

SEMANTIC AND PRAGMATIC FEATURES OF ENGLISH COLLOCATIONS: A CORPUS-BASED STUDY**Malikabonu Imonberdieva**

Teacher, Uzbekistan University of Journalism and Mass Communications

Nargiza Abdusamatovna Fuzaylova

Teacher, Tashkent state university of Oriental studies, Uzbekistan

Abstract: This article explores the semantic and pragmatic characteristics of English collocations through a corpus-based approach. Collocations, as stable combinations of lexical units, play a crucial role in natural language use and reflect both linguistic patterns and contextual meaning. The study aims to identify the most frequent types of collocations in English and analyze their semantic relationships and pragmatic functions in authentic discourse. The research is based on data obtained from major linguistic corpora, including the British National Corpus and the Corpus of Contemporary American English. By applying quantitative and qualitative methods, the study examines lexical patterns, frequency distribution, and contextual usage of collocations across different genres.

The findings reveal that English collocations demonstrate strong semantic cohesion and are often influenced by contextual and pragmatic factors such as register, discourse type, and communicative intention. Moreover, the study highlights the importance of collocational competence in language learning and translation, emphasizing that inadequate knowledge of collocations may lead to unnatural or inaccurate language production.

Keywords: collocation, corpus linguistics, semantic analysis, pragmatic features, English language, lexical units, discourse

Introduction: In recent decades, corpus linguistics has become one of the most influential approaches in modern linguistic research, providing empirical data for the analysis of language use in real contexts. One of the key areas within corpus-based studies is the investigation of collocations, which are defined as habitual co-occurrences of lexical items. Collocations are essential for achieving fluency and naturalness in language production, as they reflect conventional patterns of word combination in a given language.

In the English language, collocations play a significant role not only at the lexical level but also in shaping semantic meaning and pragmatic interpretation. The meaning of a collocation often extends beyond the sum of its individual components, which makes it a complex phenomenon for both language learners and translators. Misuse of collocations may result in unnatural expressions or even miscommunication, highlighting the importance of their systematic study.

The development of large-scale linguistic corpora, such as the British National Corpus and the Corpus of Contemporary American English, has made it possible to analyze collocations in authentic contexts across different genres and registers. These corpora provide valuable data for identifying frequent collocational patterns and understanding their semantic and pragmatic functions.

The aim of this study is to investigate the semantic and pragmatic features of English collocations using a corpus-based approach. The research focuses on identifying common collocational patterns, analyzing their meanings, and examining how contextual factors influence their use in discourse. The study also seeks to contribute to applied linguistics by highlighting the importance of collocational competence in language learning and translation.

Literature Review: The study of collocations has a long tradition in linguistics and is closely associated with the work of J.R. Firth, who famously stated that “you shall know a word



by the company it keeps.” This idea laid the foundation for the modern understanding of collocations as statistically significant co-occurrences of words. Later, John Sinclair further developed this concept through corpus-based research, emphasizing the importance of authentic language data. Sinclair introduced the distinction between the “idiom principle” and the “open-choice principle,” arguing that language users often rely on pre-constructed word combinations rather than generating sentences word by word. This perspective highlights the central role of collocations in natural language use. Another significant contribution to the study of collocations was made by Michael Halliday, whose systemic functional linguistics framework emphasizes the relationship between language and context. From this perspective, collocations are not only lexical phenomena but also carry functional and pragmatic meanings within discourse.

Research on collocations has also been closely linked to lexicography and language teaching. Scholars such as Michael Lewis have emphasized the importance of teaching collocations as part of the lexical approach, arguing that vocabulary acquisition involves learning chunks of language rather than isolated words. This approach has significant implications for second language acquisition and pedagogy.

In recent years, corpus linguistics has provided new insights into the study of collocations by enabling large-scale quantitative analysis. Studies based on corpora such as the British National Corpus and Corpus of Contemporary American English have demonstrated that collocations vary across genres, registers, and communicative contexts. Researchers have also explored the role of frequency, statistical measures (such as mutual information), and contextual factors in determining collocational strength.

Furthermore, the semantic and pragmatic aspects of collocations have attracted increasing attention. Scholars have shown that collocations often carry implicit meanings, connotations, and discourse functions that are not immediately apparent from their individual components. This makes them particularly challenging for non-native speakers and translators, who must consider not only literal meaning but also contextual appropriateness. Despite the growing body of research, there is still a need for more comprehensive studies that integrate semantic and pragmatic analysis within a corpus-based framework. This study aims to address this gap by providing a systematic examination of English collocations and their functions in real language use.

Materials and Methods: This study adopts a corpus-based methodological framework to investigate the semantic and pragmatic features of English collocations. Corpus linguistics provides a reliable empirical basis for analyzing language as it is used in real communicative contexts, allowing for the observation of authentic patterns of lexical co-occurrence. In this research, a mixed-methods approach is employed, combining quantitative statistical analysis with qualitative linguistic interpretation in order to achieve a comprehensive understanding of collocational behavior in contemporary English.

The primary data for the study were obtained from two widely recognized and representative linguistic corpora, namely the British National Corpus and the Corpus of Contemporary American English. These corpora were selected due to their large size, balanced structure, and inclusion of diverse genres such as spoken language, fiction, academic texts, newspapers, and online discourse. The use of both corpora ensures a broader perspective on English usage, including variation across registers and regional varieties. The selection of lexical items for analysis was based on frequency and productivity in forming collocations. The study primarily focuses on high-frequency verbs, nouns, and adjectives that commonly participate in collocational patterns, such as *make*, *take*, *have*, *strong*, *heavy*, and *significant*. These lexical items were chosen because they represent core vocabulary and exhibit a high degree of combinatory potential in natural language use. The selection process was guided by frequency lists derived from the corpora to ensure representativeness.



Collocational data were extracted using corpus query tools, applying a span window of ± 4 words from the selected node word. This range allows for the identification of relevant collocates within the immediate lexical environment while minimizing unrelated co-occurrences. To refine the dataset, several criteria were applied, including minimum frequency thresholds and statistical significance. Only those collocations that demonstrated stable and meaningful associations were included in the analysis.

To determine the strength of collocational relationships, statistical measures such as Mutual Information (MI score) and t-score were employed. The MI score was used to identify strong and exclusive lexical associations, whereas the t-score was applied to highlight frequent and reliable patterns. These measures provided a quantitative foundation for classifying collocations into different categories based on their degree of association and frequency. The analytical process was carried out in two main stages. The first stage involved quantitative analysis, focusing on frequency distribution and statistical evaluation of collocations across different corpora and genres. The second stage consisted of qualitative analysis, which examined the semantic and pragmatic features of the identified collocations. Semantic analysis focused on the relationships between lexical items, including semantic compatibility, synonymy, and semantic prosody, which refers to the attitudinal or evaluative meaning associated with certain word combinations. Pragmatic analysis, on the other hand, explored how collocations function in context, taking into account factors such as register, genre, communicative intention, and discourse function. The research procedure included several steps, beginning with the selection of target lexical items, followed by the extraction of collocational data, statistical analysis, classification of collocations, and contextual interpretation. Concordance lines were analyzed to observe the use of collocations in authentic discourse and to identify their pragmatic roles within different communicative settings.

In order to ensure the reliability and validity of the findings, data from the British National Corpus and the Corpus of Contemporary American English were compared and cross-checked. The use of multiple corpora reduces the risk of bias and increases the generalizability of the results. Furthermore, the integration of quantitative and qualitative methods enhances the robustness of the analysis. Despite its strengths, the study has certain limitations. The analysis is restricted to a selected group of lexical items and may not cover the full range of English collocations. Additionally, while corpus data provide extensive examples of authentic language use, they may not fully capture all pragmatic nuances, particularly in spontaneous spoken interaction. Nevertheless, the methodology applied in this study provides a solid foundation for examining collocations from both semantic and pragmatic perspectives.

Results and Discussion: The analysis of corpus data demonstrates that English collocations exhibit a high degree of semantic cohesion and pragmatic functionality across various communicative contexts. The findings, based on data extracted from the British National Corpus and the Corpus of Contemporary American English, reveal that collocations are not random combinations of words but rather conventionalized patterns that occur with notable frequency and stability in authentic language use. The quantitative analysis indicates that certain verb-noun collocations, such as *make a decision*, *take a risk*, and *have an effect*, are consistently used across different genres and registers, reflecting their strong entrenchment in the English lexicon. These combinations are characterized by high frequency and statistical significance, which confirms their status as fixed or semi-fixed lexical units. Substitution of one element in such collocations often leads to unnatural or grammatically incorrect expressions, as seen in forms like *do a decision*, which further highlights the importance of conventional usage patterns.

From a semantic perspective, the study reveals that collocations are often associated with semantic prosody, meaning that they carry implicit evaluative or attitudinal meanings. For



instance, the verb *cause* typically occurs in negative contexts, as in *cause a problem*, *cause damage*, or *cause concern*, whereas verbs such as *provide* and *offer* tend to co-occur with positively connoted nouns, such as *provide support* or *offer assistance*. This demonstrates that collocational meaning is shaped not only by the individual lexical items but also by their habitual co-occurrence in specific contexts. Furthermore, the analysis identifies the presence of selectional restrictions, whereby certain words combine only with specific lexical partners. For example, the adjective *heavy* forms natural collocations such as *heavy rain* and *heavy traffic*, while alternatives like *strong rain* are rarely used and considered non-standard. Such restrictions underline the importance of collocational knowledge in achieving semantic accuracy and naturalness in language use.

In terms of pragmatic functions, the findings suggest that collocations play a crucial role in shaping meaning within discourse and are highly sensitive to contextual factors such as register, genre, and communicative intention. In formal and academic contexts, collocations such as *conduct research* and *reach a conclusion* are preferred due to their precision and stylistic appropriateness, whereas in informal spoken discourse, simpler or more flexible expressions, such as *do research* or *come to a conclusion*, are more commonly used. This variation reflects the adaptability of collocations to different communicative environments and highlights their role in achieving pragmatic effectiveness. Additionally, collocations contribute to specific discourse functions, including politeness, emphasis, and persuasion. For example, expressions like *I would like to make a suggestion* convey politeness and formality, while phrases such as *let me tell you something* signal informality and directness. These patterns indicate that collocations are integral to the organization of discourse and the expression of speaker intention.

The analysis also reveals notable variation in collocational usage across different genres. Academic texts tend to favor more formal and specialized collocations, such as *pose a question* and *draw a conclusion*, whereas spoken language often employs more general and accessible expressions, such as *ask a question*. This genre-based variation further supports the idea that collocations are context-dependent and functionally motivated. The findings of the study have important implications for language learning and translation. Non-native speakers frequently encounter difficulties with collocations, leading to errors such as *strong rain* instead of *heavy rain* or *do a decision* instead of *make a decision*. Such errors result from a lack of awareness of conventional lexical patterns and emphasize the need for teaching collocations as integral units of meaning. In translation, the accurate rendering of collocations is essential for preserving naturalness and stylistic adequacy, as literal translation may fail to capture the pragmatic and semantic nuances of the original text. The results confirm that English collocations are characterized by stability, semantic specificity, and pragmatic relevance. They function as essential building blocks of natural language and play a key role in ensuring effective communication across different contexts.

Conclusion: This study has examined the semantic and pragmatic features of English collocations through a corpus-based approach, drawing on data from the British National Corpus and the Corpus of Contemporary American English. The findings demonstrate that collocations constitute an essential component of natural language use, characterized by their frequency, stability, and conventionalized patterns of co-occurrence.

The analysis confirms that collocations are not merely combinations of words but function as meaningful lexical units with specific semantic and pragmatic properties. From a semantic perspective, collocations exhibit strong internal cohesion and are often associated with semantic prosody, which influences the evaluative meaning of expressions. From a pragmatic standpoint, their usage is closely related to contextual factors such as register, genre, and communicative intention, making them highly relevant for effective discourse production.

Furthermore, the study highlights the importance of collocational competence in both language learning and translation. A lack of awareness of typical collocational patterns may lead



to unnatural or incorrect language use, particularly among non-native speakers. Therefore, it is essential to treat collocations as integral units of vocabulary and incorporate them into language teaching practices. The research also demonstrates the value of corpus linguistics as a methodological tool, providing reliable and authentic data for linguistic analysis. By combining quantitative and qualitative approaches, the study offers a comprehensive understanding of how collocations function in real language use.

Despite certain limitations, including the restricted selection of lexical items, the study contributes to the fields of corpus linguistics, lexicology, and applied linguistics. Future research may expand the scope by examining a broader range of collocations, exploring cross-linguistic comparisons, or analyzing collocational patterns in specialized domains.

REFERENCES

1. Firth J.R. *Papers in Linguistics 1934–1951*. – Oxford: Oxford University Press, 1957. – 246 p.
2. Abdusamatovna F. N. TEACHING ENGLISH FOR SPECIFIC PURPOSES //PEDAGOGIK TADQIQOTLAR JURNALI. – 2025. – T. 3. – №. 1. – С. 792-795.
3. Sinclair J. *Corpus, Concordance, Collocation*. – Oxford: Oxford University Press, 1991. – 179 p.
4. McEnery T., Hardie A. *Corpus Linguistics: Method, Theory and Practice*. – Cambridge: Cambridge University Press, 2012. – 294 p.
5. Lewis M. *Teaching Collocation: Further Developments in the Lexical Approach*. – London: Language Teaching Publications, 2000. – 223 p.
6. Khalikova D. ANTHROPOCENTRIC PARADIGM-THE MAIN SYSTEM IN MODERN LINGUISTICS PECULIARTIES OF BRITISH AND AMERICAN WORDS //Универсальная индексная библиотека инновационных исследований в современном мире: теория и практика. – 2023. – Т. 2. – №. 18. – С. 64-65.
7. Normurodovna X. D. SYMBOLIC UNITS IN ENGLISH DIGITAL TEXTS: THEIR LINGUISTIC FUNCTIONS IN SOCIAL MEDIA COMMUNICATION //KONFERENSIYA. – 2025. – Т. 1. – №. 2. – С. 436-442.
8. Xalikova Dildora Normurodovna. (2025). MISINTERPRETATION OF SYMBOLS IN THE ERA OF DIGITAL TRANSFORMATION. «Raqamli transformatsiya davrida pedagogik ta'limni rivojlantirish istiqbollari» (ASR), Tashkent, Uzbekistan. Academic Scientific Research Group. <https://doi.org/10.5281/zenodo.17391337>
9. Kuryozova G. A., Fuzaylova N. A. TRANSLATING WORDS AND PHRASES TAKING INTO ACCOUNT CULTURAL CONTEXT //SHOKH LIBRARY. – 2025.
10. Fuzaylova N. A. USE TECHNOLOGY IN MATERIAL DESIGN, LESSON PROCEDURES, AND ASSESSMENT //INNOVATIVE DEVELOPMENTS AND RESEARCH IN EDUCATION. – 2024. – Т. 3. – №. 33. – С. 141-143.

