JOURNAL OF MULTIDISCIPLINARY SCIENCES AND INNOVATIONS
VOLUME 5, ISSUE 03
MONTHLY JOURNALS
ISSN NUMBER: 2751-4390
IMPACT FACTOR: 9,08

ANALYSIS OF LEXICAL-SEMANTIC GROUPS IN CREATING THE NATIONAL CORPUS OF THE UZBEK LANGUAGE

Xolmurotova Iroda
Student of Termiz State Pedagogical Institute
Scientific Supervisor: **Raimnazarova Nasiba**
Termiz State Pedagogical Institute,
Doctor of Philosophy in Philology (PhD), Associate Professor

Abstract: This article discusses the analysis of lexical-semantic groups in the process of creating the national corpus of the Uzbek language. The study emphasizes the importance of organizing lexical units into semantic groups for ensuring the systematic structure, functional accuracy, and scientific value of the corpus. Lexical-semantic classification plays a significant role in identifying semantic relations among words, improving corpus annotation, and facilitating linguistic research based on corpus materials. The article examines the theoretical and practical aspects of grouping lexical items according to meaning, usage, and contextual relevance in the Uzbek language. Special attention is given to the role of lexical-semantic analysis in corpus linguistics, digital lexicography, and the development of modern language technologies.

Keywords: Uzbek language, national corpus, corpus linguistics, lexical-semantic groups, semantic analysis, lexical units, corpus annotation, digital linguistics, language resources, lexicography

In modern linguistics, corpus creation has become one of the most important directions for studying language systematically and scientifically. A national corpus serves as a large, structured, and electronically accessible collection of texts that reflects the real use of a language in different communicative contexts. It provides valuable material for linguistic analysis, lexicography, language teaching, translation studies, and the development of digital language technologies. In the case of the Uzbek language, the creation of a national corpus is especially relevant, as it contributes to the preservation, standardization, and modernization of the language in the age of digital communication.

One of the key issues in building a national corpus is the classification and analysis of lexical units according to their semantic properties. Words in a language do not exist in isolation; they are connected through meaning, usage, and contextual function. Therefore, the study of lexical-semantic groups is essential for organizing vocabulary systematically within the corpus. Lexical-semantic grouping allows linguists to identify relationships among words, determine semantic fields, and reveal how lexical items function in different textual environments. This approach makes corpus annotation more accurate and increases the scientific value of the language database.

In the Uzbek language, lexical-semantic groups reflect the richness of vocabulary and the diversity of cultural, social, and historical meanings embedded in words. The correct classification of lexical units into thematic and semantic groups is especially important for corpus design because it ensures that the corpus represents not only the formal structure of the language but also its semantic organization. Such analysis helps distinguish polysemy, synonymy, antonymy, hyponymy, and other semantic relations that are crucial for corpus-based studies. Without semantic grouping, the corpus may remain a simple text archive rather than a fully functional linguistic resource.

The importance of lexical-semantic analysis has increased further with the development of corpus linguistics, computational linguistics, and natural language processing. Modern digital systems require clearly classified and annotated lexical material in order to perform search,

tagging, automatic translation, speech recognition, and semantic analysis. In this process, lexical-semantic groups play an important role by helping structure lexical data in a way that is accessible both for researchers and for technological applications. Thus, semantic classification becomes one of the core foundations for creating an efficient and informative national corpus.

Another important aspect of this issue is that lexical-semantic groups reflect the worldview, national identity, and conceptual system of the people who speak the language. In the Uzbek national corpus, the analysis of semantic groups can reveal how language represents everyday life, traditions, social relations, professional activity, and cultural values. Therefore, studying lexical-semantic groups is not only a technical stage of corpus construction but also an important means of understanding the internal semantic system of the Uzbek language.

The relevance of this topic lies in the fact that the Uzbek national corpus requires a scientifically grounded lexical-semantic framework for its effective development and practical use. The analysis of lexical-semantic groups ensures better corpus organization, facilitates semantic annotation, and supports future research in linguistics and language technologies. This issue is particularly significant for the Uzbek language, which is actively developing in both national and digital contexts.

The purpose of this article is to analyze lexical-semantic groups in the process of creating the national corpus of the Uzbek language, to identify their role in corpus organization, and to determine their significance for linguistic research and digital language development.

The development of corpus linguistics has significantly changed the methods of studying language in modern linguistics. A corpus is no longer viewed merely as a collection of texts; rather, it is considered a structured linguistic resource that provides reliable data for lexical, grammatical, semantic, and pragmatic analysis. In this context, the creation of a national corpus requires not only the accumulation of texts but also their systematic organization according to linguistic principles. One of the most important of these principles is the classification of lexical units into lexical-semantic groups.

In general linguistic theory, lexical-semantic groups are understood as sets of words united by common semantic features and related by meaning within a certain conceptual field. Such groupings help reveal how vocabulary is organized in the mental and communicative system of a language. Scholars have emphasized that semantic classification is necessary for understanding lexical relations such as synonymy, antonymy, hyponymy, polysemy, and semantic compatibility. Therefore, lexical-semantic analysis serves as an important basis for lexicology, semantic theory, lexicography, and corpus construction.

Researchers in corpus linguistics have noted that semantic classification improves the quality of corpus annotation and allows users to work with texts more effectively. When lexical units are grouped according to semantic relations, it becomes easier to search for patterns of meaning, analyze contextual usage, and study the functional behavior of words in different genres and discourse types. This is especially important for national corpora, which are expected to reflect the vocabulary of a language comprehensively and systematically. Without semantic grouping, a corpus may provide textual material, but its analytical capacity remains limited.

In Turkic linguistics, including Uzbek linguistics, the issue of lexical-semantic grouping has become increasingly relevant with the growing interest in digital language resources. Uzbek scholars have studied semantic fields, thematic groups, lexical layers, and the semantic structure of words in relation to lexicology and lexicography. These studies provide an important theoretical foundation for corpus creation because they show that Uzbek vocabulary is organized through complex semantic relationships that should be reflected in digital language databases. The correct identification of lexical-semantic groups is therefore essential for building an Uzbek corpus that is both scientifically grounded and practically useful.

Another important point discussed in the literature is the relationship between corpus linguistics and natural language processing. Modern digital systems require semantically organized lexical material for tasks such as automatic tagging, machine translation, information retrieval, semantic search, and language modeling. In this respect, lexical-semantic groups perform not only a descriptive function but also a technological one. They help transform lexical data into a more structured and machine-readable form, which increases the applicability of the national corpus in computational linguistics and language technologies.

The literature also shows that lexical-semantic grouping is closely connected with cultural and conceptual representation. Words reflect not only objects and actions but also the worldview, traditions, and social experiences of a speech community. Therefore, the analysis of lexical-semantic groups in the Uzbek national corpus is important for preserving the cultural and conceptual richness of the language. Through semantic classification, the corpus can more accurately represent the internal organization of Uzbek vocabulary and its relation to national identity.

Overall, previous studies confirm that lexical-semantic analysis is one of the key theoretical and practical components of corpus creation. However, in the context of the Uzbek national corpus, this issue still requires further systematic investigation. In particular, more attention should be paid to the principles of semantic grouping, the identification of lexical relations, and the role of lexical-semantic classification in corpus annotation and digital language development. This article contributes to that task by examining the significance of lexical-semantic groups in creating the national corpus of the Uzbek language.

The creation of the national corpus of the Uzbek language requires a scientifically organized system in which lexical units are not only collected from texts but also classified according to their semantic and functional characteristics. In this process, lexical-semantic groups play a central role because they allow vocabulary to be arranged in a structured and meaningful way. A corpus built without semantic classification may remain only a technical archive of texts, whereas a semantically organized corpus becomes a powerful linguistic resource for research, lexicography, education, and computational applications.

Lexical-semantic groups unite words that share a common conceptual or semantic feature. In corpus creation, such grouping helps identify the internal organization of vocabulary and the relationships between lexical items. For example, words related to human activity, nature, education, family relations, emotions, movement, profession, and technology may each form specific semantic fields. These fields are important because they reveal how meaning is distributed in the language and how different lexical units function in real contexts. When such groups are clearly identified, the corpus reflects not only the formal side of the Uzbek language but also its semantic system.

In the Uzbek national corpus, the analysis of lexical-semantic groups is especially significant because Uzbek vocabulary is rich in polysemantic units, synonymic series, antonymic oppositions, and culturally marked lexical items. A word may belong to more than one semantic field depending on its contextual meaning. Therefore, corpus construction requires careful semantic annotation that takes into account not only dictionary meaning but also actual usage in discourse. This makes lexical-semantic analysis a necessary condition for building a reliable and linguistically accurate corpus.

Another important issue is the relationship between lexical-semantic grouping and corpus annotation. Annotation is one of the most important stages of corpus development, as it makes the corpus searchable and analytically useful. If lexical units are annotated according to their semantic group, researchers can identify patterns of usage more effectively, compare thematic fields across genres, and study semantic changes in language over time. For example, corpus

users may investigate how words related to technology, media, or social life have expanded in modern Uzbek. This becomes possible only when lexical items are systematically grouped and tagged according to semantic principles.

The analysis of lexical-semantic groups also contributes to the study of semantic relations such as synonymy, antonymy, hyponymy, and polysemy. In the Uzbek national corpus, these relations help reveal how words interact within the lexical system. Synonymous units may show stylistic or contextual differences, while antonyms reflect semantic oppositions that structure conceptual thought. Hyponymic relations help organize vocabulary hierarchically, and the study of polysemy reveals how one lexical form can express several meanings in different contexts. All of these aspects are crucial for corpus-based lexicology and semantic analysis.

From a practical perspective, lexical-semantic grouping increases the usefulness of the national corpus for language technologies. Modern computational systems require semantically structured language data in order to perform automatic text processing, search operations, machine translation, information extraction, and natural language understanding. If the Uzbek national corpus includes carefully analyzed lexical-semantic groups, it will become more effective not only for academic research but also for digital applications. In this sense, semantic classification forms a bridge between traditional linguistic analysis and modern technological innovation.

The results of lexical-semantic analysis also demonstrate that corpus creation is closely connected with the preservation of national identity and cultural knowledge. Words in any language reflect the material life, spiritual values, customs, and worldview of the people who speak it. In Uzbek, many lexical-semantic groups are directly related to national traditions, kinship relations, moral concepts, agriculture, crafts, and social etiquette. Including such groups in the corpus in a systematic way allows the national corpus to serve not only as a linguistic tool but also as a cultural repository. This gives the Uzbek language a stronger presence in the digital world while preserving its semantic uniqueness.

At the same time, the analysis of lexical-semantic groups in corpus creation is not without difficulty. One of the main challenges is the complexity of semantic boundaries. Some words may shift meaning according to context, while others may simultaneously belong to several lexical-semantic groups. This requires careful methodological principles and a flexible annotation system. Another challenge lies in distinguishing native lexical units from borrowed terms and determining how both categories function within the semantic structure of the language. These issues show that corpus creation is a linguistically complex task that requires both theoretical knowledge and practical precision.

In general, the analysis of lexical-semantic groups is one of the key conditions for creating a complete and effective national corpus of the Uzbek language. It ensures semantic order, improves annotation quality, strengthens the scientific value of the corpus, and increases its practical importance in education, lexicography, and digital linguistics. The study shows that lexical-semantic classification is not an additional stage of corpus construction, but one of its essential foundations. Therefore, the successful development of the Uzbek national corpus depends greatly on the accurate identification and analysis of lexical-semantic groups.

References

1. Dadaboyev, H. *Uzbek Terminology* [in Uzbek]. Tashkent, 2019.
2. Ermatov, I. *Uzbek Linguistic Terminology: Monograph* [in Uzbek]. Tashkent, 2021.
3. Hojiyev, A. *Explanatory Dictionary of Linguistic Terms* [in Uzbek]. Tashkent: O‘zbekiston Milliy Ensiklopediyasi, 2002.
4. Hojiyev, A. *Criteria for Term Selection* [in Uzbek]. Tashkent, 1996. This work is cited in later Uzbek terminology literature as a source on term selection principles.

5. Shoabdurahmonov, Sh., Asqarova, M., Hojiyev, A., Rasulov, I., & Doniyorov, H. *Modern Uzbek Literary Language. Part 1* [in Uzbek]. Tashkent: O‘qituvchi, 1980. This textbook is cited in Uzbek terminology scholarship as a foundational linguistic source.