historical periods. Comparing the interpretation of the noun "*love*" provided by the Ukrainian Dictionary to the meaning characteristic of D. Dontsov, we state the entire mismatch. In UD: 1) sympathy for a person of the other sex, a synonym for love; 2) attachment to someone or something (usually to people); 3) interest in something (literature, horses) [Білодід, 2023].

Based on the solid sample of the material used in "The Spirit of our Antiquity" ("Dukh nashoi davnyny") there were found 33 usages of the root morpheme *люб-*, (*lov*-) which is divided, according to the speech-part principle, into the noun *любов* (*love*) (28 times), the adverb *любо* (*lovely*) (3 times), the verb *любити* (*love*) (37 times), the adjective *любомудрий* (*love wise*) (1time). It becomes evident, that D. Dontsov constantly gives the adverb *любо* (*lovely*) by quoting, for example, the prince Sviatoslav (*"не любо ми єсть в Києві жити"* / *"it's not lovely that I live in Kyïv"*), Hypatian chronicle dated back to 1097, where the prince Vasylko said: *"Любо налізу собі славу, любо главо свою схожу за руською землю"* (*"Lovely to gain glory for oneself, lovely to lead one's head for the Rus' land"*).

D. Dontsov uses the noun *любов* in following meanings: a feeling, which means attachment to motherland, nation, homeland, ideals, freedom, faith, bravery, pride, domination, fame, i.e. a different semantic shade than the second meaning in UD [Білодід, 2023]. For example: *честь* (*honor*) – means "love to the things which are high and abstract, compared to them, everything premature and material, not excluding one's own life and the lives of relatives, is worth nothing"; 2) along with the meaning of "*почуття до людини*" (*feeling to a man*): "love to a man with all their weaknesses, so called humanity"; 3) the type of feeling of a certain race (Nordic, Eastern, Mediterranean etc.).

As to the first meaning (11 lexemes), we have the following word combinations: *любов слави* (*love of glory*), *любов отчизни* (*love of the homeland*), *слави* (*glory*), *пошана предків* (*respect for ancestors*), *любов речей великих* (*love for great things*), *любов до речей високих і абстрактних* (*love for lofty and abstract things*), *млявою* (*feeble*), *безсилою любов'ю* (*powerless love*), *любов'ю до свободи* (*love of freedom*), *взиваючи до любови* (*calling to love*), *момент любови до таких абстрактних ідеалів щастя* (*a moment of love for such abstract ideals of happiness*), *любов до віри* (*love of faith*) (totally 11).

In the second meaning – 3 word-usages: *любов до людини* (*love for human*), *любов до людей* (*love for people*). As to the third meaning (4 word-usages), then as to love of Eastern, Dinaric, and Nordic, we have the following word combinations: *любов до державних посад остійця* (*love for state positions of an Eastern*), *любов або ненависть динарця* (*love or hate of a Dinaric*), *любов до порядку нордійця* (*love of order of a Nordic*), *любов до моря нордійця* (*love of the sea of a Nordic*). We particularly have a compound adjective (1 word-usage) to be *любомудрим* (*love wise*) and a verb *любити ідеал* (*love the ideal*) (totally 9).

In particular, it is worth quoting D. Dontsov, whose keyword is *любов* (*love*): *"любовю до рідної землі і ворогованням з поганими"* / *"Love for the native land and hostility towards the wicked"* (the quote by Maksymovych); *"Пес стереже отару день і ніч з вродженої любови"* / *"The dog guards the flock day and night out of innate love"* (the quote by Skovoroda); *"навіть нужда, хула, переслідування – любови тої у нього вгасити не зможуть"* / *"even need, mockery, persecution – cannot extinguish that love in him"* (the quote by Skovoroda); *"любов к отчизні, де героїть"* / *"love for the homeland, where the heroism is"* (used two times) (the quote by Kotliarevsky); *"слуги своя кріпостю і любов'ю "держиться"*/ *"the servants are held by servitude and love"*, *"ніколи самою любов'ю"* / *"never by love itself"* (the quote by Joaniki Galiatovski); *"за честь любої землі"* / *"for the honor of lovely land"* (the quote by Shevchenko); *"Любо налізу собі славу, любо главо свою схожу за руську землю"* / *"Lovely to gain glory for oneself, lovely to lead one's head for the Rus' land"* (the quote from Hypatian chronicle); *"любов'ю к нації"* / *"love for the nation"* (the quote from the), *"любов свободи"* / *"love for freedom"* (the quote by st.Augustine); *"не любо ми єсть в Києві жити"* / *"it's not lovely that I live in Kyïv"* (the quote from the writing monument, Sviatoslav's words) (13 quotes totally).

During the analysis of the compiled corpus "The Spirit of our Antiquity " ("Dukh nashoi davnyny"), there were no revealed phonetic variants regarding the word usage of the lexical item **люб- (lov),** it becomes evident however, that the -*И* ending is the morphological indicator everywhere in genitive singular: *любови речей великих, вживаючи до любови* (*love for great things, applying it to love*).

Regarding the orthographic variants, we have *любов'ю (love's)* written through an apostrophe and ***любовю (love)*** (without an apostrophe), which is related to the beyond-language factors – obviously, the type of publishing was without a corrector.

Thus, the semantic field of the texts of D. Dontsov concerns, first of all, the political sphere (love to fatherland, nation, fame).

Accumulation of data enables systematization of scientific studies, it is the demonstration of the creation of a linguistic norm, for example, the writing "The Spirit of our Antiquity " ("Dukh nashoi davnyny"), published by the supporters of the author in 1951, fixes the name *Драгоманів*, meanwhile in other editions (2011) it is changed to *Драгоманов.* The system of matching such facts, their bringing together, classification will allow to match a range of facts or to differentiate them.

The lexeme **'нація'** is found in the text 108 times and represented by the following results in detail:

modifiers of 'нація': *український* (українській нації) / *Ukrainian* (Ukrainian nation), *місія* (місії нації) / *mission* (mission of the nation), *вивести* (виведе націю) / *lead* (will lead the nation), *пануючий* (панує нації) / *ruling* (rules the nation), *панівний* (панівної нації) / *predominant* (predominant nation), *цілий* (цілої нації) / *whole* (whole nation), *окремий* (окрема нація) / *separate* (separate nation), *того* (того нацією) / *of that* (of that nation), *історичний* (історичної місії) / *historical* (historical mission), *кожний* (кожної нації) / *each* (each nation), *власний* (власної нації) / *own* (own nation), *собі* (собі нація) / *to itself* (to itself nation), *самий* (самої нації) (same (same nation)), *коли* (коли нація) (when (when the nation)), *той* (тої самої нації) / *that* (of that same nation), *вони* (ними нація) / *they* (by them nation), *який* (якими нація) / *which* (by which nation), *наш* (нашої нації) / *our* (our nation), *свій* (свою націю) / *its own* (its own nation);

nouns modified by 'нація': *призначення нації (the destination of the nation)*, *політичне положення нації* (*the political position of the nation*), *існування нації* (*the existence of the nation*), *тілі нації* (*the body of the nation*), *справа нації* (*the affairs of the nation*), *любов'ю к нації* (*love for the nation*), *страждання нації* (*the suffering of the nation*), *середині нації* (*within the nation*), *будівничі нації* (*builders of the nation*), *знищенням нації* (*destruction of the nation*), *служення нації* (*service to the nation*), *колективу нації* (*collective of the nation*), *збереження нації* (*preservation of the nation*), *племя пануючої нації* (*the tribe of the ruling nation*), *української нації* (*Ukrainian nation*), *розумінням нації* (*understanding of the nation*), *чолі нації* (*head of the nation*), *кара нації* (*punishment of the nation*), *устрою нації* (*the structure of the nation*), *поняття нації* (*the concept of the nation*), *цілі нації* (*goals of the nation*), *провід нації* (*the leadership of the nation*), *правила нації* (*rules of the nation*), *обличчя нації* (*the face of the nation*), *членів нації* (*members of the nation*), *частин нації* (*parts of the nation*), *інтереси цілої нації* (*interests of the entire nation*), *ідеї панівної нації* (*ideas of the predominant nation*), *організм нації* (*organism of the nation*), *ідеалами нації* (*ideals of the nation*), *касти нації* (*caste of the nation*);

verbs with 'нація' as object: *сцементувати* (сцементувала націю) / to *cement* (cemented the nation), *завдячувати* (завдячувала нація) / *to be grateful* (the nation is grateful), *очолювати* (очолювати націю) / *to lead* (to lead the nation), *презентувати* (презентувати націю) / *to present* (to present the nation), *провадити* (провадили націю) / *to conduct* (to conduct the nation), *об'єднати* (об'єднає націю) / *to unite* (to unite the nation), *жити* (живе нація) / *to live* (the nation lives), *бути* (є нація) / *to be* (there is a nation);

genitive modifiers of 'нація': *плем'я* (нації плем'я) / *tribe* (the tribe of the nation); **prepositional phrases:** *... в "нація"* (*in "nation"*), *"нація" в...* (*"nation" in..*), *"нація" на...* (*"nation" on...*); *... на "нація"* (*...on "nation"*), *... з "нація"* (*... with "nation"*), *... від "нація"* (*... from "nation"*), *...про "нація"* (*... about "nation"*), *"нація" із ...* (*"nation" with ...*), *"нація" над ...* (*"nation" above ...*), *"нація" перш...* (*"nation" before ...*);

'нація' and/or: *...громад*а (громади чи нації), *культура* (націю й культуру).

Due to UD the lexeme **'нація / nation'** has the following characteristics: 1) specifically historical form of human community, united by one language and territory, deep internal economic relations, certain features of culture and character; 2) a state, a country; 3) figur.,iron. About any group of people (blacksmith nation, sky nation) [Білодід, 2023].

It is worth mentioning that D. Dontsov does not scatter the notion of the lexeme *nation* into elements, i.e. almost all interpretations of the lexeme *nation* are related to the first meaning. The combination of the second and the third ironic meanings was revealed in one variant, i.e. D. Dontsov

uses the construction *Ліга націй – ліга махінацій* (*League of Nations – league of machinations*), where using irony to the organization (*League of Nations*), which lost its influence totally after the Second World War and ceased to exist as an organization.

Note that all collocations can be reviewed in context, which, undoubtedly, provides a researcher with the opportunity to select material in a more detailed way if necessary.
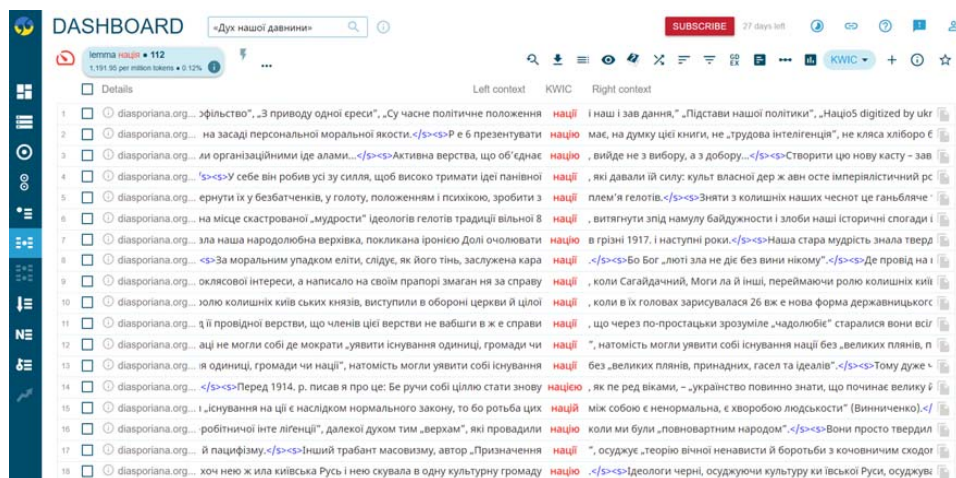

Fig. 5. The context of lexical verbalizers of the explicit concept 'нація'
in the corpus "Spirit of our Antiquity" ("Dukh nashoi davnyny")

During the compilation of the frequency list of any artist, writer, politician, etc., the presentation of the used word frequency list is a valuable function of the Sketch Engine platform. The sample of lexemes shows the distinct statistics of the lexeme usage frequency in the corpus. Even before the development of information technologies, namely applied linguistics, compilation of frequency lists required quite a troublesome and stretched effort of a researcher. At present, the program needs a few seconds to present the frequency of word usage. For example, in our produced corpus the frequency list gives the basis to determine the most used lexemes much faster and to make certain conclusions.

Table 1

The Table of lexeme frequency in the corpus

	Business online communication	Informal communication
2245 – 195	discursive words	*в (in)*, *на (on)*, *не (no/not)*, *що (what)*, *й (and)*, *до (to)*, *а (but)*, *у (in)*, *за (for)*, *коли (when)*, *від (from)*, *як (how)*
195 – 86	commonly used words	*він (he)*, *я (I)*, *бути/було (to be/was/were)*, *земля (land)*, *влада (authority/government)*, *люди (people)*
86 – 45	existential names	*життя (life)*, *каста (caste)*, *народ/ народу (nation)*, *ідеал (ideal)*, *дух (spirit)*, *душа (soul)*, *сила (strength)*, *поняття (concept)*, *мудрість (wisdom)*, *суспільність (society)*, *культура (culture)*, *маси (masses)*
45 – 25	terms, social-political vocabulary and lexemes, that reproduce moral qualities	*сила (strength)*, *суспільність (community)*, *Бог (God)*, *Вітчизна (Homeland)*, *інтелігенція (intelligentsia)*, *козацтво (Cossackdom)*, *козаки (Cossacks)*, *ідея (idea)*, *правлячий (ruling)*, *історія (history)*, *предки (ancestors)*, *церква (church)*, *Україна (Ukraine)*, *мудрість (wisdom)*, *життя (life)*, *слава (glory)*, *віра (faith)*, *старина (antiquity)*, *честь (honor)*, *відчуття (sense/feeling)*
25 – 1	solemn style vocabulary and specific idiolect of D. Dontsov	*країна (country)*, *провідний (leading)*, *думка (thought/opinion)*, *раса (race)*, *князь (prince)*, *меч (sword)*, *роль (role)*, *правда (truth)*, *культ (cult)*, *слава (fame/glory)*, *щастя (happiness)*, *воля (will/freedom)*, *держава (state)*, *традиції (traditions)*, *володар (master)*, *історія (history)*, *цілість (integrity)*, *інтелігенція (intelligentsia)*, *Україна (Ukraine)*, *серце (heart)*, *закон (law)*, *суспільство (society)*, *аристократія (aristocracy)*, *демократія (democracy)*, *відвага (courage)*, *завзяття (dedication)*, *культура (culture)*, *гріхи (sins)*, *панування (dominion/rule)*

Lexical-statistical regularity of word usage in a concordance can be depicted in the shape of a pyramid. On the top of the pyramid there are words of the highest frequency in the texts of speech, but at the base of the pyramid – thousands of words, which occur in a text once. The other words are located between the top and the base of the pyramid depending on the frequency of their usage: the higher word usage frequency, the closer to the top of the pyramid the word is.

In the process of studying the political component, motivated by word usage frequency, the "pyramid" principle is initial: at first it is necessary to master that small number of the most frequent words – the upper part of the pyramid, then to move to the words of not high usage frequency and, finally, to rarely used, placed in the low part of the pyramid.
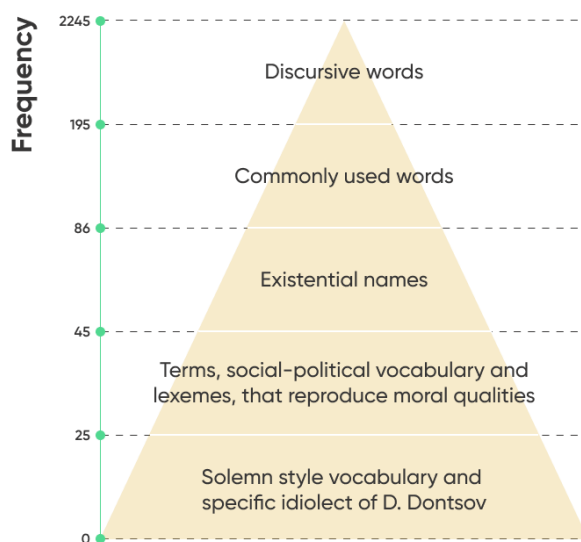

Fig. 6. Lexeme frequency in corpus

The analysis of lexeme usage frequency in text corpus of "The Spirit of our Antiquity" ("Dukh nashoi davnyny") has shown the following results:

1st group (the usage frequency from 2245 to 195) – is the upper part of the pyramid, consisting of discursive words (prepositions, conjunctions, particles, etc.). These are the original grammatical stabilizers of a language without which mastering of scientific style is impossible: *в* (*in*)*, на* (*on*)*, не* (*no/not*)*, що* (*what*)*, й* (*and*)*, до* (*to*)*, а* (*but*)*, у* (*in*)*, за* (*for*)*, коли* (*when*)*, від* (*from*)*, як* (*how*);

2nd group (the usage frequency from 195 to 86) – there are commonly used words, usually typical for all styles of speech. These lexemes can't be differentially divided due to the themes, yet their usage frequency is significant. First of all, those are nouns, pronouns, and verbs: *він* (*he*)*, я* (*I*)*, бути/було* (*to be/was/were*)*, земля* (*land*)*, влада* (*authority/government*)*, люди* (*people*);

3rd group (the usage frequency from 86 to 45) – a significant number of nouns, characterizing existential names: *життя* (*life*)*, каста* (*caste*)*, народ/ народу* (*nation*)*, ідеал* (*ideal*)*, дух* (*spirit*)*, душа* (*soul*)*, сила* (*strength*)*, поняття* (*concept*)*, мудрість* (*wisdom*)*, суспільність* (*society*)*, куль-тура* (*culture*)*, маси* (*masses*);

4th group (the usage frequency from 45 to 25). The terms, social-political vocabulary and lexemes, reproducing moral qualities, occur here the most frequently: *сила* (*strength*)*, суспільність* (*community*)*, Бог* (*God*)*, Вітчизна* (*Homeland*)*, інтелігенція* (*intelligentsia*)*, козацтво* (*Cossackdom*)*, козаки* (*Cossacks*)*, ідея* (*idea*)*, правлячий* (*ruling*)*, історія* (*history*)*, предки* (*ancestors*)*, церква* (*church*)*, Україна* (*Ukraine*)*, мудрість* (*wisdom*)*, життя* (*life*)*, слава* (*glory*)*, віра* (*faith*)*, старина* (*antiquity*)*, честь* (*honor*)*, відчуття* (*sense/feeling*);

5th group (the usage frequency from 25 to 1): it is the vocabulary of solemn style as well as the specific idiolect of D. Dontsov: *країна* (*country*)*, провідний* (*leading*)*, думка* (*thought/opinion*)*, раса* (*race*)*, князь* (*prince*)*, меч* (*sword*)*, роль* (*role*)*, правда* (*truth*)*, культ* (*cult*)*, слава* (*fame/glory*)*, щастя* (*happiness*)*, воля* (*will/freedom*)*, держава* (*state*)*, традиції* (*traditions*)*, володар* (*master*),

історія (*history*), *цілість* (*integrity*), *інтелігенція* (*intelligentsia*), *Україна* (*Ukraine*), *серце* (*heart*), *закон* (*law*), *суспільство* (*society*), *аристократія* (*aristocracy*), *демократія* (*democracy*), *відвага* (*courage*), *завзяття* (*dedication*), *культура* (*culture*), *гріхи* (*sins*), *панування* (*dominion/rule*) (see table 1);

During the processing of the concordance of the corpus copmiled it was revealed, that the author often uses proper names, in particular the surnames of famous people: *Шевченко* (*Shevchenko*) – 92, *Сковорода* (*Skovoroda*) – 60, *Франко* (*Franko*) – 36, *Платон* (*Platon*) – 30, *Куліш* (*Kulish*) – 35, *Драгоманів* (*Drahomaniv*) – 29, *Хмельницький* (*Khmel'nytskyi*) – 50, *Вишневецький* (*Vyshnevetskyi*) – 15, *Антонович* (*Antonovych*) – 13, *Глинський* (*Hlyns'kyi*) – 4, *Дашкевич* (*Dashkevych*) – 3, *Ружинський* (*Ruzhyns'kyi*), *Ляндскоронський* (*Liandskorons'kyi*), *Корецький* (*Koretskyi*), *Чернишевський* (*Chernyshevs'kyi*), *Свіргговський* (*Svirhovs'kyi*) – 2, *Кондратович* (*Kondratovych*), *Литовський* (*Lytovs'kyi*), *Самчук* (*Samchuk*) and others – 1.

During providing proper and geographical names (*Київ, Львів, Москва, Московія,* etc.), we recommend giving comments about how a person is related to D. Dontsov, what this geographical territory contributes to forming philosophical or political views of the thinker. Providing such information will help to specify a biography, relations with other personalities, forming a character, presence of a pseudonym.

Note that the program Sketch Engine has the function of calculating the usage frequency of different parts of speech: adjectives, nouns, pronouns, verbs, etc. We have revealed, that the usage frequency of every part of speech in the corpus of "The Soul of our Antiquity" is specific, since D. Dontsov associates pronouns mainly with the reproduction of the category "no stranger-stranger", where a pronoun *наш (our), наша (our (feminine)), свій (one's own), мій (my)* have an association with the lexeme *український (Ukrainian), національний (national), автохтонний (indigenous/ autochthonous), споконвічний (time-honored),* etc.

The adjectives show the original Ukrainian names, excluded from the system of the Ukrainian language under the influence of repressions. Thus, returning the adjectives to active usage is the revival of the language's originality. The most frequent adjectives are: *козацький* (*Cossack*), *великий* (*great*), *провідний* (*leading*), *демократичний* (*democratic*), *державний* (*state*), *історичний* (*historical*), *героїчний* (*heroic*), *духовий* (*spiritual*), *князівський* (*princely*), *руський* (*Rus'*), *матеріяльний* (*material*), *шляхетний* (*noble*), *володарський* (*ruling*), *відосередній* (*immediate*), *луччий* (*better*), *повинний* (*obligatory*), *гієрархічний* (*hierarchical*), *партикулярно* (*particular*), *мандрівний* (*wandering*), *осквернення* (*defilement*), *простацький* (*vulgar*).

The adverbs are pieces of evidence of being native-Ukrainian language, since this part of the language "resists" internationalization and there are no borrowings among adverbs. Revealed: *живо* (*vividly*), *нехибно* (*accurately*), *немилосердно* (*ruthlessly*), *взаїмно* (*mutually*), *органічно* (*organically*), *засадничо* (*fundamentally*), *справедливо* (*justly*), *пристрасно* (*passionately*), *безперечно* (*unquestionably*), *фізично* (*physically*), *стійко* (*steadfastly*), *пильно* (*vigilantly*), *незалежно* (*independently*), *посередньо* (*indirectly*), *одначе* (*however*).

Practical use of concordance

The produced concordance of the writings of D. Dontsov offers researchers ample opportunities. In particular:

1. Forming of electronic concordance, regarding the usage frequency of lexical items.
2. Demonstration of pragmatic filling of a given lexeme.
3. Highlighting new meanings in a word, terminologization, and determinologization of expressions.
4. Illustration of a style, where a given word is used.
5. Determination of the usage frequency of a given word.

Accordingly, to put a query on different information of a researcher's interest within the scope of his scientific investigation is believed possible by means of the modern invention Word Sketches, in particular:

• *Metadata* (the data with characteristics of a certain text (document). For example, data about an author, the style of a publication, the year of producing and the year of text publishing, etc. the query on authorship, of particular importance for D. Dontsov's writing, since he used over 20 pseudonyms or cryptonyms;

• *Positional attribute* (the information referring to the token in a corpus);

• *Tag* – a mark, attributed to the tokens in annotated corpus to designate grammatical, semantic or stylistic characteristics;

• *Absolute frequency* (the number of occurrences of a word, a phrase, a tag, etc. in a corpus, sub corpus, concordance or sample);

• *Word form* – lexeme realization in a text in any grammatical form;

• *Analyzing lemmas*, since a lemmatized corpus enables the search due to the lemma and including all of the word forms into the result (words with their phonetic variants or without them; words with word-building variants or without them.

The concordance of the writings of D. Dontsov within the scope of our study has allowed structuring of political terms, used in the language corpus. For example social-political terminology has been grouped by names: 1) political directions, movements, currents (*большевизм* (*Bolshevism*), *волюнтаризм* (*voluntarism*), *гуманізм* (*humanism*), *гуманітаризм* (*humanitarianism*), *ідеалізм* (*idealism*), *лібералізм* (*liberalism*), *матеріалізм* (*materialism*), *націоналізм* (*nationalism*), *опуртунізм* (*opportunism*), *партикуляризм* (*particularism*), *пацифізм* (*pacifism*), *провансальство* (*Provencalism*), *провінціялізм* (*provincialism*), *радикалізм* (*radicalism*), *реалізм* (*realism*), *сепаратизм* (*separatism*), *шовінізм* (*chauvinism*)); 2) social order (*демократизм* (*democratization*), *імперіалізм* (*imperialism*), *капіталізм* (*capitalism*), *комунізм* (*communism*), *соціялізм* (*socialism*), *феодалізм* (*feudalism*)); 3) student, of theories, science (*догматизм* (*dogmatism*), *інтелектуалізм* (*intellectualism*), *інтернаціоналізм* (*internationalism*), *квієтизм* (*quietism*), *космополітизм* (*cosmopolitanism*), *марксизм* (*Marxism*), *утілітаризм* (*utilitarianism*)); 4) social-political belonging of people (*більшовик* (*Bolshevik*), *демократ* (*democrat*), *комуніст* (*communist*), *махновець* (*Makhnovist*), *молодоукраїнець* (*Young Ukrainian*), *монарх* (*monarch*), *провансалець* (*Provençal* (*regionalist in the Provence region of France*)), *пролетарій* (*proletarian*), *радикал* (*radical*), *революціонер* (*revolutionary*), *соціаліст* (*socialist*)); 5) state-general terms (*влада* (*authority, power*), *держава* (*state*), *ідеологія* (*ideology*), *народ* (*people*), *нація* (*nation*), *політика* (*politics*), *самостійність* (*independence*), *суверенність* (*sovereignty*)); 6) religious, student (*буддизм* (*Buddhism*), *християнство* (*Christianity*)) etc.

Based on the concordance, it is possible to compile a dictionary of quotes from the works of D. Dontsov. For example, the lexeme '**нація**' (**nation**) is the basis for the following expressions:

*Тоді, як і тепер, боролася **нація** за національну незалежність й за віру*. (*Then, as now, **the nation** fought for national independence and faith*).

*Воля **нації** потребує свого магнета, свого ідеологічного осередку, який координував би розпорошені стремління, вказував мету*. (***The nation** will require its magnet, its ideological center that would coordinate scattered aspirations, indicating the goal*).

*Коли організм **нації** є затруєний чужою йому культурою Сходу, то жодна політична сепарація нічого нам не поможе*. (*When the body of **the nation** is poisoned by an Eastern culture foreign to it, no political separation will help us*).

*Прагнемо завернути нашу ментальність до великого минулого, коли **нація** була **нацією**, а не "**всенацією**"; коли любов до свого була любов'ю до свого, а не – водночас – до чужого й ворожого*. (*We aspire to turn our mentality back to the great past when **the nation** was **a nation**, not an "**all-nation**"; when love for one's own was love for one's own, not – at the same time – for the foreign and the hostile*).

*Твердо йти своєю життєвою дорогою може тільки **нація** людей великих характерів*. (*Only **a nation** of great characters can firmly follow its life path*).

Епопея української нації, нації героїв і войовників – ось чим є "Енеїда" для нас. (*The epic of the Ukrainian **nation**, the nation of heroes and warriors – that's what "Aeneid" is for us*).

*Партачі життя і творці життя – це дві раси. Тільки останні – єдино вони, виведуть **націю** зі стану ганьби*. (*Life-eaters and life-makers are two races. Only the latter, united, will lead **the nation** out of disgrace*).

*Активна верства, що об'єднує **націю**, вийде не з вибору, а з добору*. (*The active layer that unites **the nation** will emerge not by choice but by selection*).

*Хто хоче зробити з **нації** "велику сім'ю вольну, нову", мусить мати стільки сили душевної, щоб "кайдани порвати"*. (*Whoever wants to make **the nation** a "great, free, new family" must have enough strength of spirit to "break the chains"*).

*Голота стає **нацією** лише тоді, коли має провід.* (***A nation*** *becomes* ***a nation*** *only when it has leadership*).

*Горе **нації**, провід якої, замість того закону, кланяється ідолам матерії.* (*Sorrow for **a nation** whose leadership, instead of following the law, worships idols of materialism*).

In our view, analysis of the information out of the concordance also enables revealing the presence of lexical-semantic processes (paronymy, synonymy, antonymy), presence of dialectical forms etc.

Conclusion

Any available means of programming that facilitate the processing of large volumes of language material have been proven to be an undoubtedly productive way to conduct conceptual analysis and serve as an additional tool for studyi.

Producing the concordance of the writings of D. Dontsov will assist language historians, as the previous political achievements of the thinker can be taken into account. For philologists, the significant contribution of the author can supplement the language system, demonstrating disarchaized elements and considering altered (or still relevant) word-forming and grammatical components, stylistic word alteration analysis.

The importance of a concordance for the demonstration of a language-mental picture of the world is based on the following factors:

• compiling political dictionaries represents a new stage in the development of political discourse;

• text comprehension benefits from contextual factor that help realize certain thoughts;

• searching for specific lexeme enables the specification of utterance attributed to the author;

• the demonstration of the author's erudition, as the word usage of a specific author can be compared with commonly used word interpretations or with texts by other authors;

• analyzing the technical characteristics of a word (the frequency of its use in a text, text preparation for publication);

• studying political language reflects the evolution of social-political consciousness and presenting ideas which have or have not lost their relevance and importance;

• political vocabulary allows us to reproduce structuring of all governmental branches (legislative, judiciary and executive) according to political views and to reflect the cognitive capabilities of rulers based on usage frequency. For example, the phrase "socialism basic" indicates the frequency of usage by political leaders (such as former Romanian President N. Ceaușescu, Communist Party leader Stalin etc.) during a particular era;

• the concordance of D. Dontsov's writings enables comparison of word usage across various political or historical texts from a specific period;

• new interpretations of a particular lexeme can be realized on the contextual basis, thus serving as guidance for readers and providing heuristic function of the concordance;

• changes in the meaning of words are a result of the development of the language system; therefore, new meanings observed in concordances should be introduced into glossaries;

• the pragmatic implications of a given lexeme can be demonstrated, allowing for comparisons of the influence of "author" – "reader" during specific historical period;

• the illustration of style can be conducted through the use of a given word.

The multifaceted study of D. Dontsov's writings is believed to be of importance. The concordance serves as a new material, important for politicians, journalists, teachers, students, for it is an entrance in a system of new words, filling with new ideas, and possibility to perceive the world through the nation-state aspirations of a great thinker. Realization of his project will help to create a political-intellectual product of great importance to offer bright prospects for future linguistic study, whose tasks are supposed to use linguistic material. The concordance of Dontsov's writings opens up opportunities for creating different dictionaries (the writer language, the writing language, the concordance), which is an important achievement of lexicography.

The prospect of our future study consists in compiling text corpora (culturological, literature-scientific, political, philosophical, historical, etc.) to be processed due to the functionality of Sketch Engine.

Bibliography

Білодід, І.К. (Ред.). (2023). *Словник української мови online*. Retrieved from http://sum.in.ua/

Бук, С. (2013). Квантитативна параметризація текстів Івана Франка: проект та його реалізація. *Вісник Львівського університету. Серія філологічна*, 58, 290-307.

Бук, С., Ровенчак, А. (2010). *Онлайн-конкорданс роману Івана Франка «Перехресні стежки»*. Retrieved from http://ktf.franko.lviv.ua/users/andrij/science/Franko/concordance.html

Дарчук, Н.П. (Ред.). (2024). *Корпус української мови*. Retrieved from http://www.mova.info/corpus.aspx?l1=209

Донцов, Д. (1951). *Дух нашої давнини*. Мюнхен–Монреаль: Видано заходами прихильників автора.

Коева, С. (Ред.). (2024). *Български национален корпус*. Retrieved from http://search.dcl.bas.bg/

Кучера, Г., Нельсон Френсіс, В. (Ред.). (2024). *Браунівський корпус української мови*. Retrieved from https://github.com/brown-uk/corpus

Микитюк, О.Р. (2021). *Щастя – бути сильним. Афоризми і сентенції Дмитра Донцова*. Львів: Видавництво Львівської політехніки.

Москалевський, Б. (Ред.). (2024). *Лабораторія української*. Retrieved from https://mova.institute/

Чаплинський, Д., Дьомкін, В. (Ред.). (2024). *NER annotation corpus*. Retrieved from https://lang.org.ua/uk/corpora/

Чтиво (2024). *Корпус української мови*. Retrieved from http://korpus.org.ua/

Bouso, T., San Segundo, P.R. (2021). The British Sentimental Novel Corpus (BSNC) and the ROC-DDC alternation at the level of the individual. *NJES Nordic Journal of English Studies*, 20 (1), 215-257. DOI: https://doi.org/10.35360/njes.659

Buk, S. (2011). Slavic experience of compiling frequency dictionaries of the writer's language. *Problems of Slavic Studies*, 60, 217-224.

Buk, S., Rovenchak, A. (2019). Simple definition of distances between texts from rank-frequency distributions, A case of Ukrainian long prose works by Ivan Franko. *Glottometrics*, 46, 1-11.

Hadzalo, A. (2020). Analysis of Gender-Marked Units: Statistical Approach. In: V. Lytvyn (Ed.), *Computational Linguistics and Intelligent Systems, COLINS, CEUR workshop proceedings* (pp. 462-471). Lviv: Lviv Polytechnic National University.

Jakubíček, M., Kovář, V., Rychlý, P., Suchomel, V. (2020). Current Challenges in Web Corpus Building. In: A. Barbaresi, F. Bildhauer, R. Schäfer, E. Stemle (Eds.), *Proceedings of the 12th Web as Corpus Workshop* (pp. 1-4). Marseille: European Language Resources Association.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1, 7-36. DOI: https://doi.org/10.1007/s40607-014-0009-9

Křen, M. (Ed.). (2024). *Český národní korpus*. Retrieved from https://www.korpus.cz/

Kunilovskaya, M. (2017). Sketch Engine: A Toolbox for Linguistic Discovery. *Journal of Linguistics/Jazykovedný časopis*, 68 (3), 503-507. DOI: https://doi.org/10.2478/jazcas-2018-0006

Levická, J. (Ed.). (2024). *Slovenský národný korpus*. Retrieved from https://korpus.sk/

Lin, SY., Chen, HC., Chang, TH. et al., (2019). CLAD: A corpus-derived Chinese Lexical Association Database. *Behav Res*, 51, 2310-2336. DOI: https://doi.org/10.3758/s13428-019-01208-2

Martín, A.S., Trekker, C. (2021). Adapting Word Sketches for Specialized Knowledge Extraction. In: D. Amalia, A.D. Darnis, A. Triatna, D. Khairiah (Eds.), *Proceedings of the 14th International Conference of the Asian Association for Lexicography (ASIALEX 2021)* (pp. 64-87). Jakarta: ASIALEX.

Nahorna, O. (Ed.). (2024). *UA-GEC: Grammatical Error Correction and Fluency Corpus for the Ukrainian Language*. Retrieved from https://ua-gec-dataset.grammarly.ai/

O'Keeffe, A., McCarthy, M. (Eds.). (2010). *The Routledge Handbook of Corpus Linguistics*. London: Routledge. DOI: https://doi.org/10.4324/9780203856949

Pollak, S., Gantar, P., Holdt, Š.A. (2019). What's New on the Internetz? Extraction and Lexical Categorisation of Collocations in Computer-Mediated Slovene. *International Journal of Lexicography*, 32 (2), 184-206. DOI: https://doi.org/10.1093/ijl/ecy026

Pylypiuk, N., Ilnytzkyj, O.S., Kozakov, S. (Eds.). (2008). *Online Concordance to the Complete Works of Hryhorii Skovoroda*. Retrieved from http://www.artsrn.ualberta.ca/skovoroda/index.php?glang=

Rovenchak, A., Buk, S. (2018). Part-of-Speech Sequences in Literary Text: Evidence from Ukrainian. *Journal of Quantitative Linguistics*, 25 (1), 1-21. DOI: https://doi.org/10.1080/09296174.2017.1324601

Sass, B. (2022). Principles of corpus querying. *A discussion note Acta Linguistica Academica*, 69 (4), 599-614. DOI: https://doi.org/10.1556/2062.2022.00581

Shvedova, M., von Waldenfels, R., Yarygin, S., Rysin, A., Starko, V., Nikolajenko T. (Eds.). (2017-2022). *GRAC: General Regionally Annotated Corpus of Ukrainian*. Retrieved from http://uacorpus.org/Kyiv/en

Simpson, J. (Ed.). (2011). *The Routledge Handbook of Applied Linguistics*. London: Routledge. DOI: https://doi.org/10.4324/9780203835654

Üveges, I. (2022). Comprehensibility and Automation: Plain Language in the Era of Digitalization. *TalTech Journal of European Studies*, 12 (2), 64-86. DOI: https://doi.org/10.2478/bjes-2022-0012

Villayandre Llamazares, M. (2008). Lingüística con corpus (I). *Estudios Humanísticos. Filología*, 30, 329-349. DOI: https://doi.org/10.18002/ehf.v0i30.2847

Wójtowicz, B. (Ed.). (2024). *Narodowy Korpus Języka Polskiego*. Retrieved from http://nkjp.pl/

DMYTRO DONTSOV'S TEXTS IN THE ASPECT OF CORPUS LINGUISTICS AND LEXICOGRAPHY

Oksana R. Mykytyuk, Lviv Polytechnic National University (Ukraine)
e-mail: oksana.r.mykytiuk@lpnu.ua
Liliia V. Kharchuk, Lviv Polytechnic National University (Ukraine)
e-mail: liliia.v.kharchuk@lpnu.ua
DOI: https://doi.org/10.32342/3041-217X-2024-2-28-17

Key words: *corpus linguistics, Sketch Engine, concordance, text, linguistic worldview, Dmytro Dontsov, lexicographic practice*

Increasingly, computer technologies in linguistics offer their advanced tools to process, store and select language data, which has triggered the fast development of the actual branch of linguistic studies – corpus linguistics. Through the use of large-scale empirical data and advanced computer technologies to reach objective insights into language function, linguistic corpora have quickly become invaluable resources. Data obtained through corpus analysis facilitate the drawing of qualitatively new conclusions about language and highlight research directions that previously received little attention. Despite substantial linguistic work on Dmytro Dontsov's writings, there still exist a number of questions not studied so far. Among these are compiling a linguistic language corpus as well as a concordance of this prominent thinker, public figure, publicist utilizing modern methods to calculate and fix lexemes in order to identify their attributive collocation. Any available means of programming, facilitating processing language material of a big mass, have been realized to be an undoubtedly productive way to properly perform conceptual analysis and to serve as an additional tool for studying. *The purpose* of the study is to determine the peculiarities of the unique features of Dmytro Dontsov's maxims in the aspect of corpus linguistics and lexicography on the basis of the corpus space that forms the linguistic vision of the world and is a source of creating lexicographical pieces of work (concordances, a writing language, a writer language, etc.). We emphasize that the creation of Dmytro Dontsov's writing concordance has become the subject of corpus linguistics study for the first time. The subject of our study is the writings of Dmytro Dontsov, their lexicographical parametrization which provides all possible words with their description (phonetic, word-creating, grammatical), along with quantitative indices, i.e. it is the result of learning many linguistic disciplines, unified by the dictionary. The methodology of study is the combination of general theoretical *methods* (analysis, generalization, explanation) with the applied methods of linguistics. Analysis of the studied conception is grounded in "The Spirit of Our Antiquity" text corpus compiled by using the Sketch Engine program. On the basis of the analysis, it is found out that an electronic corpus provides the opportunity to accelerate language study and increase its effectiveness, probability and checkability significantly. The article reveals heuristic potential, practical effectiveness of the corpus and application of concordant technologies in conceptual studies. It was discovered that, the construction of the full concordance of the writings of Dmytro Dontsov will enable showing the picture of the world on the basis of learning the author's lexical wealth and reproducing his understanding of the political situation. The created concordance is a stage of forming lexicographical works about Dontsov, which provides understanding of stages, methods, principles and peculiarities of compiling Dmytro Dontsov's writing language dictionary. The multifaceted study of D. Dontsov's writings is believed to be of importance. The concordance serves as a new material, important for politicians, journalists, teachers, students, for it is an entrance in a system of new words, filling with new ideas, and possibility to perceive the world through the nation-state aspirations of a great thinker. Realization of his project will help to create a political-intellectual product of great importance to offer bright prospects for future linguistic study, whose tasks are supposed to use linguistic material.

References

Bilodid, I.K. (ed.). (2023). *Slovnyk ukrainskoi movy online* [Dictionary of the Ukrainian Language Online]. Available at: http://sum.in.ua/ (Accessed 25 November 2024).

Bouso, T., San Segundo, P.R. (2021). The British Sentimental Novel Corpus (BSNC) and the ROC-DDC alternation at the level of the individual. *NJES Nordic Journal of English Studies*, vol. 20, issue 1, pp. 215-257. DOI: https://doi.org/10.35360/njes.659

Buk, S. (2011). Slavic experience of compiling frequency dictionaries of the writer's language. *Problems of Slavic Studies*, vol. 60, pp. 217-224.

Buk, S. (2013). *Kvantytatyvna parametryzatsiia tekstiv Ivana Franka: proekt ta yoho realizatsiia* [Quantitative parameterization of Ivan Franko's texts: the project and its implementation]. *Visnyk of the Lviv University. Series Philology*, vol. 58, pp. 290-307.

Buk, S., Rovenchak, A. (2010). *Onlain-konkordans romanu Ivana Franka "Perekhresni stezhky"* [Online Concordance of Ivan Franko's novel "Cross paths"]. Available at: http://ktf.franko.lviv.ua/users/andrij/science/Franko/concordance.html (Accessed 25 November 2024).

Buk, S., Rovenchak, A. (2019). Simple definition of distances between texts from rank-frequency distributions, A case of Ukrainian long prose works by Ivan Franko. *Glottometrics*, vol. 46, pp. 1-11.

Chaplynskyi, D., Diomkin, V. (eds.). (2024). NER annotation corpus. Available at: https://lang.org.ua/uk/corpora/ (Accessed 25 November 2024).

Chtyvo (2024). *Korpus ukrainskoi movy* [Corpus of the Ukrainian language]. Available at: http://korpus.org.ua/ (Accessed 25 November 2024).

Darchuk, N.P. (ed.). (2024). *Korpus ukrainskoi movy* [Corpus of the Ukrainian Language]. Available at: http://www.mova.info/corpus.aspx?l1=209 (Accessed 25 November 2024).

Dontsov, D. (1951). *Dukh nashoi davnyny* [The Spirit of our Antiquity]. Miunkhen–Montreal: Vydano zakhodamy prykhylnykiv avtora Publ., 342 p.

Hadzalo, A. (2020). Analysis of Gender-Marked Units: Statistical Approach. In V. Lytvyn (ed.). Computational Linguistics and Intelligent Systems, COLINS, CEUR workshop proceedings. Lviv, Lviv Polytechnic National University Publ., pp. 462-471.

Jakubíček, M., Kovář, V., Rychlý, P., Suchomel, V. (2020). Current Challenges in Web Corpus Building. In A. Barbaresi, F. Bildhauer, R. Schäfer, E. Stemle (eds.). Proceedings of the 12th Web as Corpus Workshop. Marseille, European Language Resources Association Publ., pp. 1-4.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, vol. 1 pp. 7-36. DOI: https://doi.org/10.1007/s40607-014-0009-9

Koeva, S. (ed.). (2024). *Български национален корпус* [Bulgarian National Corpus]. Available at: http://search.dcl.bas.bg/ (Accessed 25 November 2024).

Křen, M. (Ed.). (2024). *Český národní korpus* [Czech National Corpus]. Available at: https://www.korpus.cz/ (Accessed 25 November 2024).

Kučera, H., Nelson Francis, W. (eds.). (2024). *Braunivskyi korpus ukrainskoi movy* [Brown Corpus of Ukrainian]. Available at: https://github.com/brown-uk/corpus (Accessed 25 November 2024).

Kunilovskaya, M. (2017). Sketch Engine: A Toolbox for Linguistic Discovery. *Journal of Linguistics / Jazykovedný casopis*, vol. 68, issue 3, pp. 503-507. DOI: https://doi.org/10.2478/jazcas-2018-0006

Levická, J. (ed.). (2024). *Slovenský národný korpus* [Slovak National Corpus]. Available at: https://korpus.sk/ (Accessed 25 November 2024).

Lin, SY., Chen, HC., Chang, TH. et al., (2019). CLAD: A corpus-derived Chinese Lexical Association Database. *Behav Res*, vol. 51, pp. 2310-2336. DOI: https://doi.org/10.3758/s13428-019-01208-2

Martín, A.S., Trekker, C. (2021). Adapting Word Sketches for Specialized Knowledge Extraction. In: D. Amalia, A.D. Darnis, A. Triatna, D. Khairiah (eds.). Proceedings of the 14th International Conference of the Asian Association for Lexicography (ASIALEX 2021). Jakarta, ASIALEX Publ., pp. 64-87.

Moskalevskyi, B. (ed.). (2024). *Laboratoriia Ukrainskoi* [Laboratory of Ukrainian]. Available at: https://mova.institute/ (Accessed 25 November 2024).

Mykytiuk, O.R. (2021). *Shchastia – buty sylnym. Aforyzmy i sententsii Dmytra Dontsova* [Happiness is To Be Strong. Aphorisms and Maxims by Dmytro Dontsov]. Lviv, Publishing House of Lviv Polytechnic, 392 p.

Nahorna, O. (ed.). (2024). *UA-GEC: Grammatical Error Correction and Fluency Corpus for the Ukrainian Language*. Available at: https://ua-gec-dataset.grammarly.ai/ (Accessed 25 November 2024).

O'Keeffe, A., McCarthy, M. (eds.). (2010). The Routledge Handbook of Corpus Linguistics. London, Routledge, 724 p. DOI: https://doi.org/10.4324/9780203856949

Pollak, S., Gantar, P., Holdt, Š.A. (2019). What's New on the Internetz? Extraction and Lexical Categorisation of Collocations in Computer-Mediated Slovene. *International Journal of Lexicography*, vol. 32, issue 2, pp. 184-206. DOI: https://doi.org/10.1093/ijl/ecy026

Pylypiuk, N., Ilnytzkyj, O.S., Kozakov, S. (eds.). (2008). Online Concordance to the Complete Works of Hryhorii Skovoroda. Available at: http://www.artsrn.ualberta.ca/skovoroda/index.php?glang= (Accessed 25 November 2024).

Rovenchak, A., Buk, S. (2018). Part-of-Speech Sequences in Literary Text: Evidence from Ukrainian. *Journal of Quantitative Linguistics*, vol. 25, issue 1, pp. 1-21. DOI: https://doi.org/10.1080/09296174.2017.1324601

Sass, B. (2022). Principles of corpus querying. *A Discussion Note Acta Linguistica Academica*, vol. 69, issue 4, pp. 599-614. DOI: https://doi.org/10.1556/2062.2022.00581

Shvedova, M., von Waldenfels, R., Yarygin, S., Rysin, A., Starko, V., Nikolajenko T. (eds.). (2017-2022). *GRAC: General Regionally Annotated Corpus of Ukrainian*. Available at: http://uacorpus.org/Kyiv/en (Accessed 25 November 2024).

Simpson, J. (ed.). (2011). The Routledge Handbook of Applied Linguistics. London, Routledge, 752 p. DOI: https://doi.org/10.4324/9780203835654

Üveges, I. (2022). Comprehensibility and Automation: Plain Language in the Era of Digitalization. *TalTech Journal of European Studies*, vol. 12, issue 2, pp. 64-86. DOI: https://doi.org/10.2478/bjes-2022-0012

Villayandre Llamazares, M. (2008). Lingüística con corpus (I). *Estudios Humanísticos. Filología*, vol. 30, pp. 329-349. DOI: https://doi.org/10.18002/ehf.v0i30.2847

Wójtowicz, B. (ed.). (2024). *Narodowy Korpus Języka Polskiego* [National Corpus of Polish]. Available at: http://nkjp.pl/ (Accessed 25 November 2024).