A Corpus-Based Comparative Study of Firstly and Initially from the Perspective of Extended Unit of Meaning in COCA

WANG Lu-han

Wuhan University of Technology, Hubei, Wuhan, China

This paper explores the differences between the near-synonyms "initially" and "firstly" using the Corpus of Contemporary American English (COCA) and the theory of extended units of meaning. The study analyzes the two words from five perspectives: word frequency, collocation, colligation, semantic preference, and semantic prosody. Results show that "initially" is more frequent, especially in academic writing, while "firstly" is common in blogs and informal contexts. "Firstly" often collocates with nouns and expresses logical sequence, while "initially" collocates with adjectives and describes initial states or scientific processes, often with negative connotations. This study highlights the effectiveness of corpora in distinguishing near-synonyms and offers insights for English vocabulary learning.

Keywords: synonym, collocation, semantic preference, semantic prosody

Introduction

Synonyms have long been a subject of interest in linguistics due to their complexity and nuanced meanings. While dictionaries are essential tools for understanding word meanings, they often fail to capture the contextual usage and cultural subtleties that differentiate near-synonyms. This is where corpora, with their extensive and authentic data, provide a valuable resource. This study examines the near-synonyms "initially" and "firstly" using the Corpus of Contemporary American English (COCA) and the theory of extended units of meaning. By analyzing these words from multiple perspectives, the study aims to highlight the subtle differences between them and demonstrate the effectiveness of corpora in distinguishing near-synonyms.

Research on Synonyms

Near-synonyms are more common than absolute synonyms, and their distinctions often involve comparing semantic fields. (Abdulkhaeva, 2020; Alduhaim, 2021; Cao, 2020; Effendi, 2020; Lin, 2021; Phoocharoensil, 2020, 2021; Shahzadi, 2019; Song, 2021). Huang and Wang (2021) pointed out that corpus-based linguistic research has achieved fruitful results in recent years, with large-scale corpora becoming the premise and main tool for dictionary compilation. Corpus-based research has proven effective in differentiating synonyms, with studies like Han (2016) using the extended unit of meaning model to compare and contrast near-synonyms. This paper builds on such research by using COCA to analyze "initially" and "firstly" from the perspectives of word

WANG Lu-han, postgraduate, School of Foreign Language, Wuhan University of Technology.

frequency, collocation, colligation, semantic preference, and semantic prosody. The goal is to provide a comprehensive understanding of how these words are used in different contexts and to offer practical insights for English learners and teachers.

Extended Unit of Meaning

Traditional semantic studies hold that a single word represents a unit of meaning, and the meaning of a whole expression can be derived by summing up the meanings of individual words. However, the meaning conveyed by a combination of words is sometimes different from the sum of the meanings of the individual words. Sinclair (1996) introduced the Extended Unit of Meaning model, which includes five elements: node word, collocation, colligation, semantic preference, and semantic prosody. This model shifts the study of meaning from individual words to phraseological levels, emphasizing the co-selection relationships between words (Wang, 2012). The node word is the central element, around which other words cluster to form fixed patterns. Collocation refers to the tendency of words to co-occur, while colligation involves the grammatical relationships between words. Semantic preference identifies the semantic field associated with the collocates of a node word, and semantic prosody reflects the overall tone or attitude conveyed by a word and its collocates.

Research Methods

The study uses the Corpus of Contemporary American English (COCA), a balanced corpus of 1 billion words, to analyze "initially" and "firstly." The research involves three steps: (1) obtaining word frequency and distribution across different domains; (2) analyzing collocates within a span of -4 to +4; and (3) manually categorizing 100 randomly sampled sentences to examine semantic preference and prosody. The Mutual Information (MI) value is used to measure the strength of collocations, with an MI value greater than 3 indicating significant collocation strength.

Results And Discussion

Word Frequency

The analysis of word frequency reveals that "initially" is more frequent overall, especially in academic writing. This is likely because academic writing requires a clear structure and logical flow, and "initially" is effective in introducing the starting point of an argument or discussion. On the other hand, "firstly" is more common in blogs, where it fits the casual and personal tone used to express opinions or organize points in order. While both words are used across various platforms, "initially" is more versatile and appears more frequently in contexts requiring accuracy and clarity, such as news reporting and magazines.

Collocation and Colligation

In terms of collocation, "firstly" has the highest proportion of noun collocates, often related to abstract concepts, roles, and processes. For example, "firstly" frequently collocates with nouns like "methodology," "internationalisation," and "councillor." In contrast, "initially" collocates more frequently with adjectives, particularly those indicating doubt, hesitation, or skepticism, such as "reticent," "hesitant," and "skeptical." Both words collocate with verbs and adjectives, but "firstly" also collocates with adverbs, such as "secondly" and "thirdly," which indicate sequence or temporal relations.

Semantic Preference and Semantic Prosody

The semantic preference of "firstly" focuses on structure and hierarchy, with its collocates often carrying neutral or slightly positive semantic characteristics. Words like "reason," "methodology," and "outline" emphasize clarity, logical structure, and prioritization. In contrast, "initially" is more descriptive and narrative, often used to describe the initial stages of states, reactions, or scientific processes. Its collocates, such as "skepticism," "misdiagnose," and "reluctant," frequently carry negative connotations, reflecting uncertainty or difficulty associated with the beginning stages of actions or processes.

Conclusion

This study demonstrates that "initially" and "firstly" differ in frequency, collocation, and semantic prosody. "Initially" is more common in academic writing and often carries negative connotations, while "firstly" is suited for informal contexts and emphasizes logical structure. The findings underscore the importance of using corpora to distinguish near-synonyms and provide practical insights for English vocabulary learning and teaching. For English learners, distinguishing near-synonyms cannot rely solely on dictionary definitions but should incorporate real-life usage contexts. Teachers should integrate corpora into vocabulary teaching, providing students with authentic examples and guiding them to master corpus analysis methods.

References

Abdulkhaeva. (2020). *The differentiation of near-synonyms*. Konferencii.

Alduhaim A. (2021). Translating near-synonyms in the Quran: A semantic analysis of three near-synonyms and their English translations. *3L: Language, Linguistics, Literature, 27*(1), 76-89.

Cao Y. (2020). A corpus-based study of synonymous intensifiers: absolutely, totally, completely. *Proceedings of the 2nd International Conference on Literature, Art and Human Development (ICLAHD 2020)* (pp. 493-498). Xiamen, China: Atlantis Press.

Effendi, I. S., Amalia, R., & Lalita, S. A. (2020). Are synonyms always synonymous? A corpus-assisted approach to announce, declare, and state. *ASIAN TEFL Journal of Language Teaching and Applied Linguistics, 5*(1), 1-20.

Phoocharoensil, S. (2020). Collocational patterns of the near-synonyms error, fault, and mistake. *The International Journal of Communication and Linguistic Studies, 19*(1), 1-17.

Phoocharoensil, S. (2021). Semantic prosody and collocation: A corpus study of the near-synonyms persist and persevere. *Eurasian Journal of Applied Linguistics, 7*(1), 240-258.

Shahzadi, A., Asghar, A., & Javed, S. (2019). Effectiveness of corpus in teaching English synonyms. *Corporum: Journal of Corpus Linguistics-CJCL, 2*(1), 51-65.

Sinclair, J. (1996). The search for units of meaning. *Textus, 9*(1), 75-106.

Song, Q. (2021). Effectiveness of corpus in distinguishing two near-synonymous verbs: Damage and destroy. *English Language Teaching, 14*(7), 8.

黄水清，王东波. (2021). 国内语料库研究综述. *信息资源管理学报, 11*(03), 4-17+87.

王均松. (2012). 扩展语义单位的意义建构模式研究. *解放军外国语学院学报, 35*(2), 10-13,127.