CURRENT RESEARCH JOURNAL OF PHILOLOGICAL SCIENCES (ISSN: 2767-3758)

VOLUME: Vol.06 Issue03 2025
Page: - 21-24
DOI: - 10.37547/philological-crjps-06-03-05

RESEARCH ARTICLE OPEN ACCESS

The Significance of Tagging Phraseological Units in The Language Corpus

Yodgorov Umidjon Saydilla oglu

Teacher, Department of Computer Linguistics and Digital Technologies, Tashkent State, University of Uzbek Language and Literature named after Alisher Navoi, Uzbekistan

Received: 27 January 2025 **Accepted:** 28 February 2025 **Published:** 31 March 2025

ABSTRACT

Corpus-based research draws scientific conclusions based on materials from existing corpora. Such research includes processes such as phraseological unit identification; statistical analysis; semantic and syntactic analysis. Corpus-based research allows us to analyze how phraseological units change over time and provides material for conducting diachronic research. Through this, the historical development and evolution of phraseological units are traced.

Types of work with corpuses are carried out in linguistic research through different approaches. Approaches such as corpus-driven (corpus-driven research); corpus-based (corpus-based research); corpus-illustrated (research that uses the corpus as an illustrative source) allow linguists to analyze data at different levels.

Keywords: Language corpus, phraseological unit, corpus-driven, corpus-based, corpus-illustrated (research that uses the corpus as an illustrative source), statistical analysis, semantic analysis, syntactic analysis.

INTRODUCTION

Phraseological units in different languages have been studied in different aspects in world phraseology, and these results serve as a theoretical source for corpus linguistics. For example, it will be possible to tag phraseological units in the corpus and search for them in the corpus. This research is also used in computer lexicography.

I. Corpus-based research. Corpus-based research (corpus-based research) - in this method, researchers make scientific conclusions based on materials taken from existing corpora. These researches include the following processes:

Identification of phraseological units: Using the corpus, the researcher identifies specific phraseological units and studies their contextual use.

1. Statistical analysis: The frequency and distribution of phraseological units in the corpus is determined using statistical methods. The extent to which these units are widespread and in what stylistic situations they are used are analyzed.

2. Semantic and syntactic analysis: The meaning and grammatical structure of phraseological units are studied based on the corpus.

This approach is useful in determining the basic and figurative meanings of phraseological units.

II. Diachronic research. Corpus-based research makes it possible to analyze how phraseological units change over time. Through this, the historical development and evolution of the phraseological units is observed.

Representation of phraseological units in the corpora makes it possible to mark phraseological units in the corpora and quickly search for them. This process consists of the following steps:

1. Identification and tagging of phraseological units: In order to include phraseological units in the corpus, it is necessary to mark them with special tags. These tags help distinguish phraseological units from other lexical units and make it easier to search for them in the corpus.

2. Automated identification algorithms: Using modern language technologies, algorithms for automatic identification of phraseological units and their tagging are developed. This method speeds up the research process and allows for analysis of large volumes of texts.

3. Storage of contextual information: Phraseological units are represented in the corpus with full context, that is, where and under what circumstances they are used. This contextual information allows researchers to determine how phraseological units are used and what stylistic characteristics they have.

4. Enrichment with additional information: In order to reflect stylistic, semantic and grammatical features of phraseological units in the corpus, it is necessary to provide them with additional information (metadata). This enables comprehensive analysis for corpus users.

III. Types of corpus work

Types of corpus work are done through different approaches in linguistic studies. These approaches allow linguists to analyze different levels of data:

Corpus-driven researches: In this approach, researchers draw conclusions by directly analyzing the data available in the corpus. In this method, there are no pre-established hypotheses before the study, that is, new scientific hypotheses are created based on the corpus data. The advantage of this approach is that it is based on real data and provides unexpected analytical results.

Corpus-based researches: In this approach, researchers use corpus data to confirm or disprove already existing hypotheses or theories. In this case, the corpus serves as the main source for the research, but the hypothesis and research questions are defined in advance. Corpus-based research is widely used in analyzing the usage, distribution and contextual meanings of phraseological units.

Corpus-illustrated researches: In this method, the examples taken from the corpus are not the main part of the research, but are used as additional visual or illustrative material. This approach is designed to enrich the theoretical parts of the research with examples from the corpus. For example, when explaining a linguistic phenomenon, concrete examples from the corpus increase the evidence of the research.

IV. Tagging of phraseological units in the Uzbek language corpus

Tagging phraseological units in the corpus greatly helps in analyzing them. Phraseological units of the Uzbek language are tagged on the basis of their lexical-semantic, morphological and grammatical characteristics.

The process of tagging phraseological units in the Uzbek language

Tagging phraseological units in the Uzbek language corpus ensures their correct identification and analysis. The tagging process consists of the following stages:

Identification and classification of phraseological units. Phraseological units included in the corpus are classified on the basis of their semantic, morphological and grammatical characteristics. They are divided into various forms such as word-phrases, compound-phrases and sentence-phrases.

Tagging technology. The achievements of computer linguistics are used to identify phraseological units. Phraseological combinations are automatically found and marked in the text using special software or algorithms.

Applying tags to the corpus. In the process of tagging, phraseological units are marked with special tags. These tags can reflect the following features:

Lexical-semantic tags: Represents the meaning of a phraseological unit.

Grammatical tags: Determines the grammatical features of the unit (morphological form, syntactic role).

Stylistic tags: Describe the stylistic properties of the unit (for example, its use in a formal or informal context).

Manual and automated tagging: In some cases, the tagging process is carried out manually by linguists, which ensures the accuracy and correctness of the results. However, automation of this process using modern technologies is also widely used.

V. The importance of tagging phraseological units in the corpus of the Uzbek language

The tagging of phraseological units increases their importance in the lexicographic process. In this process, case materials help to provide objectivity. In this way, the lexicographer will have the opportunity to choose the right context and enter objective information. To understand the importance of this process, the following aspects can be considered:

Distinguishing phraseological units: Tagging helps distinguish phraseological units from ordinary lexical units. This makes it easier to identify the phraseological units that are being searched for in the corpus and allows you to track their usage.

Objectivity of analysis and accurate results: Tagged phraseological units serve as a source of objective information in research. By tagging phraseological units in the corpus, linguistic analysis is performed in a precise and evidence-based manner.

Basis for lexicographic practices: Correct tagging of phraseological units is used in lexicographic works. It helps in lexicographic studies to determine how phraseological units are used in a real context and to enter them correctly into dictionaries.

Use in language learning and teaching: Tagged phraseological units are used as language learning and teaching materials. It provides students with additional resources for learning the richness of the Uzbek language and the methodological aspects of phraseological units.

Possibilities of contextual analysis: Phraseological units reflect not only their components, but also the context in which they are used. Tagging helps you learn more about how these units are used in different textual and stylistic contexts.

VI. The importance of the database of phraseological units in the Uzbek language

The database containing phraseological units is important not only for linguistic research, but also for various scientific and practical projects. This database covers the following aspects:

Systematic storage and organization. The database allows to systematically store phraseological units and organize them. This helps to consolidate the lexical-semantic, grammatical and stylistic features of these units in one place.

Comprehensive data. The database in the Uzbek language stores various forms of phraseological units (phraseological unions, combinations, additions) and their contextual use. It provides comprehensive information to linguists and other professionals.

Scope of application. The database serves as a useful tool for linguists, teachers, translators and lexicographers. It allows you to quickly get information on the stylistic features and contextual meanings of phraseological units.

Advantages of the database

Search and search options: In the database, it is possible to search for phraseological units according to various parameters. For example, a search can be carried out by grammatical features of the units, contextual use or stylistic features.

Statistical and analytical capabilities: The database makes it easier to determine the frequency, distribution and other statistical indicators of phraseological units. This method helps to scientifically study linguistic phenomena in depth.

Support for linguistic research: The database helps researchers to systematically study the phraseological aspects of the language. These databases ensure the objectivity and accuracy of the results of the analysis. Use in teaching the Uzbek language: The database can be used in the preparation of training manuals and teaching materials for the learners of the Uzbek language. This allows students to understand phraseological units in a correct and broad context.

Content of the database

Lexical-semantic information: Meaning, usage and synonyms of each phraseological combination.

Grammatical features: Morphological and syntactic description of phraseological units.

Stylistic information: In what stylistic contexts are phraseological units used (formal, informal, artistic, etc.).

Examples and Contexts: Examples showing how each

phraseological unit is used in real texts and their contextual analysis.

CONCLUSION

In conclusion, it can be said that the following types of work with linguistic corpuses are distinguished: corpus-oriented research; corpus-based; research using the corpus as an illustrative source. In corpus linguistics, phraseological units reflected in language corpora serve as linguistic material or research method; phraseological units are the object of analysis, and language corpora simply serve as a research tool. Research on tagging phraseological units in language corpora, reflecting phrasemes in the corpus, developing a phraseme search system, and using language corpora in searching and analyzing phrasemes has also become popular.

REFERENCES

Biber, D. (2006). University Language: A Corpus-based Study of Spoken and Written Registers. Amsterdam: John Benjamins Pages

Dekatova, Kristina. (2023). Variability of Political Phraseological Units. Vestnik of Northern (Arctic) Federal University. Series Humanitarian and Social Sciences. 47-55. 10.37482/2687-1505-V264.

Berdiyorov H., Rasulov R., Yuldashev B. Materials from Uzbek phraseology. Training manual, parts I-II-III). – Samarkand: 1976-1979-1983.

Idiyatova A.M. Using a linguistic corpus in teaching phraseology at a language university // https://cyberleninka.ru/article/n/ispolzovanie-lingvisticheskogo-korpusa-v-obuchenii-frazeologii-v-yazykovom-vuze

Nurzet C.S.-B. Problems of corpus search for German phraseological units in the Mannheim Corpus of the German language // Scientific notes of Novgorod State University. 2022. № 5 (44). C. 575-578. // https://cyberleninka.ru/article/n/problemy-korpusnogo-poiska-nemetskih-frazeologizmov-v-mangeymskom-korpuse-nemetskogo-yazyka