



ISSN: 1994-4217 (Print) 2518-5586(online)

Journal of College of Education

Available online at: <https://eduj.uowasit.edu.iq>

Rafah Mazban Badan

Misan Education  
Directorate

Email:

[rafahitwayya@gmail.com](mailto:rafahitwayya@gmail.com)*Keywords:***Quantitative analysis,  
language diversity,  
complex sentence,  
Brown and LOB  
corpus, frequency,  
style****Article info***Article history:*

Received 29.Sep.2025

Accepted 16.Nov.2025

Published 10.Febr.2026

**Frequency Analysis of Complex Sentence Structure in English  
Written Texts: A Corpus-Based Study Using Brown and LOB  
Corpora****A B S T R A C T**

Complex sentences constitute a universal feature of human language, characterised by their intricate syntactic structures and diverse types of embedded clauses, thereby establishing them as a focal point in linguistic research. This study employs the Brown and LOB written English corpora to conduct a quantitative analysis of complex sentence frequency across four dimensions: overall usage, genre, language variety (American versus British English), and the types and numbers of subordinate clauses. Key findings reveal that genre significantly influences complex sentence frequency, with the highest usage observed in religious texts, literature, biographies, and essays—attributed to their formal or serious tone—whilst fiction exhibits the lowest frequency. In contrast, language variety demonstrates no statistically significant impact. Notably, both the type and number of subordinate clauses within complex sentences significantly affect their frequency. Contrary to prior studies, relative clauses (in both American and British English) do not rank as the least frequent, thereby challenging existing linguistic assumptions.

© 2026 EDUJ, College of Education for Human Science, Wasit University

DOI: <https://doi.org/10.31185/eduj.Vol62.Iss1.4737>

"التحليل التكراري للتركيب النحوية المعقدة في النصوص الإنجليزية المكتوبة:  
دراسة قائمة على المدونة اللغوية باستخدام مدونتي براون ولوب"

م.م. رفاة مزبان بادان  
مديرة تربية ميسان

### المستخلص

تُشكّل الجمل المركّبة سمّةً عالميةً للغة البشرية، حيث إنها تتميز بتركيباتها النحوية المعقدة وأنواعها المتنوعة من الجمل المضمّنة، مما يجعلها محورًا رئيسيًا في البحث اللغوي. وتستعمل هذه الدراسة مجموعات نصوص براون وLOB المكتوبة باللغة الإنجليزية لإجراء تحليل كمي لتكرار الجمل المركّبة عبر أربعة أبعاد: الاستخدام العام، والنوع اللغوي، وتنوع اللغة (الإنجليزية الأمريكية مقابل الإنجليزية البريطانية)، وأنواع الجمل الثانوية وأعدادها. تكشف النتائج الرئيسية أن النوع اللغوي يُؤثر بشكل كبير على تكرار الجمل المركّبة، حيث يُلاحظ أعلى استعمال في النصوص الدينية والأدب والسير الذاتية والمقالات - بسبب طابعها الرسمي أو الجاد - بينما يُظهر الخيال أقلّ تواتر. في المقابل، لا يُظهر تنوع اللغة أيّ تأثير ذي دلالة إحصائية. والجدير بالذكر أن نوع وعدد الجمل الثانوية في الجمل المركّبة يُؤثران بشكل كبير على تكرارها. وخلافًا للدراسات السابقة، لا تُصنّف الجمل النسبية (في كلٍّ من الإنجليزية الأمريكية والبريطانية) ضمن الأقلّ تواترًا، مما يُشكّل تحديًا لافتراضات اللغوية القائمة.

الكلمات المفتاحية: التحليل الكمي ، التنوع اللغوي ، الجملة المعقدة ، مجموعة براون وLOB ، التردد ، الأسلوب.

## 1. Introduction

Language is inherently hierarchical in nature, whilst written language manifests itself in linear structures. Accordingly, English sentences can be classified into three types based upon the complexity of their linear structures: simple sentences, compound sentences, and complex sentences. A simple sentence consists of one independent clause, whilst a compound sentence comprises two or more coordinate clauses connected by conjunctions. Unlike the first two types, a complex sentence constitutes a grammatical structure consisting of one independent main clause plus one or more subordinate clauses (Radford, 2023; Niklaus, 2022). Subordinate clauses represent the typical feature of complex sentences. Syntactic complexity relates to the compositional and hierarchical organisational properties of linguistic units (Kim, et al., 2021; Hirvela, 2022), such as clause embedding and subordination (Fan, et al., 2023). In summary, complex sentences are generally considered the most complex type of sentence.

Due to the complexity of syntactic structures and the diversity of embedded clause types, complex sentences have long constituted a research focus in academia. Nevertheless, systematic and holistic research on English complex sentences appears to be somewhat insufficient over the past 70 years.

Understanding complex sentences represents an important aspect of children's language development (Niklaus, et al., 2023). Children begin producing various complex sentences during the preschool period, but when using causal complex sentences, the order of causal

relationships is often incorrect (Song, et al., 2024), which may relate to children's limited working memory (Shah, 2024).

The usage frequency of English complex sentences can serve as a quantitative indicator for exploring significant differences between different genres. Based upon this, the difference in usage frequency of English complex sentences across different genres constitutes the second question this study aims to answer.

In terms of language communities, British English and American English represent the two largest English variants. Related research demonstrates that British English and American English share commonalities but also have many differences. These differences primarily involve grammar, vocabulary, spelling, punctuation, and idiom usage. Although British English sometimes adopts certain American English expressions, the two variants seem to maintain their respective characteristics to the maximum extent (Liu & Sun, 2022).

Genre represents an abstract, socially recognised way of using language in a written and identifiable type of written text, such as novels, news reports, and love stories (Yang & Kim, 2020); Yang & Geng, 2021). In summary, the difference in usage frequency of English complex sentences in written British English and American English constitutes the third question this study aims to answer. English complex sentences possess significant syntactic complexity, and sentences with complex syntactic structures bring about comprehension difficulties to some extent.

## **2. Related Literature**

Previous related research has primarily focused on children's language acquisition, the teaching of English complex sentences and syntactic complexity, and machine translation in natural language processing. Furthermore, another study testing 12 children's understanding of complex sentences demonstrated that even sixth-grade students have not yet reached adult-level comprehension of complex sentences (Lu, 2020), indicating that the complexity of English complex sentence structures indeed exists. Therefore, in machine translation, to address the processing difficulties caused by such structural complexity, scholars proposed text simplification techniques in the late 1990s, using easily understood vocabulary and simple syntactic structures to improve the efficiency of natural language processing. The shorter the sentence, the better the machine processing effect (Shah, 2024). For example, Open IE can split a complex sentence into several simple sentences centred around conjunctions (Liu, et al., 2025).

Similarly, researchers designed a novel sentence simplification method called Split-and-Rephrase, whilst some researchers have divided complex sentences into semantic layers containing several simple sentences in sequence, each layer displaying the most common syntactic structures, thereby facilitating machine translation and information extraction to achieve accurate semantic recognition (Yang & Geng, 2022). The basic approach and pathway of the aforementioned natural language processing methods all involve converting complex sentences and other long, difficult sentences into short sentences, thereby promoting the operation of machine translation (Park, 2021).

Wang et al., (2020) once studied the usage frequency of subordinate clauses in writing among 54 students from fourth grade, second intermediate grade, and sixth grade. The research results demonstrated that the older the students, the more inclined they were to incorporate subordinate clauses in their written texts (Yang & Geng, 2022). This may indicate that the older the students, the more inclined they are to use complex sentences until they reach adult-level language use. Regarding adults' use of English complex sentences, some researchers have proposed that simple sentences and compound sentences tend to be more common than complex sentences (Bi & Jiang, 2020).

Research indicates that British English and American English reached their "maximum divergence" before World War II, and since then, the two have become increasingly similar, perhaps both hoping to be better understood by each other's language users (Yang, W., & Kim, 2020).

### **3. Research Questions**

This study aims to answer these questions:

1. What is the exact usage frequency of complex sentences?
2. How do children over six years old learn and master English complex sentences?
3. In written language, are there significant differences in the usage frequency of English complex sentences across different genres?
4. Do the number and types of embedded clauses in English complex sentences in written language affect the usage frequency of complex sentences?

### **4. Objectives**

This study aims to:

- 1- To determine the exact usage frequency of complex sentences in English written language by analyzing combined data from the Brown and LOB corpora, establishing what proportion of sentences in written English are complex sentences relative to simple and compound sentences.
- 2- To examine genre-based variations in complex sentence usage by comparing the frequency of complex sentences across 15 different genres (such as religion, literature, academic writings, and fiction) in both American and British English written texts, thereby identifying whether genre type significantly influences complex sentence frequency.
- 3- To investigate potential differences in complex sentence usage frequency between language varieties by comparing American English (Brown Corpus) and British English (LOB Corpus) to determine whether language variant represents a significant factor affecting the proportion of complex sentences in written discourse.
- 4- To analyze the impact of clause types and numbers on complex sentence frequency by examining how different types of embedded clauses (relative clauses, noun clauses, and adverbial clauses) and the number of clauses within complex sentences affect their usage patterns in written English across both language variants.

## 5. Corpus and Methods

Entering the era of big data, the use of corpora often brings unexpected surprises to research. If researchers still rely solely or primarily on introspective methods, those surprises might be inadvertently overlooked (Yang & Kim, 2020). Data sample size and representativeness are equally important in research (Liu, & Wu, 2025; Yang, Y., & Geng, 2022). Based on this, Brown Corpus representing American English and the LOB Corpus representing British English were selected, both of which constitute adult English written language corpora.

In this study, for the delimiting criteria of English complex sentences, the definition proposed by (Quirk & Crystal, 2010) and widely accepted in academia was adopted, namely, sentence types that contain at least one main clause and embed one or more subordinate clauses as sentence components. There are two types of embedded clauses: one type serves as syntactic components of the main clause, such as subjects and objects, called complement or noun clauses; the other type modifies components of the main clause, namely adjective clauses or relative clauses, whilst clauses that modify the entire main clause or the verb phrase in the main clause are called adverbial clauses (Wang & Zhang 2020). Additionally, it should be particularly noted that in this study, sentences containing only non-finite verbs are not considered complex sentences.

According to the above criteria, fifteen English language teachers from government secondary schools, including 7 males and 8 females with various specializations in English grammar and English as a Second Language teaching, and with different academic degrees such as BA and MA in linguistics, were invited to conduct a grammatical analysis of sentences in the Brown and LOB sets using a structured explanation task requiring clear instructional guidance, the use of a coding guide with predefined criteria based on Quirk & Crystal, (2010) definitions, model sentences to calibrate comprehension, and a categorical coding based on spreadsheets (simple, compound, or complex; relative, nominal, or adverbial sentences) along with a quality assurance protocol that includes a one-third repetition to assess the reliability of the assessment among assessors, and three rounds of linguistic review by researchers to identify inconsistencies and ensure accuracy.

Considering reliability issues, one-third of these complex sentences would randomly reappear in the assigned tasks distributed to teachers. This study compared and identified the repeated sentences in the finally returned annotated complex sentences, ensuring that only the correct one was retained. After all annotation tasks were completed, this study randomly selected annotated sentences and conducted three rounds of proofreading to ensure a relatively low error rate. During the complex sentence annotation process, relative clauses, noun clauses, and adverbial clauses were sequentially marked as "X", "Y" and "Z". The calculation methods for clause density and embedding depth both referred to Karlsson's standards (Yang & Kim, 2020). The complex sentence annotation format is shown in Table 1.

Table 1: English Complex Sentence Explanation

No.	Clause Density	Embedding Depth	Clause Identification	English Complex Sentence
1	1	1	y	You see what I mean.
2	2	1	xy	All I know is that they are gone.
3	2	2	yy	You thought I did not mean what I said.
4	3	2	yyz	I replied that I hoped it would not, unless he ran the way of trouble.
5	4	3	zyzy	While I was drinking it, I wondered what Peter Rakosi would say when I told him I wanted to marry his daughter.
6	7	4	zyxyxz	While they were told that there were some normal people who reacted differently than they had, they were also informed that there were other normals who reacted as they had.

## 6. Analysis and Discussion

### 6.1 Total Usage Frequency of English Complex Sentences in Written Language

American English and British English constitute the two major variants of English. To obtain the total usage frequency of English complex sentences, the Brown Corpus and LOB Corpus were combined into one entity and the usage frequency differences among English complex sentences, simple sentences, and compound sentences were systematically analysed.

Statistical data demonstrate that the total usage frequency of sentences in the two corpora is 108,454 (including 55,945 sentences in the Brown Corpus and 52,509 sentences in the LOB Corpus), of which the usage frequency of English complex sentences is 42,655 (including 20,575 complex sentences in the Brown Corpus and 22,080 complex sentences in the LOB Corpus), significantly lower than the sum of usage frequencies of simple sentences and compound sentences, which is 65,799. The usage frequency proportion of English complex sentences is 39.33%. In other words, without considering the two variables of genre and language variant, in English written language, when producing 10 sentences, approximately 4 would be complex sentences. This value is consistent with the research conclusion that in natural language use, simple sentences and compound sentences are often more common than complex sentences (Yang & Kim, 2020; Song, et al., 2024).

### 6.2 Impact of Different Genres on English Complex Sentence Usage Frequency

In the exploration process of this section, to avoid the possible impact of different variants and make the research as detailed and objective as possible, the data from the Brown Corpus and LOB Corpus were separated or combined, examining from three perspectives: the American English dimension, the British English dimension, and the macro dimension.

#### 6.2.1 American English Dimension

When examining the usage frequency of English complex sentences from the American English dimension, it is necessary to consider that the text sizes of the 15 types in the Brown Corpus (see Table 2) are different, meaning each sub-corpus contains different numbers of sentences. To effectively compare differences in complex sentence usage frequency, the data

were standardised, converting complex sentence usage frequency into proportions of complex sentence usage frequency for comparison.

Table 2: Specific Classification of 15 Types

	English Type	Corresponding Type		English Type	Corresponding Type
<b>A</b>	Press: reportage	News reportage	<b>B</b>	Press: editorial	News editorial
<b>C</b>	Press: reviews	News reviews	<b>D</b>	Religion	Religion
<b>E</b>	Skills, trades and hobbies	Skills, trades and hobbies	<b>F</b>	Popular lore	Oral narrative literature
<b>G</b>	<u>Belles lettres</u> , biography, essays	Literary works, biography and essays			
<b>H</b>	Miscellaneous (government documents, foundation reports, industry reports, college catalogue, industry house organ)	Others (government documents, foundation reports, industry reports, college catalogues, industry house organs)	<b>J</b>	Learned and scientific writings	Academic and scientific writings
<b>K</b>	General fiction	General fiction	<b>L</b>	Mystery and detective fiction	Mystery and detective fiction
<b>M</b>	Science fiction	Science fiction	<b>N</b>	Adventure and Western fiction	Adventure and Western fiction
<b>P</b>	Romance and love story	Romance and love story	<b>R</b>	<u>Humour</u>	Humorous discourse

From Figure 1, it can be observed that type N (Adventure and Western fiction) has the smallest proportion of complex sentence usage frequency (20.18%), whilst genres H (government documents and other materials), K (general fiction), L (mystery and detective fiction), M (science fiction), and P (romance and love story) all have proportions within 40% (from 30.93% to 32.82%). Genres D (religion) and G (literary works, biography and essays) have the highest proportions of English complex sentence usage frequency, reaching 51.90% and 46.59% respectively and indicating that the difference in complex sentence usage frequency proportions between the two groups is extremely significant. These data all indicate that light, lively genres such as novels and romance stories differ significantly from serious, solemn genres such as religion and essays in the usage frequency of English complex sentences.



Figure-1: The frequency of use of complex sentences in different styles of English in the Brown Corpus

### 6.2.2 British English Dimension

When examining the usage frequency of English complex sentences from the British English dimension, it is similarly necessary to consider the text size differences among the 15 genres in the LOB Corpus. Complex sentence usage frequency was converted into proportions (Figure 2) for comparison.

From Figure 2, it can be observed that genre M (science fiction) has the smallest proportion of complex sentence usage frequency (30.70%), whilst genres L (mystery and detective fiction), M (science fiction), and N (adventure and Western fiction) all have proportions within 40% (from 30.70% to 35.65%). Genres K, L, M, N, P, and R can be divided into two major categories compared to the other 9 genres, with the former having frequency proportions all below 40%, whilst the latter are all above 40%. Meanwhile, genres J (academic and scientific writings) and G (literary works, biography and essays) have the highest proportions of English complex sentence usage frequency, reaching 49.13% and 50.20% respectively. The above frequency proportion situation is basically equivalent to the Brown Corpus, which also indicates that although these complex sentences belong to different language variants, there is little difference in the proportion of complex sentence usage frequency. These data similarly indicate that light, lively genres such as novels and humorous discourse differ significantly from serious, solemn genres such as academic writings and biographical essays in the proportion of English complex sentence usage frequency.

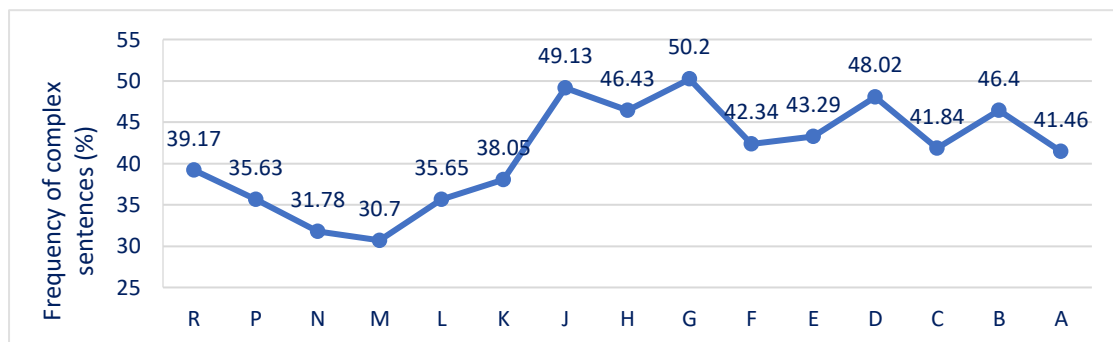


Figure -2: The frequency ratio of English complex sentences in different styles in the LOB

Comparing the proportions of complex sentence usage frequency between the Brown Corpus and LOB Corpus reveals that the two are almost overlapping, indicating that regardless of which language variant, the proportion of English complex sentence usage frequency is basically constant, which also proves the universality of language from another perspective.

From Figure 3, it can be observed that the overall change trends of the two lines are basically consistent. However, there is some fluctuation in genre N (adventure and Western fiction), which may be caused by the difference in total sentence numbers in genre N (adventure and Western fiction) in the Brown Corpus.

Data analysis of Figure 3 demonstrates that regardless of text type, except for genre N (adventure and Western fiction) in the LOB Corpus, the usage frequency of complex

sentences always accounts for more than 30% of the total sentences. The highest proportion of complex sentence usage frequency reaches 51.90%, found in genre D (religion). These results may reveal a threshold range (20.18% to 51.90%), namely the fluctuation range of complex sentence usage frequency in English written language. It is noted that the difference between 20.18% and 51.90% is quite large (over 30%), which may indicate that genre type significantly affects the proportion of English complex sentence usage frequency.

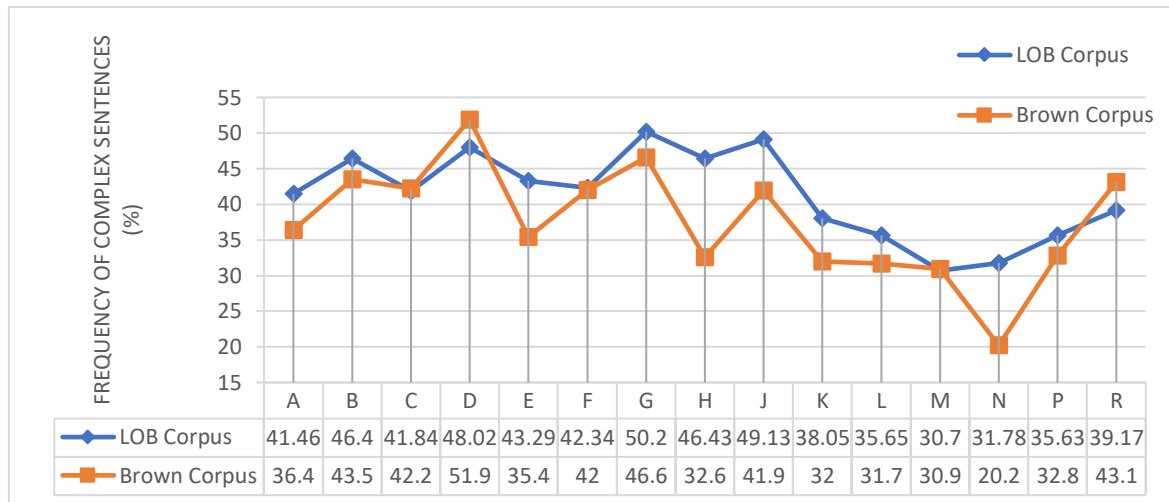


Figure -3: Comparison of the frequency of use of complex sentences in different styles of English in the Brown and LOB corpora

### 6.2.3 Macro Dimension

Combining the relevant data on complex sentence usage frequency proportions from the Brown Corpus and LOB Corpus, the frequency proportion chart for English complex sentences was obtained, as shown in Figure 4.

From Figure 4, it can be observed that genre N (adventure and Western fiction) has the smallest proportion of complex sentence usage frequency (25.98%), whilst genres D (religion) and G (literary works, biography and essays) have the highest proportions of English complex sentence usage frequency, reaching 49.96% and 48.40% respectively.

This indicates that complex sentences are often used to express more complex information, but understanding these sentences is more difficult for readers or listeners. In this sense, writers are very likely to use complex sentences as little as possible in their works to make it easier for readers to understand the stories they narrate. Furthermore, the proportion of complex sentence usage frequency may constitute a quantitative classification indicator for fiction and non-fiction.

In summary, this study found that the proportion of English complex sentence usage frequency is higher in genres D (religion), G (literary works, biography and essays), and J (academic and scientific writings), which is closely related to the seriousness or formality of the genre.

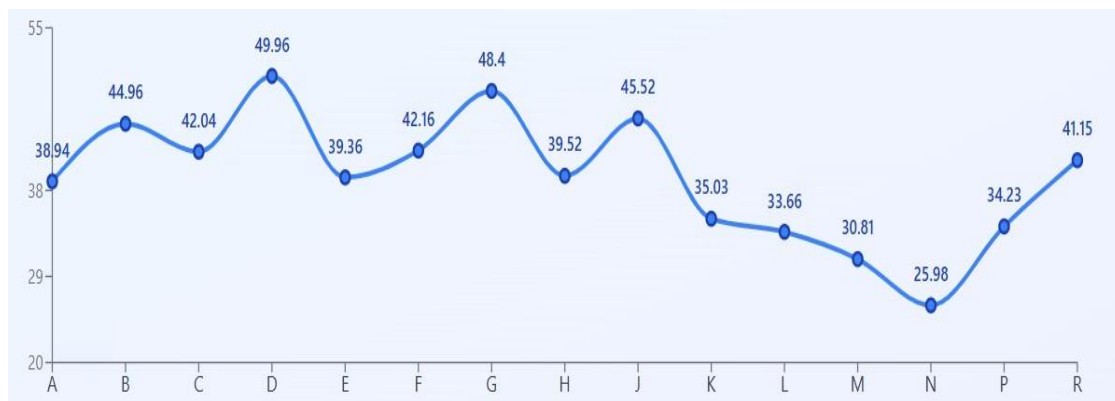


Figure -4: The frequency of use of complex sentences in different styles of English.

### 6.3 Impact of Clause Types and Numbers on English Complex Sentence Usage Frequency

Previous research has found that English complex sentences used by preschool children contain noun clauses, adverbial clauses, and relative clauses. From the perspective of usage frequency of different types of clauses, the number of complex sentences containing relative clauses is the least, accounting for 15% (Liu & Sun, 2022). Some researchers, such as Yang, et al., (2022), used the Brown Corpus to examine the diachronic usage changes of English relative clauses in British English and American English variants. The results demonstrated that whether in American English or British English, the usage frequency of relative clauses in written texts showed a sharp declining trend. An empirical study of 112 non-native English-speaking university students and 43 English native-speaking university students demonstrated that non-native English-speaking university students used English relative clauses far less frequently than English native-speaking university students (Yang, Y., et al., (2022).

Relatively speaking, there is more research on the usage frequency of English relative clauses, whilst research on the usage frequency of noun clauses and adverbial clauses is relatively scarce. Based upon this, the usage frequency of complex sentences containing different types of clauses in both Brown and LOB corpora was specifically examined to have a more comprehensive understanding of the usage of various types of clauses. In the Brown Corpus, the proportion of complex sentences containing 1 to 2 clauses (18,544) in the total complex sentence usage frequency (20,575) is 90.13%, whilst in the LOB Corpus, the proportion of complex sentences containing 1 to 2 clauses (19,683) in the total complex sentence usage frequency (22,080) is 89.14%. Both proportions are close and relatively high. Therefore, the usage frequency of complex sentences containing only 1 to 2 clauses (including both cases where clause types are the same or different) was next examined.

#### 6.3.1 Usage Frequency of Single-Clause Complex Sentences

Complex sentences containing only 1 clause were subjected to statistical analysis (see Figure 5), and it was found that the usage frequencies of relative clauses and adverbial clauses are basically consistent in both corpora, whilst the usage frequency of noun clauses demonstrates a large difference, being 862 sentences more in British English (5,086) than in

American English (4,224). Overall, whether in American English or British English, the usage frequency of relative clauses is not the lowest among the three types of clauses.

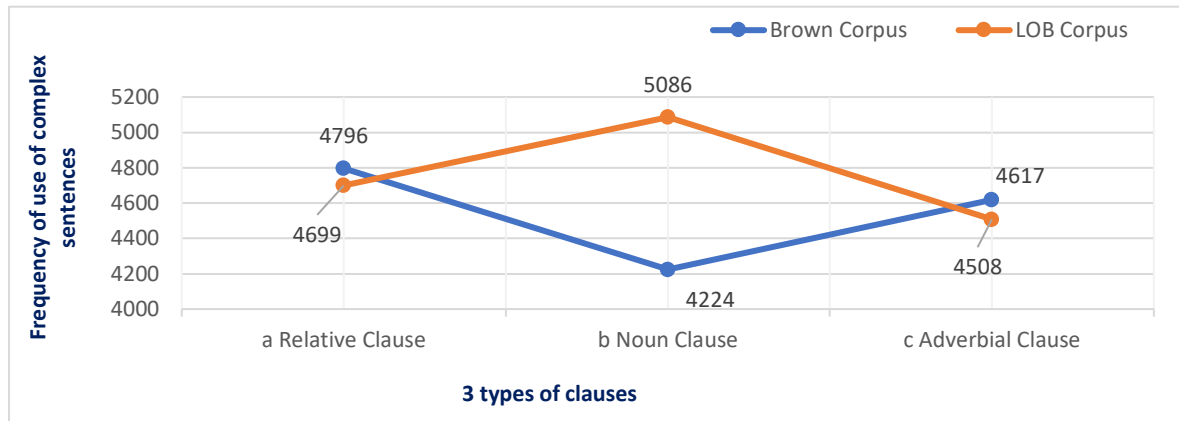


Figure -5: Comparison of the frequency of use of simple and complex clauses in British English and American English

6.3.2 Usage Frequency of Double-Clause Complex Sentences

Statistics were conducted on complex sentences containing only 2 clauses, where "11" represents that the clause combination in the complex sentence consists of two relative clauses, "22" represents that the clause combination consists of two noun clauses, "33" represents that the clause combination consists of two adverbial clauses; whilst "12" represents that the clause combination consists of one relative clause and one noun clause with the relative clause preceding the noun clause, and the meanings of "13" and "23" follow accordingly. For example: "He was copious in his praise of those that served food that was good to eat." (11); "Bang-Jensen never explained what the documents or papers were that he had in his possession." (22); "If they have trouble exporting, international bill for their support will grow larger than it otherwise would." (33).

Figure 6 demonstrates that in both corpora, the change trends of complex sentence usage frequency containing double clauses are basically consistent, but the complex sentence usage frequency in American English is slightly lower than in British English.

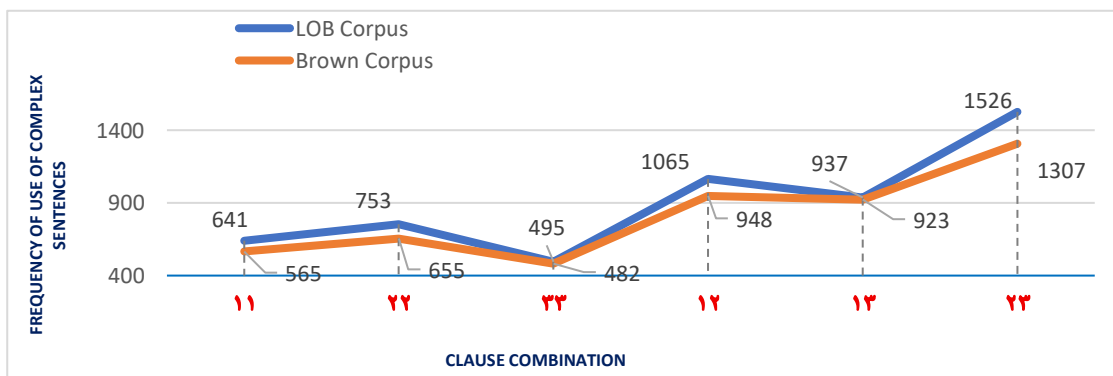


Figure -6: Comparison of the frequency of use of double-clause complex sentences in British English and American English

Among complex sentences containing double clauses, the complex sentence usage frequency when double clauses are of the same type (e.g., 11) is lower than when double clauses are of different types (e.g., 12). Among them, the complex sentence usage frequency of double adverbial clauses is the lowest, with 495 and 482 complex sentences in the Brown Corpus and LOB Corpus respectively (see Figure 6). In both American English and British English, the complex sentence usage frequency of noun clauses co-occurring with adverbial clauses is the highest.

## 7. Conclusion

1- This paper, using the Brown Corpus and LOB Corpus, conducted a quantitative investigation of English complex sentence usage frequency from four dimensions: total complex sentence usage frequency, genre, language variants, and the types and numbers of clauses contained in complex sentences. In human language use, complex sentences constitute an indispensable linguistic means for expressing more complete and complex information. However, influenced by various factors such as sentence length and types of embedded clauses, complex sentence usage is not entirely arbitrary.

2- Data analysis indicates that British English and American English demonstrate no significant difference in complex sentence usage frequency. However, differences in genre types significantly affect complex sentence usage frequency. The proportion of complex sentence usage frequency is highest in religion, literature, biography and essays, and academic and scientific writings, whilst it is lowest in the four types of fictional texts. This linguistic phenomenon relates to the seriousness or formality of texts. Other variables are also worth further research.

3- The types and numbers of clauses contained in complex sentences also affect the usage frequency of English complex sentences. Among complex sentences containing single clauses, the usage frequencies of relative clauses and adverbial clauses demonstrate little difference between the two corpora, whilst the usage frequency of noun clauses demonstrates a large difference. Whether in American English or British English, the usage frequency of relative clauses is not the lowest among the three types of clauses.

4- In complex sentences containing double clauses, the complex sentence usage frequency when double clauses are of the same type is lower than when they are of different types, with double adverbial clauses having the lowest complex sentence usage frequency and noun clauses co-occurring with adverbial clauses having the highest complex sentence usage frequency.

Finally, these tests are reliable among evaluators through sentence repetition for one-third of the sentence, three rounds of linguistic review by researchers to identify errors, and the application of predefined criteria (defined by Quirk et al., 2010) that minimize subjectivity. The consistency demonstrated by 15 English language teachers from diverse geographical regions, each with expertise in grammar, language teaching, and linguistic analysis, reinforces confidence that the identified quantitative patterns reflect the actual characteristics of written English, rather than being mere methodological constructs. This rigorous approach to explanation and verification provides a solid empirical foundation for the study's contributions to linguistic research and challenges previous assumptions about the use of complex sentences.

## Reference

- Bi, P., & Jiang, J. (2020). Syntactic complexity in assessing young adolescent EFL learners' writings: Syntactic elaboration and diversity. *System*, 91, 102248. Retrieved from <https://doi.org/10.1016/j.system.2020.102248>
- Fan, Y., Li, B., Sataer, Y., Gao, M., Shi, C., Cao, S., & Gao, Z. (2023). Hierarchical clause annotation: building a clause-level corpus for semantic parsing with complex sentences. *Applied Sciences*, 13(16), 9412. Retrieved from <https://doi.org/10.3390/app13169412>
- Hirvela, A., (2022). *Connecting reading & writing in second language writing instruction*. University of Michigan Press. Retrieved from <https://2cm.es/1fL3L>
- Kim, J., Maddela, M., Kriz, R., Xu, W., & Callison-Burch, C. (2021). BiSECT: Learning to split and rephrase sentences with bitexts. *arXiv preprint arXiv:2109.05006*. Retrieved from <https://doi.org/10.48550/arXiv.2109.05006>
- Liu, J., Yang, N., & Wu, M. (2025). A Quantitative and Comparative Study of Syntactic Complexity of Subclasses of English Nominal Clauses. *Journal of Quantitative Linguistics*, 1-34. Retrieved from <https://doi.org/10.1080/09296174.2025.2498797>
- Liu, H., & Sun, M. (2022). A review of syntactic complexity in second language writing (2010-2020). *Foreign Studies*, 10(1), 28-32. Retrieved from <https://doi.org/10.1016/j.jslw.2015.06.008>
- Lu, T. (2020, August). *Analysis on linguistics research directions in the age of big data*. Vol. 1606, No. 1, p. 012008). IOP Publishing. Retrieved from [DOI 10.1088/1742-6596/1606/1/012008](https://doi.org/10.1088/1742-6596/1606/1/012008)
- Niklaus, C. (2022). *From complex sentences to a formal semantic representation using syntactic text simplification and open information extraction*. Springer Nature. Retrieved from <https://2cm.es/1fL9Q>
- Niklaus, C., Cetto, M., Freitas, A., & Handschuh, S. (2023). Discourse-aware text simplification: From complex sentences to linked propositions. *arXiv preprint arXiv:2308.00425*. Retrieved from <https://doi.org/10.48550/arXiv.2308.00425>
- Park, J. H. (2021). *Syntactic Complexity as a Predictor of Second Language Writing Proficiency and Writing Quality*. ProQuest LLC. 789 East Eisenhower Parkway, PO Box 1346, Ann Arbor, MI 48106. Retrieved From <https://2cm.es/1fLbR>
- Quirk, R., & Crystal, D. (2010). *A comprehensive grammar of the English language*. Pearson Education India.
- Radford, A., (2023). *Analysing English sentence structure: An intermediate course in syntax*. Cambridge University Press. Retrieved from <https://2cm.es/1fLcv>
- Shah, S. R. A. (2024). Investigating Syntactic Complexity in Qatari EFL Learners' Essays: A Mixed Methods Study. *Pakistan Languages and Humanities Review*, 8(3), 515-532. Retrieved from [https://doi.org/10.47205/plhr.2024\(8-III\)47](https://doi.org/10.47205/plhr.2024(8-III)47)
- Song, J., Yang, C., Sun, Y., Qu, Y., Ma, K., & Cai, H. (2024). Quantitative Research on Chinese Sentences Structure Based on Pattern Grammar. *SAGE Open*, 14(3), 21582440241268641. Retrieved from <https://doi.org/10.1177/21582440241268641>
- Wang, L., Wu, H., & Zhang, J. L. (2020). Effects of task complexity on linguistic complexity in EFL writing. *Modern Foreign Languages*, 43(4), 503-515. Retrieved from <https://doi.org/10.3390/su14084791>

- Yang, Y., & Geng, H. (2022). A Quantitative Comparison of Syntactic Complexity Between British and American University Students' ENL Writing. In *Proceedings of the 4th International Conference of Languages, Education and Tourism (ICLET) 2021* (pp. 29-43). Retrieved from <https://2cm.es/1fLgU>
- Yang, Y., Yap, N. T., & Ali, A. M. (2022). A review of syntactic complexity Studies in context of EFL/ESL writing. *International Journal of Academic Research in Business and Social Sciences*, 12(10), 441-454. Retrieved from : <http://dx.doi.org/10.6007/IJARBSS/v12-i10/14843>
- Yang, W., & Kim, Y. (2020). The effect of topic familiarity on the complexity, accuracy, and fluency of second language writing. *Applied Linguistics Review*, 11(1), 79-108. Retrieved from <https://www.degruyterbrill.com/document/doi/10.1515/applirev-2017-0017/html>