EMJ MS
EUROPEAN MULTIDISCIPLINARY JOURNAL OF MODERN SCIENCE
ISSN 2750-6274 https://emjms.academicjournal.io Volume: 21 | Aug-2023

Film Discourse: Corpus Analysis and Synchronic Perspective

Erkaeva Dilnoza Bakhtiyorovna
English literature department, Bukhara State University,
d.b.erkayeva@buxdu.uz

Abstract: This article analyzes corpus analysis in film discourse. The synchronic perspective of corpus linguistics is analyzed with the help of examples.

Keywords: Synchronic perspective, corpus linguistics, film discourse, lexicography, Brown corpus.

Introduction. Corpus linguistics is a factor to analyze film discourse and its synchronic approach.

Main part. Yu.A. Volosnova emphasizes that in 1967 G. Kuchera and N. Francis published the work "Computational analysis of modern American English" *(" Computational Analysis of Present - Day American English ")* on the material of the Brown corpus, which marked the beginning of research involving corpus (Volosnova, 2006, 45). It is noted that in the work of N. Francis and G. Kucher, an analysis of a corpus of 1 million words is carried out (Ku c era , Francis , 1967). This quantitative indicator indicates the possibility of analyzing a large volume of material, despite the fact that hull technologies at that time were still at the inception stage.

V.P. Zakharov makes an important observation, pointing out that "the appearance of the Brown Corps aroused general interest and lively discussions. First of all, they touched upon the principles of selecting texts and the composition of tasks potentially solved on such a corpus" (Zakharov, 2005, 4). M.I. Solnyshkin and G.M. Gatiyatullina draw attention to the fact that "with the development of the Brown corpus, the concept of "reference corpus" appeared, which began to characterize all of the listed corpus, as researchers tested their assumptions and theories (the so-called "intuitive *corpus"). data")* using these corpora" (Solnyshkina and Gatiyatullina, 2020, 140).

The ability to test assumptions on the basis of the reference corpus seems to us an important step, since this contributes to the objectivity of the obtained linguistic data.

Later, a similar corps was formed *" Lancaster - Oslo - Bergen Corpus " (LOB)* , containing British English and consisting of 1 million words

(https://varieng.helsinki.fi/CoRD/corpora/LOB/) . _ _ From 1953 to 1987, work was underway to compile a corpus of British colloquial English *" The London - Lund Corpus "* (http://martinweisser.org/corpora _ _ _ _ _ _ site /1 st - gen - corp . html). The projects *"the Survey of English Usage" (SEU)* and *"the Survey of Spoken English "(SSE)* (Gatiyatullina, Bereznikov, 2017). N.V. Kozlova makes an important observation and focuses on the fact that in the early 1980s, the first oral corpora appeared based on American English (Kozlova, 2013, 82). Expanding the repertoire of corpora and the selection of oral text corpora separately indicates a certain technical progress, as a result of which, already at that time, the user could choose the most suitable corpus from the available options in accordance with the

purpose of his appeal.

At the same time, special professional communities are gradually beginning to be organized, within which an active discussion of the results achieved and possible prospects is held. In 1962 , the Association for Computational Linguistics *(ACL)* was created , under the auspices of which international conferences are held annually (https://www.aclweb.org/portal/) . Later given association received separation By territorial principle : " *European Chapter of the Association for Computational Linguistics" (EACL)* (https://www.aclweb.org/portal/eacl) And *"North American Chapter of the Association for Computational Linguistics" (NAACL)* (https://www.aclweb.org/portal/naacl). The distinction contributed to the emergence of highly specialized studies, in which language features were recorded.

M.I. Solnyshkin and G.M. Gatiyatullina found that "by the mid-1970s. the first bases for the storage and distribution of electronic corpora were created: the Oxford Archive of Machine Readable Texts *OTA (Oxford Text Archive)* (1976) and the International Archive of Electronic Texts of Modern English *ICAME (International Computer Archive of Modern English)* (1977) " (Solnyshkina, Gatiyatullina, 2020, 137).

IN 1978 _ was organized association *"The Association for Computers and the Humanities" (ACH)* (http://ach.org) . It is noted that this structure is one of the two major international professional organizations in the field of studying language and literature based on corpora (Baker, Hardie, McEnery, 2006, 15).

From the 1980s to the early 1990s, *the British National Corpus (BNC) was compiled.* The next corpus that continued development in this direction is *English Lexical Studies* . It was started in Edinburgh in 1963 and completed in Birmingham (Mansour , 2013, 12). The main developer of this project was John Sinclair, who became the first scientist who applied the corpus to the study of vocabulary and used the concept of "collocation" (Mansour, 2013, 12).

It needs to be clarified that in the early 1970s the repertoire of corpus was very limited, as was the number of functions they provided (Johansson , 2008, 48). However, O.V. Nagel rightly notes that "with the growth of the capabilities of modern computer technology, since the mid-1980s. corpus linguistics is rapidly developing, corpus projects of various scales in different languages and for various purposes began to appear actively" (Nagel, 2008, 53-54). N.V. Kozlova emphasizes that later special coordination centers arose for the collection, storage, distribution and creation of oral corpora. For example , *"Linguistic Data Consortium" (LDC)* (http://www.ldc.upenn.edu), *"Center for Spoken Language Understanding" (CSLU)* (http://www.cslu.ogi.edu), *"European Language Resources Association" (ELRA)* (http://www.elra.info) (Kozlova , 2013, 82).

In the 1980s years was created morphological analyzer texts *"The Constituent Likelihood Automatic Word-tagging System" (CLAWS)* (Solnyshkina , Gatiyatullina , 2020, 143). As a result, it became possible to carry out automatic markup by parts of speech. Note that a later version of this program *"CLAWS4"* used in *"British National Corpus* " (https://www.clarin.ac.uk/claws) . _ _ _ _ _ _ _ _ _ In turn, the version of " *CLAWS 7 "* (http://www.nateorp.ox.ac.uk/docs/claws 7.html # Toc 334867959) already includes 137 tags for the lexical list, the list of suffixes and the list of phraseological units (Solnyshkina , Gatiyatullina , 2020, 150).

In 1983, the first professional conference *"The European Association for Lexicography " (Euralex),* which aims to exchange ideas and information in the field of lexicography (Baker , Hardie , McEnery , 2006, 68). Since 2009, the conference *"e - lexicography in the 21st ‑ century "(eLex)* (Krek , 2019, 115).

The Text Coding Initiative movement was launched. *Encoding Initiative » (TEI),* whose task was to form standards for the design of electronic texts (Lavrentiev, 2004, 125). Note that at

this moment the markup language *" Standard Generalized markup Language " (SGML)* developed at the University of Oxford (Crystal , 1994, 439). A.M. Lavrentiev focuses on the fact that the main product of this movement is the "Recommendations for Encoding and Interchange of Electronic Texts" using SGML or XML standards (Lavrentiev, 2004, 125).

It seems important to point out that at this stage, scientists are beginning to be interested not only in the content of the corpus itself, but also in the inclusion of metalinguistic information in it, which is a kind of key to the primary data. Metalinguistic information can be recorded in the headings of written files, including the date of publication, information about the medium of the text, its genre, level of complexity, the target audience and its size, age and gender (Baker, 2006, 40) . For colloquial corpora (regardless of their types), information about different speakers in each text can be encoded, such as an indication of their age, gender, socioeconomic status, mother tongue or region (Baker, 2006, 40) . The ability to include metadata in a corpus increases the value of corpora for lexicography (FaaB , 2018, 135).

In the 1990s, on the basis of the University of Stuttgart in Germany, the query language Corpus was *developed. Query Language "(CQL)* which is used on the *"Sketch Engine" platform* (Kilgarriff, Baisa, Busta, Jakubicek, Kovar , Michelfeit, Rychlÿ, Suchomel, 2014). It should be noted that on the basis of this platform, a study of modern English-language film discourse was conducted, the results of which are described in Chapter III, since one of the purposes of the platform is precisely in the analysis of discourse (Kilgarriff, Baisa, Busta, Jakubicek , Kovar , Michelfeit , Rychlÿ, Suchomel , 2014).

Later in 1992, the European Corps Initiative was *created. Corpus Initiative" (ECI)* is an international organization whose goal was to form a multilingual corpus for scientific purposes (Sysoev, 2010, 102). The case has been available on CD - ROM since 1994 (Baker, Hardie , McEnery , 2006, 69).

The next step was the development of standards in a number of areas of corpus linguistics and in the direction of natural language processing *" Natural language Processing " (NLP)* within the framework of the project *" Expert advisory group on language Engineering Standards " (EAGLES)* that existed from 1993 to 1996 (Baker , Hardie , McEnery , 2006, 71).

In 2009, *the Google Books Ngram Viewer ,* which includes digitized texts of books from 1500 to 2008 (https :// books . google . com / ngrams) .

S.A. Manik notes the following trend in the composition of corpora. "In the United States and Europe, the largest lexicographical houses turned to the creation of corpora of literary English *(" Brown Corpus ", " British National Corpus , Collins wordbanks Online English Corpus ", " Cambridge English Corpus ", " The Longman Learners ' Corpus ", " The Macmillan World English Corpus ", " CORPORA ", " Oxford English Corpus "* , etc.) for further compilation and publication of dictionaries. In turn, Russian linguists at the same time began attempts to systematize and digitize terminology using terminological data banks, and the corpus of the national Russian language appeared only at the beginning of the 21st century" (Manik 2016, 17). Indeed, the national corpus of the Russian language (http://www.ruscorpora.ru/new/index.html) began to function only in 2004 , but work on its creation began as early as 1985 on the initiative of Academician A.P. Ershov (http :// cfrl . ruslang . ru) . T.V. Tolstova emphasizes the contribution of specialists from the Institute of the Russian Language. V.V. Vinogradov RAS, the Institute of Linguistics RAS, the Institute for Information Transmission Problems RAS, the All-Russian Institute for Scientific and Technical Information RAS and the Institute for Linguistic Studies RAS in St. Petersburg", who worked together on the project (Tolstova, 2018, 60).

In addition, modern corpora try to record all the linguistic changes taking place in society.

From 2010 to the present, the *“ The NOW Corpus " (News on the Web)* which contains electronic versions of newspapers and magazines (https://www.english-corpora.org/now/) . ____ In May 2018, *"The iWeb Corpus" was created,* which is based on information taken from almost 95,000 websites (https://www.english-corpora.org/iweb/) . In connection with the spread of coronavirus infection *"COVID-19"* in May 2020 , The Coronavirus *Corpus* was developed with 325 million words (https://www.english-corpora.org/corona/) . _ *Sketch Engine* Platform also offers a freely available corpus on Covid - *19 with over 224 million words* (https : //app .sketchengine .eu/#open) . These corpora have significantly expanded the repertoire of existing narrow corpora and confirmed the trend that corpora appear in a timely manner and reflect a certain linguistic stage in the life of society.

It is important to emphasize that specialized corpora for the film industry are being developed, for example, *"The TV Corpus"* includes 325 million words from 75,000 television episodes from the 1950s to the present (https://www.english-corpora.org/ tv/). The *Movie Corpus* contains 200 million words from over 25,000 films from the 1930s to the present (https://www.english-corpora.org/movies/) . Let us note the existence of *the "Multimedia Russian Corpus" (MURKO)* , compiled on the basis of cinematographic material (Grishina, 2009, 175). *"The House M.D. Corpus" (HMDC)* contains data from a television drama comedy, on the basis of which neologisms, slang words, etc. are identified. (Law, 2019). Multilingual corpus of translations *"MuST-Cinema"* created on the material of the subtitles of the *TED* conference speeches (Karakanta, Negri, Turchi, 2020). The aim of the researchers was to develop an automatic method for annotating a corpus of subtitles. On the basis of subtitles in the bilingual corpus *"CORSUBIL" (Corpus de Subtítulos Bilingües ingles-español),* the translation transformations chosen in audiovisual translation are considered (Rica-Peromingo, Martín, Riaza, 2014).

The repertoire of the narrow corpora mentioned above emphasizes their practical significance for the study of linguistic aspects, including film discourse.

In addition to the diversity of existing corpus technologies, it should be noted that they are anthropocentric, that is, they are increasingly user-oriented *(user - friendliness)* . For example, on *the Sketch Engine "* available to use a variety of colors to highlight different types of information, it is also possible to set custom settings for the location of information using the options *" view options ", " change view options »* (https://www.sketchengine.eu) . _ _ _ _ _ In addition, modern cases offer users a user-friendly interface, on-screen prompts, and technical support (Kosem , 2016, 78).

Conclusion. Thus, having considered the formation of corpus linguistics in the 21st century, we conclude that scientific and technological progress has had a significant impact on the development of this science. Imperfect corpus technologies are being replaced by more modernized and user-oriented programs, the repertoire and capabilities of which allow linguistic research to be carried out at a higher scientific level.

References:

1. Cruden’s Complete Concordance to the Old and New Testaments. - URL: https://archive.org/details/crudenscompletec00crud/page/6/mode/2up - дата обращения 30.01.2020.
2. Crystal D. The Cambridge Encyclopedia of the English Language. - Cambridge: Cambridge University Press, 1995. - 491 p.
3. 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

4. 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.
5. 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.
6. 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.
7. Hazim Hazim, Ratih Puspita Anggraenni, Akhmedova Mehrinigor Bahodirovna. Altruistic Actions in COVID-19 Corpses Care: Empathy, Modeling, and More. International Conference on Advance Research in Social and Economic Science (ICARSE 2022), 2023/4/27. - P.476-484