

# Corpus-Based Research of Literary Texts: Methods, Approaches and Experiments

Unarova Dilafruz Abdimajit qizi

Independent researcher of the Uzbek-Finnish Pedagogical Institute, Uzbekistan

**Received:** 12 April 2025; **Accepted:** 08 May 2025; **Published:** 10 June 2025

**Abstract:** In linguistics, there are various methods of analyzing literary texts that allow us to study various aspects of the language and style of poets and writers. Below, we will focus on the most common and important methods of analysis aimed at analyzing the author's idiolect.

**Keywords:** lexical, grammatical and syntactic features, Poetic meter, cognitive processes and mental models, cultural context, subjective factors, collocations, lexical, grammatical, pragmatic, discourse features, methodological approach, linguostylistic, semantic, structural, cognitive, corpus, discourse analysis.

**Introduction:** The method of linguostylistic analysis is aimed at studying the language means used in a poetic/prose artistic text to create an artistic effect. In this method, lexical, grammatical and syntactic features of the artistic text, as well as tropes and figures of speech, are analyzed. The purpose of the analysis in this method is to determine the importance of language means in expressing the meaning and emotional content of the poem/prose.

The method of semantic analysis studies the semantic relationships between words and phrases in a poetic/prose text. The meanings of words, their contextual aspects and symbolic meaning are analyzed. The purpose of the analysis in this method is to deeply reveal the meaning and interpretations of a poetic work.

The method of structural analysis studies the structure of a poetic/prose text, including its composition, rhythm, rhyme and metric. The poetic meter and structure are analyzed. The goal of analysis in this method is to determine the importance of the structure of a poetic text in creating an artistic effect.

The cognitive analysis method studies how a poetic/prose text reflects cognitive processes and mental models. Metaphors, images, and symbols used to express abstract concepts and emotional states are analyzed. The goal of the analysis in this method is to analyze the interaction of the literary text with the

reader's perception.

The corpus analysis method uses large volumes of literary texts (corpora) to identify statistical patterns in language use. Word frequency, collocations, and other language features are analyzed. The goal of the analysis in this method is to identify general trends and individual stylistic features of poets, idiolects.

The discourse analysis method studies the literary text as a form of discourse, analyzes its social and cultural context. The ideological and cultural meanings expressed in the literary text are analyzed. The goal of the analysis in this method is to analyze the interaction of the poetic text with social and cultural norms.

The above methods are not mutually exclusive and are often used in combination for a complete and in-depth analysis of poetic/prose texts.

Corpus text analysis is a method of language study based on the use of large collections of texts (corpora) using computer technology. This approach allows you to identify patterns in the realization of language features in a literary text, analyze the frequency of words and phrases, study grammatical constructions and other linguistic phenomena. There are a number of main aspects of corpus analysis, which we will briefly discuss below.

Use of big data. A corpus of texts can contain millions of words and sentences, which allows for statistical analysis and the identification of stable language

trends.

**Computer analysis.** Computer programs are used to process and analyze large volumes of text data, which allows automating many routine tasks and obtaining results that are impossible with manual analysis.

**Identification of patterns.** Corpus analysis allows for the identification of patterns in the use of lexicon, grammar, syntax, and other elements of language.

**Objectivity.** The use of computer methods allows for the reduction of the influence of subjective factors on the results of research.

**Diversity of research.** Corpus analysis is used to solve a wide range of tasks, including studying the individual style of the author; analyzing the development of language over time; comparing language styles of different periods and directions; identifying general patterns in the language; studying the frequency of use of certain words and expressions. Corpus analysis opens up new opportunities for language study, allows you to obtain more objective and complete results.

The corpus approach to the study of literary works is a method based on the analysis of large volumes of literary texts (corpora) using computer technology. This approach allows you to identify patterns in the language and style of poets, as well as to study poetry as a general phenomenon. Corpus analysis of literary works is a multifaceted process that allows the researcher to identify hidden patterns and trends and delve deeper into the world of literature. The most necessary tool for analyzing literary works using a corpus is the presence of a corpus of literary works.

The steps in analyzing a literary text using a corpus based on observations are described below.

**Step 1.** Building a corpus (choosing a corpus if an analysis is being conducted on an existing corpus). A corpus of selected poets, writers, or texts is formed for a corpus of literary works. The formation of a corpus of literary works is described in the next chapter of the study. Such a corpus, which includes a large collection of texts consisting of literary works, is a powerful tool that allows you to search for various units, such as words, phrases, grammatical constructions. There are specialized literary corpora for this, which are suitable for specific studies. Corpora in the form of digital libraries and archives, including many libraries and archives, digitize literary texts and bring them into a form convenient for analysis.

**Step 2.** Determining the objectives of the analysis. Formulating specific questions, such as studying the lexical features of the poet/writer, tracing the evolution of literary language, and comparing styles of different periods, helps to focus the study.

**Step 3.** Using corpus analysis tools. At this stage, the following operations are performed on the selected literary corpus:

1) search for key words and phrases. This allows you to determine the frequency of use of certain words and phrases in poetic/prose texts.

2) analysis of collocations. The study of words that often occur side by side helps to understand the semantic relationships and features of the language of the poetic/prose text.

3) statistical analysis. The use of statistical methods provides the basis for identifying patterns in the use of lexicon, grammar, rhythm and other elements.

4) rhythmic and metrical analysis. Computer programs (especially concordances) can help in analyzing the rhythm and metric of a poem, and in identifying the structural features of the poem.

5) data visualization. Visualization of the analysis results using graphs and diagrams clearly presents the data obtained.

**Step 4.** Interpretation of results. The data obtained should be interpreted in the context of literary theory and history.

It should be taken into account that corpus analysis is a tool that helps to identify trends, but does not replace a deep understanding of the poetic text.

In corpus-based research of a literary text, attention is paid to studying the lexical uniqueness of the poet/writer (analyzing the frequency of use of certain words and phrases, identifying images and metaphors); analyzing the development of artistic language (comparing the lexicon and grammar of literary texts of different periods); comparing the style of different poets/writers (analyzing the lexical and stylistic features of modern poets/writers); studying the rhyme and rhythmic features of the poem (analyzing rhymes, identifying the dominant meter of the poem).

Among the important aspects that should be considered in corpus-based research of literary texts, it is worth noting that corpus analysis requires knowledge of computer tools and statistical analysis methods. It is important to be critical of the results obtained and to take into account the limitations of the corpus approach. Corpus analysis cannot replace traditional literary analysis, but it can enrich it with accurate, qualitative analysis results. The lack of a ready-made corpus for analyzing literary texts does not mean that this research is impossible. There are several alternative approaches and methods that can be used.

## 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/10.1057/9780230355569>)

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.

Evert, S. (2008). Corpora and collocations. In Ludeling, Anke, & Merja Kyto. Corpus linguistics. An international handbook, 2, Berlin: Mouton de Gruyter, 223-233.

<https://www.sketchengine.eu/guide/create-a-corpus-from-the-web/>

<https://www.sketchengine.eu/documentation/preparing-corpus-text/>

Toriida, Marie-Claude. (2017). Steps for Creating a Specialized Corpus and Developing an Annotated Frequency-Based Vocabulary List. TESL Canada Journal. 34. 87-105. 10.18806/tesl.v34i1.1257.