THEORETICAL AND PRACTICAL APPROACH TO CORPUS BASED ANALYSIS OF LANGUAGE DEVELOPMENT

Ataboev, Nozimjon Bobojon ugli
Doctor of Philology (DSc), associate professor
Dean of the Faculty of Foreign Languages of BukhSU
E-mail: n.b.ataboyev@buxdu.uz
anb929292@gmail.com
https://orcid.org/0000-0002-9756-6849

Abstract: In the article, the description of the progress of this branch of science to date and the role of corpus analysis methods in the research on the enrichment of the lexical layer of the language are clearly demonstrated.

Key words: corpus linguistics, linguistic corpus, language development, diachronic corpus.

INTRODUCTION

Determination of the research subjects of corpus linguistics is carried out by constantly studying and researching the possibilities of language use in real life and the ways in which language users use the language. This process is perfected as linguists look at it from different angles. Corpus linguistics has gone through several periods until its current development. It was formed as an independent branch of science in the late 70s of the last century, and the linguistic research methods, methods and analysis tools that are its basis have been known since the 12th century. [5]. In the development of corpora, the emergence of the principles of text size and their sorting spans several periods. Initially, the period before the emergence of electronic corpora covers the period from the 13th century to the beginning of the 1960s. Advances in technology have given rise to corpus linguistics, which has freed researchers and linguists from having to spend a lot of time sorting through words. This period is associated with the 1970s, that is, with the creation of the first personal computers.

When it comes to corpus linguistics, the term corpus, which is directly studied as an object of the field and a modern approach to working with texts, requires a special explanation.

The dictionary meaning of the word "corpus" was initially used in a narrow circle, that is, the composition of the works of a certain writer in the religious and literary genre was arranged in alphabetical order, i.e., in concordance-concordance lines, it was called a corpus. The use of corpus in this sense covers the period before the emergence of electronic corpus. Corpora were formed during this period for religious, literary, and lexicographical research, and this process required long and labor-intensive manual labor.

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 14 Issue: 07 in July-2025
https://www.gejournal.net/index.php/IJSSIR

In the era before corpora were formed electronically, specific symbols were used to indicate where words and phrases were used in the text. One of the first such concordance series was created based on the holy book of the Bible. In the concordance series, all words were placed in alphabetical order and the places where they were used were indicated, which, in turn, made it possible to learn the meanings of a word. A. Cruden suggests that these biblical concordance lines should be viewed as a dictionary or tagged database [6]. Agreeing with the opinion of the scientist, we recommend including any base in the series of frequency dictionaries, which is arranged in alphabetical order and shows the level of activity of words. Such experiences not only reflect a person's interest in religious and artistic works, but also allow the reader to understand the meaning of the words before reading the work, and prevent possible difficulties during the reading process.

At the beginning of the 21st century, concordance programs are distinguished by their high functionality: one program is able to quickly create word lists, collocational units, and determine the level of activity of words [21].

Thus, since the early 1990s, technical capabilities have allowed scientists to construct and develop large-scale corpora. The corpora created during this period set themselves the goal of covering a wide range of forms of the language that are manifested in written and spoken speech, thus reflecting all the diversity of the language. It was possible to automatically tag spoken corpora at the phonetic, morphological, lexical, syntactic and discursive levels. In addition, a number of programs for automated processing of concordances have appeared.

Scientists such as G. Kennedy, P. Baker, A. Hardy, T. McEnery call the period when computers were not used and when there was a need to involve human resources on a large scale, the first generation corpora, and the corpora developed since the end of the 80s are called the second generation corpora or mega-corpora, as their size covers about 100 million words. These corpora traditionally include Longman Corpus Network, Longman Corpus Network (1991), Bank of English, The Bank English, BoE (1993), British National Corpus, BNC (1994), The American National Corpus, ANC (2008). Later, the corpus of spoken English was included in the spoken part of the British National Corpus (BNC) [23, 21, 14].

Another famous project of its time was The International Corpus of English (ICE), developed in 1996 at University College London under the leadership of S. Greenbaum. The aim of the project was to collect texts of cross-regional variants of the English language. Subcorpora include spoken and written texts of cross-regional varieties of English. These regions are: Britain (ICE-GB), East Africa, India, New Zealand, Singapore, Canada, Hong Kong, Jamaica, Philippines, USA, Cameroon, Fiji, Ireland, Kenya, Malta, Malaysia, Pakistan, Sierra Leone, Sri- Lanka, Trinidad and Tobago. Respondents were individuals over 18 years of age with a high school education in an English-speaking school. In all subcorpora, it was ensured that 60% of texts consisted of written texts and 40% of spoken transcripts. The sub-corpus of dialogic speech includes the following genres of oral speech: private conversations (personal meetings and telephone conversations) and public (lessons, radio and television conversations, radio interviews, parliamentary debates, business negotiations). The

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 14 Issue: 07 in July-2025
https://www.gejournal.net/index.php/IJSSIR

subcorpus of monologue speech is divided into two parts: the first includes sources of spontaneous speech (comments, speeches at demonstrations and in court); the second part included pre-prepared speeches (television and radio news, television and radio conversations (talk shows). Only by 2006, audio recordings of speech began to be included in the corpus [14]. Since the ICE corpus is a type of normalized diachronic corpus, it needs a wider analysis.

Thus, second-generation corpora are corpora with a size of at least one hundred million tokens, whose purpose is to represent all types of written and spoken speech. Corpus builders strive to represent as many genres and styles of spoken and written discourse as possible for a diverse range of speakers. Typically, this is a concept specific to corpora collected and tagged online according to corpus building requirements. National corpora require a long time and deep attention from the compiler and are created based on the principles of representativeness of text selection and rules specific to second-generation corpora. In the 1990s, the British National Corpus was used as a model for a new national corpus. However, it was found that diachronic corpora do not meet the standards of the research topic. In our opinion, if the corpus base does not reflect a certain period of time in chronological continuity, it will not be appropriate to recognize it as a representative of the national language.

From 1987 to 2004, software for collecting corpora, compiling metamarkers, and automatically tagging texts was developed, and a new phase of the field emerged, the third era, called Third Generation Corpora, or hegacorpora. The beginning of 2010 is characterized by the appearance of great technical possibilities: the fourth generation concordances - BNCweb (2009), CQPweb (2012), SketchEngine (2013), Wmatrix (2013) were developed, which are similar to the third generation concordances.

After the transition of the corpora to the online system, the speed of finding the lexical units in search has increased and the number of users is expanding. Direct access is done in an online search through a web browser [21]. M. Davies calls the fourth-generation concordancers hybrid corpora, because their interface is a unique common space for creating a corpus at the levels of morphemic, lexical, syntactic and phraseological units and conducting an analysis of words on the levels of activity [8].

MAIN PART.

The trend of increasing the size of the cases continued even after the 2000s. A. Mauranen, S. Kubler and H. Zinsmeister describe this generation with the motto "the bigger the corpus, the better", and L. Flowerdrew was the first to call this period the generation of gigacorpus [15, 20, 9]. Meanwhile, several new corpora (COCA, Google Books Ngram) appeared and their size reached several billion tokens. The large size of the corpora allowed extensive frequency studies and the study of collocations consisting of three, four and more words. Such categories are called lexical bundles by D. Bieber and K. Hyland, where one word can be variable [4, 12]. For example, in the five word combination: in the beginning of the, in the end of the, in the form of the- the variable is the third word. Later, these collocations were called n-grams, where bigrams are collocations consisting of two words,

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 14 Issue: 07 in July-2025
https://www.gejournal.net/index.php/IJSSIR

trigrams are collocations consisting of three words, and n-grams are collocations consisting of n words. Currently, the identification of such collocations can often be found on the Internet itself, of course, this opportunity was created by the creation of large megacorpora. In addition, such corpora offer the ability to plot n-gram frequency over different time periods from 1800 to 2010. This is one of the most important requirements for creating diachronic corpora.

The emergence of mega and gigacorpora has shown that large data corpora are not suitable for studying the speech of certain professions or speech genres, because large corpora, despite their enormous size, mainly contain texts from the most common genres of spoken and written speech. Late 1990s early 2000s. it is proved that the principles of representativeness of special corpora are observed in significantly smaller volumes, as the frequency of both terms and neutral words remains stable and the same. In this context, L. Flowerdew says that written corpora of less than 250,000 characters are considered small.
Thus, this period is characterized by the merging of corpus linguistics methods with the World Wide Web: programs were created to automatically download texts from the Internet, to treat the World Wide Web as a corpus (the Google Books corpus), as in the NOW and GloWbe corpora, and to use the World Wide Web with its own tools access (SketchEngine , BNCweb). In addition, it became possible to track the use of a certain word in large data sets, for example, the change of the form and meaning of the word in written (Google Books) or spoken speech over time.

In addition, the practice of creating corpora focused on these linguistic studies has been the focus of attention of all scientific schools of the world. For example, in their research, Korean scientists developed corpus analysis methods for classifying types of emotions in the Korean language and developed the Korean Emotion Analysis Corpus [13]. Based on the structure of the created corpus, scientists have not only gained a deeper understanding of the differences between classes of emotions in the classification of emotions, but have also created a new standard data set that allows the evaluation of emotion analysis approaches. This, in turn, shows that corpus bases can enable the researcher to observe and classify based on the reasonable accumulation of data.

The great contribution of Russian linguists to the development of corpus linguistics can be demonstrated by the example of the national corpus of the Russian language [27]. The National Corpus of the Russian Language is a linguistic corpus representing the Russian language that has been partially accessible through an online search interface since April 29, 2004. The Institute of the Russian Language of the Russian Academy of Sciences is constantly researching it. The corpus now contains more than 1 billion word forms, which are automatically lemmatized and POS-/gramme-tags, i.e. covering all possible morphological analyzes for each orthographic form. Lemmas, POS, grammatical elements and their combinations can be searched. In addition, 6 million word forms are contained in subcorpora with manual homonymy. A subcorpus with additional metadata related to morphological homonymy highlighted above is also automatically tagged. The entire corpus contains lexical semantics (LS) searchable tags, which include morphosemantic POS

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 14 Issue: 07 in July-2025
https://www.gejournal.net/index.php/IJSSIR

subclasses (noun, reflexive pronoun, etc.), corresponding LS features (subject class, causative, evaluation), derivation (diminutive, adverbial adverbs, adjectives, and etc.) consists of

Arabic linguists are also conducting extensive research on this issue. For example, a new language processing tool, Arabic Corpus Processing Tools ACPTs 4.6 Version, has been developed, which is a stand-alone open/free resource for analyzing large volumes of Arabic and English texts, with a database of more than 50 million words, compatible with more than 8 gigabytes of PCs. keeps The new ACPT has sophisticated state-of-the-art features applicable to the corpus linguistics and corpus linguistic analysis literature, especially statistical packages. When compared with other tools suitable for corpus linguistic analysis, this leads ACPTs to be noted as the most efficient tool for corpus analysis. Focusing on language teaching issues, the most important tasks of appropriate tools that can be used to improve the language teaching/learning process are included in this corpus [2]. As a continuation of such works, it is possible to mention the project of creating the International Corpus of Arabic (ICA), which was created for the first time. It is intended to include 100 million analyzed tokens with an interface that allows users to interact with the corpus data in several ways. ICA is a representative corpus of Arabic, created in 2006, which is intended to cover the Modern Standard Arabic language used throughout the Arab world. ICA was analyzed by the Bibliotheca Alexandrina Morphological Analysis Enhancer (BAM-AE). BAMAE is based on the Buckwalter Arabic Morphological Analyzer (BAMA) [22].

In our opinion, it is appropriate to focus on the elements that are the basis for the creation of the above-mentioned international corpus of the Arabic language. The website Alexandrina Bibliotheca is the core of the international Arabic corpus. Bibliotheca Alexandrina (BA) is one of the leading international institutions in Egypt, with a place in the dissemination of culture and knowledge, as well as in supporting scientific research. He launched the ambitious project to create the International Arabic Corpus (ICA) as a major effort to create a representative corpus of Arabic spoken throughout the Arab world, as this is the most common way to support research on the language. . After the corpus was created, the analyzed form was the first analyzed Arabic corpus available as a linguistic resource for researchers. It is also the first systematic analysis of cross-national studies in an Arabic-speaking community, which will be a very useful resource for linguists who believe that their theories and descriptions of language should be based on real data rather than on unsubstantiated facts. [1].

Chinese linguistics is also one of the leading fields in corpus linguistics research. Because, relying on the experience gained in creating a computer-based translator, the teaching system of the Department of Translation of Guangdong University of Foreign Languages, they summarize the principles and procedures for the development of educational translation and parallel corpus. Further research was conducted by university scientists on the integration of the translation corpus into an online autonomous learning platform for the training of translators. Starting from 2003, by selecting some texts from PACCEL-S, the process of quantitative assessment of pronunciation errors, grammatical

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 14 Issue: 07 in July-2025
https://www.gejournal.net/index.php/IJSSIR

errors and pauses was studied using the ParaConc program. According to them, the most frequent mistake in translation work of Chinese students is related to pronunciation, mainly vowels. Among 732 sentences interpreted by 183 examinees, there was an average of one pronunciation error per sentence, and each examinee made 3.64 such errors. The created corpus - ParaConc shows 90 grammatical errors, of which the most frequent cases are incorrect use of singular and plural forms and inappropriate use of speech units [11]. In this way, it can be said on the example of the research of Chinese linguists that it is possible to identify and eliminate errors in pronunciation based on corpus analysis.

In addition, a Corpus of Spoken Chinese is a collection of transcripts of spoken Chinese produced by non-native speakers and native speakers intended to be publicly available for researchers [26]. This corpus is the first case of creating a hitherto undeveloped colloquial corpus of its target Mandarin Chinese language, and serves as a great resource for the study of many English loanwords found in Chinese. takes

Also Y. Liu, M. Xiaohui Qin, L. Wang and Ch. Chinese scholars such as Huang created CCAE: A Corpus of Chinese-based Asian Englishes [19]. This, in its place, can open a wide way for scientists to study the scope of cross-linguistic interference and the influence of the dominant language on other languages.

Studies have shown that the practice of creating a national language corpus for India, which consists of a multilingual population, causes a number of complications, since the constitution of this country recognizes 2 national and 22 local languages as state languages. In our opinion, the languages in the territory of this country complement each other under the influence of interaction, communication and a single media space, causing the reduction of each other's vocabulary. In our opinion, the creation of a multilingual corpus for this nation could preserve the current diversity of languages.

Indian linguist N. Dash agrees with the above points and says: "As a multilingual country, India is a language giant. It maintains a large number of existing "living" languages of various ethnic and linguistic communities. Due to the lack of corpora, these languages suffer from the fact that they cannot achieve their technological progress. Building corpora can accelerate language learning to improve literacy, preserve endangered Indian languages, and protect languages that have lost their importance from the encroachment of English. Corpora are the only linguistic bases that can help them survive the battle of linguistic imperialism. In addition, it is characterized by the fact that it can provide statistically reliable data to restore lost position" [7; P 1]. The burning of the scientist is that in the era of globalization, if research is not properly oriented, in the current development of languages, due to the increase in the flow of popularized media text elements in other languages, weak languages will disappear or will be condemned to absorb many foreign languages, such as the Mandarin dialect of the Chinese language. This is exactly the issue that is raised in our research. After all, based on the materials that are the objects of research, if we conclude, there are languages that are ahead of us in development and production, and taking into account the fact that their influence is expanding in the prism of the media

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 14 Issue: 07 in July-2025
https://www.gejournal.net/index.php/IJSSIR

space, the practice of creating a diachronic corpus of the Uzbek language is also necessary. It is the demand of the time to set it up.

Without rejecting the above points, the study of language development means the analysis of the etymology of words, phrases and, in general, any linguistic unit that has entered the lexical layer. Etymological studies can in some sense be included in ethnographic studies. The practice of using language corpora in this field is evident in the works of K. Harrington. By the scientist, the role of corpus linguistics in the study of ethnography of a closed society is to analyze the interaction of immigrants in a refugee reception center in Ireland to determine the instinctive resourcefulness of people who are faced with the problem of communicating in the absence of a common language or culture [10]. Using three years of ethnographic observation and an innovative mix of applied methodologies, particularly corpus linguistics and conversation analysis, the following was achieved:

- Based on a corpus of 98,000 words;

- It is based on learning the use of English in the process of interaction with residents and English-speaking staff of the center;

- It was found that constructions such as speech community, communicative competence and interlinguistic interference are becoming more difficult and the purity of one language is relatively declining.

We believe that any society has a need for communication with another group of people. As this communication approaches, it affects language development in the following stages and sequence:

1. Social views harmonize with each other;
2. Religious faith and secular ideological rapprochement occurs;
3. Mutual sharing of scientific, technical and household achievements and inventions;
4. Dressing, customs, ceremonies, in a word, interference of cultures;
5. As a result of the above, the lexical layer changes and enriches under the influence of language interference.

In general, the process of interference of languages in the above mentioned five-stage sequence is accelerating today. Because the factors of the current process of communication are fundamentally different from the previous person-person and society-society categories and are being built on the basis of new, modern forms of information exchange. In our opinion, the acceleration of communication in this media field is reducing the scope of the purity and diversity of cultures and languages. Therefore, as long as the study of language development, which is the object of research, is not studied statistically based on the prism of corpora, the practice of preserving the current state of languages and leaving them as a basis for future research will not emerge. And this cannot be improved with any excuse.

According to Swedish scientist M. Kytö, electronic historical corpus and corpus methodology are research methods aimed at studying and evaluating the current state of languages and allowing to consider linguistic changes that may occur in the future. With this, the scientist emphasizes that within the wide range of corpus linguistic methodology,

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 14 Issue: 07 in July-2025
https://www.gejournal.net/index.php/IJSSIR

historical corpus linguistics has emerged as a vibrant field that has significantly increased the attractiveness for the study of language history and change. Indeed, the research process of evidence-based historical linguistics would not have been completed without the methodology and new impetus of corpus linguistics. In an era of rapidly changing life and research, increasing competition for academic careers and opportunities for young scientists, there is no doubt that the methodologically easy field has a future. Historical corpora and other electronic resources have also made the study of language history attractive: working on them engages students individually and interactively [16; P 417].

Among the researches in the world, modern research occupies a special place. An example of this is the above-mentioned historical corpora. The concept of historical does not refer to the age of the corpus, but rather refers to previous samples of the language covered by the corpus, archival documents, historical manuscripts, and previously published materials. The rapid development of this process can be seen especially in the case of the English-language COHA corpus.

The Corpus of Historical American English is the largest structured corpus of historical English. COHA contains more than 475 million words of text from 1820-2010, which is 50-100 times larger than other comparable historical corpora of the English language. In addition, the corpus is balanced in terms of genres by decades, that is, it is quantitatively equalized. Creation of the corpus stems from a 2008-2010 National Endowment for the Humanities (NEH) grant.

CONCLUSION.

In our opinion, it is important to create corpora of this type in the case of the English language as well. After all, if the number of manuscripts and published literature decreases at the same time when materials are being digitized, and if they are summarized in corpora in electronic form, any linguist conducting any linguistic research can come to a certain conclusion about the present day based on the history of the language.

REFERENCES

1. Alexandrina Bibliotheca available at https://www.bibalex.org/ica/en/About.aspx [retrived on 23 september, 2023]

2. Almujaiwel S. and Al-thubaity A. Arabic Corpus Processing Tools for Corpus Linguistics and Language Teaching // The Globalization of Second Language Acquisition and Teacher EducationAt: FukuokaVolume: 2. – 2016. 4 p. available at https://www.researchgate.net/publication/309351881_Arabic_Corpus_Processing_Tools_for_Corpus_Linguistics_and_Language_Teaching

3. Baker P., Hardie A., McEnery T. Glossary of Corpus Linguistics. –Edinburg: Edinburgh University Press, 2006. – 192 p.

4. Biber D. University Language: A Corpus-based Study of Spoken and Written Registers. - Amsterdam: John Benjamins, 2006. 261 p.

5. Corpus linguistics // available from: https://en.wikipedia.org/wiki/Corpus_linguistics [Retrived on 03.08.2023]

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 14 Issue: 07 in July-2025
https://www.gejournal.net/index.php/IJSSIR

6. Cruden, A. *A Complete Concordance to the Holy Scriptures of Old and New Testament*. - London: Fleming H. Revell Company, 1937. – 756 p.

7. Dash N.S. Language Corpora: Present Indian Need. - Indian Statistical Institute, Kolkata. 2004. – 1-6 pp. available at https://www.researchgate.net/publication/2938590_Language_Corpora_Present_Indian_Need

8. Davies M. Corpora: an introduction // The Cambridge handbook of Corpus Linguistics / ed. by D. Biber, R. Reppen. Cambridge University Press, 2015. P. 11–31.

9. Flowerdew L. The argument for using English specialized corpora to understand academic and professional language // Discourse in professions: perspectives from Corpus Linguistics / ed. by U. Connor, T. Upton. 2004. P. 11–33

10. Harrington, K. The Role of Corpus Linguistics in the Ethnography of a Closed Community: Survival Communication. // International Journal of Corpus Linguistics, 24 (4). – 2018. – pp. 541-547. ISSN 1384-6655

11. Hu K. and Kim K. H. (eds.), Corpus-based Translation and Interpreting. Studies in Chinese Contexts, Palgrave Studies in Translating and Interpreting, https://doi.org/10.1007/978-3-030-21440-1_3 // written by B. Wang (*) – Guangdong University of Foreign Studies, Guangzhou, China. 2009. – 61-87 pp.

12. Hyland K. As it can be seen: Lexical bundles and disciplinary variation // English for Specific Purposes. 2008. Vol. 27. P. 4–21.

13. Jung Y. and others A corpus-based approach to classifying emotions using Korean linguistic features // Cluster Comput (2017) 20: - pp. 583-595 DOI 10.1007/s10586-017-0777-8

14. Kennedy G. An Introduction to Corpus linguistics. –London and New York: Addison Wesley Longman limited, 1998. – 315 p.

15. Kuebler S., Zinsmeister H. Corpus Linguistics and Linguistically Annotated Corpora. London: Bloomsbury Publishing, 2015.P.10. 320 p.

16. Kytö M. Corpora and historical linguistics // RBLA, Belo Horizonte, v. 11, n. 2. – Uppsala University, Uppsala: Sweden, 2011 – pp. 417-457.

17. Lamel L., Cole R. Spoken Language Corpora // Survey of the State of the Art in Human Language Technology. New York //1997. P. 338–391.

18. Laurence A. A critical look at software tools in corpus linguistics // Linguistic Research.-Tokyo, № 30 (2)// 2013. P. 141–161.

19. Liu Y. and others CCAE: A Corpus of Chinese-based Asian Englishes available at https://arxiv.org/abs/2310.05381v1 [address date: 30.10.2023]

20. Mauranen A. Speaking professionally in L2 // Variation and change in spoken and written discourse: Perspectives from Corpus Linguistics / ed. by J. Bamford, S. Cavalereri, G. Diani. 2013. P. 5–31.

21. McEnery T., Hardie A. Corpus Linguistics: Method, theory and practice.-New York: Cambridge university press, 2012.P.35. 312 p.

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 14 Issue: 07 in July-2025
https://www.gejournal.net/index.php/IJSSIR

22. Nagi M. The International Corpus of Arabic: Compilation, Analysis and Evaluation. // Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP). – January 2014. – 8-17 pp. available at https://www.researchgate.net/publication/301404178_The_International_Corpus_of_Arabic_Compilation_Analysis_and_Evaluation

23. The British National Corpus. [Online] Available from: http://www.natcorp.ox.ac.uk

24. The Corpus of Historical American English: available at https://www.english-corpora.org/coha/ [Retrived on 31.10.2023]

25. The International Corpus of English. [Online] Available from:: http://www.ucl.ac.uk/englishusage/projects/ice.htm

26. Wang J. Recent Progress in Corpus Linguistics in China // International Journal of Corpus Linguistics // 6(2). DOI: 10.1075/ijcl.6.2.05wan – China, July 2002 – 281-304 pp.

27. *Национальный корпус русского языка* https://ruscorpora.ru/

ISSN 2277-3630 (online), Published by International journal of Social Sciences & Interdisciplinary Research., under Volume: 14 Issue: 07 in July-2025
https://www.gejournal.net/index.php/IJSSIR