# The Lexical Landscape of Begin, Start, and Initiate: A Corpus-Based Analysis of English Synonyms

Anawin Khlakheang[1], Amorn Cherngroongroj[2]

[1]International College, Dhurakij Pundit University, Bangkok, Thailand

[2]Independent Scholar, Bangkok, Thailand

[1]anawin.khl@dpu.ac.th, [2]amorn@live.com

**Received :** July 23, 2024; **Revised :** Octorber 7, 2024; **Accepted :** January 29, 2025

## Abstract

Synonyms can pose challenges for language learners, as Hemchua and Schmitt (2006) observed, particularly due to misunderstandings that arise from inappropriate word choices. This study employed a descriptive corpus-based research methodology, incorporating both quantitative analyses of frequency distributions and qualitative examinations of collocational patterns, to analyze the synonymous verbs "*begin*," "*start*," and "*initiate*," which appeared in the Oxford 3000 and 5000 lists—key vocabulary for English learners. The study aimed to examine instances of these verbs in the Corpus of Contemporary American English, supplemented by data from Longman and Oxford Advanced Learner's Dictionaries. By integrating corpus findings with dictionary insights, the study examined the frequency, distribution patterns, noun collocations, and semantic differences of these verbs. The analysis revealed that although "*begin*" and "*start*" were often interchangeable, "*start*" was preferred for scheduled events, while "*begin*" suggested initiation after deliberation. The verb "*initiate*" was predominantly used in formal and proactive contexts. These distinctions emphasized the importance of context in verb choice. By applying Mutual Information scores of 3 or above, the study uncovered significant collocational patterns, offering insights that extended beyond traditional lexical descriptions. The findings highlighted the pedagogical value of teaching vocabulary within contextual frameworks and demonstrated the potential of corpus linguistics in improving English as a Foreign Language instruction in contemporary educational settings.

**Keywords:** The Corpus of Contemporary American English; Corpus-Based Study; Synonyms

## Introduction

In the realm of linguistics, mastering the nuances and usage patterns of synonyms is crucial for both theoretical insights and practical applications. English verbs like "*begin*," "*start*," and "*initiate*" are frequently interchangeable yet carry subtle differences in connotation and context. This study employs a comprehensive corpus-based analysis to explore these distinctions. Synonym usage profoundly impacts language teaching and lexicography (Barcroft, 2016; Saville-Troike, 2012; Szudarski, 2018; Wilkins, 1972). Previous studies by Miller and Charles (1991) on semantic similarity and Jurafsky and Martin (2009) on computational linguistics provided foundational insights. Khlakheang and Cherngroongroj (2024) underscored corpus analysis's capacity to reveal nuanced usage patterns, surpassing traditional dictionary definitions. However, there remains a recognized gap in research regarding the nuanced practical usage of "*begin*," "*start*," and "*initiate*" across diverse contexts.

EFL learners face challenges in selecting appropriate synonyms due to nuanced differences not fully captured by traditional dictionaries. This necessitates integrating real-life usage patterns into vocabulary instruction (Norris, 2016; Thornbury, 2002). Cherngroongroj (2023) demonstrated how corpus-based analyses enrich language learning materials by revealing nuanced differences in synonym usage. The verbs "*show*" and "*display*" highlighted significant disparities in usage frequency and contextual preferences, emphasizing the complexity of synonyms. According to the Oxford 3000 and 5000 lists, "*start*" and "*begin*" are in the former, while "*initiate*" is in the latter, underscoring their importance in essential English vocabulary. Dictionaries often define synonyms using each other, such as defining "*start*" as 'to begin happening' or 'to make something begin happening,' and "*initiate*" as 'to arrange for something important to start,' reflecting their interrelated definitions.

This study targeted EFL learners and educators, aiming to enhance language proficiency and communicative competence through a deeper understanding of synonym usage patterns (Saville-Troike, 2012; Yeh et al., 2007). Methodologically, the study employed quantitative analyses of frequency distributions and qualitative examinations of collocational patterns, drawing insights from COCA and supplementing them with comparative evaluations from the Longman and Oxford dictionaries.

The research systematically explored "*begin*," "*start*," and "*initiate*" through detailed corpus analysis, aiming to advance the understanding of lexical semantics. The findings enhanced pedagogical practices for EFL learners' vocabulary acquisition and communicative competence, with broader implications for linguistic research and language education curriculum development.

## Research Objectives

1. To explore the frequency differences and distribution patterns of the synonyms "*begin*," "*start*," and "*initiate*" across genres in the Corpus of Contemporary American English.
2. To examine the similarities and dissimilarities between the synonyms "*begin*," "*start*," and "*initiate*" in terms of meanings and collocations in the Corpus of Contemporary American English.
3. To compare and contrast data obtained from dictionaries with data derived from corpus analysis.

## Literature Review

### Corpus Linguistics and Word Studies

Corpora were integral to contemporary linguistic research (Szudarski, 2018). Crawford and Csomay (2016) proposed corpus linguistics as an invaluable approach to understanding linguistic analysis and language, examining how language was authentically employed within diverse contexts and acknowledging its variability across different situations. Moreover, Moon (2010) emphasized the interconnected nature of language, highlighting how corpora unveiled the typical contexts of word usage and illuminated the significant contributions of word associations in enhancing our understanding of individual terms.

### The Definitions of Synonyms

According to Crystal (2018), synonyms, while ostensibly sharing the same meaning, might not have had exact equivalence in all cases. These lexemes were broadly classified into two categories. Cruse (2000) described absolute synonymy as the situation where two words had identical meanings and could be used interchangeably across all contexts. In contrast, Jackson and Amvela (2000) characterized near-synonyms as terms with overlapping meanings that could be interchanged in specific contexts but lacked universal substitutability. Murphy (2010) emphasized that determining synonymy often necessitated considering context or collocation.

### Collocations

Language intricacies often lay in the associations between words and their habitual pairing within speech. Exploring these linguistic phenomena, O'Dell and McCarthy (2017, p.6) defined a collocation as "a combination of two or more words which frequently occur together." Understanding these pairings was crucial in navigating the nuances of everyday communication. For instance, while someone might have been understood saying, "She has got yellow hair," this wasn't a conventional English expression; the more standard phrasing was, "She has got blond hair." Hence, "yellow" and "hair" were not commonly associated in English. Instead, "yellow" more naturally paired with items like flowers or paint. Consequently, students equipped with a comprehensive grasp of collocations could employ their existing vocabulary more precisely and organically.

### Previous Studies

A range of scholars conducted corpus-based investigations into the nuanced usage patterns of synonyms, employing diverse methodological approaches. Phoocharoensil (2010) explored the nuances of five English verb synonyms—"*ask*," "*beg*," "*plead*," "*request*," and "*appeal*"—examining their meanings, formality, connotations, collocations, and grammatical patterns, revealing regional variations in idiom usage and distinct collocational preferences. Similarly, Chung (2011) analyzed the synonyms "*create*" and "*produce*" using the Brown Corpus and the Frown Corpus, highlighting their shared meanings and the exclusive distributional insights provided by corpora.

General corpora, such as the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), were frequently employed in synonymous research. Figes (2013) utilized the BNC alongside a dictionary, while Thamratana (2013) and Jarunwaraphan and Mallikamas (2020) combined COCA with dictionaries to investigate synonyms. Cherngroongroj (2023) conducted an in-depth exploration of the synonymous verbs "*show*" and "*display*," revealing distinctive usage patterns not apparent in dictionary definitions. The study found a significant disparity in frequency, with "*show*" being more prevalent due to its broader interpretations. Contextual analysis across genres showed that "*show*" was more natural in contexts associated with "ARRIVING AT THE PLACE," while "*display*" was more prevalent in scenarios involving a "COMPUTER" or "SCREEN."

Building on this, Khlakheang and Cherngroongroj (2024) explored the synonymous verbs "*decline*," "*decrease*," and "*reduce*" using data from COCA and dictionaries, finding that "*decline*" and "*decrease*" were common in academic contexts, while "*reduce*" had broader usage across academic, environmental, and financial domains, emphasizing the importance of context in vocabulary instruction for EFL learners. Existing literature highlighted nuanced distinctions among synonyms, including varying meanings, dialectical nuances, degrees of formality, connotations, and collocational preferences. While dictionaries and corpora offered valuable data for comparing synonyms, some studies focused on limited criteria or relied solely on corpus data. Addressing these gaps, the present study aimed to establish comprehensive criteria—including word frequency, distribution patterns across genres, meanings, and collocations—to analyze the synonymous verbs "*begin*," "*start*," and "*initiate*," bridging the gap between corpus-based insights and practical vocabulary instruction.

## Methodology

This study employed a corpus-based approach to linguistic analysis, which was recognized for its effectiveness in exploring language usage across diverse contexts (Lindquist, 2009). Drawing on insights from previous research, such as Cherngroongroj (2023) on "*show*" and "*display*," this study applied similar frameworks and methods to examine the synonyms "*begin*," "*start*," and "*initiate*." Primary data sources included the Corpus of Contemporary American English (COCA), the Longman Dictionary of Contemporary English (6th ed., 2014), and the Oxford Advanced Learner's Dictionary (10th ed., 2020). The Longman dictionary offered comprehensive entries with over 230,000 words, examples, collocations, and thesaurus notes. In contrast, the Oxford dictionary provided a detailed lexicon of over 60,000 words with meanings and practical usage examples. The study employed a systematic five-step process: analyzing distribution patterns of synonym usage across eight genres in COCA; compiling and categorizing synonym meanings based on Longman and Oxford definitions; identifying collocates using COCA's features with a Mutual Information (MI) score ≥ 3 for statistical relevance (Cheng, 2012); examining semantic preferences of collocates to refine understanding; and conducting a comparative analysis integrating corpus data and dictionary definitions to reveal similarities and distinctions among the synonyms. This methodological rigor ensured a thorough exploration of usage patterns and semantic nuances crucial for effective language teaching and learning.

## Results

The ensuing section delineates the research outcomes concerning the three synonyms, namely "*begin*," "*start*," and "*initiate*." It is structured into three principal segments to present the comprehensive findings. The initial segment provides an analysis of distribution patterns prevalent across diverse genres. The subsequent part delves into an in-depth examination of the nuanced meanings associated with these terms. Lastly, the third segment accentuates observed collocational preferences within these synonyms.

### Distribution Patterns

In alignment with Research Objective 1, the investigation aimed to examine the frequency differences and distribution patterns of three synonyms, namely "*begin*," "*start*," and "*initiate*," across diverse genres within the COCA corpus. The specific distribution patterns for these identified synonyms are explicitly presented in Table 1.

**Table 1** Distribution patterns of "*begin*," "*start*," and "*initiate*" across eight genres in COCA

| Genre | *Begin* | | *Start* | | *Initiate* | |
|---|---|---|---|---|---|---|
| | Frequency | Per million | Frequency | Per million | Frequency | Per million |
| Blog | 12,264 | 95.36 | 45,661 | 355.03 | 615 | 4.78 |
| Web | 13,228 | 106.46 | 39,558 | 318.36 | 916 | 7.37 |
| TV / Movies | 7,360 | 57.47 | 44,302 | 345.91 | 684 | 5.34 |
| Spoken | 28,486 | 225.84 | 41,596 | 329.77 | 281 | 2.23 |
| Fiction | 10,275 | 86.84 | 24,031 | 203.10 | 272 | 2.30 |
| Magazine | 15,160 | 120.23 | 35,858 | 284.38 | 785 | 6.23 |
| News | 13,100 | 107.60 | 34,324 | 281.94 | 395 | 3.24 |
| Academic | 12,537 | 104.66 | 10,506 | 87.70 | 1,881 | 15.70 |
| **Total** | **112,410** | | **275,836** | | **5,829** | |

As presented in Table 1, the analysis of "*begin*," "*start*," and "*initiate*" across various genres in the COCA corpus reveals noteworthy trends in token frequency. "*Begin*" showed outstanding prevalence in spoken discourse (28,486 tokens; 225.84 per million), highlighting its frequent use in everyday communication contexts. In contrast, "*start*" demonstrated higher token counts in digital media such as blogs (45,661 tokens; 355.03 per million) and TV/movies (44,302 tokens; 345.91 per million), suggesting its prominent role in

narrative and visual storytelling. "*Initiate*," with its lowest token count overall, was notably more common in academic writing (1,881 tokens; 15.70 per million), underscoring its formal and deliberated usage in scholarly discourse.

Comparing the usage across genres based on token frequency per million, "*begin*" showed a significant presence in spoken discourse, where its frequency of 225.84 tokens per million stands out compared to its lower usage in other genres. Conversely, "*start*" exhibited consistent high frequencies in blogs (355.03 per million) and TV/movies (345.91 per million), indicating its prevalence in digital and visual media narratives. "*Initiate*," while least frequent overall, showed a distinct preference for academic writing (15.70 per million), emphasizing its specialized usage in scholarly contexts. These observations illustrated distinct contextual preferences and usage patterns among the synonyms, highlighting their nuanced distinctions in various linguistic contexts.

**Meanings and collocations**

Aligned with Research Objective 2, this study aims to explore the similarities and distinctions among three synonyms: "*begin*," "*start*," and "*initiate*," focusing on their collocations and meanings within the COCA corpus. Two methods were employed to investigate these verbs. Initially, their definitions from the Longman and Oxford Advanced Learner's Dictionaries were scrutinized, revealing frequent overlaps that may lead to confusion. "*Start*" generally denoted commencing a new action, event, or process, including the initiation of machinery, movements, or journeys. Similarly, "*begin*" signified the commencement of actions or events, while "*initiate*" involved starting formal or significant processes or introductions. However, subtle nuances can obscure their practical distinctions.

As highlighted by Murphy (2010), relying solely on dictionary definitions is inadequate. Therefore, the study also examined the collocates associated with these verbs to discern their semantic preferences, following Sinclair's concept of "semantic preference" (2004, p. 142). This approach focused on noun collocates to analyze the objects typically associated with each synonym. The top-thirty nouns frequently co-occurring with the synonymous verbs "*begin*," "*start*," and "*initiate*" were retrieved from COCA, prioritizing those with an MI score of 3 or higher to ensure statistical significance (Cheng, 2012). These findings were detailed below.

**Table 2** Noun Collocates with "*begin*," "*start*," *and* "*initiate*"

| No. | *Begin* | | | *Start* | | | *Initiate* | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noun Collocate | Freq. | MI Score | Noun Collocate | Freq. | MI Score | Noun Collocate | Freq. | MI Score |
| 1 | Clip | 8,341 | 9.29 | Head | 4,061 | 3.30 | Process | 171 | 4.83 |
| 2 | Video | 7,799 | 7.01 | Season | 2,770 | 3.44 | Action | 122 | 5.02 |
| 3 | Process | 1,448 | 3.65 | Finish | 1,253 | 4.02 | Contact | 117 | 5.71 |
| 4 | Break | 810 | 3.61 | Scratch | 761 | 5.57 | Conversation | 107 | 5.67 |
| 5 | Construction | 489 | 4.26 | Menu | 669 | 4.20 | Sequence | 83 | 7.36 |
| 6 | Trial | 371 | 3.42 | Businesses | 550 | 3.01 | Program | 87 | 3.80 |
| 7 | Journey | 374 | 4.30 | Crying | 375 | 3.31 | Change | 88 | 3.31 |
| 8 | Caption | 281 | 6.64 | Engine | 371 | 3.12 | Proceedings | 80 | 7.94 |
| 9 | Sen | 269 | 4.32 | Treaty | 340 | 4.04 | Investigation | 74 | 5.27 |
| 10 | Negotiations | 275 | 4.62 | Button | 429 | 3.35 | Sex | 70 | 4.47 |
| 11 | Talks | 245 | 3.56 | Negotiations | 226 | 3.03 | Conversations | 64 | 6.64 |
| 12 | Classes | 219 | 3.14 | Engines | 206 | 3.51 | Discussion | 61 | 4.84 |
| 13 | Healing | 179 | 4.29 | Bidding | 159 | 4.73 | Programs | 56 | 4.05 |
| 14 | Hearings | 153 | 4.61 | Packing | 154 | 3.89 | Changes | 48 | 3.95 |
| 15 | Announcer | 139 | 4.53 | Digging | 147 | 3.40 | Dialogue | 50 | 6.13 |
| 16 | Phase | 139 | 3.16 | Basics | 165 | 4.01 | Search | 36 | 4.04 |
| 17 | Dialogue | 127 | 3.35 | Digest | 118 | 4.39 | Emergency | 34 | 4.79 |
| 18 | Preparations | 109 | 5.50 | Slate | 131 | 3.71 | Computer | 34 | 3.70 |
| 19 | Ceremony | 108 | 3.53 | Premise | 119 | 3.14 | Plans | 54 | 3.82 |
| 20 | Proceedings | 107 | 4.20 | Semester | 105 | 3.07 | Movement | 33 | 3.80 |
| 21 | Trials | 96 | 3.36 | Kindergarten | 93 | 3.22 | Efforts | 34 | 3.95 |
| 22 | Careers | 103 | 3.75 | Inning | 86 | 3.10 | Ability | 36 | 3.60 |
| 23 | Deliberations | 84 | 6.19 | Riot | 83 | 3.29 | Discussions | 31 | 5.79 |
| 24 | Impeachment | 84 | 4.34 | Rehab | 70 | 3.27 | Activities | 30 | 3.90 |
| 25 | Festivities | 83 | 6.00 | Drip | 59 | 3.93 | Attack | 31 | 3.54 |
| 26 | Gov | 76 | 3.12 | Downswing | 59 | 6.54 | Activity | 29 | 3.89 |
| 27 | Quest | 81 | 3.48 | Opener | 61 | 3.10 | Response | 33 | 3.37 |
| 28 | Playoffs | 87 | 3.77 | Countdown | 60 | 4.01 | Procedures | 29 | 5.26 |
| 29 | Descent | 78 | 4.29 | Seedlings | 48 | 4.26 | Actions | 32 | 4.21 |
| 30 | Menachem | 82 | 8.94 | Daytona | 44 | 4.32 | Investigations | 26 | 6.10 |

Researchers analyzed noun collocates from COCA with an MI score of 3 or higher to explore semantic patterns associated with "*begin*," "*start*," and "*initiate*." This analysis revealed contextual preferences and associations, which are detailed further in the following section along with tables illustrating usage in specific linguistic contexts.

**Table 3** The semantic preferences of nouns that collocate with *begin*

| No. | Semantic preference | Noun collocates of *begin* |
|---|---|---|
| 1 | Action/Process | 1. Clip, 2. Video, 3. Process, 4. Break, 5. Construction, 6. Trial, 13. Healing, 16. Phase, 20. Proceedings, 21. Trials, 23. Deliberations, 27. Quest, 28. Playoffs, 29. Descent |
| 2 | Event/Activity | 7. Journey, 8. Caption, 12. Classes, 19. Ceremony, 24. Impeachment, 25. Festivities, 30. Menachem |
| 3 | Negotiation | 9. Sen, 10. Negotiations, 11. Talks |
| 4 | Career | 22. Careers |
| 5 | Announcement | 14. Hearings, 15. Announcer |
| 6 | Preparation | 17. Dialogue, 18. Preparations |
| 7 | Government | 26. Gov |

**Table 4** The semantic preferences of nouns that collocate with *start*

| No. | Semantic preference | Noun collocates of *start* |
|---|---|---|
| 1 | Action/Process | 1. Head, 6. Businesses, 8. Engine, 11. Negotiations, 12. Engines, 13. Bidding, 14. Packing, 15. Digging, 23. Riot, 24. Rehab, 27. Opener, 28. Countdown |
| 2 | Time/Event | 2. Season, 20. Semester, 22. Inning, 30. Daytona |
| 3 | Creation/Initiation | 4. Scratch, 9. Treaty, 10. Button, 18. Slate, 19. Premise, 29. Seedlings |
| 4 | Completion | 3. Finish |
| 5 | Emotional | 7. Crying |
| 6 | Miscellaneous | 5. Menu, 16. Basics, 17. Digest, 21. Kindergarten, 25. Drip, 26. Downswing |

**Table 5** The semantic preferences of nouns that collocate with *initiate*

| No. | Semantic preference | Noun collocates of *initiate* |
|---|---|---|
| 1 | Process/Action | 1. Process, 2. Action, 6. Program, 13. Programs, 16. Search, 18. Computer, 27. Response, 28. Procedures |
| 2 | Communication | 3. Contact, 4. Conversation, 11. Conversations, 12. Discussion, 15. Dialogue, 23. Discussions |
| 3 | Legal/Official | 8. Proceedings, 9. Investigation, 30. Investigations |
| 4 | Change/Plans | 7. Change, 14. Changes, 19. Plans |
| 5 | Movement/Efforts | 5. Sequence, 20. Movement, 21. Efforts, 24. Activities, 26. Activity |
| 6 | Ability | 22. Ability |
| 7 | Emergency | 17. Emergency |
| 8 | Sexual Activity | 10. Sex, 25. Attack, 29. Actions |

## Discussions

Semantic prosody, as described by Louw (1993, p. 157), encapsulated the consistent aura of meaning imbued by collocates surrounding a lexical item. The analysis of noun collocates associated with "*begin*," "*start*," and "*initiate*" in the COCA corpus reveals distinct semantic preferences and contextual nuances. "*Begin*" and "*initiate*" shared common collocates such as '*process*,' '*dialogue*,' and '*proceedings*,' indicating their usage in contexts involving structured or formal sequences of actions. Both verbs were used in initiating formal procedures and discussions. Conversely, "*begin*" and "*start*" co-occur with the noun '*negotiations*,' suggesting their interchangeable use in the context of initiating discussions and agreements. However, no significant shared collocates were found between "*start*" and "*initiate*," highlighting their distinct semantic fields.

The semantic preferences of nouns that collocate with each verb further illustrated their contextual nuances. "*Begin*" frequently collocated with nouns related to actions or processes, events or activities, negotiations, careers, announcements, preparations, and government contexts. "*Start*" commonly paired with nouns denoting actions or processes, times or events, creation or initiation, completion, and emotional states. "*Initiate*" tended

to collocate with nouns associated with processes or actions, communication, legal or official contexts, changes or plans, movements or efforts, abilities, emergencies, and sexual activities.

Overall, while "*begin*," "*start*," and "*initiate*" exhibit overlapped in certain formal contexts, "*initiate*" stood out for its exclusive association with deliberate and formal beginnings. This analysis underscored the nuanced semantic fields of these verbs, contributing to a deeper understanding of their usage patterns and contextual associations in academic and professional discourse.

The findings of this study aligned with existing research on the contextual nuances of near-synonyms. Jackson and Amvela (2000) delineated near-synonyms as terms with overlapping meanings that cannot universally substitute each other in all contexts. Murphy (2010) underscored the necessity of considering precise contextual and collocational factors beyond mere dictionary definitions. Cherngroongroj (2023) delved into the subtle differentiations among near-synonyms, emphasizing the pivotal role of context in determining appropriate word usage. This study reinforced these observations by elucidating how "*begin*," "*start*," and "*initiate*" exhibit varied usage patterns contingent upon the specific formality and context in which they are employed.

**Contrasting Dictionary Definitions with Corpus Insights**

Aligned with Research Objective 3, this study undertakes a comprehensive comparison and contrast of the meanings and example sentences derived from dictionaries and corpus analysis to enrich our understanding of the synonyms "*begin*," "*start*," and "*initiate*." Dictionaries typically provided succinct definitions and illustrative sentences that serve as foundational references for these verbs. For instance, "*begin*" and "*start*" were often described similarly as initiating actions or events, while "*initiate*" was noted for its formal or significant connotations.

In contrast, corpus analysis offered a more nuanced perspective by revealing how these verbs were used across a wide range of authentic contexts. This approach uncovered subtle distinctions in their usage patterns that may not be immediately apparent from dictionary entries alone. For example, while dictionaries might suggest broad interchangeability between "*begin*" and "*start*," corpus data often showcased specific collocations and contextual preferences that distinguish their practical usage.

By juxtaposing dictionary definitions with corpus findings, this study enhanced our grasp of how these synonyms were employed in different linguistic registers and communicative contexts. This comparative analysis not only clarified semantic nuances but also provided valuable insights into the pragmatic functions of these verbs in everyday language use. Such insights were crucial for refining language teaching strategies and promoting more accurate and effective language acquisition among learners.

## Conclusion

The analysis of noun collocates associated with "*begin*," "*start*," and "*initiate*" provided valuable insights into their semantic preferences and contextual usage in discourse. While "*begin*" and "*start*" shared common collocates related to the initiation of actions and interpersonal interactions, "*initiate*" stood out for its specific association with formal and proactive initiation contexts. This study underscored the nuanced distinctions between these verbs, highlighting how they were employed in various linguistic contexts to convey the initiation of activities, projects, and interpersonal engagements. Regarding meaning and collocations, "*begin*," "*start*," and "*initiate*" were near-synonyms. As asserted by Jackson and Amvela (2000), a near-synonym refered to terms that can be used interchangeably where their meanings overlap but cannot be substituted in every context. Murphy (2010) added that the precise context or collocation must be considered because the definition of the word itself is insufficient. To clarify, "*begin*" and "*start*" were often interchangeable, as in (a). However, in cases of formal initiation, "*initiate*" was more natural, as in (b). Furthermore, there are contexts where only "*begin*" or "*start*" was appropriate, highlighting their distinct usage. In (c), "*start*" was preferred for scheduled events or activities, showing its specific usage compared to "*begin*." In (d), both "*begin*" and "*start*" were grammatically correct, but "*begin*" subtly implied initiation after preparation or deliberation, emphasizing the process, while "*start*" suggested a more immediate commencement of action. Thus, choosing between "*begin*" and "*start*" could reflect nuanced differences in how actions are perceived or described.

**Table 6**: Usage of Synonyms "Begin," "Start," and "Initiate" in Contextual Examples

| Context Description | Begin | Start | Initiate |
|---|---|---|---|
| (a) She decided to ___ her new project. | ✓ | ✓ | ✘ |
| (b) The company will ___ the new procedures next month. | ✘ | ✘ | ✓ |
| (c) The movie will ___ at 7 PM. | ✘ | ✓ | ✘ |
| (d) He ___ to understand the complex theory after hours of study. | ✓ | ✘ | ✘ |

### Contextual Explanation

The table summarized the contextual usage of the synonyms "*begin*," "*start*," and "*initiate*" based on specific sentence examples. The checkmarks (✓) indicated whether each verb can appropriately replace the blank in each context, while the cross marks (✘) indicated inappropriateness. This distinction highlighted the nuanced differences in verb selection that could impact meaning and perception in language use.

Understanding these semantic nuances enhanced our grasp of their usage patterns in academic, professional, and everyday communication, enriching our comprehension of how language expresses the commencement of actions and processes. This comparative analysis between corpus data and dictionary definitions revealed the importance of context in determining appropriate word choice, thereby contributing to more precise and effective language use.

## Pedagogical Implication

The research article "The Lexical Landscape of Begin, Start, and Initiate: A Corpus-Based Analysis of English Synonyms" offers significant pedagogical implications for language educators and curriculum developers. Analyzing the synonyms "*begin*", "*start*", and "*initiate*" across various genres in the Corpus of Contemporary American English (COCA) reveals their nuanced usage in different contexts, enriching teaching materials with precise word meanings and appropriate usage contexts (Davies, 2008). This understanding aids in teaching vocabulary and semantics effectively, improving learners' communicative competence by emphasizing real-world language use over static dictionary definitions. Integrating corpus-based examples into instruction supports data-driven learning principles, ensuring students engage with authentic language contexts and enhancing their lexical precision (Boulton, 2012). By incorporating these insights, educators can design more

effective exercises that help students master the subtle differences in meaning and usage of "*begin*", "*start*", and "*initiate*", thereby fostering greater linguistic competence in English language learning (Schmitt, 2000).

## Recommendation for Further Studies

Future research should explore the nuanced distinctions and usage patterns of "*begin*," "*start*," and "*initiate*" across languages and cultures, tracking their evolution over time and analyzing their roles in specialized domains, including comparative studies using corpora like the British National Corpus (BNC) for insights into cultural and linguistic variations.

## References

Barcroft, J. (2016). *Vocabulary in Language Teaching (The Routledge E-Modules on Contemporary Language Teaching)*. Routledge.

Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. *English for Specific Purposes*, *31*(1), 1-13.

Cheng, W. (2012). *Exploring Corpus Linguistics: Language in Action (Routledge Introductions to Applied Linguistics)*. Routledge.

Cherngroongroj, A. (2023). *A Corpus-Based Analysis of English Synonyms with The Sense of "SHOWING"* [Master's Thesis, Kasetsart University, Bangkok, Thailand].

Chung, S.-F. (2011). A corpus-based analysis of "create" and "produce". *Chang Gung Journal of Humanities and Social Sciences*, *4*(2), 399–425.

Crawford, W. & Csomay, E. (2016). *Doing corpus linguistics*. Routledge.

Cruse, A. (2000). *Meaning in Language: An Introduction to Semantics and Pragmatics (Oxford Textbooks in Linguistics)*. Oxford University Press.

Crystal, D. (2018). *The Cambridge Encyclopedia of the English Language* (3rd ed.). Cambridge University Press.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Retrieved from https://www.english-corpora.org/coca/

Figes, S. M. (2013). *A corpus-based study on the English verbs: want, fancy, and wish*. MA Independent study. Thammasat University, Bangkok, Thailand.

Hemchua, S. & Schmitt, N. (2006). An analysis of lexical errors in the English compositions of Thai learners. *Prospects*, *21*(3), 3–25.

Jackson, H., & Amvela, Z. E. (2000). *Words, Meaning and Vocabulary: An Introduction to Modern English Lexicology*. Cassell.

Jurafsky, D., & Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (2nd ed.). Pearson Prentice Hall.

Jarunwaraphan, B., & Mallikamas, P. (2020). A Corpus-Based Study of English Synonyms: Chance and Opportunity. *rEFLections*, *27*(2), 218–245.

Khlakheang, A., & Cherngroongroj, A. (2024). A corpus-based analysis of English synonym usage problems in EFL learning: Decline, decrease, and reduce. In *Proceedings of the 4th International Conference on Education 2024: Directions for Thai Education to Meet Labor Market Needs in the Digital Age* (pp. 33-47). Chateau de Khao Yai Hotel & Resort, Pakchong Nakhonratchasima, Thailand: Association of Private Higher Education Institution of Thailand.

Lindquist H., (2009). *Corpus linguistics and the description of English.* Croydon, England: Edinburgh University Press.

Longman Dictionary of Contemporary English (6th ed.). (2014). Pearson Education.

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 157–75). Amsterdam, Netherlands: John Benjamins.

Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes, 6*(1), 1-28. https://doi.org/10.1080/01690969108406936

Moon, R. (2010). What can a corpus tell us about lexis?. In O'Keefe, A. and McCarthy, M. (Eds.). *The Routledge handbook of corpus linguistics*. (pp.197–211). Abingdon, England: Routledge.

Murphy, M. L. (2010). *Lexical meaning*. Cambridge Cambridge University Press

Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge University Press.

Norris, C. B. (2016). *Academic writing in English*. University of Helsinki.

O'Dell, F., & McCarthy, M. (2017). *English collocations in use* (2nd ed.). Cambridge University Press.

Oxford Advanced Learner's Dictionary (10th ed.). (2020). Oxford University Press.

Phoocharoensil, S. (2010). A corpus-based study of English synonyms. *International Journal of Arts and Sciences*, *3*(10), 227-245.

Saville-Troike, M. (2012). *Introducing Second Language Acquisition (Cambridge Introductions to Language and Linguistics)* (2nd ed.). Cambridge University Press.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.

Sinclair, J., & Carter, R. (2004). *Trust the text: Language, corpus and discourse*. Routledge.

Szudarski, P. (2018). *Corpus Linguistics for Vocabulary: A Guide for Research (Routledge Corpus Linguistics Guides)*. Routledge.

Thamratana, P. (2013). *A corpus-based analysis of reduce, decrease, diminish, dwindle and decline*. Independent study. Language Institute, Thammasat University, Bangkok, Thailand.

Thornbury, S. (2002). *How to teach vocabulary*. Harlow.

Wilkins, D. A. (1972). *Linguistics in Language Teaching*. Arnold

Yeh, Y., Liou, H. C., & Li, Y. H. (2007). Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning,20*(2), 131–152.