

DEVELOPMENT AND CURRENT STATE OF CORPUS LEXICOGRAPHY

Erkaeva Dilnoza Bakhtiyorovna

English literature department, Bukhara State University

Annotation. This article gives full overview of analysis of corpus lexicography in film discourse. The development and current state of corpus lexicography is analyzed with the help of examples.

Key word: Corpus lexicography, film discourse, lexicography, applied linguistics.

Introduction. Corpus lexicography is a factor to analyze film discourse and its development and current state.

Main part. Advances in the development of computer technology have contributed to the emergence of a new area of lexicography - corpus lexicography (" *corpus lexicography* ").

S.A. Manik focuses on the fact that corpus lexicography (previously unfamiliar to most) currently occupies a leading position due to the emergence of electronic databases, "banks of national languages", which are becoming a reliable source for creating various kinds of reference publications (Manik, 2010, 224). M. Rundell and A. Kilgarriff note that the main task of the 21st century lexicographer is to evaluate the information already prepared for him (obtained as a result of computer technology) and, if necessary, correct it (Rundell , Kilgarriff , 2011, 278).

Lexical database or lexicographic database (" *lexical database* " / " *lexicographic database* ") is created by experts, based on corpus data and intended as a source for creating dictionaries (Horak , Rambousek , 2018, 184). W. Atkins and M. Randell introduce the concept of lexicographic corpus « *lexicographic corpus* " and note that this is a collection of language material designed specifically for use in creating dictionaries (Atkins , Rundell , 2008, 54). *Corpus - based approach _ approach* " is successfully used in this area, since the appearance of a lexicographic publication is preceded not only by the identification and study of the requests of potential users and the compilation of principles for the lexicography of language units, but also the creation of a linguistic corpus, which is the necessary basis for the formation of a future authoritative lexicographic publication. In this regard, we note the study of I.I. Sazhenin, which is devoted to the description of corpus methods in lexicography (Sazhenin, 2013). I. Kosem singles out the " *corpus - informed* " approach (often equated with the " *corpus - based* " approach), according to which sentences from the corpus are added to the dictionary entry (Kosem , 2016, 79). In this regard, we note the option " *Good Dictionary Example* » (*GDEX*) on the *Sketch platform Engine* " , which allows you to automatically select the best examples from the concordance for inclusion in the dictionary. O.M. Karpova points out that computer data processing allows not only obtaining good quality examples, but also verified definitions (Karpova, 2018, 26).

Despite the slow pace of development of case technology in the 20th century, in the 1970s the " *COBUILD project* " , which was 8 million words by the early 1980s (Kilgarriff , Grefenstette , 2003, 334). P.V. Sysoev emphasizes that new dictionaries of the " *COBUILD* " type (" *Collins Birmingham university International language Database* ") (*corpus - built* - created by analyzing

the linguistic corpus), reflecting current trends in the use of the language (Sysoev, 2010, 102). Of course, when working with corpus data, critical analysis is necessary when choosing lexical units that will be included in the dictionary.

In 1987, "Collins" was released. *COBUILD English language Dictionary* , which became the first dictionary compiled on the basis of corpus data (Rundell , Jakubicek , Kovar , 2020, 20). A. Ludeling and M. Walter note that educational English dictionaries like " *COBUILD* " systematically use corpus data and their analysis (Ludeling , Walter , 2009, 8). It is true that " *The COBUILD project* " has set a new standard for dictionary production, whereby the systematic use of corpora has become an important element in lexicography (Johansson , 2008, 48).

I. Kozem and R. Krishnamurti identify several groups of dictionaries developed on the basis of corpus data:

- General Dictionaries: Dictionaries for Teaching English as a Foreign Language (*EFL dictionaries*) , dictionaries for native speakers (*native - speaker dictionaries*) , bilingual dictionaries (*bilingual dictionaries*) ;
- specialized dictionaries: collocation dictionaries (*dictionaries of collocations*) , production dictionaries (*production dictionaries*) , dictionaries of idioms (*dictionaries of idioms*) , phrasal verb dictionaries (*dictionaries of phrasal verbs*) ;
- thesauri (Kosem and Krishnamurthy, 2007).

I. Kosem and R. Krishnamurthy emphasize that dictionaries designed for teaching English as a foreign language (*EFL*) and dictionaries for academic purposes (*EAP*) use corpus resources for teaching and learning in different ways (Kosem , Krishnamurthy , 2007). Attention is focused on the fact that lexicographers play an important role in filling *EFL* dictionaries, whose task is to include information in dictionaries based on corpus analysis. *EAP* Developers dictionaries receive little help from lexicographers, preferring to work directly with the corpus (Kosem and Krishnamurthy, 2007).

The classification of online dictionaries is interesting, in which dictionaries are divided into 5 categories depending on the degree of technology use:

- 1) Copycats (photographed or scanned copies of already existing printed dictionaries, which are uploaded to the Internet more often in pdf format).
- 2) Faster Horses (new technologies are being used, but in a very limited amount, for example, more efficient search and following links are available).
- 3) Stray Bullets (the ability to get more information, increase quantitative indicators).
- 4) Model T Fords.
- 5) Rolls Royces (Fuertes-Olivera, Tarp, 2014).

The last two categories provide not only faster access, but also dynamic content "*dynamic articles with dynamic data*" (Fuertes-Olivera, Tarp, 2014). According to scientists, the category of dictionaries "*Rolls Royces*" includes customized lexicographic tools that help meet the information needs of a particular user (Fuertes-Olivera, Tarp, 2014). In our opinion, it is possible to fully satisfy the needs of users of modern electronic dictionaries only when using corpus data or referring to corpus.

S.A. Manik points out that "mass computerization of lexicography, the use of electronic catalogs and data banks for processing and storing information provide new opportunities and meet the

requirements of the modern user" (Manik, 2010, 223).

ON THE. Sivakova believes that "at present, truly scientific descriptions of the grammatical structure of languages, as well as authoritative academic dictionaries, are compiled on the basis of the corpora of these languages" (Sivakova 2004, 12). As the scientist quite rightly emphasizes, one of the purposes of the corpus in recent years is to become a resource for lexicography. O.O. Boriskina focuses on the fact that "the corpus nature of dictionaries and grammar increases their reliability and verifiability, avoids the subjectivity and incompleteness that often suffer from descriptions based solely on the introspection of a linguist" (Boriskina, 2015, 26).

So, "*The bank of English Corpus*" was originally compiled for lexicographic purposes. Thus sections of this corpus were used to create the "*BBC English Dictionary*" and "*Collins COBUILD English Dictionary*" (Meyer, 2002, 15). In turn, the research project "*The Cambridge language Survey*" developed 2 cases: "*The Cambridge International Corpus*" and "*The Cambridge Learners' Corpus*", on the material of which the dictionary "*Cambridge International Dictionary of English*" (Meyer, 2002, 15).

Longman has compiled a large corpus of spoken and written American English that has served as the source for the *Longman Dictionary. Dictionary of American English*, and the corpus "*British National Corpus*" resource for "*Longman Dictionary of Contemporary English*" (Meyer, 2002, 15). Thus, the 2nd edition of the dictionary "*Longman Dictionary of Contemporary English*" was not based on corpus data, and the 3rd edition already included them. As a comparison, an illustrative example with the verb "*know*" is given. The 2nd edition of the above dictionary gives 20 meanings for this lexical unit, and the 3rd edition provides more than 40 meanings, which highlights the fact that more complete data is obtained when relying on corpus material (Baker, Hardie, McEnery, 2006, 108).

As T.V. Tolstov, in the basis of the dictionary "*Macmillan English Dictionary*" (2002), which included frequency information for the first time, was put "*World English Corpus*" (Tolstova, 2018, 58).

In domestic lexicographic practice, the dictionary "*The Basic Russian Dictionary*" (BRD), which contains a description of frequently used lexical units of the Russian language through a description of their semantic and grammatical relationships. When creating the dictionary, the material "*Russian National Corpus*" (Shvedova, Sitchinava, 2016).

It seems necessary to single out dictionaries compiled on the basis of the national corpus of the Russian language. For example, the grammatical dictionary of new words of the Russian language, ed. E.A. Grishina, O.N. Lyashevskaya (<http://dict.ruslang.ru/gram.php>), a new frequency dictionary of Russian vocabulary, ed. HE. Lyashevskaya, S.A. Sharova (<http://dict.ruslang.ru/freq.php>), Dictionary of Russian Idioms, ed. G.I. Kustovoy (<http://dict.ruslang.ru/magn.php>), Dictionary of verbal compatibility of non-objective names of the Russian language, ed. O.L. Biryuk, V.Yu. Guseva, E.Yu. Kalinina (http://dict.ruslang.ru/abstr_noun.php).

Thus, corpus technologies, which have come a long way in development, are becoming more and more technically advanced, as a result of which they are actively used in corpus lexicography, contributing to the development of authoritative lexicographic publications that include valuable authentic material.

References:

1. An American Dictionary of the English Language: Intended to Exhibit, 1828. -

2023: International Conference on Agriculture Sciences, Environment,
Urban and Rural Development (Morocco)

<https://www.conferenceseries.info/index.php/morocco>

URL: <https://archive.org/details/americanfiction01websrich/page/1/mode/2up>.

2. Baker P., Hardie A., McEnery T. A Glossary of Corpus Linguistics. - Edinburgh University Press, 2006. - 192 p.
3. Cambridge Dictionary. - URL: <https://dictionary.cambridge.org/ru/> - дата обращения 10.08.2020.
4. Collins Dictionary. - URL: <https://www.collinsdictionary.com> - дата обращения 07.08.2020.
5. Cruden's Complete Concordance to the Old and New Testaments. - URL: <https://archive.org/details/crudenscomplet00crud/page/6/mode/2up> - дата обращения 30.01.2020.
6. Crystal D. The Cambridge Encyclopedia of the English Language. - Cambridge: Cambridge University Press, 1995. - 491 p.
7. ISLOMOV ELDOR YUSUPOVICH, AHMEDOVA MEHRINIGOR BAHODIROVNA. THE ESSENCE OF SPIRITUALITY IN THE UZBEK LANGUAGE. XIII МЕЖДУНАРОДНАЯ НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ " ЯЗЫК И КУЛЬТУРА" Челябинск, 26 апреля 2018 года. - P.12-15
8. Akhmedova Mekhrinigor Bahodirovna. "ANALYSIS AND DIFFERENT INTERPRETATIONS OF THE CONCEPT OF SPIRITUALITY". Indonesian Journal of Innovation Studies, Vol. 18, May 2022, doi:10.21070/ijins.v18i.590.
9. Magdalena NGONGO, Akhmedova Mehrinigor. A Systemic Functional Linguistic Analysis of Clauses Relationship in Luke Gospel Text, Janji Baru Using Kupang Malay. Studies in Media and Communication Journal. Vol.11, 2023. - P. 33-40.
10. Fitria Nur Hasanah, Rahmania Sri Untari, Shofiyah Al Idrus, and Akhmedova Mehrinigor Bahodirovna. Excel in Critical and Creative Thinking in Object-Oriented Programming. H. Ku et al. (Eds.): ICARSE 2022, ASSEHR 748, 2023. - P. 301–305.