

A Corpus-based Study on Usage Characteristics of “But” as a Discourse Marker in Chinese English Learners’ Writing

Linfang Zhang

School of Foreign Languages, Shandong University of Science and Technology, Qingdao, 266590, Shandong, China

Abstract: In English writing, discourse markers play a crucial role in the construction of texts that are coherent, clear, and logically structured. Based on the corpus-based research method, this study explores characteristics of the use of the discourse marker “but” in Chinese English learners’ writing from four aspects: usage frequency, syntactic position, collocation pattern, and pragmatic function. The research finds that Chinese English learners use “but” significantly more frequently in their writing than native English speakers. Chinese English learners tend to use “but” most often at the beginning of sentences, whereas native English speakers prefer to use it in the middle of sentences. Chinese English learners also exhibit certain limitations in using collocations of “but”, which is less rich and diverse compared to native English speakers. In terms of pragmatic functions, Chinese English learners are inclined to overuse “but” for stance expression but underuse it in other pragmatic functions. This study also proposes some targeted English teaching suggestions and attempts to provide valuable references and insights for corpus-based discourse marker studies and related English teaching research.

Keywords: Discourse marker; “But”; Chinese English learners; English writing; Corpus

1. Introduction

In English writing, constructing coherent and logically clear discourse is crucial, and discourse markers play an indispensable role in this process. Discourse markers, also known as discourse connectives or pragmatic markers, refer to a common category of words or phrases in daily language use, such as “but” “so” “you know” and “I mean” in English. They are frequently used in communication to guide and constrain the listener’s understanding of the discourse, playing a key role in cohesion and coherence. “But” as a fundamental and frequently used discourse marker in English, is primarily used to express logical relationships such as contrast and comparison, which can clearly demonstrate semantic changes between sentences or paragraphs and make the author’s intent more explicit. Its usage can largely reflect the language proficiency and pragmatic competence of learners. For Chinese English learners, in-depth research on their usage characteristics of “but” in writing holds significant practical importance.

This study aims to explore the usage characteristics of the discourse marker “but” in writing by Chinese English learners from four aspects: frequency of use, syntactic position, collocation pattern and pragmatic function, which attempts to provide effective suggestions for English learning and teaching and provide useful references for

Received: April 16, 2026

Revised: April 28, 2026

Accepted: May 11, 2026

Published: May 13, 2026

Copyright: © 2026 by the authors. Licensee Axon Academic Publishing Institute, Hong Kong, China. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

corpus-based research on discourse markers.

2. Literature Review

Discourse markers constitute a pivotal area of inquiry within linguistic research, serving as essential linguistic devices employed by speakers to guide and modulate listeners' comprehension of discourse (He & Ran, 1999)[1]. These markers, encompassing words, phrases, or structures, fulfill specific pragmatic functions, facilitating the smooth flow and coherence of communication. The study of discourse markers has a rich history, with foreign research initiating earlier and yielding abundant findings. Foreign research on discourse markers, particularly focusing on "but" began with an emphasis on its semantic and pragmatic functions. Schiffrin (1987)[2], in her seminal work *Discourse Markers*, conducted a comprehensive qualitative and quantitative analysis of eleven discourse markers, including "but". She noted that "but" primarily serves a contrastive semantic function, signaling a shift in the speaker's turn-taking and maintaining discourse continuity. The information following "but" typically contrasts with the preceding information, thereby introducing a new perspective or counterpoint. Lakoff (1971)[3] further refined this understanding by defining "but" as a contrastive discourse marker, suggesting it represents a comparative relationship between the preceding and following elements. This definition implies the speaker's anticipation of a negative response or counterargument from the listener, thereby setting the stage for a dialectical exchange. Subsequent research expanded the scope of inquiry to include the syntactic characteristics and register-based variations of "but". Biber et al.[4] analyzed multiple corpora to investigate the frequency and syntactic positions of "but" across different registers, revealing notable differences between spoken and written language. Their findings indicated that "but" is more frequently employed in spoken language, where it exhibits greater syntactic flexibility. This observation underscores the dynamic nature of discourse markers in oral communication, where they play a crucial role in managing turn-taking and maintaining conversational flow.

In contrast to the early foreign research, domestic studies on discourse markers initially focused on introducing and presenting foreign theories, as well as examining Chinese discourse markers. However, with the growing emphasis on English teaching in China, scholars began to shift their attention to the use of discourse markers by Chinese English learners. Zhao (2018)[5], for instance, investigated the use of discourse markers in the spoken English of Chinese learners, identifying existing problems and proposing pedagogical suggestions for improving spoken English proficiency. This study highlighted the importance of discourse markers in oral communication and underscored the need for targeted instruction to enhance learners' pragmatic competence. More recently, domestic scholars have employed corpus research methods to compare and analyze the differences in the use of discourse markers between Chinese English learners and native English speakers. Chen (2021)[6], based on corpus data, compared the use of discourse markers in written language between Chinese English majors and native speakers, revealing that Chinese learners tend to overuse and misuse "but" in their writing. This finding suggests that Chinese learners may lack a nuanced understanding of the pragmatic functions and syntactic constraints of "but" leading to inappropriate or excessive use. Other scholars have also explored "but" from various

perspectives, including semantics, pragmatics, and cognition, further enriching the theoretical landscape of discourse marker research.

Despite the significant contributions of previous research, several limitations persist, particularly concerning the usage characteristics of “but” in the writing of Chinese English learners. Firstly, existing studies predominantly focus on the analysis of “but” in terms of discourse coherence and pragmatic functions, often neglecting its syntactic and collocational aspects. This narrow focus limits our understanding of the multifaceted nature of “but” as a discourse marker. Secondly, related research tends to emphasize dialogue in corpus selection, with few in-depth systematic investigations into the use of “but” in English writing. This bias towards oral communication may stem from the assumption that discourse markers are more salient and frequent in spoken language. However, writing, as a distinct mode of communication, presents unique challenges and opportunities for the use of discourse markers. Therefore, the exploration of written language is essential for gaining a holistic understanding of “but” usage.

In light of the aforementioned limitations, this study attempts to conduct a comprehensive corpus-based analysis of the usage characteristics of “but” in the writing of Chinese English learners. This study innovatively explores the usage features of “but” from four dimensions: frequency of use, syntactic positions, collocation patterns, and pragmatic functions. By adopting a corpus-based approach, this study aims to provide a nuanced and in-depth understanding of how Chinese learners use “but” in their writing, thereby contributing to the field of discourse marker research and provide practical implications for English teaching and learning.

3. Research Method

This study selected the Chinese Learner English Corpus (CLEC) as the research corpus and the Louvain Corpus of Native English Essays (LOCNESS) as the reference corpus. CLEC contains approximately 1.07 million words of written language from Chinese learners, covering essay genres such as argumentative, expository, and narrative writings. LOCNESS comprises about 320,000 words of written language from native English speakers, consisting of three categories: advanced-level essays from British high school students, course essays from British university students, and course essays from American university students. The topics of the writings in reference corpus are diverse and its style is formal. Therefore, this reference corpus is representative to some extent.

Firstly, for frequency analysis, the total number of occurrences of “but” in CLEC and LOCNESS are obtained by the wordlist function of WordSmith Tools 8.0, which provides a clear picture of its overall frequency. At the same time, the number of occurrences of “but” in five subcorpora of CLEC are also retrieved. Regarding the analysis of syntactic positions, The concord function of WordSmith Tools 8.0 was employed to search for the word “but” in both two corpora and retrieve the surrounding context for each occurrence. The classification of syntactic positions are identified as occurring at the beginning, in the middle, and at the end of a sentence. To ensure the accuracy and reliability of this classification, a manual verification process has been carried out. For collocation pattern extraction, WordSmith Tools 8.0’s collocate function is utilized, which is set to identify words that frequently co-occurred with “but” within a

specified span. The collocates are then ranked based on their frequency of co-occurrence. In terms of pragmatic function annotation, a set of predefined pragmatic function categories is established based on previous research. Each instance of “but” in the corpora is then manually annotated according to these categories. To ensure reliability, the manual verification process has also been carried out. By compiling statistical data obtained through these procedures and combining it with the corpus data, this research explores the usage characteristics of “but”, analyzes its causes, and finally provides meaningful insights for relevant English teaching.

4. Analysis and Discussion

4.1 Frequency Analysis of “But”

Table 1 shows the frequency and normalized frequency per thousand words (ptw) of “but” in CLEC and LOCNESS. The study indicates that the normalized frequency of “but” in the writing of Chinese English learners is significantly higher than that of native English speakers by the chi-square test ($X^2=1524.28$, $p<0.001$). To further analyze the frequency of “but” usage among Chinese English learners at different learning stages, this study conducts separate statistical analyses on the frequency of “but” in the five writing sub-corpora from CLEC. The results are presented in Table 2.

Table 1 Frequency of “but” used in CLEC and LOCNESS

	CLEC	LOCNESS
Frequency of “but”	5653	691
Normalized Frequency of “but” (ptw)	5.28	2.13

Through analysis of variance, it is found that there are significant differences in the normalized frequency of “but” usage among Chinese English learners at different learning stages ($F=8.35$, $p<0.001$). Specifically, for non-English majors, the normalized frequency of “but” usage shows an upward trend as the learning stage progressed. During the College English Test Band 4 and Band 6 stages, the normalized frequency of “but” usage was relatively high, which may be because learners at these stages begin to pay more attention to expressing logical relationships in English writing, but their proficiency in using discourse markers like “but” is not yet sufficient, and their diversity of usage remained limited, which leads to the overuse of “but”. For English majors, the normalized frequency of “but” usage increases from high school to the first and second years of the major, but decreases in the third and fourth years as learners’ language skills gradually improved. They gain more knowledge and proficiency in other lexical expressions for adversative relationships, such as “nevertheless” “whereas” and “conversely”.

Table 2 Frequency of “but” by Chinese English learners at different learning stages

	Frequency	Normalized Frequency (ptw)
High School (ST2)	595	5.93
College English Test Band 4 (ST3)	816	6.80
College English Test Band 6 (ST4)	1296	7.20
The first and second years of undergraduate English majors (ST5)	915	6.10
The third and fourth years of undergraduate English majors (ST6)	1010	5.05

4.2 Analysis of Syntactic Positions of “But”

Discourse markers serve functions of local coherence and overall coherence (Li, 2003)[7]. In English writing, the syntactic position distribution of “but” has a significant impact on the semantic expression of sentences and discourse coherence. By retrieving data from CLEC and LOCNESS respectively, it is found that the syntactic positions of “but” mainly include the beginning, middle, and end of sentences. Moreover, there are certain differences in the syntactic position distribution of “but” between Chinese English learners and native English speakers, as shown in Table 3.

Table 3 Distribution of the syntactic positions of “but” in CLEC and LOCNESS

		At the beginning of the sentence	In the middle of the sentence	At the end of the sentence
CLEC	Frequency	2938	2430	285
	Normalized Frequency (ptw)	2.74	2.27	0.27
	Percentage	52.0%	43.0%	5.0%
LOCNESS	Frequency	193	449	49
	Normalized Frequency (ptw)	0.59	1.38	0.15
	Percentage	27.9%	65.0%	7.1%

This study finds that Chinese English learners and native English speakers show significant differences in the use of “but” at the beginning and middle of sentences ($p=0.001<0.05$, $p=0.002<0.05$), but no significant difference in the use of “but” at the end of sentences ($p=0.12>0.05$). Native English speakers tend to use “but” less frequently at the beginning of sentences and prefer to use it in the middle, possibly because they are more inclined to achieve semantic transition and cohesion at the beginning through other means, such as using words like “however” or “nevertheless” or by altering sentence structure to express the contrast. Additionally, in academic writing, words like “however” and “nevertheless” are more formal than “but” when placed at the beginning of a sentence, which is why native English speakers use “but” less frequently at the beginning of their writing. In contrast, Chinese English learners have the highest proportion of using “but” at the beginning of sentences, which may be related to

negative transfer from their native language. Chinese English learners may not have a deep understanding of the function of “but” in different syntactic positions, possibly believing that using “but” at the beginning of a sentence can more directly express contrast. In Chinese expressions, “但是” is usually placed at the beginning, and this negative transfer from their native language influences their choice of syntactic position for “but” in English writing.

Additionally, “but” in the middle of a sentence indicates semantic connections between adjacent clauses, serving the function of local coherence. When “but” is at the beginning of a sentence, it often signifies the start of discourse and constructs the overall framework of the statement, serving the function of overall coherence. English, originating from low-context cultures, places greater emphasis on local coherence, hence the more frequent use of “but” in the middle of sentences. In contrast, Chinese, derived from high-context cultures, prioritizes overall coherence in meaning (Lian, 1993)[8]. Influenced by Chinese culture, Chinese learners tend to prefer using “but” at the beginning of sentences.

Example 1: But, I think we should go to the park tomorrow. (CLEC)

Example 2: The weather was supposed to be sunny, but it started to rain in the afternoon. (LOCNESS)

Example 1 is a specific example from the CLEC corpus to further analyze the potential issues with Chinese learners’ sentence-initial use of “but”. The use of “but” at the beginning of the sentence seems rather abrupt and unnecessary. There is no clear contrast established with a previous statement in this single sentence context. This could lead to semantic ambiguity as the reader may struggle to understand what the “but” is contrasting with. Besides, it fails to create a smooth connection with the surrounding text. If this sentence were part of a larger passage, the lack of a proper preceding context for the “but” would disrupt the overall flow and coherence of the discourse. This indicates that Chinese learners’ over-reliance on the sentence-initial position of “but” may sometimes result in inappropriate textual construction. For comparison, in example 2, “but” is placed in the middle of the sentence, clearly connecting the two clauses “The weather was...” and “it started to...”. It effectively shows the contrast between the expected and the actual situation, creating a smooth local coherence within the sentence. This usage is in line with the typical pattern of native English speakers, who are more adept at using “but” in the middle of sentences to achieve precise semantic transitions and maintain proper textual cohesion.

4.3 Analysis of Collocation Patterns of “But”

Through searching for “but”-related collocations in CLEC and LOCNESS, it is found that Chinese English learners exhibit certain regularities and characteristics in their collocation patterns of “but” in writing. The main types of collocations include those indicating contrastive, adversative, and exclusive relationships, with the results shown in Table 4.

Table 4 Distribution of collocation patterns of “but” in CLEC and LOCNESS

Collocation Patterns	CLEC		LOCNESS	
	Frequency	Normalized Frequency (ptw)	Frequency	Normalized Frequency (ptw)
but not	1126	1.05	98	0.30
not only...but also...	872	0.81	32	0.10
but on the other hand	58	0.05	19	0.06
but(used alone)	2420	2.26	328	1.01
but rather	42	0.04	28	0.09
but yet	18	0.02	16	0.05
but actually	105	0.10	9	0.03
all but	12	0.01	23	0.07
anything but	8	0.01	27	0.08
everyone but	5	0.005	14	0.04
do nothing but	32	0.03	11	0.03

In contrastive collocations, common ones include “but not”, “not only...but also...”, and “but on the other hand”. For adversative relationships, “but” used alone is the most basic form, with additional types such as “but rather”, “but yet”, and “but actually”. Collocations expressing exclusion include “all but”, “anything but”, “everyone but”, and “do nothing but”.

According to Table 4, the study figures out that among the 11 collocation patterns listed, the normalized frequency of Chinese English learners differs significantly from that of native English speakers, with only “but not” and “but actually” are overused, while the rest are underused. Chinese English learners exhibit certain limitations in collocation usage, whereas native English speakers have a richer and more diverse range of collocation choices and can flexibly employ various collocations according to context and communicative needs. When expressing contrastive relationships, native English speakers, in addition to commonly used collocations, also use some more distinctive expressions such as “yet and still”, which are relatively less used by Chinese English learners. In some complex contexts, native English speakers can skillfully use collocations to convey subtle semantic differences accurately, whereas Chinese English learners may struggle to express themselves naturally and accurately due to poor command of “but”-related collocations.

4.4 Analysis of Pragmatic Functions of “But”

Multifunctionality is a hallmark of discourse markers. Discourse markers serve functions such as stating facts, expressing stance (Wang & Tsai, 2007)[9], conveying beliefs, making suggestions, shifting topics, reinforcing propositions, and supplementing information (Müller, 2005)[10]. This study statistically analyzed the pragmatic functions of “but” in CLEC and LOCNESS, with the results presented in Table 5.

Table 5 Distribution of Pragmatic Functions of “but” in CLEC and LOCNESS

Pragmatic Functions	CLEC		LOCNESS	
	Frequency	Normalized Frequency (ptw)	Frequency	Normalized Frequency (ptw)
Expressing stance	2208	2.06	101	0.31
Making suggestions	142	0.13	58	0.18
Stating facts	1865	1.74	263	0.81
Conveying beliefs	123	0.11	97	0.30
Shifting topics	47	0.04	63	0.19
Strengthening propositions	623	0.58	132	0.41
Supplementing information	495	0.46	215	0.66
Other Functions	150	0.14	62	0.19

The research shows that Chinese English learners and native English speakers exhibit significant differences in the use of “but” to convey various pragmatic functions such as expressing stance, making suggestions, stating facts, and expressing beliefs. In addition to the significant overuse of Chinese English learners in the pragmatic function of expressing stance, there is a marked underuse in other pragmatic functions. In terms of the function of expressing stance, as seen in Example 1, Chinese learners mechanically use “but” to connect opposing viewpoints, forming a templated structure of “opponent+but+supporter”. This phenomenon can be deeply attributed to the instructional contexts and textbook influences. In many Chinese English textbooks, contrastive sentence patterns are often emphasized as a key language point. For instance, exercises frequently require students to create sentences that present two opposing ideas connected by “but”. This repetitive training in a controlled environment leads students to internalize this structure as a standard way of expressing contrast. Moreover, the literal translation transfer of the native language word “但是” also plays a role. Since “但是” is commonly used in Chinese to connect contrasting statements, learners tend to directly apply this pattern when using “but” in English, without fully understanding the nuances and flexibility of “but” in different English contexts.

Example 1 : Some claim technology isolates people, but I insist it connects us.
(CLEC)

Example 2: These findings support the theory. But shifting the lens: what if cultural variables intervene? (LOCNESS)

Example 2 is an example of the function of shifting topics whose usage rate among Chinese English learners is extremely low. In Example 2, native English speakers naturally transition to a new perspective using structures like “but + topic marker” while compositions in the Chinese learner English corpus generally lack such strategies, resulting in stiff paragraph transitions in English writing. This difference may stem from learners’ limited

understanding of “but” being confined to semantic opposition, while neglecting its function in discourse organization. Native English speakers, however, utilize compound structures such as “but to reframe” or “but consider” to achieve natural topic shifts. Textbooks may also lack sufficient examples that show the diverse pragmatic functions of “but”. They mainly focus on the basic semantic contrast function, neglecting to introduce how “but” can be used for topic shifting, making suggestions, or stating facts in a more natural and context-appropriate way. As a result, Chinese English learners have a limited understanding of the full range of “but”’s functions, leading to the overuse in expressing stance and underuse in other areas.

Besides, the impact of developmental features of learners’ interlanguage can not be neglected. At the initial stages of English learning, Chinese learners’ interlanguage is mainly influenced by their native language. They tend to rely on the simple and direct contrast function of “but” that is similar to “但是” in Chinese. Their interlanguage development may get stuck in a stage where they are only comfortable with the basic usage of “but” and are reluctant to explore its other functions. The lack of exposure to authentic English materials in the learning process also hinders the development of their interlanguage. Chinese English learners may have limited access to such authentic materials, especially in the early stages of learning, which restricts the development of their interlanguage and their ability to use “but” flexibly.

5. Suggestions for Related English Teaching

In terms of teaching content, teachers play a crucial role in enhancing students’ understanding of the word “but”. They ought to offer an all-encompassing explanation of both the semantic and pragmatic knowledge associated with “but”. This involves delving into its various meanings in different contexts and how it functions in practical communication. Additionally, teachers should systematically present its common collocations and alternative words such as “however”, “yet” and “whereas”. By doing so, students can significantly enrich their vocabulary and gain a more nuanced grasp of how to use “but” effectively.

Regarding teaching methods, the contrastive teaching approach is highly recommended. Teachers can use a multitude of examples to showcase the differences in the use of “but” between Chinese English learners and native English speakers. Through this comparison, students can actively identify the problems and shortcomings in their own usage of “but”. Moreover, it helps them become aware of and overcome the negative transfer of their native language in English learning, enabling them to use “but” more in line with native-speaker norms.

As for pedagogical strategies, it is absolutely essential to significantly increase students’ exposure to a wide variety of authentic English materials. These materials encompass not only well-known English novels that offer rich language contexts but also daily English newspapers which present real-world usage and current expressions, as well as English movies that provide natural spoken language and cultural insights. Simultaneously, incorporating structured writing practice with targeted feedback is of great benefit. Teachers can assign regular writing tasks that specifically focus on the proper use of “but”. By providing detailed feedback on these assignments, students can gradually correct their mistakes and develop a better understanding, ultimately

enhancing their ability to use “but” correctly in both written and spoken English.

From the perspective of evaluation methods, teachers should design diverse assessment tasks. Besides traditional written tests, oral presentations and group discussions can be included. In oral tasks, students’ use of “but” in real-time communication can be observed, and teachers can offer on-the-spot guidance. Group discussions also encourage students to interact and use “but” naturally, helping them improve their practical application ability.

6. Conclusion

This study, utilizing the corpus method, has conducted an in-depth investigation into the usage characteristics of the discourse marker “but” in the writing of Chinese English learners. In terms of usage frequency, Chinese English learners’ overall use of “but” was significantly higher than that of native English speakers. There are significant differences in “but” usage frequency among Chinese English learners at different learning stages. For non-English majors, the usage frequency of “but” increases with advancing learning stages, whereas for English majors, it follows an upward then downward trend. Regarding syntactic positions, Chinese English learners employ “but” most frequently at the beginning of sentences, while native English speakers tend to use it more within sentences. In terms of collocation patterns, the main collocation types of “but” involve contrastive, adversative and exclusive relationships. Chinese English learners exhibit certain limitations in the collocational use of “but”, which is less rich and diverse compared to native English speakers. As for pragmatic functions, Chinese English learners and native English speakers show significant differences in the pragmatic functions of “but”. Chinese English learners are inclined to overuse “but” to express stance but underuse it in other pragmatic functions, which reflects a rigid, formulaic application and incomplete acquisition of its pragmatic versatility.

Future related research can be expanded and deepened in multiple directions, for example, by comparing Chinese English learners with those from other Asian countries in their use of “but”, thereby revealing the general patterns and special phenomena in the use of discourse markers among second language learners from a broader perspective.

References

- [1] He, Z. R., & Ran, Y. P. (1999). Pragmatic constraints of discourse connectives. *Foreign Language Teaching and Research*, (03), 3-10+82.
- [2] Schiffrin, D. (1987). *Discourse markers*. Cambridge University Press.
- [3] Lakoff, R. (1971). *Ifs, ands, and buts about conjunction*. In C. J. Fillmore & D. T. Langendoen (Eds.), *Studies in linguistic semantics* (pp. 114-149). Holt, Rinehart and Winston.
- [4] Biber, D. et al. (1999). *Longman grammar of spoken and written English*. Longman.
- [5] Zhao, Y. (2018). A study on the acquisition of discourse markers in Chinese English learners’ oral English. *Survey of Education*, 7(14), 100-101.
- [6] Chen, R. (2021). A study on the use of discourse markers in written English by English majors. *Overseas English*, (11), 250-251.
- [7] Li, Z. W. (2003). Discourse Connectives as indicators of coherence relations. *Shandong Foreign Language Teaching*, (01), 32-36.

-
- [8] Lian, S. N. (1993). *Contrastive studies of English and Chinese*. Higher Education Press.
- [9] Wang, Y. F., & Tsai, P. H. (2007). Textual and contextual contrast connection: A study of Chinese contrastive markers across different text types. *Journal of Pragmatics*, 39(10), 1775-1815.
- [10] Müller, S. (2005). *Discourse markers in native and non-native English discourse*. John Benjamins Publishing Company.