Nataliia M. Bober

Kyiv National Economic University
named after Vadym Hetman,
Kyiv, Ukraine

METHODOLOGICAL CORPORA TOOLKIT AND ITS POSSIBILITIES FOR MODELLING COGNITIVE AND SEMANTIC MATRICES

Abstract

The article substantiates the necessity and effectiveness of involvement of corpus tools for studying the semantics of a word from the standpoint of interpretation of its cognitive nature, whose representatives have defended the encyclopaedic nature of meaning in general, unlike the views of scholars of classical structural semantics. In this connection, the correctness of Plungyan's hypothesis that linguistics "outlines the contours of a new model of language, which is significantly and fundamentally different from the former models postulated in the last quarter of the XX century," is commented on. Given this understanding of linguistic meaning and its role in presenting a new model of language, it has been suggested that it is important to study it in a broad and narrow context, in particular in terms of the combinatorial potencies of words – their lexical and grammatical compatibility, closely linked in corpus linguistics with such concepts such as collocations and colligations. The definitions of both terms have been clarified, and convincing arguments have been made in favour of the fact that collocations are conditionally free combinations of words used to characterize stereotypical situations and are displayed in the language of the native speakers in the form of ready phrases with inherent semantics, while colligations are limited by the morphological-syntactic frame of a certain structure. The methodological experience of corpus studies of colligations and collocations is analysed and proposed to be used to construct cognitive-semantic matrices of phrasal verbs in English. The main focus is on the capabilities of the Sketch Engine corpus system, in particular the availability of tools (Collocations, Word sketch, Thesaurus, Clustering, Sketch diff, etc.) that allow to integrate the classical (structural) method of distribution-statistical analysis of phrase-verbal collocations and colligations, and the method of lexico-semantic clustering, and the method of combinatorial syntagmatics. A hypothetical conclusion has been formulated that these and other procedural methods together will facilitate the disclosure of cognitive-

semantic connections between the units under study with quantitative and statistical calculations of their performance. It is proved that the corpus-oriented principle of combinatorial syntagmatics becomes the leading methodological principle of modern cognitive-interpretative semantics.

Key words: *modelling, corpus tools, cognitive-semantic matrix, phrasal verbs complexes, combinatorial syntagmatics, collocations, colligations.*

1. Introduction.

Cognitive semantics (J. Lakoff, R. Lenecker, L. Talmi, J. Taylor, C. Phillmore et al.) as a modern sphere of linguistic research has matured on the powerful ground of works performed within the framework of transformational (N. Chomsky and his school) grammar and, more broadly speaking, the structural one, developed in particular by American generativists, whose radical representatives (L. Bloomfield and others), according to Kucher, generally "ignored the problem of meaning because it did not fit into the rigid format of the then analysis of language forms" (Kucher, 2015: 6).

According to Toporov (2004), today the development of general semantics as an interpretative sphere of knowledge of native speakers has an "intensive character" as never before, "and it is in semantics that a "breakthrough" in new directions related to experimental study of meaning formation should be expected" (p. 7), as well as to experimental study of their motivational and pragmatic changes. In our deep conviction, such a "linguistic revolution" has already taken place when the *corpus linguistics* declared itself with its powerful methodological capabilities which opened up new perspectives for studying the meaning of a word on the basis of lexical-statistical fixation in concordances of corpora of its synchronous-systemic relations: paradigmatic, syntagmatic, word formation-motivational (associative), etc., as well as with the help of new methodological corpus tools. We assume that the corpus approach (Hvishiani, 2008, Rykov, 2002) will bring scholars closer to answering the questions of what the processes of perception, categorization, classification of phenomena of objective reality are, and how knowledge is accumulated, which systems provide for processing of information about various types of human activity (of course, reflected through the meaning of a word in its close relationship with other words).

It is clear that the study of word compatibility, first of all, its syntagmatic relations (distributive, valence, which are called *combinatorial* in the new cognitive semantics – see M. V. Vlavatska, N. M. Bober, S. G. Ter-Minasova, A. V. Korolyova et al.) requires taking into account the notion of context (situation), wide and narrow, against which the meaning of a word, that is, its interpretation given in explanatory dictionaries or other lexicographic sources, is clarified. At present, the linguistic corpus, which is a collection of texts of different forms (oral or written) of discourses and genres, compiled according to established principles and standardized, can be considered to be a wide context. But the main advantage of the linguistic corpus is that it is equipped with specialized search engines (Corpus linguistics URL: http://corpora.iling.spb.ru/theory.htm).

And despite some skeptical views on corpus capabilities (discussed in one of our publications (Bober, 2018)), it is still worth agreeing with the statement that "the national corpus of language today is both a base and a tool for linguistic research, and the third compulsory format for presenting linguistic knowledge and language in general, following traditional grammars and dictionaries" (Komarova, 2010: 15).

Based on this methodological formulation of the question, we should agree with Sosnina (2012), whose views are in agreement with the ideas of Dronova (2009: 120) in that "[…] the considerable advances of modern semantics make it possible to solve the problems of expanding the field of research (methods and procedures) in the study of the meaning of a word from the standpoint, first of all, of interpretive linguistics (as linguocognitive studies position themselves) [...] (p. 19) in its alliance with corpus technological capabilities.

2. Aim.

The **aim** of the article is to characterize the methodological tools of text corpora and their possibilities for modelling cognitive-semantic matrices.

3. Cognitive Modelling of Colligations and Collocations.

In linguistics, the end of the XX c. was generally marked by the reorientation of research attention from grammar (morphology and syntax) to the lexical-semantic system, and the beginning of the twenty-first century, by the active involvement of corpus technologies that have helped to revise traditional views of the language as a whole. As a consequence, it has been suggested that linguistics "outlines the contours of a new language model, which is significantly and fundamentally different from the earlier models postulated in the last quarter of the twentieth century" (Plungyan, 2008: 7–20).

In order to correctly understand in what scientific interpretation Plungyan uses the term "new language model", we briefly formulate our own vision of the concept of "language model", as well as try to understand the cognitive nature of such a model and its mechanisms, which are obviously based the ability of a native speaker of a particular language to reflect the ontology of that entity, using its multiple levels, and, to be more precise, to perform various combinations of linguistic characters (both in terms of expression and in terms of content) to provide efficiency of communication.

And precisely for the solution of such linguistic problems in the dynamic peak of the development of structuralism, when the procedures of modelling process in phonology and in syntax were actively developed (and semantics, on the contrary, remained the most debatable sphere at that time), Z. Harris introduces the term "model" in the scientific apparatus of linguistics, borrowing it from philosophical works devoted to the problems of mathematical logic (cit. Apresyan, 1966: 99–100). These priorities are fundamental for the development of new directions in structuralism – *mathematical linguistics and quantitative linguistics*.

Among the many definitions of the term "model" that exist today, the most optimal is philosophical interpretation of this term, which means an "object, artifact (artificially created scheme), which serves to present reality, reflects a set of features of existing or hypothetical objects and depends on the tasks formulated and solved by a researcher (its designer)" (Lukach, 2013: 144). The above-mentioned definition does not contradict the views of Losev (2004) on the methodological process of modelling, which in his understanding involves the construction, first, of simulation models of really existing objects and phenomena, and, second, of analytical models of hypothetical objects for predicting their functional signs.

A deeper study of modelling problems in the field of artificial intelligence has directed the attention of scholars to discovering the mechanisms of human consciousness. And designing models of virtual realities has changed in general scientific views about the limits of reality and led to a revision of the principles of construction of phenomena of model nature. As a result of these observations, the term "model / modelling", and most importantly, the role of language in the processes of constructing model objects were redefined.

Actually, the hypotheses, which assumed that language is the mediator between the world and man, were discussed in the XIX c., beginning with the works of Humboldt and Potebnya, but they gained theoretical justification only in the XX century, and first of all, when a new understanding of language in cognitive science appeared in connection with the study of *cognitive models* – structures that reflect the perception, storage and transmission of information about the world (Kubryakova, 1996: 56–57).

Directly in science, a cognitive model is studied as an epistemological construct – a hypothesis about the structure of human consciousness and thinking. But in case of empirical confirmation, this model can acquire ontological status.

Representatives of cognitive semantics (J. Lakoff and others) distinguish two types of basic models: models of identification (or models of categorization) and models of mentalisation (i.e. conceptualisation) (Korolyova, 2012). Cognitive models of the first type reflect the process of segmentation from the holistic image of the world of individual most significant objects in order to further categorize them. Mentalisation models are closely linked to linguistic structures because they are a direct reflection of cognitive and purely linguistic semantics.

The question of how knowledge is represented in human memory is related to both processes: categorization and conceptualization, where the latest symbolic models provide the storage of knowledge of language, while the first models allow us to re-explain the essence of lexical and grammatical phenomena (Piaget, 1983).

Taking into account the new methodological provisions of the theory of categorization as a process of modelling the phenomena of objective reality, Lewis's statements that the language is a grammaticalised lexis, not a lexicalised grammar, seem to be correct (Lewis, 1999). In this case, cognitive-grammatical ability is considered, in particular, as an opportunity to form unique combinations of lexical units to describe and interpret non-frequent situations / cases, while the ready-made vocabulary already stores in the memory lexical combinations to designate standard and highly probable events (see more: Horina, 2014: 20). Of course, grammatical and lexical combinatorial models are reduced copies of two basic cognitive models.

These issues have recently been actively discussed by representatives of interpretive or cognitive semantics, who suggest that the combinatorial nature of character compatibility is most effectively studied with the involvement of corpus managers / corpus tools.

Compatibility of words, as Horina (2014) writes in her thesis (p. 17), citing Hvishiani (1979), should be studied on the basis of the cognitive unity of collocation and colligation, since the correct construction of speech is impossible only with the observance of grammatical valence, or, in terms of the corpus linguistics, – the colligative correctness.

Hvishiani (1979) explains his assumption by the fact that "the idea of the areas and rules of word usage is realized by a speaker under the influence of functioning of a certain number of phrases with that word. And in terms of a speech community, these phrases meet all the criteria of the language norm (including the sociolinguistic aspect), that is, their language status is recognized" (p. 172).

This methodological principle has led to a new vector in the study of problems of semantics: from the analysis of the meaning of an individual word (in explanatory dictionaries) to its compatibility with other words (both content words and function words) in phrases (including phrasal verbal complexes in English), called multiword expressions (MWE) in the corpus linguistics.

Due to such a turn, studies of the action of two processes in their unity were updated: *colligative* (a set of morphological-syntactic conditions that ensure the compatibility of linguistic units) and *collocational* (lexical-phrase and lexical-phraseological) combinatorics (semantic syntagmatics). Combinatorial syntagmatics becomes the leading methodological principle of cognitive-interpretive semantics. Ter-Minasova (2007; 2008) points it out in her writings and states that at present "the study of a word in only colligative manner, that is, mainly from the standpoint of its grammatical valence, is insufficient" in order to explain communicative failures and even the phenomenon of cognitive shock.

Sinclair (1996), the developer and manager of the first corpus programs for the presentation of speech samples, which were part of COBUILT (Collins Birmingham University International Language Database) project, was of the same opinion. We fully share his view that units of meaning are already embedded in ready phrases, meaning that mini-models of potential combinations of characters due to their cognitive character have the nature of ready phrasal formations reflected in consciousness of native speakers of every language (such as phrase verbs in English). That is why corpus tools include a specially designed mechanism for the automatic separation of collocations and colligations from texts.

It has long been proven, and mentioned above, that words do not function on their own, out of context, unless, of course, they are very rare names or terms. The vast majority of words are used in an already similar environment and are therefore schematised / stereotyped in the mind of a speaker and await the recurring events of their further connectivity, which are called collocations in the corpus linguistics (Sinclair, 1996). And, as experience shows, native speakers much more commonly use ready-made, commonly used templates, stencils, clichés, fixed phrases in everyday speech, rather than generate unique, unprecedented (in lexical and grammatical structure) phrases (Zymnya, 1972).

In order to understand the cognitive nature of combinatorics as a phenomenon of compatibility in the system of each language, it is necessary to find out the scientific scope of the terms "colligation" and "collocation".

The first term "colligation" usually refers to the grammatical rules of compatibility, and therefore is used to designate the grammatical environment to which the word / sequence of words refers (grammatical company the wordkeeps) (Vlavatska, 2011, Flowerdew, 2012), or vice versa (avoids keeping). At that time, it was introduced exactly in this sense into the scientific circulation by the representatives of logical grammar to study the combinatorial capabilities of any part of language (this is what Katsnelson called valence (see: Bober, 2018)). In the most recent works on combinatorial syntagmatics, it was used the aforementioned Vlavatska (2011) when performing a valence analysis of verb compounds in English. As a result of this analysis, she developed a classification of combinatorial models of English verbs based on their valencies.

In the methodological tradition of the London Structural School, its founder, Firth (1957), actively used two terms: collocation and colligation, and believed that any meaning of a word depends on its contextual environment. However, he mostly studies the meaning of a word in collocations.

Since then, the concept of collocation has been linked to the textual influence on the choice of vocabulary to form a compound or a phrase. From this consideration, it follows that the very context of a particular situation in lexical terms can also be referred to as "collocation" in the sense of the typical and constant environment of a particular word. As a rule, collocation is considered through the relationship between individual lexical elements, and collocation is considered as the relationship between elements of a particular construction. According to Vlavatska (2011), collocations, that is, word compounds occupy an intermediate position and are at the intersection between lexicology and phraseology, as they are peripheral units both "for lexicology, which deals with mostly free lexical compatibility (syntagmatics), and for phraseology", the object of which are connected units, that is, idioms" (p. 19–26).

But there is another opinion on this matter by Ter-Minasova (2009), who is worth agreeing, and who considers that all phrases in languages are motivated sociolinguistically, and denies the existence of absolutely free phrases that mechanically form a productive abstract construction, i.e. a superficial grammatical model of their creation, or a block diagram.

4. *Corpus Methods and Their Effectiveness for Construction of Cognitive –*
Semantic Matrices.

This discussion motivates and actualizes the development of new methods for analysing combinatorial compatibility rules for units such as phrasal verbs in the English language, which until recently have been considered a mechanical combination of a verb and a postpositive component, resulting in a certain construct with a constant meaning. For this purpose, classical structural techniques (distributive analysis, component analysis) were mainly used, accompanied by the statistical method. At present, the corpus methodological experience makes it possible to study the combinatorics of phrasal verbs as collocations and simultaneously colligations in the aspect of their cognitive-motivational relations.

We would like to draw attention to the methodological approach developed in the works of Horina (2014: 23), who proposes to involve for the analysis of such entities, first and foremost, translation techniques, including transformations, and corpus tools. Of course, translation transformations are effective for studying the combinatorics of phrasal-verb collocations by non-native speakers. The scholar suggests that collocations are more quickly fixed in the memory of a representative of another linguistic and cultural environment through the adequate translation into their native language. At the same time, those phrasal collocations the meaning of which substantially and fundamentally differs in the source language and the language of translation should be semantised by means of a corpus contextual environment, which will facilitate the disclosure of combinatorial (syntagmatic and sometimes conceptually-integrative) relations of a word in the most typical situations of its use with the highest frequency. Obviously, the corpus methodical algorithm will be a relevant tool for studying collocations and colligations of phrasal verbs and for English-speaking scholars.

The complex corpus method developed by Horina (2014) makes it possible to analyse the combinatorial potencies of words in two directions: internal collocation (as a phrase, word combination) and external collocation – in the corpus context (p. 32). The main tool in such a procedure is corpus concordance per word, that is, the word valence strings arranged in the right and left context of the query. Such a graphical representation allows for vertical reading or scanning of information, and immediately allows to see patterns in word connections, grammatical preferences and other important information.

But for revealing the cognitive-matrix relations between collocations and colligations, for example, of phrasal verbal complexes in English, the Sketch Engine corpus system seems to be especially productive, first of all, due to the presence in it of tools that implement both the classical method of distributive-statistical analysis of them and the methodology of lexico-semantic clustering, and the technique of combinatorial syntagmatics. These include such tools as Collocations – automatic collocation search (the so-called broad context when potential links between phrase complexes throughout the corpus are measured); Word sketch – automatic colligation search (within a given phrasal verb formation – its morphological-syntactic mini-model/formula-template; it is this tool that creates a list of the most established compounds calculated on the basis of the logDice statistical measurement separately for each morphological-syntactic formula of a phrasal verbal complex, and also issues estimates of the total number of such compounds / combinations throughout the corpus); Thesaurus – to establish system relationships; Clustering – to group selected thesaurus units into appropriate clusters (i.e. LSG) with subsequent profile building; Sketch diff – to identify similarities and differences in the mechanisms and methods of combinatorial compatibility of word pairs.

Taken together, these and other procedural methods make it possible to reconstruct cognitive-semantic connections between the units under study with quantitative and statistical calculations of their performance (Zakharov, 2015: 128).

5. *Steps of Working with Corpus Tools in the Process of Analysis of Phrasal Verbs Complexes.*

Direct work with the corpus toolkit involves performing the following steps: 1) processing of concordance, 2) calculating of absolute frequency, 3) analysing of the left and right valence (in this case, verb collocate), 4) modelling of clusters to construct cognitive-semantic profiles of units under study.

The corpus concordance processing tools traditionally include generating its string or combinatorial string. In other words, when you type a word in the search string, the corpus manager generates a so-called automatic rendering like KWIC (key word in context), i.e. the ordered string of the left-side and right-side context of the inquiry of word combinatorics. Such organization of information about the word under study allows to trace the frequency of its use in a variety of contexts. In addition, the corpus tools allow to measure not only the frequency of word use, but also to build a frequency chart by genres of the corpus. Vertical reading/scanning of concordance allows to find left-side and right-side verb collocations in the right contexts, as well as to construct cognitive-semantic profiles of each verbal-phrase complex. Further involvement of the methods of conceptual integration of mental spaces will contribute to the presentation of a cognitive-matrix model of phrasal verbs with all their profiles.

In the process of finding collocations, it is possible to set to the program an operation to determine parts of speech of colligations, to calculate the number of words between, for example, the phrasal verb under study and its environment. As a result, words that occur in the closest environment more often than others are automatically displayed in the Concordance window. When making an inquiry, it is also possible to choose a type of genre in which the desired word works, as well as to select the search area, that is, to specify at what distance from the word collocates are searched. As a rule, corpus software capabilities provide the following search area: one to five words to the left and right of the word under study.

And the available "extended context" option will not only give a string of compatibility, but also a few sentences where the word, phrase, and chunk are used. An important step in the corpus procedure is to analyse the ratio of word forms to the number of tokens (type / token ratio). This is an indicator of lexical variability of a text.

In order to use the capabilities of corpus tools, it is necessary to get familiar with the syntax of the corpus inquiry and generally master the corpus competence. At the same time, it seems to us, that a scholar who is already trained and aware of laws not only of corpus linguistics, but, above all, of settings and laws of mathematical linguistics. For example, a scholar should already have knowledge of the regularities of linear organization of text syntax, in particular that words closely related to the syntactic dominant tend to be placed next to it. This principle (of saving language effort) (Zipf, 1949: 309) was formulated in quantitative linguistics by Zipf, the American scientist, in 1949, and was named as Zipf's Law (an empirical law that relates word rank in a frequency dictionary to word frequency) (Zipf, 1949: 484–490), which is in line with the equilibrium principle of Vilfredo Pareto, according to which any resources are self-organized so as to minimize efforts for the work done. Accordingly, 20–30% of the resource form 70–80% of the aggregate result.

For example, according to the observations of Martynenko (2015), 20% of the most frequently used words usually make up 80% of word usage of a text. Zipf's law, like Pareto's law, formulated for rank distribution, has the form of a non-equilibrium hyperbola formula:

$$k_{r} = \frac{k_{max}}{r^{\gamma}}$$

where r is word rank, k_r means frequency of a word of rank r, k_{max} means frequency of the most frequent word, and γ means a factor that characterizes the unevenness (scattering and concentration) of the frequency distribution (p. 22–23).

6. Conclusions.

In general, when evaluating the prior performance of corpus tools for constructing cognitive-semantic matrices, we can see its advantages in detecting tendencies for combinatorial word compatibility in real and stylized speech situations as compared to the existing lexical and grammatical norms of compatibility. The Sketch Engine corpus system demonstrates the greatest potential, because it has tools that allow for the processing of material on the basis of classical methods of distributive-statistical analysis, methods of lexico-semantic clustering, and methods of combinatorial syntagmatics. These include such corpus tools such as Collocations, which provide for automatic collocation search (the so-called broad context when measuring potential links between, for example, phrasal verbs throughout the corpus); Word sketch – automatically searches for a given phrasal verbal formation – its morphological-syntactic mini-model/formula-template; Thesaurus – allows to construct systemic relationships between the units under study; Clustering – grouping of selected thesaurus units into appropriate clusters (i.e. LSG) with subsequent modeling of respective profiles; Sketch diff – allows to identify similarities and differences in the mechanisms and methods of combinatorial word pair combinability.

Taken together, these and other procedural techniques make it possible to reconstruct cognitive-semantic connections within the matrix with quantitative and statistical calculations of performance of the units under study using mathematical formulas, in particular, based on the laws of Zipf and Pareto.

Prospects for further research are approbation of the developed corpus methodology for the analysis of phrasal verb complexes in English and construction of their cognitive-semantic matrix.

R e f e r e n c e s

Apresyan, Yu. D. (1966). *Idei i metody sovremennoy strukturnoy lingvistiki (kratkiy ocherk)* [Ideas and methods of modern structural linguistics (short essay)]. M. : Prosveshcheniye.

Barlow, M. (1996). Corpora for theory and practice. *International Journal of Corpus linguistics*, 1(1). P. 1–37.

Bernadini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In Sinclair J. McH. (ed.) *How to use corpora in language teaching*. Amsterdam [u.a.]: Benjamins.

Bober, N. M. (2018). English Phrasal Verbs as Cognitive and Semantic Complexes and Fragment of Multilateral Knowledge of Matrix Format. *Scientific Journal of National Pedagogical Dragomanov University. Series 9. Current Trends in Language Development*, 18. P. 22–32. DOI: https://doi.org/10.31392/NPU-nc.series9.2018.18.02

Dronova, L. P. (2009). Sinkhroniya i diakhroniya: otlozhennaya vstrecha [Synchrony and diachrony: delayed meeting]. *Vestn. Tom. gos. un-ta. Filologiya*, 3(7). P. 116–123.

Fillmor, Ch. (1988). Freymy i semantika ponimaniya [Frames and the semantics of understanding]. *Novoye v zarubezhnoy lingvistike*, 23. S. 52–92.

Firth, J. (1957). *Papers in Linguistics: 1934–1951*. Oxford: Oxford Univ. Press.

Flowerdew, L. (2012). *Corpora and Language Education*. S. l.: Palgrave Macmillan.

Gorina, O. G. (2014). Ispolzovaniye tekhnologiy korpusnoy lingvistiki dlya razvitiya leksicheskikh navykov studentov-regionovedov v professional'no-oriyentirovannom obshchenii na angliyskom yazyke [The use of corpus linguistics technologies for the development of lexical skills of area-studies students in professionally oriented communication in English]: Thesis. Sankt-Peterburg.

Gvishiani N. B. (1979). *Polifunktsional'nyye slova v yazyke i rechi* [Multifunctional words in language and speech]. Moskva: Vyssh. Shkola.

Gvishiani, N. B. (2008). *Praktikum po korpusnoy lingvistike* [Workshop on Corpus Linguistics]. Moskva: Vysshaya shkola.

Komarova Z. I. (2010). Problemy yazyka nauki [Problems of Language of Science]. *Aktualnyye problemy germanistiki, romanistiki i rusistiki*. Yekaterinburg: Ural. gos. ped. un-t.

Korolyova, A. V. (2018). Combinatorial Syntagmatics: from the Theory of Valency to the Theory of Conceptual Integration // Scientific Journal of National Pedagogical Dragomanov University. Series 9. Current Trends in Language Development. V. 17. 99–111.

Korolyova, A. V. (2012). Protsesy kontseptualizatsiyi i katehoryzatsiyi yak rezultaty piznavalnoyi i klasyfikatsiynoyi diyalnosti lyudskoyi svidomosti i myslennya [Processes of conceptualization and categorization as results of cognitive and classification activities of human consciousness and thinking]. *Movna systema yak rezultat vidobrazhennya protsesiv kontseptualizatsiyi i katehoryzatsiyi navkolyshn'oho svitu*: kolektyvna monohrafiya / za red. A. V. Korolovoyi. Kyiv : Hileya.

Korpusnaya lingvistika (2008) [Corpus Linguistics]. *In-t lingv. issledovaniy RAN*. Retrieved from: http://corpora.iling.spb.ru/theory.htm

Kubryakova, Ye. S. i dr. (1996). *Kratkiy slovar kognitivnykh terminov* [Short Dictionary of Cognitive Terms]. Moskva : filol. f-t MGU imeni M. V. Lomonosova.

Kucher, I. A. (2015). *Linhvokohnityvne modelyuvannya leksyko-semantychnoho polya diyesliv rukhu u finskiy i ukrayinskiy movakh* [Linguistic-cognitive modeling of lexical-semantic field of movement verbs in Finnish and Ukrainian]: Thesis. Kyiv.

Lakoff, G. (1990). The invariance hypothesis: is abstract reason based on image-schemas? *Cognitive linguistics*, 1(1). P. 39–74.

Langacker, R. W. (2008). *Cognitive grammar. A basic introduction*. N. Y. : Oxford University Press.

Lewis, M. (1999). *The Lexical Approach: The state of ELT and a way forward*. Hove / England: Language Teaching Publications.

Losev, A. F. (2010). *Vvedeniye v obshchuyu teoriyu yazykovykh modeley* [Introduction to the general theory of language models]. 3-ye. izd. M. : Editorial URSS.

Lukach, M. O. (2013). Typy linhvistychnykh modeley ta yikh zastosuvannya dlya rozvyazannya linhvistychnykh zadach [Types of linguistic models and their application for solving linguistic problems]. *Visnyk Natsionalnoho universytetu "Lvivska politekhnika"*, 770. Informatsiyni systemy ta merezhi.

Martynenko, G. Ya. (2015). Stilemetriya: vozniknoveniye i stanovleniye v kontekste mezhdistsiplinarnogo vzaimodeystviya [Stylometry: Emergence and Formation in the Context of Interdisciplinary Interaction]. *Strukturnaya i prikladnaya lingvistika*, 11.

Piazhe, Zh. (1983). Skhemy deystviya i usvoyeniye yazyka [Schemes of action and language acquisition]. *Semiotika*. Moskva : Raduga.

Plungyan, V. A. (2008). Korpus kak instrument i kak ideologiya: O nekotorykh urokakh sovremennoy korpusnoy lingvistiki [Corpus as a tool and as an ideology: On some lessons of modern corpus linguistics]. *Russkiy yazyk v nauchnom osveshchenii*, 2(16). P. 7–20.

Rykov, V. V. (2002). Korpus tekstov kak realizatsiya oyektno-oriyentirovannoy paradigmy [Corpus of Texts as an Implementation of the Object-Oriented Paradigm]. *Kompyuternaya lingvistika i intellektualnyye tekhnologii*, 1. P. 390–393.

Sinclair, J. (1991). *Corpus: Concordance and Collocation*. Oxford : Oxford Univ. Press.

Sinclair, J. (1996). The search for units of meaning. *Textus*, 9(1). P. 75–106.

Sinclair, J. M., Mauranen, A. (2006) *Linear Unit Grammar: Integrating Speech and Writing: Studies in Corpus Linguistics*. Amsterdam / Philadelphia : John Benjamins.

Sosnina, A. A. (2013). Ideosemantika kak odin iz komponentov etimologicheskogo analiza [Ideosemantics as one of the components of etymological analysis]. *Kritika i semiotika*, 2(19). P. 14–30.

Sternin, I. A. (2013). *Metody opisaniya semantiki slova* [Methods for describing the semantics of the word]. M. : "Istoki".

Strukturnaya i prikladnaya lingvistika [Structural and applied linguistics], 11. SPb. : Izd-vo S.-Peterb. un-ta, 2015.

Talmy, L. (1985). Lexicolazation patterns: semantic structure in lexical forms. In Shopen T. *Language typology and syntactic description*. Cambridge : Cambridge University press. P. 57–149.

Ter-Minasova, S. G. (2007). *Slovosochetaniye v nauchno-lingvisticheskom i didakticheskom aspektakh* [Word combination in scientific-linguistic and didactic aspects]. 3-ye izd. M. : Izd-vo LKI.

Ter-Minasova, S. G. (2008). *Voyna i mir yazykov i kultur* [War and the world of languages and cultures]. M. : Slovo.

Ter-Minasova, S. G. (2008). *Yazyk i mezhkulturnaya kommunikatsiya* [Language and intercultural communication]. M. : Slovo.

Ter-Minasova, S. G. (2009). Sintagmatika rechi: ontologiya i evristika [Syntagmatics of speech: ontology and heuristics]. *Obshchaya i angliyskaya sintagmatika sostavnykh nominativnykh grupp*. 2-ye izd. M. : Knizhnyy dom "Librokom".

Toporov, V. N. (2004). *Issledovaniya po etimologii i semantike* [Studies in etymology and semantics]. Tom I. Teoriya i nekotoryye chastnyye yeye prilozheniya. Moskva : Yazyki slavyanskoy kultury.

Vlavatskaya, M. V. (2011). Ponyatiya kollokatsii i kolligatsii v diakhronicheskom rassmotrenii [The concepts of collocation and colligation in diachronic consideration]. *Aktualnyye problemy filologii i metodiki prepodavaniya inostrannykh yazykov: mater. mezhdunar. nauch.-prakt. konf*. P. 19–26.

Zakharov, V. P. (2015). Korpusno-oriyentirovannyy podkhod k postroyeniyu tezaurusov i ontologiy vzaimodeystviya [A corpus-oriented approach to the construction of thesauruses and ontologies of interaction]. *Strukturnaya i prikladnaya lingvistika*, 11. P. 124–141.

Zimnyaya, I. A. i dr. (1972). *Osnovnyye napravleniya v metodike prepodavaniya inostrannykh yazykov v XIX–XX c.* [Main directions in the methodology of teaching foreign languages in the XIX and XX centuries]. Moskva : Pedagogika.

Zipf G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, (Mass.): Addison‑Wesley.

Бібліографічний опис:

Анотація

У статті обґрунтовано необхідність й ефективність залучення корпусного інструментарію для вивчення семантики слова з позицій витлумачення його когнітивної природи, представники якої обстоювали енциклопедичний характер значення загалом, на відміну від поглядів учених класичної структурної семантики. У зв'язку з цим прокоментовано коректність гіпотези В. О. Плунгяна про те, що в лінгвістиці "окреслюються контури нової моделі мови, яка значно і принципово відрізняється від колишніх моделей, постульованих в останній чверті XX ст.". З огляду на таке розуміння мовного значення та його ролі у представленні нової моделі мови, зроблено припущення щодо важливості його вивчення у широкому й вузькому контекстах, зокрема у плані комбінаторних потенцій слів – їх лексичної і граматичної сполучуваності, тісно пов'язаної у корпусній лінгвістиці з такими поняттями, як колокації й колігації. Уточнено визначення обох термінів й наведено переконливі аргументи на користь того, що колокації є умовно вільними комбінаціями слів, що вживаються для характеристики стереотипних ситуацій й відображаються у свідомості носіїв мови у вигляді готових фраз з властивою для них семантикою, тимчасом колігації обмежені морфолого-синтаксичними рамками певної конструкції. Проаналізовано методичний досвід корпусних досліджень і колігацій, і колокацій й запропоновано застосувати його для побудови когнітивно-семантичних матриць фразових дієслів англійської мови. Основну увагу зосереджено на можливостях корпусної

системи Sketch Engine, зокрема на наявності в ній тих інструментів (Collocations, Word sketch, Thesaurus, Clustering, Sketch diff та ін.), які дають змогу комплексно поєднати і класичну (структурну) методику дистрибутивно-статистичного аналізу фразово-дієслівних колокацій і колігацій, й методику лексико-семантичної кластеризації, і методику комбінаторної синтагматики. Сформульовано гіпотетичний підсумок про те, що ці та інші процедурні методики разом сприятимуть розкриттю когнітивно-семантичних зв'язків між досліджуваними одиницями з кількісними й статистичними обчисленнями їхньої продуктивності. Доведено, що корпусно-орієнтований підхід комбінаторної синтагматики стає провідним методологічним принципом сучасної когнітивно-інтерпретаційної семантики.

Ключові слова: моделювання, корпусний інструментарій, когнітивно-семантична матриця, фразово-дієслівні комплекси, комбінаторна синтагматика, колокації, колігації.