

Corpus Linguistics in Analyzing Film Discourse: A Diachronic Perspective

Erkaeva Dilnoza Bakhtiyorovna

English literature department, Bukhara State University

d.b.erkayeva@buxdu.uz

Abstract: This article gives full overview of analysis of corpus linguistics in film discourse. The diachronic perspective of corpus linguistics is analyzed with the help of examples.

Keywords: Diachronic perspective, corpus linguistics, film discourse, lexicography, applied linguistics.

Introduction. The need for systematization and generalization of the found empirical material existed long before the advent of corpus linguistics. Researchers actively developed special cards with data, which were drawn up in file cabinets. In order to organize knowledge printed concordances, glossaries, encyclopedias (Cambridge world, 2019). The first concordances were made to the texts of the Bible in order to fix the lexical layer of the Holy Scriptures. The processing of information for lexicographic publications without the use of technical means took a lot of time. P. Hanks notes that before the use of corpus technologies until 1987, there were two ways to collect information for compiling dictionaries: collecting quotations from literature using a special program that allows you to copy information from texts; using the intuition of lexicographers (Hanks , 2020, 219).

With the advent of technically more advanced tools, the process of developing reference literature began to go faster due to process automation. The formation of corpus linguistics as a separate science, which allows even more saving of human resources and time, was preceded by various ideas about the language, the change of which indicates a change in the scientific paradigm and scientific knowledge.

Literature Review. Structural linguistics, which developed in 1920-1930 (represented by F.F. Fortunatov, L. Bloomfield, F. de Saussure), is a special direction in which language is understood as a sign system "with clearly distinguished structural elements", which tends to strict description of the language (<https://bigenc.ru/linguistics/text/4169948>). Thus, the followers of structural linguistics strove for maximum objectivity in describing the structure of the language by means of speech analysis. However, O.V. Lukin notes that "when studying language as a formalized static system, most schools of structuralism ignored fundamental general theoretical problems such as language and thought, language and society, the origin of language, and many others" (Lukin, 2015, 17).

In the late 1950s, generative linguistics (Chomskian linguistics) emerged, the founder of which was the American linguist N. Chomsky. Much attention in this direction was given to the linguistic personality of a person, his thinking and intuition. In accordance with this approach, "language is understood as a model that can generate an infinite number of correct sentences" (Lukin, 2015, 19). N. Chomsky draws a fundamental distinction between competence (knowledge of one's language by the speaker-hearer) and use (actual use of the language in specific situations) (Chomsky, 1972, 9). The scientist believes that "the grammar of the language tends to be a description of the competence inherent in the ideal speaker-hearer" (Chomsky, 1972, 9). This idea did not find

support within the scientific community, which believed that in this case the important context was outside the scope of linguistic analysis. However, the context plays a significant role in the study of the language of a particular subject area, therefore, in our work, much attention is paid not only to a separate lexical unit, but also to the features of the communicative situation in which it functions.

At the same time, N. Chomsky proclaimed the idea that language is infinite. Indeed, the presence of linguistic diversity is beyond doubt, so linguists are trying to identify, describe and, if possible, systematize the ongoing changes. Invaluable help in ordering a large amount of information is provided by corpus technologies that have appeared in the framework of corpus linguistics.

Further, it seems necessary for us to consider the definitions of "corpus" (in linguistics), "corpus linguistics", "computer linguistics", starting from the more general direction "applied linguistics" in the works of domestic and foreign scientists.

B.Yu. Gorodetsky makes important observations about the status of applied linguistics (applied linguistics) and notes that it “encompasses a wide field of joint activity of representatives of various sciences engaged in solving extremely important (but at the same time very heterogeneous) practical problems” (Gorodetsky, 1983, 6).

Applied linguistics is broadly defined as a practical activity aimed at solving language problems in society (Dobric, Graf, Onysko, 2016).

T.V. Tolstova rightly focuses attention on the fact that “in the USSR, the term “applied linguistics” became widespread in the 1950s in connection with the appearance of the first computer systems for automatic processing of text information (machine translation, automatic abstracting, etc.). That is why the terms “computational linguistics”, “computational linguistics”, “automatic linguistics” and “engineering linguistics” are often used in the Russian-language literature to this day instead of the term “applied linguistics” in the same sense (Tolstova, 2018, 62) .

In turn, A.S. Syrchina points out that “corpus linguistics is a branch of applied linguistics that develops general principles for constructing and using linguistic corpora (linguistic corpus/text corpus) ” (Syrchina, 2016).

V.P. Zakharov and S.Yu. Bogdanova believe that "corpus linguistics is a branch of computational linguistics that develops general principles for the construction and use of linguistic corpora (corpus of texts) using computer technology" (Zakharov, Bogdanova, 2011). Scientists further refine the definition, emphasizing that corpus linguistics is a branch of applied linguistics, the object of study of which is a corpus of texts, and the subject of study is “the theoretical foundations and practical mechanisms for creating and using representative arrays of linguistic data intended for linguistic research in the interests of a wide range of users » (Zakharov, Bogdanova, 2011, 10).

Thus, applied linguistics includes computational and corpus linguistics. In the following, it is necessary to consider these components in more detail.

Analysis. A.V. Guslyakova identifies the following areas of computational linguistics:

- 1) natural language processing (NLP - natural language processing);
- 2) corpus linguistics, creation and use of electronic text corpora;
- 3) creation of electronic dictionaries, thesauri, ontologies;
- 4) automatic translation of texts by means of specialized programs;
- 5) automatic extraction of facts from the text;
- 6) autoreferencing;

- 7) building knowledge management systems;
- 8) creation of question-answer systems;
- 9) optical character recognition;
- 10) automatic speech recognition and synthesis (Guslyakova, 2016, 47).

A.V. Guslyakova comes to the conclusion that computational linguistics (computational linguistics) is gradually becoming an important component of artificial intelligence technology (Guslyakova, 2016, 46).

Having considered the definitions of scientists, we conclude that computational linguistics, formed within the framework of applied linguistics, can be divided into some sections, one of which is corpus linguistics (Fig. 1).

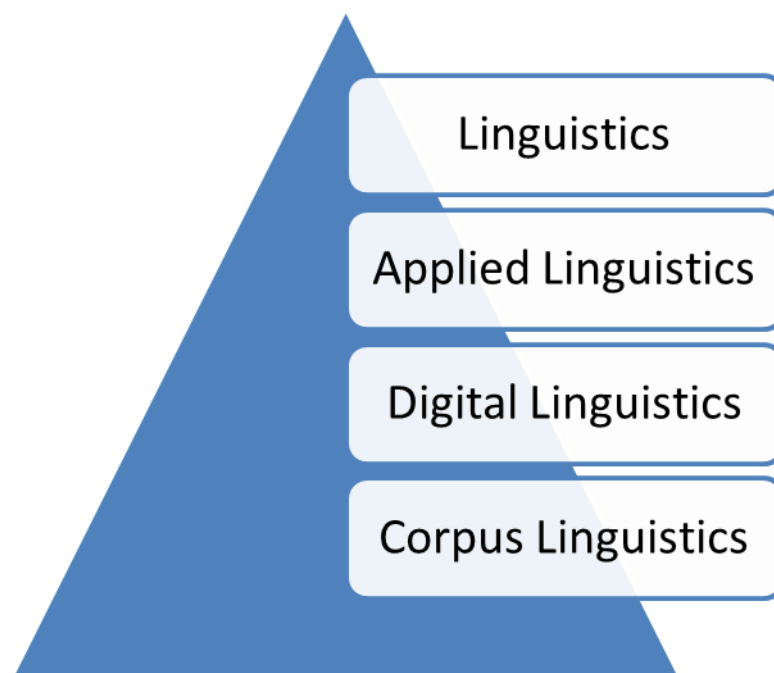


Figure 1. Structural division of linguistics.

M.V. Kopotev and A. Mustajoki identify several meanings of corpus linguistics. Thus, they believe that this term includes the theory and methodology of creating corpora, as well as corpus research, that is, “language research using corpus methods” (Kopotev, Mustajoki, 2008, 11). Scientists emphasize that there is no clear distinction between the two interpretations, however, corpus linguistics, in their opinion, should be understood as corpus studies (Kopotev, Mustajoki, 2008, 11). This observation confirms the predominance of works aimed more at describing the corpus analysis itself, and not at the corpus creation algorithm.

Thus, corpus linguistics is a special branch of applied linguistics that deals with the solution of theoretical and practical problems associated with the development, creation and use of corpora. In our opinion, it is fair to single out corpus linguistics as a separate discipline, since it has its own subject and object of study, as well as specific goals and objectives.

Discussion. In turn, the main object of corpus linguistics is the linguistic corpus. A corpus is a collection of texts in writing or transcribed speech that can serve as the basis for linguistic analysis and descriptions (Kennedy , 1998, 1). Y. Ding notes that a corpus is a collection of authentic texts that represent linguistic diversity (Ding , 2018, 665). The inclusion of authentic texts in the corpus

allows the scientist to work with real speech samples created by native speakers, which emphasizes their research value. Thus, the corpus of authors, which will be considered in more detail in Chapter III, is developed on the material of the periodicals "American Cinematographer" and "Total Film", reflecting the modern language of the studied subject area «cinema».

D. Crystal believes that corpora can include both whole texts and certain sections, for example, dialogues, magazine articles, brochures, newspapers, lectures, sermons, a recording of a broadcast or chapters of a literary work (Crystal, 1994). Our study will analyze entire texts of selected issues of periodicals, such as interviews with movie stars, reviews of cinematic products, etc.

V.A. Plungyan emphasizes that "the corpus is not only a powerful tool for studying the language, but also a new ideology that orients the researcher to the text as the main object of theoretical reflection" (Plungyan, 2008, 14).

G. Kennedy proposes the following classification of buildings in accordance with the evolutionary stages:

- ✓ preelectronic corpora (biblical and literary studies, early dictionaries, etc.);
- ✓ the first generation of large hulls;
- ✓ the second generation of mega-corps (Kennedy, 1998, 13).

M.I. Solnyshkin and G.M. Gatiyatullina expand the above classification and distinguish third-generation cases, which they also call giga cases (Solnyshkina, Gatiyatullina, 2020, 150). This name seems apt to us, because it emphasizes the significant amount of information that modern corpora are able to include and process.

Examples of preelectronic corpora are: Alexander Cruden's complete Old and New Testament concordance ("Cruden's complete Concordance of the old and New Testaments"), created in 1800 (<https://archive.org/details/crudenscompletec00crud/page/6/mode/2up>);

American dictionary English language ("An American Dictionary of the English Language") 1828 (<https://archive.org/details/americandictionary01websrich/page/1/mode/2up>);

dialectal dictionary English language ("The English Dialect Dictionary") 1905 (<https://archive.org/details/englishdialectdi06wriguoft/page/n5/mode/2up>); Modern English grammar based on historical principles ("A Modern English Grammar on Historical Principles") (<http://arrow.latrobe.edu.au/store/3/4/3/2/5/public/B11775361V1.pdf>) And t.p.

First generation major buildings includes V yourself: "The Brown Corpus", "FROWN" (Freiburg-Brown Corpus of American English), "AmE06 Corpus", "B-BROWN", "The Lancaster-Oslo/Bergen (LOB) Corpus", "Freiburg LOB Corpus of British English" (FLOB), "BE06 Corpus", "The Kolhapur Corpus of Indian English", "The Wellington Corpus of Written New Zealand English", "The Australian Corpus of English" (ACE), "The Corpus of English-Canadian Writing", "The London-Lund Corpus" (LLC) (<http://martinweisser.org/corpora/site/1st-gen-corp.html>).

Open American National Corpus (ANC), British National Corpus (BNC), International Corpus of English (ICE), Corpus of American English (COCA) (<http://martinweisser.org/corpora/site/2nd-gen-corp.html>) are corps second generations.

M.I. Solnyshkina And G.M. Gatiyatullina believe that To corps third generations or gigacorporuses Can attributed "Corpus of Contemporary American English" (COCA) And "Google Books" (Solnyshkina, Gatiyatullina, 2020, 154).

P.V. Sysoev writes that the creation of the first corpus of texts "Brown Corpus" associated with scientists N. Francis (Nelson Francis) and G. Kuchera (Henry Ku c era), who carried out

developments at Brown University in the USA (Sysoev, 2010, 101), after which this corpus got its name. Complete Name corpus sounds next image : "Standard Corpus of Present-Day Edited American English for Use with Digital Computers" (<http://icame.uib.no/brown/bcm.html>).

Speaking about the terminology of corpus linguistics in Russian, V.P. Zakharov and S.Yu. Bogdanova indicate that the plural term "corps" can have the variants "corps" or "corps" (Zakharov, Bogdanova, 2020, 16). However, scientists prefer the "corps" option (Zakharov, Bogdanova, 2020, 16). This observation explains the variability in the use of the plural form in the works of Russian scientists. For the purpose of uniformity, following the scientists V.P. Zakharov and S.Yu. Bogdanova in this study, the plural form "corpuses" is used.

Conclusion. In the framework of this study, it is necessary to mention the possibility of creating narrow buildings, the content of which varies depending on the goals and objectives set by the scientist.

Based on large corps, for example, " British National Corpus ", a variety of highly specialized studies are carried out, in particular, the frequency of inclusion of biblical idioms in the British National Corpus is revealed (Pinnavaia , 2012). However, we emphasize that for this kind of research, highly specialized cases are increasingly being created, which can be filled with material at your discretion. For example, the case " Rapid Dialogue Games " (RDGs) contains player dialogue recordings (Paetzel , Racca , DeVault , 2014, 4189), " Air Traffic Control " (ATC) consists of recordings of conversations between pilots and US airport controllers (Baker , Hardie , McEnery , 2006, 8). Scholars have attempted to create a corpus (consisting of students' opinions) to evaluate the curriculum (Kotani , Yoshimi , Uchida , 2018).

References:

1. An American Dictionary of the English Language: Intended to Exhibit, 1828. -URL: <https://archive.org/details/america01websrich/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.
 11. 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