

## THE USAGE OF COCA AND INVESTIGATIONS OF VOCABULARY FORMING

*Rakhimov Mubin*

*Teacher of Bukhara State University*

*Mustaqimova Ozoda*

*Rahimova Mashhura*

*Hoshimova Ozoda*

*Students of Bukhara State University*

**Abstract.** This article explores the utilization of the Corpus of Contemporary American English (COCA) as a valuable resource for investigating vocabulary formation in contemporary English. The COCA database provides researchers with access to a vast collection of authentic written and spoken texts, allowing for the analysis of lexical patterns, word frequency, and usage trends across different contexts and genres. Through empirical studies and quantitative analysis, researchers have gained insights into various aspects of vocabulary formation, including the emergence of new words, the evolution of meanings, and the influence of linguistic factors on lexical development. This article reviews recent research conducted using COCA data, focusing on its contributions to understanding lexical creativity, semantic change, and sociolinguistic variation in English. Additionally, the article discusses methodological considerations and challenges associated with using COCA for vocabulary research, such as corpus representativeness, data selection, and computational techniques. This article highlights the importance of COCA as a valuable tool for investigating the dynamic nature of vocabulary in contemporary English and offers suggestions for future research directions in the field of corpus linguistics and lexical studies.

**Key words:** *Corpus of Contemporary American English (COCA), vocabulary formation, lexical patterns, word frequency, usage trends, lexical creativity, semantic change, sociolinguistic variation, corpus linguistics.*

**Introduction.** The Corpus of Contemporary American English (COCA) is a large and comprehensive database of English language usage, compiled and maintained by Mark Davies, a professor of corpus linguistics at Brigham Young University. COCA is one of the most extensive corpora available for the study of contemporary English, containing over 560 million words of text collected from a diverse range of sources.

COCA is designed to represent a wide variety of written and spoken language as it is used in the United States from the early 1990s to the present day. The corpus is

continuously updated to reflect changes in language usage over time, making it a valuable resource for researchers, linguists, educators, and language learners.

The texts included in COCA come from a range of sources, including newspapers, magazines, books, academic journals, transcripts of spoken language (such as interviews, speeches, and conversations), blogs, TV shows, and movie subtitles. This diverse collection of texts provides a comprehensive and representative sample of contemporary American English in various contexts and registers.

COCA allows users to search for specific words or phrases and analyze their frequency, usage patterns, collocations, and contexts of occurrence. Researchers can use COCA to study language variation, linguistic features, stylistic differences, and changes in language over time. Additionally, COCA provides tools for conducting quantitative analysis, generating frequency lists, and comparing usage across different genres and time periods.

One of the key strengths of COCA is its size and representativeness, which allows researchers to draw reliable conclusions about language usage in contemporary American English. By analyzing the vast amount of data available in COCA, researchers can gain insights into language trends, sociolinguistic patterns, lexical creativity, semantic change, and many other aspects of language variation and change.

Another valuable feature of language corpora is their ability to reveal common collocations – words that frequently appear together. By studying these word combinations, you gain insight into natural language patterns and idiomatic expressions, enhancing your fluency and communicative competence.

Furthermore, language corpora facilitate contextual learning by providing rich examples of word usage in diverse genres and registers. Whether you're interested in news articles, academic papers, fiction, or spoken discourse, corpora offer a wealth of authentic texts for exploration. Engaging with different genres exposes a learner to specialized vocabulary and language conventions, broadening linguistic horizons.

In the sphere of corpus linguistics and vocabulary formation, various scholars and researchers have made significant contributions to our understanding of language usage and development. Mark Davies<sup>1</sup>, Professor of Corpus Linguistics at Brigham Young University and creator of the Corpus of Contemporary American English (COCA), which has provided researchers with a vast database for studying contemporary English usage. John Sinclair<sup>2</sup>, Pioneer in corpus linguistics and author of "Corpus, Concordance, Collocation," which introduced the concept of using computerized corpora for linguistic analysis. Michael Stubbs<sup>3</sup>, linguist known for his

<sup>1</sup> Davies, Mark. "The Corpus of Contemporary American English (COCA): 560+ million words, 1990-present." Available online at: <https://www.english-corpora.org/coca/>

<sup>2</sup> Sinclair, John. "Corpus, Concordance, Collocation." Oxford University Press, 1991.

<sup>3</sup> Stubbs, Michael. "Words and Phrases: Corpus Studies of Lexical Semantics." Blackwell, 2001.

work on lexical semantics and author of "Words and Phrases: Corpus Studies of Lexical Semantics," which examines how words acquire meaning in context. Michael Oakes<sup>4</sup> is linguist known for his work on statistics for corpus linguistics, provides methodologies for analyzing and interpreting corpus data.

**Main part.** Utilizing the Corpus of Contemporary American English (COCA) for forming new vocabulary can be deemed an effective strategy for language learners and enthusiasts. There are several ways in which COCA can be employed for this purpose.

*Identification of new words and phrases.* One can search for unfamiliar words or phrases in COCA to uncover how they are employed in authentic contexts. Attention can be paid to their meanings, collocations, and grammatical structures. For example: Should one encounter the word "*ubiquitous*," it can be searched for in COCA to find examples such as "*Technology has become ubiquitous in our daily lives.*"

*Exploration of synonyms and antonyms.* Synonyms and antonyms of known words can be looked up in COCA to expand vocabulary. Analysis can be conducted on how these words are used in different contexts and their nuances noted. For instance: synonyms of "*happy*" in COCA might include "*joyful*," "*content*," and "*elated*."

*Study of word families.* Investigation into related words sharing the same root or stem can deepen understanding of vocabulary. Exploration can be conducted on how these words are formed and how their meanings may vary. For example: Should one study the word "*environment*," related words like "*environmental*," "*environmentalist*," and "*environmentally*" can be searched for in COCA.

*Analysis of collocations.* Examination of common word combinations and collocations can elucidate how words are typically used together in natural language. Attention can be given to fixed expressions and idiomatic phrases. For instance: Collocations of "*take*" might include phrases like "*take a break*," "*take into account*," and "*take advantage of*" in COCA.

*Exploration of specialized vocabulary.* COCA can be employed to explore specialized vocabulary pertinent to specific topics or domains of interest. One can look for terms used in academic disciplines, industries, or hobbies. For example: If one is interested in psychology, terms like "*cognitive psychology*," "*behavioral therapy*," or "*neuroscience*" can be searched for in COCA to learn how they are used in scholarly texts and discussions.

*Analysis of word usage trends.* Investigation into word usage trends and frequency distributions in COCA can aid in identifying emerging words or lexical patterns. Monitoring of how language evolves over time can guide vocabulary learning. For instance: Tracking the usage of technology-related terms like "*smartphone*,"

<sup>4</sup> Oakes, Michael. "Statistics for Corpus Linguistics." Edinburgh University Press, 1998.

"*selfie*," and "*streaming*" in COCA can provide insights into contemporary language trends. By employing these strategies and actively engaging with COCA, language learners can expand their vocabulary, deepen their understanding of word usage, and become more proficient communicators in English.

Investigating vocabulary formation through the use of COCA involves several methodologies and approaches. Researchers typically analyze the corpus to identify patterns, trends, and factors that contribute to the creation and evolution of words and phrases. Researchers examine the frequency of words and phrases within the corpus to determine their prevalence in contemporary usage. By tracking changes in frequency over time, they can identify emerging vocabulary and lexical trends.

Collocations refer to the tendency of words to occur together frequently. Researchers analyze collocations within COCA to understand how words form associations and clusters in natural language usage. This provides insights into the contexts in which new vocabulary emerges and how it integrates with existing linguistic structures.

Semantic analysis involves examining the meanings and connotations of words and phrases within different contexts. Researchers use COCA to study semantic shifts, polysemy (multiple meanings), and semantic extensions, which shed light on the processes of vocabulary formation and evolution. Moreover, etymology is the study of the origins and history of words. Researchers investigate the etymology of words found in COCA to trace their linguistic roots and historical development. This helps identify borrowing from other languages, word formation processes, and semantic changes over time.

Another positive side of it that discourse analysis focuses on the use of language in social and cultural contexts. Researchers analyze discourse patterns within COCA to understand how vocabulary is used in specific domains, genres, or communities. This provides insights into the social factors influencing vocabulary formation and variation. Furthermore, corpus-based lexicography involves compiling dictionaries and lexical resources based on corpus data. Researchers use COCA to extract examples of word usage, definitions, and collocations, which are then used to create comprehensive dictionaries and reference materials.

**Conclusion.** In summary, the Corpus of Contemporary American English (COCA) is a powerful and comprehensive resource for the study of contemporary English language usage. Its extensive collection of texts, diverse sources, and advanced search and analysis tools make it an invaluable resource for researchers and language enthusiasts seeking to explore the intricacies of language in modern American society. Investigating vocabulary formation through the use of the Corpus of Contemporary American English (COCA) offers valuable insights into the dynamic nature of language. By employing various methodologies such as frequency analysis, collocation

analysis, semantic analysis, etymological analysis, discourse analysis, corpus-based lexicography, and comparative studies, researchers can uncover patterns, trends, and factors that shape the creation and evolution of words and phrases. Through COCA, researchers can track the frequency of words over time, analyze how words form associations and clusters in natural language usage, understand semantic shifts and historical development, explore the use of language in social and cultural contexts, compile dictionaries and lexical resources, and conduct comparative studies to identify regionalisms, dialectal variations, and cultural influences.

**References:**

1. Davies, Mark. "The Corpus of Contemporary American English (COCA): 560+ million words, 1990-present." Available online at: <https://www.english-corpora.org/coca/>
2. Biber, Douglas et al. "Longman Grammar of Spoken and Written English." Pearson Education, 1999.
3. Sinclair, John. "Corpus, Concordance, Collocation." Oxford University Press, 1991.
4. Stubbs, Michael. "Words and Phrases: Corpus Studies of Lexical Semantics." Blackwell, 2001.
5. Oakes, Michael. "Statistics for Corpus Linguistics." Edinburgh University Press, 1998.
6. Rakhimov, M. "The Grammatical Meaning and Form Correlations in Uzbek and English." *Excellencia: International Multi-disciplinary Journal of Education* (2994-9521) 2.3 (2024): 179-182.