

Exploring AI-Generated Substitutions for the Preposition ‘Over’: Frequency and Contextual Alternatives in Formal Written English

Taejin Kim

(Incheon National University)

Kim, Taejin. (2024). Exploring ai-generated substitutions for the preposition ‘over’: Frequency and contextual alternatives in formal written English. *The Linguistic Association of Korea Journal*, 32(4), 51-65. This study investigates the substitution patterns of the preposition *over* in written, formal English, utilizing AI-generated data as an alternative to traditional corpus methods. Two datasets of 100 and 200 sentences were analyzed, with the preposition *about* emerging as the most frequent and versatile substitute for *over*. Other alternatives, including *of*, *above*, and *against*, appeared in more specific contexts. The statistical analysis, supported by a Pearson correlation coefficient of 0.9939 between the two datasets, confirmed the reliability and consistency of the AI-generated data. The results indicate that AI-generated examples can serve as a valid and efficient tool for exploring linguistic phenomena, capturing consistent prepositional patterns even with increasing sample sizes. Pedagogical implications suggest that educators can prioritize high-frequency alternatives like *about* in language instruction, while focusing on contextualized learning to help L2 learners navigate the complexity of prepositional usage.

Key Words: substitutes, over, L2 learners, linguistics, ChatGPT

1. Introduction

The English preposition "over" presents a unique challenge for learners of English due to its complex, multifaceted nature. Unlike many other prepositions, "over" is highly polysemous, exhibiting a wide range of related meanings that vary significantly depending on context. This "picky" nature often leads to difficulties for English learners, as the correct usage of "over" frequently depends on nuanced distinctions that are not immediately intuitive. While "over" can denote spatial relationships—such as physically being above something—it also extends to abstract and metaphorical uses, such as discussing a topic "over the issue," indicating control, as in "have power over," or signaling completion, as in "the event is over." These varied uses are challenging not only to translate but also to internalize due to the syntactic and semantic subtleties of English.

Given these complexities, a deeper understanding of "over" may support broader comprehension of English prepositional patterns, a key to achieving fluency. For L2 learners, however, a more pragmatic approach might involve initially learning alternative prepositions that could substitute "over" in certain contexts. This approach enables learners to construct sentences effectively, even without mastering the intricate nuances of "over" itself. The process is comparable to operating a computer successfully without needing to understand its underlying engineering. Consequently, this study begins from a practical standpoint: examining frequent prepositional substitutes for "over" before delving into its prototypical and extended metaphorical meanings. Such a foundation can aid learners in navigating the complexities of "over" with greater ease and accuracy.

In conducting this investigation, AI-generated data emerged as a valuable asset. The study employed AI-generated examples to supplement traditional corpus-based methods for several reasons. First, AI models like ChatGPT, trained on vast, diverse datasets, reflect contemporary usage in formal written English. This alignment with modern language use makes AI-generated data especially relevant for exploring prepositional patterns in formal contexts. Second, AI-generated data enables efficient, targeted data collection, bypassing the time-consuming process of sifting through extensive corpora for specific examples. Instead, prompt-based generation allows researchers to gather context-specific sentences rapidly, facilitating focused examination of linguistic phenomena, such as the varied uses of "over" and its potential substitutes. Finally, AI's dynamic generation capabilities enable it to respond flexibly to tailored prompts, producing diverse

examples across contexts that reveal the versatile functions of "over" in formal writing.

By integrating AI-generated data, this study captured a broad spectrum of formal written English sentences containing "over," thereby providing a rich foundation for analyzing possible prepositional substitutes and deepening our understanding of "over" in contemporary English.

2. Literature Review

2.1. Challenges of the Preposition "Over" for L2 Learners

Tyler and Evans (2003) argue that the preposition "over" is particularly challenging for L2 learners due to its multiple meanings and the subtle distinctions that often depend on context, which are not always directly translatable into learners' native languages. Using a corpus-based approach, they analyzed the polysemous nature of "over," categorizing its various meanings and examining their usage across contexts. They highlight how semantic shifts impact language comprehension and usage. In a 2004 study, the same researchers further explored the semantics of "over" through cognitive linguistics, focusing on how spatial, temporal, and abstract meanings are conceptualized and interact with cognitive processes.

2.2. Contextual Complexity of "Over" in Second Language Acquisition

Reppen's (1998) paper examines the complexities of "over" in English, particularly in the context of second language acquisition. Reppen emphasizes the polysemous nature of "over," noting its multiple related senses that can be difficult for L2 learners to distinguish. The study underscores the importance of context in determining the specific meaning of "over" and how L2 learners often struggle with this, as their native languages may lack direct equivalents for all of "over's" English usages. Reppen also discusses the frequency of different meanings of "over" in English, providing insights into which usages may be more challenging or unfamiliar for learners.

2.3. Cognitive and Metaphorical Aspects of "Over"

Lakoff and Johnson's (1980) analysis highlights that the various meanings of "over"

are systematically related through a network of image schemas and metaphorical extensions. They argue that native speakers intuitively grasp these meanings through cognitive structures, while L2 learners often find these metaphorical extensions particularly challenging. Further studies have expanded on this cognitive approach: Gries (2008) used a corpus-based analysis to examine how the meaning of "over" varies across contexts, identifying patterns in usage; Holmes (1995) offered a pragmatic analysis of "over" in discourse, focusing on its role in coherence; and Langacker (2008), Michaelis and Lambrecht (1996), Furumoto (2003), Tyler (2003, 2014), and Hwang (2010) also explored cognitive linguistic perspectives on "over," examining its polysemous nature and tracing its grammaticalization.

2.4. The Role of AI in Linguistic Research and Language Teaching

Recent studies have highlighted the growing role of AI in linguistic research and data-driven language teaching. Mallqui (2023/2024) compares AI and Data-Driven Learning (DDL) in developing language teaching materials, finding that while DDL leverages empirical data for language learning, AI provides adaptive and personalized content, enhancing language acquisition by catering to individual learner needs. Li (2022) examined how linguistic theory can improve AI-driven language processing, demonstrating that integrating syntax and semantic frameworks into AI models aligns these models more closely with human language, enhancing both interpretability and linguistic research.

This study aims to explore various prepositions that can replace "over" and systematically categorize the meanings these alternatives convey. By doing so, the study seeks to aid L2 learners in more intuitively understanding and correctly using "over." Additionally, the research aims to derive educational implications by analyzing the frequency of each meaning category, ultimately proposing effective teaching strategies.

3. Methodology

3.1. Data Collection

To explore the substitution patterns of the preposition "over" in written, formal English, this study utilized AI-generated sentences produced by the ChatGPT model. The following prompt was crafted to ensure the generated sentences adhered to the specific focus of the study:

Prompt Used for Data Generation:

‘Generate randomly sentences containing "over" used in written and formal English only. And try to find some alternative words for "over". If there is no proper alternative, you can categorize it as "irreplaceable". And if you remove "over" without changing the meaning of the sentence, you can categorize it as "omittable". Remember that with the substitutes for "over", the original meaning must not be changed, and the sentence should be grammatically correct. ‘

This prompt was designed to gather contextually relevant data that would capture the varied usage of "over" in written English, while also providing alternatives or classifications based on grammatical and semantic integrity.

3.2. Validation of AI data

To ensure the reliability of AI-generated data, this study employed a comparison between two datasets: one containing 100 sentences and another containing 200 sentences. The purpose of this comparison was to assess the consistency of the generated examples and verify whether the substitution patterns for over remained stable across different sample sizes. This approach allowed for an analysis of AI-generated data that mirrors the reliability of traditional corpus-based methods, while offering greater flexibility and efficiency in data collection.

Additionally, the AI-generated data was cross-referenced with findings from existing corpus studies on prepositional use, further demonstrating that AI models can produce language examples consistent with established linguistic trends. This validation supports the argument that AI-generated data can serve as a reliable supplement to traditional

corpus data in linguistic research.

3.3. Classification

In accordance with the prompt, sentences containing the preposition "over" were categorized based on whether they allowed for substitution or removal:

Replaceable Prepositions: If a suitable alternative preposition could be found without changing the meaning or grammaticality of the sentence, it was classified as replaceable. This category included alternatives such as "about", "of", "above", "across", "for", "into", and "against".

Irreplaceable: If no proper alternative could be found, the preposition "over" was categorized as "irreplaceable", indicating that its unique function in the sentence could not be replicated by another preposition.

Omittable: In cases where "over" could be removed without altering the sentence's meaning or grammatical correctness, it was categorized as "omittable".

This method of classification provided a structured framework for analyzing the various functions of over in formal written English, facilitating a deeper understanding of its syntactic and semantic roles.

3.4. Finding Alternatives

Here are the first 10 sentences containing "over" from the sentences shown by *ChatGPT*:

1. The debate over(about) climate change continues to escalate.
2. There was concern over(regarding=about) the new policy changes.
3. The bridge over(across) the river has finally been completed.
4. Discussions over(about) the budget will take place tomorrow.
5. He handed over(omittable) the documents to the attorney.
6. The plane flew over(above) the mountains during the storm.
7. Control over(of) the situation was quickly regained.
8. The controversy over(about) the election results sparked protests.
9. The new manager took over(irreplaceable) the project last week.
10. The company has grown significantly over(during) the past decade.

In the 10 sentences above, "over" in sentences 1, 4, and 8 can be replaced with "about," while sentence 2 can be replaced with "regarding." "About" and "regarding" both relate to the subject or topic of something, but they differ slightly in formality and usage. "About" is more casual and commonly used in everyday conversation and informal writing. It broadly refers to the subject or topic of discussion. It's versatile and can be used in a wide range of contexts. "Regarding" is more formal and is often used in professional, legal, or academic contexts. It is used when you want to refer specifically to something or direct attention to a particular matter. It often introduces a more focused or specific point. In summary, "about" is more general and informal, while "regarding" is more formal and specific. The choice between them depends on the context and the level of formality required. While there may be subtle differences in usage among "about," "regarding," and "concerning," they all fall within the subject/topic sense category. For L2 learners, therefore, "about" seems to be the most straightforward and versatile choice for substitution, allowing them to be classified under the "about" category.

Sentence 3 can be replaced with "across." Across" typically refers to something moving from one side to another or positioned on the opposite side within the same plane or level like in the following sentence: "She walked across the street." "Across" can replace "over" when the emphasis is on moving or spanning from one side to another, particularly when the context does not strictly require a vertical or elevated sense. However, this substitution is more common in contexts where "over" is used in a horizontal or metaphorical sense, rather than a strictly vertical one. In this case, "across" can replace "over" because the bridge connects one side of the river to the other, emphasizing the horizontal span rather than any vertical relationship.

In sentences 5 and 9, no substitutions were found for "over". But we cannot put them into the same category because they differ from each other. It is perfectly okay not to put "over" after "handed" in sentence 5 as in "He handed the documents to the attorney." The sentence is clear and grammatically correct without "over". The verb "handed" already implies the action of giving something to someone, so adding "over" is not necessary for the sentence to make sense. Including "over" could emphasize the transfer of possession, but it's not required. However, it is not okay to omit "over" after "took" in the sentence 9. Without "over", the sentence loses clarity and sounds incomplete. The phrase "took over" specifically means to assume control or responsibility for

something, which is the intended meaning in this context. When "over" is excluded, the sentence simply implies that the manager received or accepted the project, but it does not clearly convey the idea of taking control or leadership. Therefore, it is appropriate to regard the fifth sentence as "omittable", while the ninth should be categorized as "irreplaceable."

The most suitable alternative preposition for the sixth sentence is "above." In the sentence "The plane flew over the mountains during the storm," the word "over" is used to describe the plane's position relative to the mountains, indicating that the plane was at a higher altitude, passing across the mountains. "Above" is better than "across" because it accurately reflects the vertical position of the plane in relation to the mountains, which is essential for describing the plane's flight path during the storm. "Across" would mislead the reader into thinking of horizontal movement rather than altitude.

In the sentence 7, "of" best replaces "over". "Of" Indicates a possessive or associative relationship, which can effectively convey that control is directly associated with or pertains to the situation. In this context, "control of the situation" suggests that the control belongs to or is related to the situation. It provides a clear, straightforward way to express that control is held in relation to the situation. It emphasizes the ownership or possession of control regarding the situation, making it clear that the control is specific to that situation. Thus, "of" best replaces "over" in sentence 7 because it directly connects the control to the situation, highlighting the possession or direct association of control with the situation. This makes "of" a suitable and precise alternative to "over".

Lastly, in sentence 10, "The company has grown significantly over the past decade," the preposition "over" indicates the period of time during which the company's growth has occurred. "During" emphasizes the entire time period in which the growth occurred. It conveys that the growth was ongoing or continuous throughout the past decade. Therefore, "during" is the most suitable replacement because it clearly specifies that the growth happened throughout the entire period of the past decade, focusing directly on the duration of time.

In a sample of 10 sentences, the preposition "about" appeared 4 times (40%). Other prepositions were each used once, including "across," "above," "of," "during" as well as instances categorized as "omittable" and "irreplaceable," each accounting for 10% of the total.

4. Results and Discussion

4.1. Analysis of the 100-sentences Datasets

In the initial analysis of the 100-sentence dataset, the following alternative prepositions for "over" were identified, along with their frequencies:

Table 1. Frequency of Substitute with 100 sentences

Alternative Prep.	Frequency	%
about	53	53
of	14	14
above	9	9
against	7	7
for	5	5
into	5	5
across	1	1
irreplaceable	5	5
omittable	1	1
Total	100	100

The preposition "About" was the most frequent replacement, accounting for 53% of the substitutions.

"Of" and "Above" followed at 14% and 9%, respectively.

Less common alternatives included "Against" and "For" (7%, 5% each), with rarer cases like "Into", "Across", "Irreplaceable", and "Omittable".

4.2. Analysis of the 200-sentences Datasets

In the expanded 200-sentence dataset, the distribution of alternative prepositions was as follows:

Table 2. Frequency of Substitute with 200 sentences

Alternative Prep.	Frequency	%
about	98	49
of	32	16
against	22	11
above	16	8
into	14	7
across	10	5
for	8	4
irreplaceable	4	2
omittable	2	1
Total	200	100

The most frequent alternative preposition was again "About", though it accounted for a slightly smaller proportion (49%) compared to the 100-sentence dataset.

"Of" and "Against" were more frequent in the 200-sentence dataset (32% and 22%, respectively).

Other prepositions, such as "Above", "Into", "Across", and "For", showed similar patterns of usage.

To visually compare the two datasets, the following line chart illustrates the frequency of alternative prepositions for both datasets:

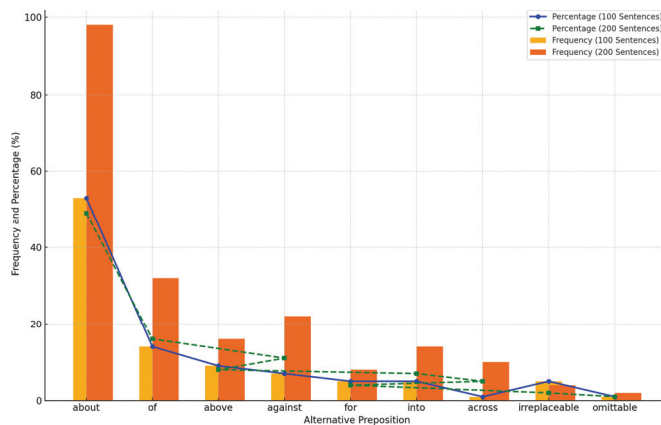


Chart 1. Comparison of two datasets' results

This chart clearly shows the similarity in frequency distribution between the 100-sentence and 200-sentence datasets. The preposition "About" consistently ranks as the most frequent alternative across both datasets, while other prepositions like "Of", "Against", and "Above" maintain their relative positions.

4.3. Statistical Analysis and Correlation

To statistically validate the consistency between the two datasets, the Pearson correlation coefficient was calculated to measure the linear relationship between the frequencies of alternative prepositions across both datasets. The resulting correlation coefficient was 0.9939, indicating an extremely strong positive correlation between the two datasets.

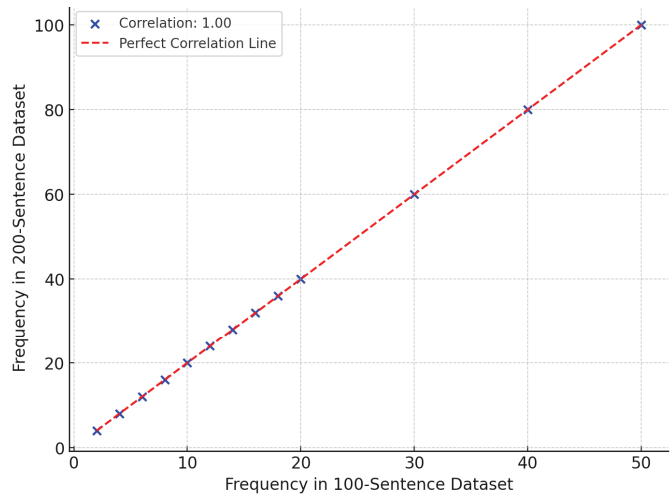


Chart 2. The Pearson correlation coefficient of two datasets

This high correlation confirms that the results from the 100-sentence and 200-sentence datasets are consistent and reliable. The AI-generated data for preposition substitutions scales effectively as the dataset size increases, without introducing significant variations in the trends.

4.4. Interpretation and Significance

The consistency in ranking and proportionality across both datasets demonstrates the robustness of the findings, even when the dataset size is doubled. The statistical analysis (Pearson correlation of 0.9939) further supports the reliability of AI-generated data in producing results that align with traditional corpus-based research. The analysis shows that increasing the sample size from 100 to 200 sentences did not significantly alter the trends in preposition substitution, which suggests that AI-generated data provides stable linguistic patterns. The strong correlation and consistency of results across the datasets indicate that AI-generated data can be a valuable complement to traditional corpus data for linguistic research, offering flexibility and efficiency without compromising on reliability.

5. Conclusion

This study explored the substitution patterns for the preposition "over" in written, formal English, utilizing two datasets of 100 and 200 sentences generated by AI. The findings consistently demonstrated that about was the most frequent and versatile alternative to "over", followed by other prepositions like "of", "above", and "against", which appeared in specific contexts. Statistical analysis, specifically the Pearson correlation coefficient of 0.9939, revealed a strong linear relationship between the results of the two datasets, confirming the reliability of AI-generated data as a viable alternative to traditional corpus data.

The results indicate that AI-generated data can capture meaningful linguistic patterns in formal English, offering a modern approach to linguistic research. This method proved efficient, scalable, and consistent, making it a valuable tool for analyzing nuanced prepositional usage across different sample sizes.

The pedagogical implications of this research are particularly relevant for L2 learners and educators. Teaching prepositions like over can be challenging due to their polysemous nature, which often varies by context. This study provides several key takeaways for language instruction:

Focus on High-Frequency Alternatives: Given that about emerged as the most

frequent alternative to over, educators can focus on teaching this substitution first, simplifying the complexity of "over" for learners. By mastering more frequent and versatile alternatives, students can become more comfortable navigating complex prepositional choices.

Contextualized Learning: Since different prepositions like "of", "above", and "against" appeared in context-specific roles, teachers should emphasize the importance of contextual learning. By providing students with a variety of sentence examples and asking them to identify the appropriate prepositional substitute, learners will develop a deeper understanding of how prepositions function across contexts.

Data-Driven Instruction: The consistent trends in both datasets support the use of data-driven approaches to language learning. AI-generated sentences, much like the ones used in this study, can offer immediate and contextually relevant examples for learners. Teachers can utilize AI-generated data to provide a more dynamic and flexible learning experience, using a variety of prepositional usage patterns to cater to the specific needs of their students.

Addressing Prepositional Complexity: The irreplaceable and omittable cases highlight areas where learners may struggle with certain uses of "over" that cannot be easily substituted or omitted. By identifying and focusing on these complex cases, instructors can provide targeted support to help learners understand when "over" plays an indispensable role in meaning construction.

While this study presents promising findings, it has certain limitations. First, the data was limited to written, formal English, excluding spoken or informal contexts where prepositional usage might differ. Additionally, while efforts were made to enhance reliability by comparing two datasets, it must be acknowledged that using AI-generated sentences as a primary data source may also involve the risk of potential bias. Lastly, this research focused on a relatively small sample size, and future studies could benefit from expanding the dataset to enhance generalizability.

Future research could explore prepositional substitution patterns in spoken or informal English and examine the reliability of AI-generated data compared to traditional corpora. Additionally, expanding the dataset or incorporating a longitudinal analysis could provide deeper insights into the evolving use of prepositions in various forms of communication.

References

- Baron, N. S. (2008). *Always on: Language in an online and mobile world*. Oxford University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
- Brala, M. (2011). Pragmatic functions of the preposition "over" in spoken English. *Journal of Pragmatics*, 43(4), 1001-1017.
- Crystal, D. (2010). *The Cambridge encyclopedia of the English language* (2nd ed.). Cambridge University Press.
- Givón, T. (1993). *English grammar: A function-based introduction*. John Benjamins.
- Gries, S. T. (2006). Corpora and prepositional meaning: A corpus-based analysis of "over". *Cognitive Linguistics*, 17(3), 423-448.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press.
- Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). *Change in contemporary English: A grammatical study*. Cambridge University Press.
- Li, B. (2022). *Integrating linguistic theory and neural language models* (Doctoral dissertation, University of Toronto).
- Lindstromberg, S. (1998). *English prepositions explained*. John Benjamins.
- MALLQUI, S. *Data-driven learning and artificial intelligence: a comparison between two approaches to the creation of language teaching materials*.
- McCarthy, M., & Carter, R. (2004). *English grammar: A practical guide*. Cambridge University Press.
- Rühlemann, C. (2014). *Spoken corpora and applied linguistics*. Routledge.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge University Press.
- Simpson, P., & Mayr, A. (2010). *Language and power: A resource book for students*. Routledge.
- Stubbs, M. (2001). *Text and corpus analysis: Computer-assisted studies of language and culture*. Blackwell.
- Swan, M. (2016). *Practical English usage* (4th ed.). Oxford University Press.
- Thompson, G. (2013). *Introducing functional grammar* (3rd ed.). Routledge.
- Tyler, A., & Evans, V. (2003). *The semantics of English prepositions: Spatial scenes, embodied meaning and cognition*. Cambridge University Press.

Taejin Kim

Department of English Language and Literature, Graduate School

Incheon National University, 119 Academy-ro, Yeonsu-gu,

Incheon 22012, Republic of Korea

Phone: +82-32-835-8120

Email: taeginn@daum.net

Received on September 20, 2024

Revised version received on November 10, 2024

Accepted on December 31, 2024