# A New Approach to Creating an Independent Corps

Unarova Dilafruz Abdimajit qizi

Independent researcher of the Uzbek-Finnish Pedagogical Institute, Uzbekistan

**Abstract:** In the field of linguistics, there exist multiple techniques for examining literary works that enable us to explore different dimensions of the language and stylistic choices of poets and authors. Below, we concentrate on the most prevalent and significant methods of examination aimed at investigating the author's unique linguistic style.

**Introduction:** Creating an independent corpus. To create a corpus of a specific work, the following should be considered:

Collecting texts. The State Library, the National Electronic Library, and other digital libraries and archives can be used. Texts are collected from poetry collections, anthologies, and literary journals. It is advisable to use online resources dedicated to works of art, since these resources store texts in electronic form.

Digitization and processing. If the texts are available only in printed form, they should be digitized using a scanner and text recognition programs (OCR). The texts are cleaned of non-textual characters, familiar errors are corrected, and brought into a uniform format. Metadata such as author, title, year of writing, and genre are added, and the texts are extralinguistically tagged.

Creating a database. Database creation programs such as SQLite or MySQL can be used to store and manage the collected texts. Tables are created to store text, metadata, and analysis results.

Use of separate text analysis tools. Text analysis programs allow you to determine the frequency of words, analyze collocations, and perform other types of analysis. There are online tools for text analysis that can be used to analyze literary texts. Even without using a corpus and computer programs, you can manually analyze poetic/prose texts by identifying lexical, stylistic, and rhythmic features.

Creating an independent corpus takes time and effort, but it allows you to obtain information that meets specific research goals. Online resources and separate text analysis tools can be used, but it is important to take into account their limitations. Manual analysis is labor-intensive, but it allows you to understand poetic/prose texts in depth. Below, we will focus on studies conducted based on the corpus approach and their results.

There is a large experience of corpus-based research in the world. A number of works can be cited on the study of lexical, grammatical, pragmatic, discourse properties of the language and language education based on the corpus. A.R. Arellano is devoted to the problem of teaching English as a second language using the corpus, V. Bhatia chooses the corpus as an object for the analysis of literary genres. D. Bieber, R. Reppen use the corpus to study the syntactic properties of the language. P. Baker uses the corpus in discourse analysis, while S. Adolfs focuses on the use of electronic text analysis in language and literature education. P. Baker et al. analyzed the use of the word "Muslim" in British media publications in 1998-2009 based on the corpus. V. Brezina et al. focused on the study of collocations in the text based on the corpus. S. Evert also studied the meanings of collocations based on the

corpus. Of particular note is McKenry's textbook, which describes the methods, theory and practice of corpus linguistics, and his monograph on corpus-based linguistic research. The textbook on corpus linguistics by A. Ludeling and M. Cueto also describes in detail the mechanism and importance of conducting corpus-based research. The article on corpus-based discourse analysis by E. Semino et al. highlights the use of the corpus in discourse analysis of text and its specific results.

"Corpus-based discourse analysis" sarlavhali maqolada korpusga asoslangan diskurs tahlili bo'yicha tadqiqotlar quyidagi uchta asosiy aspektdan muhokama qilinadi. Bu yondashuvlar o'zaro bir-biriga bog'liq, chunki aslida har bir yondashuv boshqa yondashuvlarning aspektlaridan ham foydalanadi.

1. Textual: approaches that focus on the language choices, meanings, and expressions in texts, as well as various phraseological elements operating at the discourse level.

2. Critical: approaches that bring a critical attitude, such as critical discourse analysis, also use systemic functional linguistics.

3. Contextual: analyses that adopt a sociolinguistic approach to corpus data, taking into account the situational factor in such an analysis. This approach uses conversation analysis, speech act theory, and pragmatics.

L. Flowerdi "Is Corpus Linguistics a Theory, Methodology, or Approach?" raises the question and describes this issue in his article as follows: "Twenty years ago, corpus linguistics was considered a methodology for analyzing data because it could speed up and systematize the study of lexis or grammar. Later, T. Bonelli says that some linguists began to consider corpus linguistics as a more evolutionary important method, combining linguistic theory and data, and realizing what Halliday called "probabilistic grammars". It seems that L. Flowerdi views corpus linguistics as both a methodological approach and a way of drawing theoretical conclusions.

In the article "An Approach to Corpus-based Discourse Analysis: The Move Analysis as Example", he analyzes a seven-step approach to corpus-based discourse analysis, which begins with a detailed analysis of each individual text in the corpus, and then can be generalized to all texts in the corpus. This approach is applied to the methodology used to analyze texts in terms of the functional/communicative structures that constitute the genre. A corpus-based approach to analysis significantly enhances the value of the methodology often used, while also providing much-needed guidance. The article "Text at Scale: Corpus Analysis in Technical Communication" reviews approaches to corpus analysis. Corpus analysis supports analytical research on lexis, grammar, and discourse. These linguistic layers of analysis can be viewed as a measure of the shift from more objective, grammatical units to more interpretive, but still observable units. Since the goal of corpus analysis is to quickly analyze more content than can be read and analyzed manually, the first two layers of analysis (lexicon and grammar) are primarily quantitative. As mentioned above, these quantitative assessments can answer research questions or serve as a foundation for further research. Using lexical or grammatical analysis to guide further research, the researcher may use more interpretive types of analysis, such as discourse analysis.

According to S. Carradini, discourse analysis uses the results of lexical and grammatical approaches as a way to identify areas that are worth further study at the discourse level. Discourse is concerned with the function of words in the context of their use. The researcher seeks to understand how word(s) function in a text and how the text contributes to a particular type of speech act. Discourse analysts can understand the function that individual passages in a text perform by assessing the use of high-frequency words, their use with other words, and their use in context. Another method of text discourse analysis is the corpus approach. By identifying high-frequency, common words, or other word usage cases, the corpus shows what kind of discourse the corpus represents.

Collocation is a technique that identifies words that occur frequently in a corpus near a target word or phrase. The goal of collocation analysis is to identify quantitative relationships between words, which can be further analyzed to understand qualitative relationships between words. By observing collocations in a corpus, it is possible to more accurately analyze the meaning of the context.

The study, titled "A Corpus-Based Analysis of "For Example" and "For Instance," used both quantitative and qualitative research paradigms to interpret and analyze the correspondences. Quantitatively, the number of occurrences of the units "For Example" and "For Instance" in the Brown Corpus was counted. Qualitatively, the quantitative data collected was analyzed to interpret the various uses of the compounds in sentences. Integrating quantitative methods used in corpus analysis with qualitative methods in linguistic research is a strategy adopted by various linguists and researchers to bring more objectivity to the results of such studies and research, as well as to obtain reliable results by using a more scientific method of analyzing the emergence of

language capabilities. Integrating quantitative methods used in corpus analysis with qualitative methods in linguistic research is a strategy adopted by various linguists and researchers to bring more objectivity to the results of such studies and research, and thereby adopt a scientific method of analyzing language use, which leads to reliable results.

The article "A Corpus-based Multidimensional Analysis of Linguistic Features of Truth and Deception" divides the data into two parts: word count, word count per sentence, and multidimensional analysis. This study shows that the number of words in false speech and the number of words per sentence are significantly smaller than those of truthful speech. The results of the multidimensional analysis show that false speech performs worse on the narrative, interpersonal, and cognitive measures.

The article "Corpus-based critical discourse analysis of reporting practices in English news reports on public health events in China and the United States" also uses the corpus-based text analysis method. It was found that Chinese and American news reports used approximately the same high-frequency reporting verbs in reporting on the COVID-19 pandemic. The Chinese and American news corpora showed differences in the distribution of high-frequency reporting verbs across semantic categories. Chinese and American news The reports used the most speech report verbs, which indicated an objective attitude towards the reported event, and used speech report verbs and speech act report verbs to present the reported speech with a relatively high degree of accuracy. While American news reports often use mental report verbs to indicate an attitude of uncertainty towards the reported speech, Chinese news reports may need to be more aware of the use of mental report verbs to express the opinions and attitudes of ordinary people or the government. The results of this study may provide insights into research on the reporting strategies of emergency news reports in China for foreign audiences.

The article entitled "Argument structure, animacy, syntax and semantics of passivization in Turkish: A corpus-based approach" analyzes the animacy of arguments, their presence as indirect subjects of impersonal passive constructions with the participation of various predicates. Three different sources of animacy effects in the language are proposed. Only units denoting animate beings by nature can be used as indirect subjects of impersonal passive constructions. The article also carries out a syntactic analysis of the verb constructions of the state of mind in Turkish.

As can be seen from the results of the above studies,

corpus-based analysis is not only a quantitative analysis, but also a qualitative analysis is carried out through quantitative analysis. The experience of these studies shows that the results of corpus-based research and analysis are relatively accurate and objective. Therefore, in this dissertation, the object of study is studied from the point of view of corpus-based analysis.

## REFERENCES

Zheng Fanghua. A Corpus-based Multidimensional Analysis of Linguistic Features of Truth and Deception // https://aclanthology.org/Y18-1097.pdf

https://www.futurelearn.com/info/courses/improve-ielts-speaking/0/steps/98854

Carradini, Stephen, & Jason Swarts. (2023). Text at Scale: Corpus Analysis in Technical Communication. The WAC Clearinghouse; University Press of Colorado. https://doi.org/10.37514/TPC-B.2023.2104

Flowerdew, Lynne. (2012). How is Corpus Linguistics Related to Discourse Analysis? / Corpora and Language Education. 10.1057/9780230355569_4. (Электрон ресурс: мурожаат санаси: 25.03.2025. https://link.springer.com/book/10.1057/97802303555 69)

Baker, P., Gabrielatos, C., & McEnery, T. (2013). Sketching Muslims: A corpus driven analysis of representations around the word 'Muslim' in the British press 1998– 2009. Applied Linguistics, 34(3), 255-278.

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. International Journal of Corpus Linguistics, 20(2), 139-173.