

ТЕОРЕТИЧЕСКАЯ, ПРИКЛАДНАЯ И СРАВНИТЕЛЬНО-СОПОСТАВИТЕЛЬНАЯ ЛИНГВИСТИКА /
THEORETICAL, APPLIED AND COMPARATIVE LINGUISTICS

DOI: <https://doi.org/10.18454/RULB.2023.45.27>

COLLOCATIONAL COMPETENCE FORMATION SOFTWARE

Research article

Arkhipova E.I.^{1,*}

¹ORCID : 0000-0001-9992-3443;

¹Novosibirsk State Technical University, Novosibirsk, Russian Federation

* Corresponding author (elena1503[at]inbox.ru)

Abstract

This article discusses the available software as a means for developing the collocational competence of people studying the language and as a tool for linguists. The purpose of the study is to demonstrate the advantages of using the concordancing software: it is easy to find words frequently used together, the program gives an idea of the collocations that include the desired unit, you can analyze the grammatical features and differences in connotation. The use of concordance resources while studying language and in linguistic studies in general contributes to the objectivity of the selection of lexical material, as well as taking into account the peculiarities of its use. Particular attention is paid to different tools designed for the users.

Keywords: collocations, collocational competence, concordance, corpus, concordancing software.

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ ФОРМИРОВАНИЯ КОЛЛОКАЦИОННОЙ КОМПЕТЕНЦИИ

Научная статья

Архипова Е.И.^{1,*}

¹ORCID : 0000-0001-9992-3443;

¹Новосибирский государственный технический университет, Новосибирск, Российская Федерация

* Корреспондирующий автор (elena1503[at]inbox.ru)

Аннотация

В данной статье рассматривается программное обеспечение как средство развития коллокационной компетенции изучающих язык и инструмент для лингвистических исследований. Цель работы – продемонстрировать преимущества использования компьютерной программы для автоматического составления конкорданса: легкость поиска слов, часто употребляемых вместе, возможность получить представление о коллокациях, в состав которых входит искомая единица, можно проанализировать грамматические особенности и различия в коннотации. Использование данных ресурсов при изучении языка и в лингвистических исследованиях в целом способствует объективности отбора лексического материала, а также учету особенностей его использования. Особое внимание уделяется различным инструментам, предназначенным для пользователей.

Ключевые слова: коллокации, коллокационная компетенция, конкордансы, корпусы текстов, программное обеспечение.

Introduction

Transition from handwritten lexicographic practice to digital dictionaries happened due to an applied scientific discipline – computational lexicography [6]. It deals with the investigation of methods and software tools for processing text. Analyzing the needs of the users, scientists developed computer technologies for the compilation and operation of dictionaries. Print versions of dictionaries were thoroughly studied in order to present all the necessary aspects of information: spelling, morphology, pronunciation, subject matter, definition, syntax, etymology, examples, the run-ons, synonyms, antonyms, etc. For the dictionary creation, computational lexicologists built large corpora with the tags [3]. The issues of corpus organization, organization of material, inclusion of extralinguistic information were of great importance.

Quick and convenient access, unlimited space for posting information, advanced information retrieval, links to additional sources – are among the advantages of online dictionaries over printed ones. Since the main aim of electronic dictionaries is teaching the up-to-date speech, lots of them include modern English. Upon user request, dictionaries display a selection of contexts, showing the collocability: the headword, its distributors, examples from the sources of different genres [4], [7]. Along with the idea of teaching, cyberlexicography – is one of the main sources for the linguists, studying language development and its inner peculiarities.

The following methods formed the basis of the study: descriptive research method of language units' analysis. The collection of factual material required the use of the method of continuous sampling.

Main results

The development of Internet technologies has led to the emergence of corpus linguistics – an interdisciplinary field aimed at studying the possibilities of using language corpora in the process of language learning. The analysis of data collected in the corpus began to be used to solve various applied problems.

There are several concepts of corpus teaching: a direct concept (data-driven learning) involves direct work with a corpus of texts, an indirect one (corpus-based) includes an intermediary in the learning process – a teacher who builds the learning process using the corpus, selecting the necessary material and developing exercises [8, P. 641-658].

Developing learning tools using corpus data is a time-consuming task, but thanks to automated data analysis, it is possible to build training on up-to-date information about new words, word forms and shades of meaning, as well as using illustrative examples.

“AntConc” corpus manager designed by Dr. Laurence Anthony is designed to obtain vocabulary minimums associated with a given subject area, lists of collocations, including terminological ones, samples for thematic groups of words. Using this concordancing software you can search for contexts, evaluate their typicality [1].

The work principles of this corpus-manager can be considered by the example of an article “Combinatorial Linguodidactics as a New Direction in the Methodology of Foreign Language Teaching” [2]. First of all, open TXT file from the top left corner of the menu (“Open file as Quick Corpus”). The name of the file will appear in the left window (under “Target corpus”).

Click the “Collocate” button and insert any word you want to analyze, e.g. *linguodidactics*, press the “Start” button. The program will show all collocations with this word in order of frequency (see fig.1).

Collocate	Rank	Freq(Scaled)	FreqL	FreqR	Range	Likelihood	Effect
1 combinatorial	1	310	21	3	1	66.158	3.174
2 fundamentals	2	40	4	1	1	18.125	3.865
3 theoretical	3	60	3	2	1	14.185	3.280

Figure 1 - Analysis of the word “linguodidactics” in the AntConc software. Collocate inlay
DOI: <https://doi.org/10.18454/RULB.2023.45.27.1>

The user can click on any word from the list, it will be automatically searched in the Concordance window. If the Concordance window is open, you can enter the search term, it will be searched in contexts.

In order to get the information about the combinability of the node, the user can insert the word in the inlay “N-Gram”, the concordancing software will arrange the results according to the frequency of appearance in the text. Moreover, you can set the parameters, for example, the n-gram size – the amount of words after the search term (see fig. 2).

Type	Rank	Freq	Range
1 linguodidactics is	1	5	1
2 linguodidactics in	2	3	1
3 linguodidactics aimed	3	2	1
4 linguodidactics as	3	2	1
5 linguodidactics can	3	2	1
6 linguodidactics collocation	3	2	1
7 linguodidactics it	3	2	1
8 linguodidactics a	8	1	1
9 linguodidactics are	8	1	1
10 linguodidactics contains	8	1	1
11 linguodidactics develops	8	1	1

Figure 2 - Analysis of the word “linguodidactics” in the AntConc software. N-Gram inlay
DOI: <https://doi.org/10.18454/RULB.2023.45.27.2>

KWIC inlay sorts and aligns the left and the right contexts of the sought-for word, e.g. the request word *linguistics* gives us six hits, shown in red colour, moreover, the user can set up the amount of tokens on the left and on the right – the context size needed for the research.

There are nine tools for carrying out corpus linguistics research:

- it shows the results in a concordance format;
- the search tool shows the position of the search results in the particular text;
- it presents information on the contexts of individual texts;
- it shows word patterns based on the search word;
- the tool shows N-word clusters;
- it allows the user to study the co-occurrence of the words;
- the tool shows all the words from the text in order of their frequency of occurrence;

- one of the tools presents information on the unusually frequent words, that are characteristic of its corpus – the keywords;
- a wordcloud tool [1].

The advantage of this software program is open access – the user can download the official release of the software from the homepage.

The use of concordance software [5], [10] provides the following opportunities: you can get an idea of the collocations and phrases that include the desired unit (collocational competence [9]); this type of presentation of words gives a visual representation of semantic prosody. It shows the way in which words can get positive or negative associations, e.g. the verb *to inflict* – *to force someone to experience something very unpleasant*, is mostly used in a negative context, which is clearly seen in the concordance materials. The program arranges keywords in context, arranges phrases in alphabetical order, and thus, hidden associations can be revealed.

Conclusion

Concordancing software is a great source of information and a powerful language learning tool due to extensive data for the study. These programs are relatively simple and effortless to use, so learners can use them on their own.

Language corpora and concordances are resources that are not yet fully used in the learning process, but professional linguists frequently use the existing tools in the research. A careful study of word combinations and expressions in dictionaries and concordances will provide students with important information about collocations in English. The use of concordances can develop learners' "lexical vision and understanding" of the language.

Constant attempts to optimize the process of language learning force specialists to look for and develop new ways of selecting and applying lexical material. The use of corpus resources can be of great help in learning, so the skills of forming a query, using tools, interpreting the results will contribute to independent study and autonomous work. Authentic illustrative examples contribute to the formation of lexical skills of correct word usage and give an idea of the combinatorial possibilities of the units being sought.

Authentic illustrative examples contribute to the formation of lexical skills of correct word usage and give an idea of the combinatorial possibilities of the units being sought.

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы на английском языке / References in English

1. Laurence A. AntConc Corpus Manager / A. Laurence. — 2022 — URL: <https://www.laurenceanthony.net/software.html> (accessed: 27.07.2023)
2. Arkhipova E. Combinatorial linguodidactics as a new direction in the methodology of foreign language teaching / E. Arkhipova, M. Vlavatskaya // *Advances in Intelligent Systems and Computing*. — 2019. — 1. — p. 91-97. — URL: <https://link.springer.com/book/10.1007/978-3-030-11473-2> (accessed: 20.07.2023) DOI: 10.1007/978-3-030-11473-2_11.
3. Association for Computational Linguistics. — URL: <https://www.aclweb.org/portal/> (accessed 10.06.2023).
4. British National Corpus. — URL: <https://www.english-corpora.org/bnc/> (accessed: 10.06.2023).
5. Cloudnine Concordance. — URL: <https://cloudnine.com/software/concordance/> (accessed: 10.06.2023).
6. Gibbon D. *Computational Lexicography* / D. Gibbon — Bielefeld: ELSNET, 1997. — 106 p. — URL: http://www.homes.uni-bielefeld.de/gibbon/Dafydd_Gibbon_Publication_PDFs/1997_Gibbon_Computational_Lexicography_for_Speech_and_Language_ELSNET_SummerSchoolDg_lex.pdf (accessed: 10.06.2023)
7. Google Books (American). — URL: <https://www.english-corpora.org/googlebooks/x.asp> (accessed: 10.06.2023).
8. Levinzon A.I. *Corpus Teaching in the Russian School* / A.I. Levinzon // *Proceedings of the Institute of the Russian Language*. — 2015. — 6. — p. 641-658.
9. Lontas J.I. *Collocational Competence and Electronic Corpora: Implications for English Language Teaching. Teaching and Technology* / J.I. Lontas, S. Li, I. Bangun // *Teaching and Technology*. — 2019. — 1. — URL: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/9781118784235.eelt0982> (accessed: 03.07.2023) DOI: 10.1002/9781118784235.
10. Simple Concordance Program. — URL: <http://www.textworld.com/scp/> (accessed: 10.06.2023).