INNOVATIVE WORLD
Ilmiy tadqiqotlar markazi

INNOVATSION TALABALAR AXBOROTNOMASI

htttps://innoworld.net
+998335668868

ILMIY JURNAL

JOURNALS MASTER LIST
ISSN
doi
zenodo
OpenAIRE
Academia Resource Index ResearchBib
Google Scholar
open access.nl

INNOVATSION TALABALAR AXBOROTNOMASI

2-JILD, 3-TO'PLAM
2025

Jurnal quyidagi xalqaro bazalarda indekslanadi:

Google Scholar | doi digital object identifier | ResearchGate | zenodo

ADVANCED SCIENCE INDEX | OpenAIRE | Academic Resource Index ResearchBib | DRJI Directory of Research Journals Indexing

Ilmiy jurnalning rasmiy sayti:
www.innoworld.net

O'ZBEKISTON-2025

Theme: How to Find Common Words in a Text

Abdullajonova Hakima
Senior teacher of JSPU
Mirhamidova Madina Quvanova Mohinur
Jizzakh State Pedagogical University
Foreign Languages

Annotation. The process of finding common words in a text is one of the fundamental tasks in corpus linguistics and computational language analysis. Identifying frequently occurring words helps researchers understand the lexical profile of a text, determine its key topics, and compare it with other texts or registers. This paper discusses practical methods for finding common words using both manual and computational approaches, explains their significance for linguistic research, and highlights applications in teaching, stylistics, and lexicography.

Key words: corpus, frequency, word list, text analysis, concordance, lexis, computation, token, type, language

Аннотация. Определение наиболее часто встречающихся слов в тексте является одной из основных задач корпусной лингвистики и компьютерного анализа языка. Идентификация частотных слов помогает исследователям понять лексическую структуру текста, выявить его основные темы и сравнить с другими типами текстов. В статье рассматриваются методы определения частотных слов, их значение для лингвистических исследований и применение в преподавании, стилистике и лексикографии.

Ключевые слова: корпус, частотность, список слов, анализ текста, конкорданс, лексика, вычисление, токен, тип, язык

Main Part. Finding common words in a text involves analyzing its lexical frequency — the number of times each word appears. This process reveals which words dominate a text and provides insights into its style, topic, and communicative purpose. In corpus linguistics, such analysis is crucial for building frequency lists, developing dictionaries, and studying stylistic variation.

The simplest way to find common words is through manual counting or using word frequency functions in text processing tools. However, large texts require computational methods that can automatically tokenize, count, and rank words. Tools such as AntConc, WordSmith Tools, and Python's Natural Language Toolkit (NLTK) allow researchers to perform this efficiently.

For example, using Python, a researcher can process a text file with the following simple commands:

```
from collections import Counter
text = open("sample.txt").read().lower().split()
freq = Counter(text)
print(freq.most_common(10))
```

This code returns the ten most frequent words, such as "the," "and," "of," and "to," which are common in English. To make results more meaningful, linguists often remove stop words — high-frequency function words that carry little content meaning — using pre-defined lists.

Once common words are identified, researchers can compare frequency data across different texts or genres. For instance, academic texts often contain words like research, method, and analysis, while fiction tends to include said, looked, and thought. These lexical differences reflect genre conventions and communicative goals.

Frequency analysis is also valuable in language teaching. Teachers can use word lists to focus on the most common and useful vocabulary for learners. For example, the General Service List (GSL) and Academic Word List (AWL) are based on corpus frequency data and guide curriculum design for English learners.

In stylistics, identifying frequent words helps describe an author's style. Repetition of certain lexical items can indicate thematic focus or emotional intensity. In lexicography, frequency counts ensure that dictionary entries reflect real usage, not intuition.

Moreover, frequency analysis supports computational applications such as search optimization, keyword extraction, and authorship attribution. For example, comparing word frequency profiles of different authors can help determine authorship of anonymous texts — a method widely used in forensic linguistics.

In summary, finding common words in a text is not merely a counting exercise but a gateway to understanding linguistic patterns and communicative functions. It connects language data with theory, pedagogy, and technology.

Conclusion. Finding common words in a text is a simple yet powerful linguistic technique. It provides quantitative evidence about vocabulary distribution, genre characteristics, and stylistic tendencies. Through corpus tools and computational analysis, researchers can handle large datasets and obtain objective insights into language use. The method has wide applications in teaching, lexicography, stylistics, and computational linguistics, making it an essential component of modern corpus-based research.

References (Books)

1. Biber, D., Conrad, S., & Reppen, R. (1998). Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press.
2. Hunston, S. (2002). Corpora in Applied Linguistics. Cambridge University Press.
3. McEnery, T., & Hardie, A. (2012). Corpus Linguistics: Method, Theory and Practice. Cambridge University Press.
4. Sinclair, J. (1991). Corpus, Concordance, Collocation. Oxford University Press.
5. Stubbs, M. (1996). Text and Corpus Analysis. Blackwell.
6. Tognini-Bonelli, E. (2001). Corpus Linguistics at Work. John Benjamins Publishing.