

Corpus Analysis of Word Semantics

Alla Taran

Cherkasy Bohdan Khmelnytskyi National University, 81 Shevchenko Boulevard, 18031, Cherkasy, Ukraine

Abstract

New computer technologies for acquiring knowledge, including knowledge about language, open attractive prospects for solving practical and theoretical linguistic problems. The article examines the possibilities of the language corpus as a tool for researching the semantics of words. It was found that the linguistic corpus provides more reliable material for researching the meaning of words. The analysis of collocations obtained with the help of the corpus proves that words show different combinations, in which changes in meaning can be traced. Options for using the text corpus for research purposes are illustrated by the example of semantic innovations. Neosemantism contributes to the formation of new areas in the cognitive space of the modern Ukrainian language, deepening the verbalization of concepts already known to the Ukrainian language.

Keywords

Language corpus, language practice, collocation, concordance, meaning, neosemantism.

1. Introduction

The orientation of Ukrainian studies to the use of new computer technologies for obtaining knowledge, a new way of research work marks a qualitatively new stage of its development. In the field of scientific research of linguists, databases about language objects are existing, and a new subject of communication – a computer – needs a new apparatus for describing language, new procedures for its modeling, analysis and synthesis. At the same time, the language system, the products of its implementation in the language activity of modern Ukrainian society – various oral and written texts – appear in the computer environment in other ways, turn to the researcher with their new facets, which forces him to think about hitherto unimaginable problems, to look for new, other methods of representation of language information, to develop means of communication with a new participant in communication – a computer.

Any linguistic research begins with the selection of material and the arrangement of the file. Qualitative and quantitative parameters of the collected material largely determine the nature of scientific work.

The creation of texts corpora of different national languages significantly simplified the preparatory, but extremely significant stage for the further development of the research. It is enough for the user to enter a word in the search engine of the corresponding corpus to get a complete index of its word forms in the nearest and extended contexts with appropriate addressing. Creating an empirical research base in this way ensures a minimum expenditure of effort and time for data collection, the volume of which will increase significantly at the same time. On the one hand, it is simplified and facilitated, but on the other hand, it complicates the process of analyzing and comprehending so many examples.

The Internet and its quantitative data, which accompany the results of searches for individual words and phrases, testify to a fairly objective picture of the functioning of the language, since in these data and their "quantitative coefficients" of each word, a sample of source material for

COLINS-2023: 7th International Conference on Computational Linguistics and Intelligent Systems, April 20–21, 2023, Kharkiv, Ukraine

EMAIL: alla_taran@ukr.net (A. Taran)

ORCID: 0000-0001-8091-1477 (A. Taran)



© 2023 Copyright for this paper by its authors.

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

CEUR Workshop Proceedings (CEUR-WS.org)

multifaceted generalizations is revealed that deserves attention. A standard operation of searching for a text or a single word in the hands of a philologist can become a useful tool for lexicographic research, which will most likely end up with a lexicographic product – a dictionary article.

In research practice, the main approaches to the study of linguistic data available in corpus linguistics are widely used – corpus-driven (CDA) and corpus-based (CBA). The first of these approaches allows you to put forward scientific hypotheses, in another way, its task is to support mainly deductive research, the second has a pronounced inductive nature and fully corresponds to the directions of experimental and evidential linguistics. The best result, obviously, makes it possible to simultaneously use the advantages of both approaches, for example, to determine the word-forming or semantic potential of lexical items.

2. Related Works

The reference to corpus data is associated with the name of J. Sinclair, who formulated the well-known principle of idiomacity: "The principle of idiom is that language user has available to him a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments" [15]. Corpus material is used to study linguistic phenomena as a source of language study (according to F. de Saussure – langue). Corpus reflect real facts of word usage. Corpus data are used as proof of the existence of a particular phenomenon in the language or to test a particular hypothesis. Ch. Fillmore rightly noted the value of corpus data: "I don't think there can be any corpora, however large, that contain information about all the areas of English lexicon and grammar that I want to explore, ... every corpus I have had the chance to examine, however small, has taught me facts I couldn't have imagined finding out in any other way" [18].

New methods of semantic analysis based on corpus materials are presented in the monographs of S. Th. Gries and A. Stefanowich [17], as well as D. Glynn and K. Fischer [20]. British linguists proved that the frequency of use of a word is closely related to the structure of the conceptual structure underlying its meaning [16].

3. Methods and Materials

The article uses the method of corpus research and the lexicographic method (analysis of explanatory dictionaries). All word usages are interpreted taking into account the meanings available in the dictionaries. The meaning of neosemanticism is recorded in dictionaries, so we offer our own definition. In this way, we obtain a semantic grid covering regular and single interpretations, which can be characterized by chronology. In this way, we obtain the chronological dynamics of the lexical meaning.

The interpretation of language as a code led to the use of statistical methods of text research to identify certain linguistic regularities in them. According to G. E. Miram, the main principle of distributive theory is to study the textual behavior of language elements for their further comprehensive study [1]. The distributive technique, combined on the basis of statistical techniques, stimulated the formation of theoretical linguistics, where at this time they began to apply modeling, the construction of models that explain the operation of language laws or check the operation and efficiency of devices that reproduce the language process. This technique involves the study of large arrays of source texts to obtain reliable data. As G. E. Miram rightly notes, the study of distribution makes it possible to determine the meaning model, that is, the composition of the main components that together form the meaning of this lexical unit; establish a model of lexeme compatibility with other lexical units; determine the formal structure of a lexical unit [1]. At the level of content understanding, the distribution of a language element is a fixation in certain contexts, a set of neighboring language units. At the same time, the depth of the context can be set taking into account the possibilities of the research.

As the object of discussion of the raised problems, we chose neosemanticisms associated with the further aspectualization of certain concepts in the Ukrainian language. The appearance of new meanings in the semantic structure of a word is one of the powerful mechanisms of replenishing its

lexicon in the Ukrainian language today. The term neosemantism is broader in terms of content than the generally accepted secondary nomination, semantic, or intraword derivation, since it covers both the formation of new meanings on their own ground in the structure of already existing words, and the appearance of new borrowed words-meanings (primary and secondary), and independent interword derivation with repeated use of the same model of word formation. Therefore, the term neosemanticization can be used to denote all the results of updating the semantics of a word at a certain stage of language development. In general, active semantic changes are stimulated by the need of the Ukrainian language to acquire new fields of hitherto unknown concepts.

4. Experiment

Information about the peculiarities of the use of a word, its frequency and typical semantic and grammatical environment opens up for researchers the opportunity to determine its place in the structure of a certain linguistic register and to see a separate linguistic unit in the language system in general.

The goal is to investigate the mechanisms of the appearance of new meanings in words with an already fixed form, using data from the GRAK corpus (General Regionally Annotated Corpus of the Ukrainian Language).

Among the new applications of corpus linguistics that deepen the examination of traditional concepts in the field of lexical semantics, the problem of distinguishing between polysemy and homonymy occupies an important place. At the same time, for linguists who use the corpus on a regular basis, the focus of research gradually shifts, covering an ever-increasing range of questions: from collocation and phraseology to collimation and problems related to grammatical word classes.

The data obtained on the basis of the corpus show that one of the main factors is the word belonging to a certain lexical-grammatical class. It has long been known that the belonging of a word to a specific part of the language reflects the difference in the semantics of signs of different types and demonstrates the difference in the behavior of words in the text. The belonging of a lexical unit to this or that part of the language largely determines the semantics, and accordingly, the peculiarities of its semantic structure.

The problem of the specificity of polysemy of different parts of the language arises in connection with the consideration of certain regularities of the interaction of lexical and grammatical meaning. It has been repeatedly noted that the definition of the lexical meaning of a word already implies an indication of its grammatical characteristics.

On the one hand, the word is a unit of the lexical system, and on the other hand, as a grammatical unit, it has the imprint of both systems, the word is a part of the language, the lexical-grammatical category and those obligatory grammatical meanings that are characteristic of a part of the language.

The experiment consists in clarifying the semantic potential of nouns, adjectives and verbs on the material of the corpus of the Ukrainian language GRAK.

Nouns, adjectives and verbs are carriers of categorical meaning, realize different meanings of the semantic structure. Verbs and nouns, for example, realize the tendency to form a number of independent nominative meanings connected by parallel polysemy, adjectives, on the contrary, are characterized by the "diffusion" of the semantic structure, where the connection between LSV (Lexico-semantic variant) derivatives and the main meaning of the word can be clearly traced. This is characteristic of the radial type of semantic connection, where the radial-metaphorical meaning is important.

5. Results

Let's pay attention to the semantic potential of adjectives and nouns. One of the indicators that the language is in the field of emotional tension of its speakers are new nuances in the meanings of adjectives that concentrate relevant, socially significant thoughts. The semantic potential of the words *нафталін* [naftalin] (*Naphthalene*) / *нафталіновий* [naftalinovyi] (*Naphthalene*), *тефлоновий* [teflonovyi] (*Teflon*) is indicative.

The GRAK corps provided the following statistics:

Нафталін [*naftalin*] (*Naphthalene*)

Found 796 documents.

According to the Dictionary of the Ukrainian Language in 11 volumes, **нафталін** [*naftalin*] (*Naphthalene*) is "a white crystalline substance with a pungent odor, used to combat collar mites, weevils, etc., to protect woolen products and fur from moths, as well as in technology and medicine" [2], and **нафталіновий** [*naftalinovyi*] (*Naphthalene*) is an adjective for **нафталін** [*naftalin*] (*Naphthalene*).

In sources from 1929 to the end of the 20th century, we trace the use of the word in its literal sense. However, from the beginning of the XXI century. contexts realize the figurative use and meaning of "which has gone into the past, is irrelevant."

Нафталіновий [*naftalinovyi*] (*Naphthalene*)

Found 264 documents.

We record only a few sentences where the adjective *naphthalene* has a direct meaning. All other contexts of use have a negative emotional and evaluative shade of the meaning "irrelevant, which has gone into the past".

Contexts of use in **Figure 1** *naphthalene* army, *naphthalene* hero, *naphthalene* "Regions` Party ", *naphthalene* smell, *naphthalene* team, *naphthalene* glamor of the film represent a new meaning of the adjective.

38	2006 • NONE • [Интернет-газета «Високий замок»]	<input type="checkbox"/> <> Це ж про оту саму брежневську стабільність склали анекдоти, глузуючи з нафталінові армії керманців і партбосів, середній вік яких перевалив за 70 (аббревіатуру СССР у народі розшифрували як " Страна Самых Старых Руководителей "). </>
39	2007 • NONE • [Онлайн-ЗМІ «UNIAN.NET»]	<input type="checkbox"/> <> По-четверте, на думку депутата, " нафталіновий герой " Микола МЕЛЬНИЧЕНКО, який поспішив прокоментувати вирвані з контексту слова Ю. ЛУЦЕНКА, тільки таким чином і може привернути до себе увагу. </>
40	2007 • NONE • [Онлайн-ЗМІ «UNIAN.NET»]	<input type="checkbox"/> <> Лідер блоку також закликав людей" викинути на смітник історії нафталінову Партію регіонів ". </>
41	2007 • NONE • [Онлайн-ЗМІ «UNIAN.NET»]	<input type="checkbox"/> <> З одного боку, я хотів би повернутися, аби почитати міністерство від корупціонерів, які знову принесли нафталіновий запах Кучми в коридори МВС масово, беспардонно і цинічно продовжуючи ті недобрі часи. </>
42	2007 • NONE • [Онлайн-ЗМІ «UNIAN.NET»]	<input type="checkbox"/> <> На думку Ю. ЛУЦЕНКА, в країні відбувся різкий стрибок кількості ДТП із смертельними випадками саме через недолугу політику" нафталінові команди " МВС під керівництвом В. ЦУШКА. </>
43	2007 • NONE • [Онлайн-ЗМІ «UNIAN.NET»]	<input type="checkbox"/> <> На думку Ю. ЛУЦЕНКА, в країні відбувся різкий стрибок кількості ДТП із смертельними випадками саме через недолугу політику" нафталінові команди " МВС під керівництвом В. ЦУШКА. </>
44	2008 • NONE • [Интернет-газета «Високий замок»]	<input type="checkbox"/> <> Виходить, увесь той жах, яким протягом двох тижнів тортувала мільйони українців нафталінова команда УТ-1, є нічим іншим, як державним тероризмом... </>
45	2008 • NONE • [Онлайн-ЗМІ «Zaxid.net»]	<input type="checkbox"/> <> Там величають нафталіновий гламур фільму«Секс і місто» та клянуться бути схожими на сентиментальних старушенцій. </>

Figure 1: Concordance of the adjective *naphthalene* in the GRAK corpus

Тефлоновий [*teflonovyi*] (*Teflon*)

Found 166 documents.

We trace social factors in the formation of a new meaning of the adjective **тефлоновий** [*teflonovyi*] (*Teflon*), which denotes a politician who avoids criticism and maintains a political position and a good reputation, despite shortcomings in his work. This meaning is recorded in the Cambridge Dictionary with text illustration Teflon The president survived the crisis with his reputation intact. Reveals a direct motivational connection of the meaning "resistance to external influences, the ability to survive" with its source word "stos. to Teflon, made of it" [7], teflon (chemical term) - "a high-molecular plastic substance (artificial resin), characterized by the greatest resistance to the action of concentrated acids, alkalis and solvents" [2] The specialized meaning of the adjective is not yet fixed by the Ukrainian standard general language dictionaries, but the lexeme is quite active in modern Ukrainian general linguistic practice, cf. the following contexts of its use.

Правильний [pravylnyi] (Correct)

Found 87304 documents.

This adjective illustrates radial-chain polysemy using the example of the semantic structure of the word **правильний** [pravylnyi]: the meanings "true" (**правильний шлях** [pravylnyi shliakh] right way), "just" (**правильне розв'язання конфлікту** [pravylne rozv'iazannia konfliktu] correct resolution of the conflict), "good, just" (**правильний батько** [pravylnyi batko] trueparent) have a common component "which corresponds to reality"; the meaning "unerring" (**правильний напрям** [pravylnyi napriam] correct direction) in turn has a connection with the meaning "which corresponds to the established rules, norms" (**правильні сівозміни** [pravylni sivozminy] correct crop rotations), and it – with the meaning "regular, rhythmic" (**правильне биття серця** [pravylne byttia sertsia] correct heartbeat). All these meanings are fixed in dictionaries. Instead, we observe a change in the meaning and popularity of the adjective **правильний** [pravylnyi] in glossy publications and in certain circles of society. Phrases such as **правильна їжа** [pravylna yizha] (the right food), **правильний ресторан** [pravylnyi restoran] (the right restaurant), **правильний одяг** [pravylnyi odiah] (proper clothing) were not used before. Let's illustrate with more vivid examples of the use of the adjective **правильний** [pravylnyi] in this new evaluative meaning: *Правила **правильної** жінки: пора зайнятися собою, а не пошуком принців!* [Pravyla pravylnoi zhinky: pora zainiatysia soboiu, a ne poshukom pryntsyv!] (The rules of a correct woman: it's time to focus on yourself, and not on the search for princes!) This usage is close in meaning to Galician *comilfo*– French *comme il faut*–it should be, as it should be. And not without the influence of English *right* meaning "suitable, proper". Through the relevant literature and its not always successful translation, such colloquial, subcultural tracings come, because they are used by a certain stratum of society, which is "planted" on such literature, with such interests and instructions.

With the help of the **correct** word, publications try to form a new style of behavior: how to behave, what clothes to wear, what to eat, what to read. This whole system of rules is hidden behind the new use of the **correct** word. Another example: *На платформі ZOOM проходило навчання на тему: «Сучасний процес рекрутингу. «Правильні люди на правильні посади»* (<https://tax.gov.ua> 29.04.2021) [Na platformi ZOOM prokhodylo navchannia na temu: «Suchasnyi protses rekrutynhu. «Pravylni liudy na pravylni posady»] (Training was held on the ZOOM platform on the topic: "Modern recruiting process. "The right people for the right positions"), where the right positions are positions in state bodies. In such contexts, there is manipulation of public consciousness.

The adjective conjugation is **correct** in the GRAK corpus in **Figure 2**: right diet, correct diagnostics, right date, correct shadows, right way, proper food ration, correct selection, correct oval, correct face, right life, proper nutrition.

The combination of the *right diet, the right date, the right way, the proper food ration, and the proper nutrition* reflect the formation of a new meaning of "suitable".

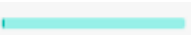
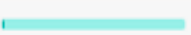
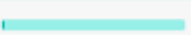
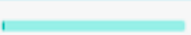
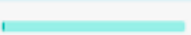
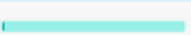
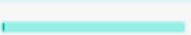
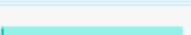
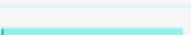
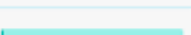
925	правильна дієта	4	< 0.01		...
926	правильна діагностика	4	< 0.01		...
927	правильна дата	4	< 0.01		...
928	Правильні тіні	4	< 0.01		...
929	Правильний спосіб	4	< 0.01		...
930	Правильний раціон	4	< 0.01		...
931	Правильний підбір	4	< 0.01		...
936	Правильне обличчя	4	< 0.01		...
937	Правильне життя	4	< 0.01		...
938	Правильне годування	4	< 0.01		...

Figure 2: The adjective conjugation is **correct** in the GRAK corpus

Маніпуляція [manipuliatsiia] (manipulation)

Found 26910 documents.

The noun acquired a new special meaning in political discourse. Ordinary native speakers may not guess the origin of this word: it comes from the Latin *manipulus* – "handful", which comes from *manus* "hand" and *pleo* "fill" and has the direct meaning "complicated action on something" and figuratively "machination". In the Dictionary of Foreign Words, edited by O. S. Melnychuk, this word is recorded with the explanation "1) movements of the hand or both hands to perform a certain task (e.g., in working with a telegraph key). 2) any complex action. 3) figuratively – a fraudulent prank." With changes, it was recorded in the later Modern Dictionary of Foreign Words by O. I. Skopnenko and T. V. Tsymbalyuk: "French. manipulation, manipulus – "handful" 1. The movement of the hand or hands associated with the performance of a certain task, e.g. M. doctor; 2) demonstration of tricks, which is based mainly on the dexterity of the hands, able to divert the attention of the audience from that; what should be hidden from them; 3) transferred Frauds, fraudulent pranks, e.g. M. with securities" [9].

Since manipulation is not a physical influence on a person, its object is psychological structures, then the new metaphorical meaning is "programming the thoughts and aspirations of the masses, their moods and even mental state to ensure the behavior that is needed by those who possess the means of manipulation. It is the art of controlling people's behavior by purposefully influencing their consciousness and instincts". The lexeme manipulation expands the scope of use by increasing the volume of semantics, which is characterized by a higher degree of generalization, for example: *маніпуляції в темах* [manipuliatsii v temakh] (manipulations in topics), *маніпуляції фактами* [manipuliatsii faktamy] (manipulations of facts), *маніпуляції влади з виборами* [manipuliatsii vlady z vyboramy] (manipulations of the authorities with elections), *тиск та маніпуляції* [tysk ta manipuliatsii] (pressure and manipulations), *договорняки та маніпуляції* [dohovorianky ta manipuliatsii] (negotiations and manipulations).

Changes in the denotative component of the meaning of the lexeme manipulation consist in the loss of several differential sems and the appearance of the "socio-political" and "psychological" sems. Oksana Zabuzhko aptly defined the result of the development of a new meaning of this lexeme: The manipulator presses the buttons not of our strengths, but of our weaknesses. He takes advantage of our sins. I hate people who exploit weaknesses. They know what they are doing. But those *who are manipulated* refuse rationality. They are driven by subconscious impulses. This is what is called the crisis of rationalism, which exists in our time (Gazeta po-ukrainsky, 19.09.2019).

Concordance of the noun *manipulation* in the GRAK corpus in **Figure 3** indicates the emergence of the terms "social and political", "psychological": *The election manipulation of the 2004 made visible the legislation in force at the time, and I consider it to be quite legislation. This is not a serious approach from the point of view of responsible politics, but it is very convenient for those who will later resort to manipulation. "Perhaps some have rightly decided that Lutsenko, despite all his activity, has not been of any real use for quite some time, while his bright image of a martyr languishing in prison is a more playful an object for campaign manipulation," Zaitsev reflects.*

195	2009 • NONE • [Онлайн-ЗМІ «Табло ID»]	<input type="checkbox"/> <s> " Метою її дешевих маніпуляцій була спроба списати свої борги на місто чи організаторів Етапу Кубка Європи ", - заявив він. </s>
196	2010 • Аліна Пастухова • Діява демократія	<input type="checkbox"/> <s> – Маніпуляції на виборах 2004 року зробило видимими чинне на той час законодавство, і я вважаю це його неабияким досягненням. </s>
197	2010 • Аліна Пастухова • Діява демократія	<input type="checkbox"/> <s> Це несерйозний підхід із погляду відповідальної політики, але дуже зручний для тих, хто потім вдаватиметься до маніпуляцій . </s>
198	2010 • Андрій Лаврик &amp; Богда Буткевич • Жорстка «посадка» екс-міністра	<input type="checkbox"/> <s> «Можливо, дехто справедливо вирішив, що від Луценка, незважаючи на всю його активність, вже досить давно не було жодної реальної користі, у той час як його світлий образ мученика, який знемагає у в'язниці, являє собою більш «ігровий» об'єкт для агітаційних маніпуляцій , – розмірковує Зайцев. </s>

Figure 3: Concordance of the noun *manipulation* in the GRAK corpus

According to the Dictionary of the Ukrainian language in 11 volumes, the lexeme *маніпулятор* [*manipulator*] (*Manipulator*) is "1. A person doing various manipulations (in 1 sign.); // A circus magician skillfully manipulating various objects 2. A device for transmitting telegraphic signals. 3. special The device on the control panel, in the control room, etc. to regulate complex production processes" [2]; in the Modern Dictionary of Foreign Words, the meaning "French" is recorded. *manipulateur, manipulus* – 1. A mechanism that, under the control of an operator, performs actions similar to the actions of a human hand; 2) telegraph key; 3) circus performer, illusionist" [9]. Therefore, we observe the formation of a new meaning of the lexeme manipulator "a person who is able to influence another, knows his weak points, affects feelings and emotions."

Синдром [*syndrom*] (*Syndrome*)

Found 15929 documents.

In such a generalized deterministic meaning, the verb *to articulate* becomes a synonym of the verbs *to express, to reveal, to define, to outline, to formulate, to form, to demonstrate* with the qualifiers *exactly, expressively, clearly*, and therefore, expands its lexical and syntactic compatibility.

Variants of using the corpus of texts for research purposes will be illustrated on the example of neosemantism, a **синдром** [*syndrom*] fixed with the meaning: "complex of symptoms characteristic of a certain disease" [2]; *syndrome* [French. *syndrome* Greek. *syndrome* – the one that runs together] – med. a combination of signs (symptoms of a disease) [9]; *CHІД* [*SNID*] (*AIDS*) – acquired immunodeficiency syndrome. Today, this word began to be used more often in the language of politics with the meaning "complex of features distinctive for a certain phenomenon, object."

Similar "syndromes" distinguish phenomena undesirable for speakers of the Ukrainian language, which they assess as a threat to their peaceful, stable life.

Conjugation of the noun *syndrome* in the GRAK corpus

The collocation of the noun *syndrome* in **Figure 4** reflects the new meaning of "complex of features distinctive for a certain phenomenon, object": *desacralization syndrome, war syndrome, rejection syndrome, vagrancy syndrome, indifference syndrome, Chernobyl syndrome*.

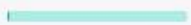
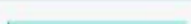
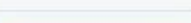
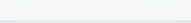
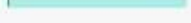
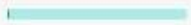
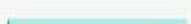
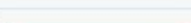
161	Синдром десакралізації	2	< 0.01		...
162	Синдром війни	2	< 0.01		...
163	Синдром відмови	2	< 0.01		...
164	Синдром бродяжництва	2	< 0.01		...
165	Синдром байдужості	2	< 0.01		...
166	Синдром Чорнобиля	2	< 0.01		...
167	Синдром Свіржа	2	< 0.01		...
168	Синдром Рубінштейна	2	< 0.01		...

Figure 4: Conjugation of the noun *syndrome* in the GRAK corpus

Collocation of the noun *syndrome* in **Figure 5** distinguishes phenomena undesirable for the carriers, which they assess as a threat: *Othello syndrome, Mariupol syndrome, Berlin syndrome, Moscow syndrome*.

593	Отелло	3	957	1.73	8.53	2.54	...
594	маріупольський	3	981	1.73	8.49	2.54	...
595	Берлінський	3	1,010	1.73	8.45	2.54	...
596	московським	4	6,665	1.97	6.14	2.54	...

Figure 5: Collocation of the noun *syndrome*

Артикулювати [artykuliuvaty] (to articulate)

Found 1114 documents.

The verb *артикулювати [artykuliuvaty] (to articulate)* is recorded in standard dictionaries only in the sense of "pronounce speech sounds." However, the contexts of use in the corpus testify that it is capable of expressing a new, figurative meaning "to clearly define, outline something, express one's attitude towards something." In this sense, there are reasons to consider it a hidden borrowing from modern Western European languages, cf. English *articulate* – in the sense of "express clearly, explain; formulate". The words *артикуль [artykl]* and the Polish and old Ukrainian *артикул [artykul]* (*article*) have a common etymology with *articulate* (lat. *articulus* – member, section). So, you can logically explain, formulate, emphasize, clearly outline anything, for example, position, interests, nation, as evidenced by the information from the GRAK corpus.

Conjugation of the verb *to articulate* in the GRAK corpus in **Figure 6: articulate interests, articulate opinion, articulate this, articulate questions, articulate self, articulate to the world, articulate things, articulate problems, articulate position, articulate constancy.** All phrases represent a new, figurative, meaning "to clearly define, outline something, express one's attitude towards something."











10	артикулювати інтереси	4	< 0.01		...
11	артикулює думку	3	< 0.01		...
12	артикулюють це	3	< 0.01		...
13	артикулюють питання	3	< 0.01		...
14	артикулювати себе	3	< 0.01		...
15	артикулювати світові	3	< 0.01		...
16	артикулювати речі	3	< 0.01		...
17	артикулювати проблеми	3	< 0.01		...
18	артикулювати позицію	3	< 0.01		...
19	артикулювати незмінність	3	< 0.01		...

Figure 6: Conjugation of the verb *to articulate* in the GRAK corpus

In such a generalized deterministic meaning, the verb *to articulate* becomes a synonym of the verbs *to express, to express, to reveal, to define, to outline, to formulate, to form, to demonstrate* with the qualifiers *clearly, distinctly, clearly, and therefore, expands its lexical and syntactic compatibility.*

A dictionary of contexts of use, or concordance, is an important research tool in corpus linguistics. In it, the word is represented in textual contexts of use. Ye. A. Karpilovska singles out fundamental concordances, which describe the spectrum of word usage in a separate work or in the works of certain authors, and research concordances subordinated to the solution of a specific task [21]. Unlike dictionaries, the concordance comprehensively displays all shades of the meaning of each word.

The organization of the corpus can be very diverse. Depending on the purpose of its creation, there may be texts written in a specific language, by one or more authors and literary genres written in a certain historical period. The entire array of texts in the corpus is systematized. This means that it records the position of each word in a sentence relative to other words, and also takes into account the frequency of use. A concordance is a list of all words of any text with an indication of the contexts of their use [21].

The fundamental difference between concordances and dictionaries lies in the principles of selection, presentation and description of lexical units, in setting the purpose of the description. In addition, the dictionary approach to vocabulary description is focused on representativeness and normativeness, while the corpus version (concordance) is focused on exhaustive description; the dictionary seeks to find an invariant, concordance – a variant, and primarily reflects all cases of word use; the completeness of the dictionary is determined by the desire for an exhaustive description of the

meanings, and the completeness of the concordance is determined by the exhaustive nature of the description of the corresponding corpus. Because of this, there is a fundamental need for grammatical (morphological) information in the concordance, which helps to characterize and distinguish forms, and the need for semantic analysis (description or interpretation of meanings) in the dictionary.

The object and unit of dictionary description is the meaning and the dictionary article, and for the concordance it is the word usage and the corpus. Its completeness requires that it be a register-type dictionary.

In addition, the dictionary approach seeks to highlight some essence and interpret it. This is a normative explanatory dictionary, which contains a large amount of semantic information, semantic and stylistic markings, and the concordance reproduces the completeness of the thesaurus type (a dictionary that maximally but fully presents the words of the language with examples of their use in the text), it is the first derivative of the corpus, from list of words (dictionary), it is distinguished by the presence of morphological characteristics of word usage.

It is worth noting that all modern dictionaries are based on corpora, so there is also a feedback relationship – lexicographic and grammatical knowledge is reflected in the corpus in the process of creating a system of linguistic annotation of the corpus, but the dictionary and concordance are derivatives of the text – a "chopped test" [3].

The concordance makes it possible to analyze large arrays of text to identify patterns in the use of words or phrases in the language. Based on the obtained results of the concordance, it is possible to understand the meaning of the word from the context and carry out an analysis of its usage. Let's analyze the concordance of the word form on zero and the adjective zero.

На нулі [*na nuli*] (*At zero*)

Found 694 documents.

Of these, 50 documents cover the period up to 2022, all others for 2022, and most of them have the meaning "located closest to the front line, to the enemy". In the Dictionary of the Ukrainian Language in 11 volumes, the ambiguous word zero has the meaning: "1. The number 0, which means no value, and when it is substituted for another number on the right, it means increasing it tenfold.

Нуль уваги [*nul uvahy*] (*Zero attention*) – no attention.

2. A conditional value from which the deduction of values similar to it (time, temperature, pressure, etc.) begins.

3. In the pre-revolutionary school, the lowest score for the evaluation of knowledge and behavior.

4. transferred That which has no value and meaning» [2].

Concordance of the noun **At zero** (frontline) in **Figure 7** testify to the formation of a new meaning of the noun **at the zero** (frontline) of the "leading position":

*Guys are **at zero** (frontline), we need two Motorolas.*

*My friend, the major, was **at zero** (frontline) somewhere for 3 months, now he has moved to Lviv and is actually teaching paratroopers how to fight here.*

*For 102 brigades from Ivano-Frankivsk, which are "**at zero**"(frontline), it was possible to collect 80 thousand hryvnias quickly.*

*Many of them voluntarily joined the Armed Forces, spent months "**at zero**"(frontline) and during vacation have the opportunity to meet their families who are abroad.*

*Critical thinking is **at zero** (frontline), as are causal relationships.*

*"After six months **at zero** (frontline), it's hard to find something I'm not ready for," says the singer.*

*Let all troubles, bad mood and evil intentions of enemies be **at zero** (frontline) in the daily reporting.*

*The fact that everything is calm in Kyiv region now, there is no sense of war in the city, people live as usual in everyday life – go to the beach, shops, cinemas – this is all thanks to the guys who are **at zero** (frontline) now in hellish conditions and are taking the fight.*

*Thanks to our volunteers once again, who organize special courses, including online, where you can learn and master the big control of drones, learn some life hacks, how to use them **at zero** (frontline).*

	<input type="checkbox"/>	<> Хлопці на нулі , треба дві мотороли. </s>	
69	<input type="checkbox"/>	2022 • NONE • [Онлайн-ЗМІ «24 канал»] <> Мій друг майор десь був на нулі 3 місяці, тепер він переїхав у Львів і тут по суті навчає десантників, як воювати. </s>	
70	<input type="checkbox"/>	2022 • NONE • [Онлайн-ЗМІ «24 канал»] <> Для 102 бригади з Івано-Франківська, які перебувають " на нулі " , вдалося швидко зібрати 80 тисяч гривень. </s>	
71	<input type="checkbox"/>	2022 • NONE • [Онлайн-ЗМІ «24 канал»] <> Для 102 бригади з Івано-Франківська, які перебувають " на нулі " , вдалося швидко зібрати 80 тисяч гривень. </s>	
72	<input type="checkbox"/>	2022 • NONE • [Онлайн-ЗМІ «24 канал»] <> Багато хто із них добровільно пішов у ЗСУ, місяцями перебував " на нулі " і під час відпустки має отримати можливість зустрітися зі своїми родинами, які перебувають за кордоном. </s>	
73	<input type="checkbox"/>	2022 • NONE • [Онлайн-ЗМІ «24 канал»] <> Багато хто із них добровільно пішов у ЗСУ, місяцями перебував " на нулі " і під час відпустки має отримати можливість зустрітися зі своїми родинами, які перебувають за кордоном. </s>	
74	<input type="checkbox"/>	2022 • NONE • [Онлайн-ЗМІ «24 канал»] <> Критичне мислення – на нулі , як і причинно-наслідкові зв'язки. </s>	
75	<input type="checkbox"/>	2022 • NONE • [Онлайн-ЗМІ «24 канал»] <> " Після шести місяців на нулі важко знайти те, до чого я не готовий ", – каже співак. </s>	
76	<input type="checkbox"/>	2022 • NONE • [Онлайн-ЗМІ «24 канал»] <> Нехай в щоденній звітності всі біди, поганих настроїв і злі наміри недругів будуть на нулі . </s>	
77	<input type="checkbox"/>	2022 • NONE • [Онлайн-ЗМІ «Апостроф»] <> Те, що на Київщині зараз все спокійно, у місті не відчувається війна, люди живуть як завжди у повсякденному житті – ходять на пляж, у магазини, у кінотеатри – це все завдяки хлопцям, які на нулі зараз в пекельних умовах приймають бій. </s>	
78	<input type="checkbox"/>	2022 • NONE • [Онлайн-ЗМІ «Апостроф»] <> Велика подяка, знову ж таки, нашим волонтерам, які організують спеціальні курси, зокрема онлайн, де можна навчитися й опанувати саме керування дронами, дізнатися якісь лайфхаки, як їх використовувати на нулі . </s>	

Figure 7: Concordance of the noun *At zero* in the GRAK corpus

Нульовий [nulovyj] (Null)

Found 20179 documents.

The concordance of the adjective makes it possible to trace the formation of the new meaning "located closest to the front line." But the documents reflecting this meaning are much less than with the word form at zero in the same meaning.

According to the Dictionary of the Ukrainian Language in 11 volumes, zero has the meaning "which is equal to zero". [2]

6. Discussion

In the process of studying the development of polysemy, one cannot limit oneself to only the symbolic aspect of a word. In the process of studying the semantics of a word, it is impossible to be limited only to its symbolic and onomasiological aspect, that is, to the analysis of the logical and subject content of the word.

It is necessary to investigate how the experience fixed in the meanings of words is the basis of the words generated by speech and how a new reality enters the language as a fact of its system.

The relationship between language and speech is so important that one necessarily implies the other. Language is the source of all the subjective that penetrates the objective language system, it is the channel through which human practice is realized in speech. In this sense, speech is primary to language.

That is why corpus linguistics primarily reflects speech and is of great importance for the identification of lexical units.

The development of the meaning of a word includes the following indicators: frequency in the original corpus of texts; the number of lexical meanings, the sequence of meanings in the normative dictionary. And this is significant; the most important values are first recorded in the dictionary article, and then all the others in order of decreasing usage. The synchronous vector of a word makes it possible to obtain the general distributional and statistical characteristics of a word in a certain language field and to find trends in the development of its meanings.

The mechanism of determining the meaning of a word consists precisely in assigning a word to a certain paradigmatic series, to a conceptual field. This is evidenced by the fact that in many cases the lexeme is diagnostic, not connected to the analyzed word by a syntactic connection, but capable of signaling one of several possible paradigmatic series for this word.

For example, in the sentence *Війську потрібна свіжа кров, або Ротацію тим, хто «на нулі»* (<https://www.pravda.com.ua/columns/2022/12/20/7381565/>) [*Viisku potribna svizha krov, abo Rotatsiiu tym, khto «na nuli»*] (*The army needs fresh blood or Rotation to those who are "at zero"*) military vocabulary, although not syntactically related to the word *на нулі* [*na nuli*] (at zero), indicates that the last word should be understood not as "number", but as "to be on the front lines."

The specificity of the keyword has a great influence on the diagnosis of the meaning of the word. Not every word connected with another can indicate the meaning of the latter. Words with very wide connecting possibilities (lexemes with extremely developed polysemy and words with unclear meaning) have no diagnostic power at all. The appearance of a new meaning is determined by the context of the word's functioning and its internal semantic possibilities, because the core of meaning - a set of signs of something – remains in all phrases. The social context provides objects with such sets of features.

As Sidney Landau points out, a common problem when working with a citation index is the lack of examples of specific vocabulary; a common problem when working with the corpus is an excess of examples ("hits") [4].

The results of using corpora of language bases in lexical semantics demonstrate that the significance of this approach is not limited to corpus resources, that is, the material used by researchers. Today, corpus linguistics is acquiring more and more signs of a new linguistic paradigm. Despite the fact that the data of corpus linguistics mainly clarify the existing descriptions of linguistic phenomena, without radically changing them, the novelty and importance of this approach in lexical semantics should be seen in the fact that an empirical research base is formed and new opportunities for the analysis of paradigmatic and syntagmatic qualities of linguistic units.

7. Conclusions

Obtaining a modern model of the Ukrainian language system involves not only the analysis and description of the form and semantics of lexical units, their functioning in texts, but also the creation of reliable tools for monitoring the course of its dynamics. Internet and language corpora are the main source of information about a language in our days.

As we can see, an important feature of Internet data as a linguistic source is the possibility of direct access to contexts in which the use of the desired word is recorded. This is important for the transition from quantitative to qualitative characteristics, from an abstract invariant represented in the quantitative coefficient of a word to its semantic, stylistic, and other variants.

8. References

- [1] G. Miram. Translation Algorithms. Twin inter, Kyiv, 1998, 153 p.
- [2] Dictionary of the Ukrainian Language. URL: <http://sum.in.ua>
- [3] A. Taran. The Role of Keyword Language in the Database of World Slavic linguistics "iSybislaw", in: Computational Linguistics and Intelligent Systems. Proc. 6th Int. Conf. COLINS 2022. Vol. I (2022), pp. 266–276.
- [4] Sidney I. Landau. Dictionaries: The Art and Craft of Lexicography, K.I.S., Kyiv. 2012.
- [5] M. Shvedova, A. Rysin, V. Starko, Handling of Nonstandard Spelling in GRAC. 2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), Vol. 2, Lviv, Sept. 22–25 2021, pp. 105–108. doi:10.1109/CSIT52700.2021.9648834.
- [6] M. Shvedova, R. von Waldenfels, S. Yarygin, A. Rysin, V. Starko, T. Nikolajenko et al. GRAC: General Regionally Annotated Corpus of Ukrainian. Electronic resource: Kyiv, Lviv, Jena. 2017–2023. URL: <http://uacorpus.org>.
- [7] V. Busel (Ed.) A Large Explanatory Dictionary of Modern Ukrainian, Perun, Kyiv, 2005.

URL:<http://irbis-nbuv.gov.ua/ulib/item/UKR0000989>.

- [8] V. Starko. The Ukrainian Semantic Lexicon. 2022. URL: https://github.com/brown-uk/dict_uk/tree/master/data/sem.
- [9] O. I. Skopnenko, T. V. Tsybaliuk. Modern dictionary of foreign words, Dovira, Kyiv, 2006, 709 p.
- [10] J. Zhan, H. Zhao, Span Model for Open Information Extraction on Accurate Corpus. Proceedings of the AAAI Conference on Artificial Intelligence, 34(05), 2020, P. 9523-9530.
- [11] H. Hoherchak, N. Darchuk, S. Kryvyi, Representation, analysis, and extraction of knowledge from unstructured natural language text, Cybernetics and Systems Analysis, 2021, Volume 57, № 3., P. 164–183. <https://doi.org/10.1007/s10559-021-00373-7>
- [12] H. Hoherchak, Knowledge Based and Description Logics Applications to Natural Language Texts Analysis, Proceedings of the 12th International Scientific and Practical Conference of Programming (UkrPROG 2020), 2021, Volume 2866, P. 259–269.
- [13] D. Rothman, Transformers for Natural Language Processing (2nd addition), publishing Packt, 2021, pp. 384.
- [14] W. Che, Y. Lui, Y. Wang, B. Zheng, T. Liu, Towards Better UD Parsing: Deep Contextualized Word Embeddings, Ensemble, and Treebank Concatenation. – Proceedings of the {CoNLL} 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. – Association for Computational Linguistics, 2019, P. 55–64.
- [15] J. Sinclair. Lexical Grammar, in: Njujži Metodologija. 24 (2000). pp. 191–203.
- [16] D. Divjak. Mapping Between Domains. The Aspect-Modality Interaction in Russian, in : Russian Linguistics. 33 (3), 2009. pp. 249–269.
- [17] S. Th. Gries. Corpus-based methods and cognitive semantics: the many meanings of to run, in : S. Th. Gries, A. Stefanowich (Ed.), Corpora and cognitive linguistics: corpus-based approaches to syntax and lexis, Berlin, 2006. pp. 57–99.
- [18] Ch. Filmore. Corpus linguistics or computer-aided armchair linguistics, in: Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82. 1991. pp. 35–38.
- [19] M.-W. Devlin, K. Chang Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, NAACL-HLT, 2019, P. 4171–4186.
- [20] D. Glynn. Corpus-driven Cognitive Semantics Introduction to the field, : Dylan Glynn, Kerstin Fischer (Ed.), Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches, 2010, pp. 1–42. doi: [10.13140/RG.2.1.2325.1365](https://doi.org/10.13140/RG.2.1.2325.1365)
- [21] Ye. Karpilovska. Introduction to Computational Linguistics, TOV Yuho-Vostok, Ltd, Donetsk, 2003, 264 p.