Zhytomyr Ivan Franko State University Journal. Philological Sciences. Vol. 1 (99)

Вісник Житомирського державного університету імені Івана Франка. Філологічні науки. Вип. 1 (99)

ISSN (Print): 2663-7642
ISSN (Online): 2707-4463

УДК 81'324.36.42
DOI 10.35433/philology.1(99).2023.93-104

QUANTITATIVE CORPUS-BASED METHODS FOR CONSTRUCTION GRAMMAR RESEARCH

V. V. Zhukovska*

With the rapid proliferation of computerized language corpora, modern linguistics is becoming more data-driven and corpus-based, and quantitative linguistic research is gaining popularity. The methodology for linguistic analysis is being improved, and statistically reliable tools are being actively employed to test scientific hypotheses and theories. Cognitive-quantitative construction grammar is a new interdisciplinary study field that results from the integration of qualitative and quantitative linguistic approaches. Drawing on research from various linguistic fields, this framework comprehensively studies general and idiosyncratic features of language constructions – 'form-meaning' pairings. This article provides an overview of a number of quantitative corpus-based methods that can be used in cognitive-quantitative construction grammar research to establish distinctive linguistic parameters (factors) of form and meaning of complex clause-level constructions. Clause-level constructions are distinguished by a set of linguistic parameters and factors that determine the specificity of their form (morphosyntactic, relational, referential, positional, syntactic-functional, distributive, etc.) and meaning/ function (collostructional, semantic, communicative, etc.). To determine distinctive parameters of the form of a construction, specialized quantitative corpus methods can be employed: 1) one-factor analysis of variance (ANOVA), 2) multifactor analysis of variance (MANOVA) and 3) post-hoc test by Tukey's multiple comparison method. To identify distinctive parameters of construction's semantics, the method of collostructional analysis can be utilized. By integrating corpus linguistics and quantitative studies, a new level of comprehension can be attained, which makes it possible to analyze linguistic data more accurately. Quantitative corpus-based methods should be considered an effective tool for construction grammar research, allowing linguists to move beyond intuition and gain a deeper understanding of the complexity of language phenomena.

Keywords: *language corpus, quantitative methods, linguistic parameter, grammatical construction*

* Candidate of Sciences (Philology), Professor
(Zhytomyr Ivan Franko State University),
victoriazhukovska@gmail.com
ORCID: 0000-0002-4622-4435

КВАНТИТАТИВНІ КОРПУСО-БАЗОВАНІ МЕТОДИ В ДОСЛІДЖЕННЯХ ІЗ КОНСТРУКЦІЙНОЇ ГРАМАТИКИ

Жуковська В. В.

Зі стрімким поширенням комп'ютеризованих мовних корпусів сучасна лінгвістика стає більш керованою даними та корпусоорієнтованою, а квантитативні дослідження мови набувають усе більшої популярності. Методологія лінгвістичного аналізу вдосконалюється, а статистично надійні інструменти активно застосовують для перевірки наукових гіпотез і теорій. Когнітивно-квантитативна граматика конструкцій – це новий міждисциплінарний напрям досліджень, який є результатом інтеграції квалітативного та квантитативного лінгвістичних підходів. Інтегруючи теоретичні та методологічні настанови різних лінгвістичних галузей, цей напрям комплексно вивчає загальні та ідіосинкратичні особливості мовних конструкцій – пар "форма-значення". У цій статті подано огляд низки квантитативних корпусо-базованих методів, які можуть бути використані в когнітивно-квантитативних конструкційних дослідженнях для встановлення дистинктивних лінгвістичних параметрів (факторів) форми й значення складних конструкцій рівня клаузи. Конструкції рівня клаузи вирізняються набором лінгвальних параметрів і факторів, які визначають специфіку їхньої форми (морфосинтаксичні, реляційні, референційні, позиційні, синтаксично-функційні, дистрибутивні тощо) та значення / функції (колострукційні, семантичні, комунікативні тощо). Для визначення дистинктивних параметрів форми конструкції можуть бути застосовані спеціалізовані квантитативні корпусобазовані методи: 1) однофакторний дисперсійний аналіз (ANOVA), 2) багатофакторний дисперсійний аналіз (MANOVA) та 3) постхок-тест за методом множинного порівняння Тьюкі. Для виявлення відмінних параметрів семантики конструкції може бути використаний метод колострукційного аналізу. Поєднання корпусної та квантитативної лінгвістики дає змогу досягти нового рівня розуміння, що уможливлює точніший аналіз лінгвістичних даних. Квантитативні корпусобазовані методи потрібно розглядати як ефективний інструментарій граматики конструкції, що дає змогу лінгвістам вийти за межі інтуїції й глибше зрозуміти сутність мовних явищ.

Ключові слова: *корпус мови, квантитативні методи, лінгвальний параметр, граматична конструкція*

Introduction. The development of contemporary cognitive linguistics, and construction grammar in particular, is accompanied by a discussion on enhancing the objectivity of data and a search for methods to improve the precision of research. As a result, the methodology for linguistic analysis is being refined, and statistically reliable tools are being actively employed to validate scientific hypotheses and theories [5; 12; 19]. Traditional methods of linguistic analysis are supplemented by quantitative methods developed and tested by corpus and experimental linguistics [7; 10].

Analysis of previous research. Obviously, the synthesis of qualitative and quantitative analyses of linguistic phenomena is becoming a reliable approach for obtaining objective data about language. This trend has become especially prominent in recent years, with the publication of a large number of linguo-cognitive studies based on a broader and more balanced empirical foundation, involving large amounts of natural language data [3; 30] and employing sophisticated methods of inferential statistics [9; 27; 28; 32].

The integration of qualitative and quantitative approaches gives rise to the new interdisciplinary area of linguistics, which combines cognitive linguistics (Langacker (1987; 1991); Janda, (2013)), cognitive construction grammar (Goldberg (1995; 2006; 2019)), usage-based construction grammar (Hoffmann (2016)*;* Hilpert (2019)), quantitative linguistics (Levytskyi (2007), Perebyinis (2013)), quantitative corpus linguistics (Gries, Stefanowitsch (2004, 2013); Brezina (2018)), and automatic speech processing (Darchuk (2013) into one study field – cognitive-quantitative construction grammar (CQCxGr). By drawing upon research from various fields, this new interdisciplinary framework brings together a variety of different methods to comprehensively explore general and idiosyncratic features of language *constructions*[5] – 'form-meaning' pairings, holistic semiotic models of varying degrees of schematicity and complexity [15; 24; 26]. CQCxGr aims at providing an explanation for the semiotic phenomena of language and speech on their mental basis, with special focus on usage-based methodology, extensive exploitation of linguistic corpus data, and application of quantitative methods and specialized computer data processing programs [1; 9; 20; 27]. The quantitative parameterization of the linguistic properties of form and meaning of a *construction* is among the most researched areas in the field.

The aim of the article. The purpose of this article is to discuss the essence and practical application of quantitative corpus-based methods for determining the distinctive linguistic parameters (factors) of form and meaning of complex *clause-level constructions*.

Results. With the rapid spread of computerized language corpora, modern linguistics is becoming more data-driven and corpus-based, and the quantitative study of language is gaining popularity. Depending on the proportion of qualitative and quantitative aspects in data interpretation, corpus-based research can be mainly qualitative, proportionally combine qualitative and quantitative aspects, or be entirely based on quantitative principles, merging methodologically with computational linguistics and automated information retrieval. The usage-based approach to language acknowledges the frequency of a linguistic unit in a corpus as one of the operationalizations of this unit's conventionality (*cognitive entrenchment* [4: 44; 11: 292–293; 36]) in the language community. On the basis of language unit frequency, the degree to which observed data deviate from expected data and the statistical significance of deviations are determined. Generally, the degree of deviation is determined using mono-factor/multi-factor, or multivariate analysis.

Complex clause-level *constructions* are distinguished by a set of linguistic parameters and factors that determine the specificity of their *form* (plane of expression) (morphosyntactic, relational, referential, positional, syntactic-functional, distributive, etc.) and *semantics / function* (plane of meaning) (collostructional, semantic, communicative, etc.). To establish statistically significant linguistic parameters and factors that determine the variability of a specific *construction* in a constructional network, specialized quantitative corpus methods are employed [8; 21; 32]. The most widely utilized methods are multivariate quantitative data analysis models (principal component method, multiple regression analysis, analysis of variance (single-factor ANOVA, multifactor MANOVA), linear discriminant analysis, cluster analysis, factor analysis, etc.), which permit simultaneous operation with multiple variables [2: 337] These methods are employed to establish relationships between two or more attributes and to determine which factors are significant and to what extent. Multivariate probabilistic methods overcome the limitations of conventional approaches to the study of differences between *constructions* by simultaneously analyzing a number of factors (parameters) that determine the linguistic characteristics of a particular *construction*.

In order to establish statistically significant indicators that determine the distinctive linguistic features of a *construction*, its functioning in contemporary English, and can serve as indicators of speakers' choices among the nodes of the constructional network for categorizing their experience, a three-phase linguistic quantitative procedure involving the sequential use of a number

[5] The italicized variant of the term *'construction'* is part of the terminology of a specific grammar theory – construction grammar

of methods 1) one-factor analysis of variance (ANOVA), 2) multifactor analysis of variance (MANOVA) and 3) post-hoc test by Tukey's multiple comparison method. The application of the chosen statistical methods is justified by their heuristic potential for establishing statistically significant differences between *constructions*.

ANOVA (*ANalysis Of VAriance*) is a parametric statistical method for comparing multiple samples on a metric scale. Developed by R. Fisher in 1925 [14: 32], ANOVA aims to detect dependencies in experimental data by measuring the significance of differences in mean values [37: 12]. This method is intended to test hypotheses regarding the relationship between a particular attribute and the factors under study, as well as to determine the degree of influence of the factors and their interaction [31: 23]. ANOVA is employed to reduce the possibility of the first type of error when multiple comparisons are made.

Notably, the ANOVA method is recognized as a special case of regression analysis [31: 171]. Regression analysis is used to analyze observational data such as corpus data, questionnaires, and surveys, whereas ANOVA is primarily employed in experimental research (pedagogy, psychology, and sociology). Nevertheless, we believe that ANOVA as a type of regression analysis has considerable potential for analyzing the variation in the manifestation of linguistic features of grammatical *constructions* in a corpus.

ANOVA methods include 1) one-factor ANOVA, which allows testing the hypothesis that the independent variable under study has an effect on the dependent variable; 2) multivariate ANOVA, which is designed to study the effect of multiple independent variables on the dependent variable; 3) repeated measures ANOVA, used when different values of the independent variable correspond to different groups of objects; and 4) multivariate ANOVA, which is applied to study the effects of independent variables on several dependent variables rather than one.

One-factor ANOVA is utilized to estimate the impact of a particular factor on a *construction*. The *F* statistics is calculated to determine the ratio of variance attributable to the factor and "random" variance. One-factor ANOVA calculates the significance level *p*, which is used to conclude that the analyzed samples are homogeneous, i.e., that the factor has no effect, or vice versa. The preliminary stage is to test the null hypothesis that there is no effect of the factor(s) under study, i.e., the hypothesis that differences in the attribute values of compared samples are due to random chance and all data come from the same general population.

MANOVA (*Multivariate ANalysis Of VAriance*) is a multivariate analysis of variance similar to ANOVA, except that instead of testing group differences with respect to a single dependent variable, MANOVA tests group differences with respect to multiple dependent variables [37: 385].

When statistically significant differences ($p<0.05$) exist, it is essential to determine the groups that differ. For this purpose, a post-hoc comparison procedure is conducted. The Tukey test, the Kruskal-Wallis H-test, and the Mann-Whitney test are utilized for post hoc comparisons [5: 195; 37: 680]. In our research, posterior testing is performed using the Tukey's multiple comparison method

The outlined three-phase linguistic quantitative procedure is the most efficient for determining distinct parameters of the form of a *construction*, such as morphosyntactic, positional, relational, referential, distributional, syntactic-functional, etc. For identification of distinctive parameters of construction's meaning, the procedure of collostructional and frame analysis is suggested.

The lexical and semantic properties of the lexemes that fill the constructional slots are defined as a *construction's* collostructional parameter. The semantic (conceptual) properties of the lexemesfillers of a *construction's* core slots are defined, using of highly specialized quantitative corpus methods, with *the method of collostructional analysis* being the most influential.

The method of collostructional analysis is based on inferential statistical measures and is intended to provide a more refined set of procedures for investigating the interaction between lexis and grammar (the blend *"collostruction"* combines the nominations *"collocation"* and *"construction"*). The studies conducted by the method's creators, A. Stefanowitsch and St. Gries [16; 17; 34; 35] provide a comprehensive explanation of the procedure and implementation details.

Collostructional analysis combines a number of corpus-based quantitative techniques for determining the meaning of a grammatical construction by quantifying the degree of mutual attraction and repulsion between the construction and the lexemes that fill a specific constructional slot [16; 17; 22; 34; 35]. The lexemes that show statistically significant indicators of attraction to a particular *construction's* slot are *construction's collexemes*. Based on a usage-based approach to language study, this corpus-driven, quantitative, and mathematically appropriate method aims to: 1) improve the quality of grammatical description by providing an objective means of establishing the meaning of a grammatical construction and determining the degree of preference (attraction) or vice versa restriction of the use (repulsion) of certain lexical items in certain slots of this construction; and 2) provide competent data for building a theory of language [34: 209].

Collostructional analysis techniques can be applied to linguistic units of any level of abstraction (words, fixed expressions, argument structures, grammatical categories of tense, aspect, or mood) in order to solve a variety of linguistic problems by establishing the attraction, repulsion, and mutual distribution of units in the context of different constructions. In other words, collostructional analysis is applied to linguistic phenomena that open up grammatically defined slots for the use of specific lexemes. The method is based on the principle of semantic consistency, which states that a lexeme can be used in a construction if it is semantically compatible with the meaning of the construction itself (or, more precisely, with the meaning that the construction ascribes to a specific slot in which this lexeme appears [34: 213]. The method of collostructional analysis does not rely on pre-formulated hypotheses and is entirely based on corpus and statistical data. The method has already been proven successful by linguists from other countries [6; 12; 25; 39; 40], but domestic linguistics has not yet given it the attention it deserves.

Collostructional analysis includes three interconnected quantitative techniques: 1) *simple collexeme analysis* [34] measures the degree of attraction/ repulsion of a lexeme in relation to a construction's slot; 2) *distinctive collexeme analysis* [16] determines which of the two constructions a lexeme prefers; and 3) *co-varying collexeme analysis* [22; 35: 7] measures the interdependence between lexemes filling two different slots of the same construction [5; 16; 22: 7; 35:] (Fig. 1).

The *simple collexeme analysis* is the most commonly used technique of the collostructional analysis. The technique is applied to establish the attraction/ repulsion between a lexeme and a slot in a *construction* by comparing the frequency of the lexeme in the slot to its frequency in other contexts outside the *construction* (i.e., in the corpus). This technique provides objective evidence as to whether a lexeme that is frequently used in a *construction's* slot is actually attracted to it, or whether this lexeme is a high-frequency unit in the corpus and thus frequently occurs in the given *construction*. The results of the simple collexeme analysis provide important insights into the meaning of a construction by identifying the semantic properties of lexemes that exhibit statistically significant attraction to a specific slot.

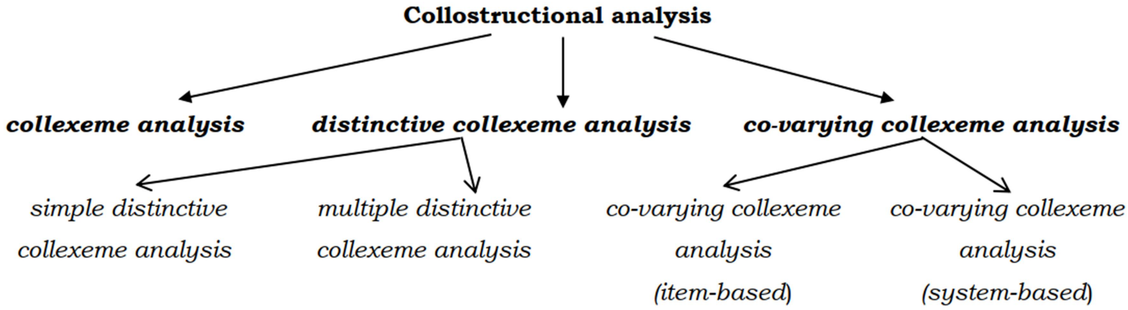

Fig.1 Techniques of collostructional analysis (adopted from: [17: 506])

The *simple collexeme analysis* is based on the frequency of occurrence of language units in a corpus, as determined by the statistical evaluation of the observed frequencies of lexemes/ constructions in relation to the total frequency of occurrence of other lexemes/ constructions in the corpus. For the collexeme analysis to be applied, the corpus must yield four types of frequency indices for the lexical item under study (L) and the *construction* under study (C): 1) the frequency of the lexical item (L) in the analyzed *construction* (C), 2) the frequency of the lexical item (L) in all other *constructions*, 3) the frequency of the analyzed *construction* (C) with lexemes other than the lexical item under study, 4) the frequency of all other with lexemes other than the lexical item under study. The obtained indices are entered into a contingency table or a four-field table for calculating x^2 (chi-square) (see Table 1).

Table 1

Data for the simple collexeme analysis

	+ lexeme (L)	**– lexeme (L)**	
+ construction (C)	frequency of the analyzed lexeme (L) in in the analyzed *construction* (C)	frequency of the analyzed *construction* (C) with lexemes other than the analyzed lexeme	*total (= frequency of the analyzed construction (C) in the corpus)*
– construction (C)	frequency of the analyzed lexeme (L) in other *constructions* (C) in the corpus	frequency of all other *constructions* with other lexemes than the analyzed lexeme (L)	*total*
	total (=frequency (L) in the corpus)	*total*	***Total***

Table 1 provides the data for statistical calculations that measure the degree of relationship between constructions and lexemes. The Fisher's exact test or the Fisher-Yates exact test is the most suitable for this purpose. In fact, a number of association measures can be used to calculate the strength of the association, such as MI (Mutual Information), t, z) [17: 508; 34: 217] (for more information on the advantages and disadvantages of different methods of calculating association measures, see [38]), but for practical purposes, the Fisher's test is accepted by default, since it provides accurate results even on small samples and for low-frequency tokens and is recognized as the most accurate collocation test [20]. The probability value (*p-value)* indicates the degree of collocation strength. Fisher's test and x^2 (chi-square) both test the null hypothesis that the mutual occurrence of L and C is independent. This can be interpreted as follows: the smaller the *p-value*, the higher the probability that the observed frequency distribution of the studied units is not random, and the stronger the force of attraction between the lexeme and the *construction*. The lexemes that showed a statistically significant attraction to a particular slot of the *construction* exhibit semantic coherence with it, based on general knowledge about the surrounding reality, which is reflected in semantic frames [17; 29; 33: 26; 35: 23]. The semantic properties of filler lexemes serve as indicators of the construction's conventional use and determine its semantics [6: 280; 23].

The need to analyze large amounts of linguistic data is no longer possible without specialized software, which in turn encourages linguists to use complex computer-assisted quantitative methods that open up new ways of studying language and have great potential for solving numerous theoretical and practical problems in language research. The statistical data analysis platform R can be used to calculate the statistical methods for linguistic parameterization of clause-level constructions (one-factor analysis of variance (ANOVA), multifactor analysis of variance (MANOVA), post-hoc test by Tukey's multiple comparison method, collostructional analyses).

The statistical data analysis system *R* is one of the most widely used analytical tools for quantitative processing of empirical data in corpus-oriented linguistics. It is an extremely powerful, freely distributed statistical software environment for data analysis that provides the researcher with the entire set of methods necessary for qualitative linguistic and statistical analysis, as well as visualization of the results. The software environment enables the handling of large amounts of multidimensional data and the application of various methods of processing them, such as visualization methods and primary data analysis, the creation of matrix plots, scatter plots, and icons, along with classification methods including data ordering and ordination, statistical verification, and mathematical modeling. Today, support and development of new versions of *R* is provided by a united team of developers from all over the world. *R* software is freely available as open source from the *R* project website (https://www.r-project.org/). Typically, additional specialized software, such as *R Studio*, is used to facilitate the writing of *R* codes, installation and connection of additional modules, creation and storage of graphs of statistical distributions, etc. The main advantages of *R* are the ability to repeat and reproduce the analysis data, the availability of a large number of ready-made packages for various statistical data analyses, and the capability to generate high-quality graphics. The algorithm for applying R packages to identify distinguishing linguistic factors of clause-level *constructions* using ANOVA, MANOVA, and Tukey's post-hoc test is described in the article [43], and the procedure for applying collostructional analysis using the R script written by St. Th. Gries [18] to measure the degree of attraction/ repulsion between lexemes and core slots of a construction is presented in the studies [41; 42].

Conclusions. In summary, it can be concluded that the application of quantitative corpus-based methods for establishing distinctive linguistic parameters/ factors of clause-level constructions provides reliable data on their internal dynamics and variability, which cannot be obtained through intuition or analysis of raw corpus frequencies. By combining the techniques of corpus linguistics and quantitative studies, a new level of understanding can be attained, allowing for more precise analysis in construction grammar studies. At the same time, quantitative methods cannot completely replace other linguistic methods, but they can supplement them with reliable and objective quantitative data, thereby enhancing the descriptive and explanatory reliability and validity of the obtained results. Consequently, quantitative corpus-based methods should be regarded as a valuable tool for construction grammar research, helping grammarians go beyond intuition and better understand the complexities of language phenomena.

REFERENCES

1. Жуковська В. В. Англійська абсолютна конструкція: досвід колострукційного аналізу. *Збірник наукових праць "Актуальні проблеми філології та перекладознавства"*. 2017. Випуск 12. С. 80–85.
2. Baayen H. R. Multivariate Statistics. *Research methods in Linguistics* / Eds. R.J. Podesva, D. Sharma. Cambridge: Cambridge University Press, 2013. P. 337–372.
3. Barlow M., Kemmer S. Usage-Based Models of Language. Stanford, Calif.: CSLI Publications, 2000. 384 p.
4. Blumenthal-Dramé A. Entrenchment in Usage-based Theories: What Corpus Dara Do and Do Not Reveal About the Mind. Berlin/Boston: De Gruyter Mouton, 2012. 282 p.
5. Brezina V. Statistics In Corpus Linguistics. A Practical Guide. Cambridge: Cambridge University Press, 2018. 296 p.
6. Colleman T. Beyond the Dative Alternation: The Semantics of the Dutch aan-dative. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches* / Eds. D. Glynn, K. Fischer. Berlin, New York, 2010. C. 271–303.
7. Corpus Linguistics On the Move. Exploring and Understanding English Through Corpora / Eds. M. J. López-Couso, B. Méndez-Naya, P.Núñez-Pertejo, I. M. Palacios-Martínez. Leiden, Boston: Brill Rodopi, 2016. 367 p.
8. Corpus Methods for Semantics. Quantitative Studies in Polysemy and Synonym / Eds. D. Glynn, J. A. Robinson. Amsterdam/ Philadelphia: John Benjamin's Publishing Company, 2014. 545 p.
9. Corpus-Based Approaches to Construction Grammar / Eds. Yoon J., St. Th. Gries Amsterdam/ Philadelphia: John Benjamin's Publishing Company, 2016. 274 p.
10. Corpus-Based Approaches to Grammar, Media and Health Discourses. Systemic Functional and Other Perspectives / Eds. D. Yang, W. Li. Singapore: Springer, 2020. 396 p
11. Croft W., Cruse A. D. Cognitive Linguistics. Cambridge: Cambridge University Press, 2004. 374 p.
12. Desagulier G. Corpus Linguistics and Statistics with R. Introduction to Quantitative Methods in Linguistics. Cham: Springer International Publishing, 2017. 353 p.
13. Desagulier G. Visualizing Distances in a Set of Near-Synonyms: Rather, Quite, Fairly, and Pretty. *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy* / Eds. D. Glynn, J. A. Robinson. Amsterdam/Philadelphia: John Benjamin's Publishing Company, 2014. P. 145–178.
14. Field A. P. Analysis of Variance (ANOVA). *Encyclopedia of Measurement and Statistics*. Vol.1 Ed. N. J. Salkind. Thousand Oaks, London, New Delhi: Sage Publications, 2007. P. 32–35.
15. Goldberg A. E. Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions. Princeton/ Oxford: Princeton University Press, 2019. 195 p.
16. Gries S. T., Stefanowitch A. Extending Collostructional Analysis: A Corpus-Based Perspective on "Alternations". *International Journal of Corpus Linguistics*. 2004. № 9. P. 97–129.
17. Gries S. Th. More (Old and New) Misunderstandings of Collostructional Analysis: On Schmid and Küchenhoff (2013). *Cognitive Linguistics*. 2015. № 26 (3). P. 505–536.

18. Gries S. Th. Coll.analysis 4.0. A script for R to compute collostructional analyses. URL: https://www.stgries.info/teaching/groningen/index.html (дата звернення: 23.01.2023).
19. Gries St. Th. Quantitative Corpus Linguistics with R. A Practical Introduction. New York, London: Routledge, Taylor & Francis Group, 2017. 272 p.
20. Gries St. Th. The Role of Quantitative Methods in Cognitive Linguistics: Corpus and Experimental Data on (Relative) Frequency and Contingency of Words and Constructions. *Change of paradigms – new paradoxes: Recontextualizing language and linguistics* / Eds. P. Daems, E. Zenner, K. Heylen, D. Speelman, H. Cuyckens, Berlin, New York: De Gruyter Mouton, 2015. P. 311–325.
21. Gries St. Th., Ellis N.C. Statistical Measures for Usage-Based Linguistics. *Language Learning*. Volume 65, Suppl. 1. 2015. P. 228–255.
22. Gries St. Th., Stefanowitsch A. Covarying Collexemes in the *Into*-Causative. *Language, Culture, and Mind.* Stanford, CA: CSLI Publications, 2004. P. 225–236.
23. Herbst Th. Constructions, Generalizations and the unPredictability of Language: Moving Towards ColloConstruction Grammar. *Construction Grammar across Borders* / Eds. T. T. Torrent, E. E. Silva Matos, N. Sathler Sigiliano. Amsterdam/Philadelphia: John Benjamin's Publishing Company, 2020. P. 56–95.
24. Hilpert M. Construction Grammar and its Application to English. Edinburgh: Edinburgh University Press, 2014. 219 p.
25. Hilpert M. Germanic Future Constructions: A Usage-Based Approach to Language Change. Amsterdam/Philadelphia: John Benjamin's Publishing Company, 2008. 217 p.
26. Hoffmann T. Construction Grammar. Cambridge: Cambridge University Press, 2022. 336 p.
27. Janda L. A. Cognitive Linguistics: The Quantitative Turn. Berlin/Boston: Mouton de Gruyter, 2013. 321 p.
28. Levshina N. How to do Linguistics with R. Data Exploration and statistical analysis. Amsterdam/ Philadelphia: John Benjamin's Publishing Company, 2015. 443 p.
29. Lyngfelt B. Introduction. *Constructions and Constructicography. Constructicography. Constructicon Development Across Languages* / Eds. B. Lyngfelt, L. Borin, K. H. Ohara; T. T. Torrent. Amsterdam: John Benjamin's Publishing Company, 2018. P. 1–18.
30. Nordquist D. Comparing Elicited Data and Corpora. Language, Culture, and Mind. Stanford, Califf.: CSLI, 2004. P. 211–224.
31. Oakes M.P. Statistics for Corpus Linguistics. Edinburgh: Edinburgh University Press, 1998. 287 p.
32. Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches / Eds. D. Glynn, A. Fischer. Berlin/ New York: De Gruyter Mouton, 2010. 387 p.
33. Schönefeld D. Things Going Unnoticed – A Usage-Based Analysis of *GO*-Constructions. *Frequency Effects in Language Representation* / Eds. D. Divjak, S.T. Gries. Vol. 2, Berlin/ New York: De Gruyter Mouton, 2012. P. 11–49.
34. Stefanowitsch A., Flach S. The Corpus-Based Perspective on Entrenchment. *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge* / Ed. H.-J. Schmid, Berlin, Boston: De Gruyter Mouton, 2017. P. 101–128.
35. Stefanowitsch A., Gries S. Th. Collostructions: Investigating the Interaction of Words and Constructions. *International Journal of Corpus Linguistics.* 2003. Vol. 8. P. 209–243.
36. Stefanowitsch A., Gries S. Th. Covarying Collexemes. *Corpus Linguistics and Linguistic Theory.* 2005. Vol. 1. P. 1–43.
37. Tavakoli H. A Dictionary of Research Methodology and Statistics in Applied Linguistics. Tehran: Rahnama Press, 2012. 753 p.

38. Wiechmann D. On the Computation of Collostruction Strength: Testing Measures of Association as Expressions of Lexical Bias. *Corpus Linguistics and Linguistic Theory*. 2008. № 4–2. P. 253–290.
39. Wiliński J. A Covarying-Collexeme Analysis of the German Caused-Motion Construction in the Soccer Domain. *Acta Philologica*. № 46. 2015. P. 17–24.
40. Wiliński J. To Move Theme to Locative Goal vs. to Move Theme to Recipient in the Soccer Domain: A Distinctive-Collexeme Analysis of two Near-Equivalent Constructions. *Beyond Words: Crossing Borders in English Studies: Language, Culture and Translation* / Eds. J. Leśniewska, M. Urban, Vol. 2, Kraków: Tertium. 2016. P. 41–55.
41. Zhukovska V. English detached adjectival constructions with an explicit subject: A quantitative corpus-based analysis. *Journal of Linguistics (Jazykovedný časopis)*, ROČNÍK 72 (2), 2021. P. 465–477.
42. Zhukovska V. Quantitative Corpus-Driven Approach to Disambiguation of Synonymous Grammatical Constructions. *Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020).* Volume I: Main Conference, Lviv, Ukraine, April 23-24, 2020. CEUR Workshop Proceedings 2604, CEUR-WS.org 2020. P. 507–522.
43. Zhukovska V.V., Mosiyuk O. O. Statistical software R in corpus-driven research and machine learning. *Information Technologies and Learning Tools*. 2021. Vol. 86, № 6. P. 1–18.

REFERENCES (TRANSLATED & TRANSLITERATED)

1. Zhukovska, V. V. (2017). Anhliiska absoliutna konstruktsiia: dosvid kolostruktsiinoho analizu [English Absolute Construction: A Case of Collostructional Analysis]. *Zbirnyk naukovykh prats "Aktualni problemy filolohii ta perekladoznavstva". Vypusk* 12. P. 80–85 [in Ukrainian].
2. Baayen, H. R. Multivariate Statistics. (2013). *Research methods in Linguistics* / Eds. R.J. Podesva, D. Sharma. Cambridge: Cambridge University Press. P. 337–372 [in English].
3. Barlow, M., Kemmer, S. (2000). Usage-Based Models of Language. Stanford, Calif.: CSLI Publications. 384 p [in English].
4. Blumenthal-Dramé, A. (2012). Entrenchment in Usage-based Theories: What Corpus Dara Do and Do Not Reveal About the Mind. Berlin/Boston: De Gruyter Mouton. 282 p [in English].
5. Brezina, V. (2018). Statistics In Corpus Linguistics. A Practical Guide. Cambridge: Cambridge University Press. 296 p [in English].
6. Colleman, T. (2010). Beyond the Dative Alternation: The Semantics of the Dutch aan-dative. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches* / Eds. D. Glynn, K. Fischer. Berlin, New York. P. 271–303 [in English].
7. Corpus Linguistics On the Move. Exploring and Understanding English Through Corpora (2016). / Eds. M. J. López-Couso, B. Méndez-Naya, P.Núñez-Pertejo, I. M. Palacios-Martínez. Leiden, Boston: Brill Rodopi. 367 p [in English].
8. Corpus Methods for Semantics. Quantitative Studies in Polysemy and Synonym (2014). / Eds. D. Glynn, J. A. Robinson. Amsterdam/ Philadelphia: John Benjamin's Publishing Company. 545 p [in English].
9. Corpus-Based Approaches to Construction Grammar (2016). / Eds. Yoon J., St. Th. Gries Amsterdam/ Philadelphia: John Benjamin's Publishing Company. 274 p [in English].
10. Corpus-Based Approaches to Grammar, Media and Health Discourses. Systemic Functional and Other Perspectives (2020). / Eds. D. Yang, W. Li. Singapore: Springer. 396 p [in English].
11. Croft, W., Cruse, A. D. (2004). Cognitive Linguistics. Cambridge: Cambridge University Press. 374 p [in English].

12. Desagulier, G. (2017). Corpus Linguistics and Statistics with R. Introduction to Quantitative Methods in Linguistics. Cham: Springer International Publishing. 353 p [in English].
13. Desagulier, G. (2014). Visualizing Distances in a Set of Near-Synonyms: Rather, Quite, Fairly, and Pretty. *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy* / Eds. D. Glynn, J. A. Robinson. Amsterdam/Philadelphia: John Benjamin's Publishing Company. P. 145–178 [in English].
14. Field, A. P. (2007). Analysis of Variance (ANOVA). *Encyclopedia of Measurement and Statistics*. Vol.1 Ed. N.J. Salkind. Thousand Oaks, London, New Delhi: Sage Publications. P. 32–35 [in English].
15. Goldberg, A. E. (2019). Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions. Princeton/ Oxford: Princeton University Press. 195 p [in English].
16. Gries, S. T., Stefanowitch A. (2004). Extending Collostructional Analysis: A Corpus-Based Perspective on "Alternations". *International Journal of Corpus Linguistics*. № 9. P. 97–129 [in English].
17. Gries, S. Th. (2015). More (Old and New) Misunderstandings of Collostructional Analysis: On Schmid and Küchenhoff (2013). *Cognitive Linguistics*. № 26 (3). P. 505–536 [in English].
18. Gries, S. Th. (2022). Coll.analysis 4.0. A script for R to compute collostructional analyses. URL: https://www.stgries.info/teaching/groningen/index.html (date of access: 23.01.2023) [in English].
19. Gries, St. Th. (2017). Quantitative Corpus Linguistics with R. A Practical Introduction. New York, London: Routledge, Taylor & Francis Group. 272 p. [in English] .
20. Gries, St. Th. (2015). The Role of Quantitative Methods in Cognitive Linguistics: Corpus and Experimental Data on (Relative) Frequency and Contingency of Words and Constructions. *Change of paradigms – new paradoxes: Recontextualizing language and linguistics* / Eds. P. Daems, E. Zenner, K. Heylen, D. Speelman, H. Cuyckens, Berlin, New York : De Gruyter Mouton. P. 311–325 [in English].
21. Gries, St. Th., Ellis N.C. (2015). Statistical Measures for Usage-Based Linguistics. *Language Learning*. Volume 65, Suppl. 1. P. 228–255 [in English].
22. Gries, St. Th., Stefanowitsch, A. (2004). Covarying Collexemes in the *Into*-Causative. *Language, Culture, and Mind.* Stanford, CA: CSLI Publications. P. 225–236 [in English].
23. Herbst, Th. (2020). Constructions, Generalizations and the unPredictability of Language: Moving Towards ColloConstruction Grammar. *Construction Grammar across Borders* / Eds. T. T. Torrent, E. E. Silva Matos, N. Sathler Sigiliano. Amsterdam/Philadelphia: John Benjamin's Publishing Company. P. 56–95 [in English].
24. Hilpert, M. (2014). Construction Grammar and its Application to English. Edinburgh: Edinburgh University Press. 219 p [in English].
25. Hilpert, M. (2008). Germanic Future Constructions: A Usage-Based Approach to Language Change. Amsterdam/Philadelphia: John Benjamin's Publishing Company. 217 p [in English].
26. Hoffmann, T. (2022). Construction Grammar. Cambridge: Cambridge University Press. 336 p [in English].
27. Janda, L. A. (2013). Cognitive Linguistics: The Quantitative Turn. Berlin/Boston: Mouton de Gruyter. 321 p [in English].
28. Levshina, N. (2015). How to do Linguistics with R. Data Exploration and statistical analysis. Amsterdam/ Philadelphia: John Benjamin's Publishing Company. 443 p [in English].

29. Lyngfelt, B. (2018). Introduction. *Constructions and Constructicography. Constructicography. Constructicon Development Across Languages* / Eds. B. Lyngfelt, L. Borin, K. H. Ohara; T. T. Torrent. Amsterdam: John Benjamin's Publishing Company. P. 1–18 [in English].
30. Nordquist, D. (2004). Comparing Elicited Data and Corpora. Language, Culture, and Mind. Stanford, Califf.: CSLI. P. 211–224 [in English].
31. Oakes, M.P. (1998). Statistics for Corpus Linguistics. Edinburgh: Edinburgh University Press. 287 p [in English].
32. Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches (2010) / Eds. D. Glynn, A. Fischer. Berlin/ New York: De Gruyter Mouton. 387 p [in English].
33. Schönefeld, D. (2012). Things Going Unnoticed – A Usage-Based Analysis of *GO*-Constructions. *Frequency Effects in Language Representation* / Eds. D. Divjak, S.T. Gries. Vol. 2, Berlin/ New York : De Gruyter Mouton. P. 11–49 [in English].
34. Stefanowitsch, A., Gries, S. Th. (2003). Collostructions: Investigating the Interaction of Words and Constructions. *International Journal of Corpus Linguistics.* Vol. 8. P. 209–243 [in English].
35. Stefanowitsch, A., Gries, S. Th. (2005). Covarying Collexemes. *Corpus Linguistics and Linguistic Theory.* Vol. 1. P. 1–43 [in English].
36. Stefanowitsch, A., Flach, S. (2017). The Corpus-Based Perspective on Entrenchment. *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge* / Ed. H.-J. Schmid, Berlin, Boston: De Gruyter Mouton. P. 101–128 [in English].
37. Tavakoli, H. (2012). A Dictionary of Research Methodology and Statistics in Applied Linguistics. Tehran : Rahnama Press. 753 p [in English].
38. Wiechmann, D. (2008). On the Computation of Collostruction Strength: Testing Measures of Association as Expressions of Lexical Bias. *Corpus Linguistics and Linguistic Theory*. № 4–2. P. 253–290 [in English].
39. Wiliński, J. (2015). A Covarying-Collexeme Analysis of the German Caused-Motion Construction in the Soccer Domain. *Acta Philologica*. № 46. P. 17–24 [in English].
40. Wiliński, J. (2016). To Move Theme to Locative Goal vs. to Move Theme to Recipient in the Soccer Domain: A Distinctive-Collexeme Analysis of two Near-Equivalent Constructions. *Beyond Words: Crossing Borders in English Studies: Language, Culture and Translation* / Eds. J. Leśniewska, M. Urban, Vol. 2, Kraków: Tertium. P. 41–55 [in English].
41. Zhukovska,V. (2021). English detached adjectival constructions with an explicit subject: A quantitative corpus-based analysis. *Journal of Linguistics (Jazykovedný časopis)*, ROČNÍK 72 (2). P. 465–477 [in English].
42. Zhukovska, V. (2020). Quantitative Corpus-Driven Approach to Disambiguation of Synonymous Grammatical Constructions. *Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020).* Volume I: Main Conference, Lviv, Ukraine, April 23–24, 2020. CEUR Workshop Proceedings 2604, CEUR-WS.org. P. 507–522 [in English].
43. Zhukovska, V. V., Mosiyuk, O. O. (2021). Statistical software R in corpus-driven research and machine learning. *Information Technologies and Learning Tools.* Vol. 86, № 6. P. 1–18. [in English]

Стаття надійшла до редколегії: 02.02.2023
Схвалено до друку: 28.02.2023