

A Comparative Study of Discourse Markers Usage between Chinese and English Native Speaker College Students in English Expository Writing Utilizing Corpus-Based Techniques

**Jiahuang Chen^{1*}, Cholthicha Sudmuk²,
Suwaree Yordchim³, Montarat Rungruangthum⁴ and Behrad Aghaei⁵**

^{1,2,3} Suan Sunandha Rajabhat University, Thailand

⁴ Rajamangala University of Technology Phra Nakhon, Thailand

⁵ University of Michigan, USA

Corresponding Author, E-mail: homeley77@163.com

Abstract

This study compares the use of discourse markers (DMs) in expository writing between Chinese EFL learners and native English speakers, aiming to identify overused and underused DMs and inform teaching strategies. Using corpus-based techniques and Fraser's (1999) taxonomy, the research analyzed DMs in Chinese college students' writing from the Chinese Learners English Corpus (CLEC) and matched samples from the British Academic Written English (BAWE) corpus. Results indicated that Chinese learners overused DMs like 'so,' 'but,' 'and,' and 'because,' while underusing 'thus,' 'in conclusion,' 'therefore,' and 'hence.' These findings suggest that explicit instruction on DMs is crucial for improving Chinese EFL learners' academic writing proficiency.

Keywords: Discourse Markers; Corpus-Based Techniques; Expository Writing; Comparative Study

Introduction

Discourse markers (henceforth DMs) are pivotal in academic and professional communication, yet their use by non-native English speakers in formal writing, especially expository writing, has been understudied. As English becomes the lingua franca, the demand for second language learners to engage in formal writing is increasing (Jalilifar, 2008). The proficient use of DMs in expository writing is essential, with overuse and underuse often indicating novice-level writing skills (Al-khazraji, 2019).

Expository writing is a prevalent genre in EFL education, emphasized in international assessments such as IELTS and TOEFL, and Chinese university English exams like TEM (Test for English Majors) and CET4/6 (College English Test Band 4 or 6). The proficient use of DMs is a hallmark of effective expository writing, which is defined by the logical and straightforward presentation of a topic using factual evidence (Bruffee, 1993). However, EFL learners often struggle with the appropriate frequency and functional application of DMs, leading to errors that can impede clarity and coherence (Kizil, 2017; Dumlao & Wilang 2019; Alahmed et al., 2020; Jakupčević, 2019; Wang & Modehiran, 2023).

Despite the acknowledged importance of DMs in academic writing, several research problems persist. There is a lack of comprehensive studies comparing DM usage between Chinese EFL learners and native speakers, particularly in the context of expository writing. Additionally, the specific DMs that are overused or underused by Chinese EFL learners remain unclear. Addressing these gaps is crucial for developing targeted instructional strategies to enhance DM usage in EFL writing.

This study investigates the DM usage of Chinese college students by analyzing a corpus of their expository writings and comparing it with a similar corpus from British university students. Using Fraser's (1999) taxonomy, the study identifies and compares the types and frequencies of DMs used by Chinese college students. The findings aim to contribute to the understanding of DM usage in EFL expository writing and inform pedagogical approaches to teaching DMs in academic writing contexts.

Research Objective

To delve deeper into DM usage among Chinese EFL learners, a thorough corpus-based analysis within a structured taxonomic framework is essential, particularly focusing on DM usage within a specific writing genre by a diverse group of Chinese EFL learners. This study employed corpus linguistic techniques to analyze large datasets, aiming not only to minimize research bias but also to generalize the findings. Additionally, there is a need for research that scrutinizes the potential overuse, underuse, and misuse of specific DMs by Chinese EFL learners in comparison with native English speakers. Consequently, this study aims:

1. To explore what types and frequencies of DMs, according to Fraser's (1999) taxonomy, are used in Chinese college students' expository writing, as revealed through a corpus-based analysis?
2. To reveal How DMs are used in Chinese college students' expository writing compared to that of native speakers, especially which markers are overused or underused.

Literature Review

1. Discourse Markers and Fraser's Taxonomy

DMs are linguistic items prevalent in spoken or written discourse, serving to structure and connect discourse units. Initial studies, such as Schiffrin (1987), identified DMs like "well," "now," and "so," highlighting their role in signaling relationships between utterances. Redeker (1990) expanded on this, viewing DMs as expressions that link discourse units and indicate coherence and ideational relationships. Blakemore (2002) furthered this understanding by arguing that DMs contribute to procedural meaning, guiding the interpretation of subsequent discourse.

Scholars (Briton, 1996; Schourup, 1999; Dér, 2010; Furkó, 2020; Badan & Cenni, 2021) generally agree on the characteristic features of DMs, which can be categorized into three domains.

Lexical and Phonological Features: DMs are multicategorical, including conjunctions (e.g., "but," "and"), adverbials (e.g., "frankly," "honestly"), interjections (e.g., "oh," "well"), phrases (e.g., "as a result," "in addition"), and clausal structures (e.g., "you know," "I mean"). Phonologically, they are marked by reduction and independence.

Semantic and Pragmatic Features: DMs are non-propositional and multifunctional, encoding procedural rather than conceptual meaning. They influence utterance interpretation without altering truth-conditional content.

Formal and Syntactic Features: DMs exhibit syntactic optionality and can be syntactically independent, often positioned initially in sentences but not exclusively.

DMs serve multifunctional roles, providing contextual coordinates (Schiffrin, 1987), signaling speaker intentions (Fraser, 1996), and indicating interpersonal relationships (Brinton, 1996). Their functions can be broadly categorized into textual and interactional dimensions. Textual functions organize discourse, promoting coherence and elucidating relationships between ideas, while interactional functions manage social and communicative aspects, facilitating meaning negotiation.

In conclusion, DMs encode procedural meaning and are defined by semantic non-propositionality, syntactic optionality, and phonological independence. Their multicategorical and multifunctional nature presents analytical challenges, necessitating a taxonomic framework for focused examination and linguistic inquiry.

Fraser (1999) proposed a comprehensive taxonomy of discourse markers (DMs), widely recognized as one of the most thorough classifications in the analysis of written discourse (Rahimi, 2011, p. 71). Fraser's taxonomy distinguishes between two primary types of DMs: those that connect the topic of the subsequent discourse segment (S2) with the previous one (S1), and those that link the explicit interpretation of S2 with some aspect of S1. A summary of Fraser's 1999 classification, which includes 97 different DMs, is presented in Table 1.

Table 1 Fraser's (1999) Taxonomy of DMs

Type	Content
Elaborative Markers	and, above all, also, besides, better yet, for another thing, furthermore, in addition, moreover, more to the point, on top of it all, to cap it all off, what is more, I mean, in particular, in any event, namely, parenthetically, that is to say, analogously, by the same token, correspondingly, equally, likewise, similarly, be that as it may, or, otherwise, well
Inferential Markers	so, of course, accordingly, as a consequence, as a logical conclusion, as a result, because of this, because of that, consequently, for this reason, for that reason, hence, it can conclude that, therefore, thus, in this case, in that case, in any case, under these conditions, under those conditions, then, all things considered
Contrastive Markers	but, however, although, though, in contrast to this, in contrast to that, in contrast with this, in contrast with that, in comparison to this, in comparison with this, in comparison to this, in comparison to that, on the contrary, contrary to this, contrary to that, conversely, instead of this, instead of that, rather than this, rather than that, despite, on the other hand, despite this, despite that, in spite of this, in spite that, nevertheless, nonetheless, still
Causative Markers	after all, because, for this reason, for that reason, since
Topic Relating Markers	back to my point, before I forget, by the way, incidentally, just to update you, on a different note, speaking of, that reminds me, to change to topic, to return to my topic, while I think of it, with regards to

2. Usage of DMs across Various Writing Genres

DMs are pivotal in organizing written discourse by establishing coherence, interlinking ideas, and steering readers through the logical progression of arguments or narratives. Variations in the frequency and types of DMs across genres such as argumentative, expository, and narrative writing have been well-documented.

In narrative texts, DMs are utilized to facilitate storytelling, with evaluative markers like "interestingly" and "surprisingly," and temporal markers such as "then" and "afterwards" being frequently employed, temporal markers serving to frame the narrative and signify shifts in perspective (Fludernik, 2003), and elaborative markers being as the most prevalent in narrative essays to elaborate stories (Rabab'ah et al., 2022).

Argumentative essays demonstrate a high usage of DMs due to the need to clearly present and interconnect arguments, counterarguments, and evidence. Commonly used DMs in this genre include contrastive markers (e.g., "nevertheless," "however"), causal markers (e.g., "because," "since"), and additive markers (e.g., "furthermore," "moreover") (Rahimi, 2011). Contrastive DMs are essential for introducing opposing views and constructing balanced arguments, while inferential DMs like "hence" and "thus" are used to draw conclusions from evidence, thereby bolstering the text's persuasiveness (Dumlao & Wilang, 2019).

Expository writing utilizes DMs to present information systematically, with a focus on markers that clarify conceptual relationships and enhance textual coherence. Elaborative DMs such as "for example" and "in addition" provide supporting details and examples (Adewibowo et al., 2018), while inferential markers like "thus" and "so" create logical connections between pieces of information (Frápolli & Assimakopoulos, 2012). In addition, elaborative markers (e.g., "in other words," "that is to say"), additive markers (e.g., "furthermore," "moreover"), and sequential markers (e.g., "first," "second") are crucial for maintaining textual cohesion and guiding the reader through the text's structure (Meyer & Rice, 1982).

The variation in DM usage across genres can inform writing instruction and assessment, enabling writers to employ DMs effectively and appropriately within different contexts. This understanding is vital for enhancing writing skills and adapting to the demands of various writing genres.

3. Usage of DMs among Chinese EFL Learners

Chinese learners of English as a Foreign Language often encounter significant challenges in the appropriate use of DMs. Research has identified a tendency among these learners to overuse certain DMs. Chen (2006) found that additive markers such as "and" and "also" were frequently overused by Chinese EFL learners, a phenomenon attributed to the transfer of similar structures from Chinese and the perception of these markers as "safe" choices. Liu (2013) observed that Chinese EFL learners tended to overuse linking adverbs like "first(ly)," "for example," "so," "then," "of course," and "in fact," while underusing markers such as "instead," "yet," and "anyway," suggesting a reliance on more familiar DMs due to L1 transfer.

Conversely, underusing DMs is also a notable pattern among Chinese EFL learners. Wei (2011) noted that Chinese students underused contrastive markers such as "nevertheless" and "nonetheless," particularly in academic writing where these markers were crucial for argument construction. Zhao (2013) reported that even advanced Chinese EFL learners underused inferential markers like "thus" and "hence," due to a lack of awareness of subtle differences between similar markers and limited understanding of their appropriate contexts.

Research Methodology

1. Population and sample

The study's population comprises Chinese college students learning English as a Foreign Language (EFL) and native English-speaking university students in the United Kingdom. The sample for this comparative analysis was drawn from two distinct corpora: the Chinese Learners English Corpus (CLEC) (Gui & Yang, 2003) and the British Academic Written English (BAWE) corpus (Nesi et al., 2004). Specifically, the CLEC sample included 1,499 expository texts from the ST3 and ST4 levels, written by both English majors and non-majors during their initial college years, totaling 243,163 tokens. These texts were sourced from standardized nationwide College English Tests, ensuring a broad and authentic representation of Chinese students' expository writing abilities and their application of discourse markers. For the native English speakers, a subset of 110 expository compositions from the BAWE corpus was extracted, also totaling 243,160 tokens, authored by native English speakers in their first and second years of university. This sample allowed for a direct comparison of discourse marker usage between the two groups, providing insights into the differences in DM usage patterns between non-native and native speakers in the context of expository writing.

The population and the sample from the two corpora could be presented in table 2 and table 3 respectively.

Table 2 Descriptive Data in CLEC and BAWE

Type of corpus	Size of the corpus	Description
CLEC	1,070,602 tokens	The corpus involves thousands of English writings by Chinese EFL learners of five stages labeled separately as ST2, ST3, ST4, ST5, and ST6, varying from middle school students to college students.
BAWE	6,504,356 tokens	The corpus consists of 2,761 passages written by three British university students in 35 disciplines over 3 undergraduate and 1 master year, grouped into 4 disciplinary areas.

Table 3 Comparative Overview of Expository Compositions from CLEC and BAWE Corpora

Feature	CLEC (ST3 & ST4)	BAWE (Native English Speakers)
Selected Number of Expository Texts	1,499	110
Total Tokens	243,163	243,160
Authors	Chinese EFL college students, both English majors and non-majors	Native English university students
College Years Represented	First and second-years	First and second-years
Composition Source	Nationwide college English tests	Writing assignments
Composition Genre	Exposition	Exposition

2. Research tools

The research utilized corpus-based techniques and AntConc, a concordance and corpus analysis tool, to examine the usage of DMs in expository writing by Chinese EFL learners and native English speakers. Fraser's (1999) taxonomy served as the analytical framework to categorize and identify DMs within the corpora. The study employed an online log-likelihood calculator to measure the statistical significance of observed frequency distributions, which was crucial in corpus linguistics research for determining patterns of overuse and underuse of specific DMs.

3. Data collection

Data was collected from two corpora: the Chinese Learners English Corpus (CLEC), comprising compositions by Chinese college students, and the British Academic Written English (BAWE) corpus, consisting of compositions by native English speakers. A balanced number of tokens focusing on expository writing samples were extracted from each corpus. The selected samples underwent data cleansing to eliminate irrelevant information such as HTML tags, metadata, and formatting codes, ensuring consistency and UTF-8 encoding.

The selected expository texts from both corpora were refined by removing irrelevant elements such as HTML tags, metadata, and non-textual formatting codes. Subsequently, the classification of DMs was adjusted according to Fraser's (1999) taxonomy to ensure an accurate assessment of their types and frequencies. While Fraser's framework was comprehensive, it required contextual adaptation, as not all DM variants served the same function in every situation. For instance, "and," categorized as an elaborative marker, did not function as a DM in the phrase "and so on".

The canonical form was represented as <S1./, DM + S2>, where S1 and S2 were typically separated by a DM with a period or a comma. The DM would be excluded from the study if it was concatenated with S1 and S2 without such punctuations.

A DM should link a complete segment rather than a word, a phrase, or an elliptical expression. Otherwise, the DM would be excluded from the study.

DMs in the form of adverbs or prepositional phrases were typically separated from the S2, indicated by their direct adjacency to a comma within S2. Adverbs or prepositional phrases not conforming to this pattern were excluded from the study.

Words classified as DMs in Fraser's (1999) taxonomy were excluded from the study if they were part of a larger phrase that did not function as a DM.

4. Data analysis

The analysis involved the identification and comparison of the types and frequencies of DMs used by Chinese college students and British university students. AntConc was used to quantify the DMs within the selected expository writings, and manual verification was conducted to ensure accuracy. The log-likelihood test was applied to identify which DMs were significantly overused or underused by Chinese EFL learners in comparison to native English speakers. This comparative analysis aimed to reveal discrepancies in the usage of DMs and to inform pedagogical approaches to teaching DMs in academic writing contexts.

5. Conceptual framework

To address the research questions, a methodical investigation was conducted to ascertain the types and frequencies of DMs employed by Chinese college students in expository writing and to contrast this usage with that of British university students. The study involved the following processes:

Compilation of two corpora: the Chinese Learners English Corpus (CLEC) for compositions by Chinese college students and the British Academic Written English (BAWE) corpus for compositions by native English speakers.

Extraction of a balanced number of tokens from each corpus, focusing on expository writing samples.

Data cleansing of the selected samples to eliminate irrelevant information such as HTML tags, metadata, and formatting codes, and to ensure all texts were formatted consistently and encoded in UTF-8.

Adaptation of Fraser's (1999) DMs for analysis using AntConc, a tool for concordance and corpus analysis, to determine the types and frequencies of DMs.

Comparative analysis of the frequency of DM variants across the two corpora using a log-likelihood test to identify patterns of overuse and underuse by Chinese college students in comparison to British university students.

According to the processes, a conceptual framework could be drawn in figure1:

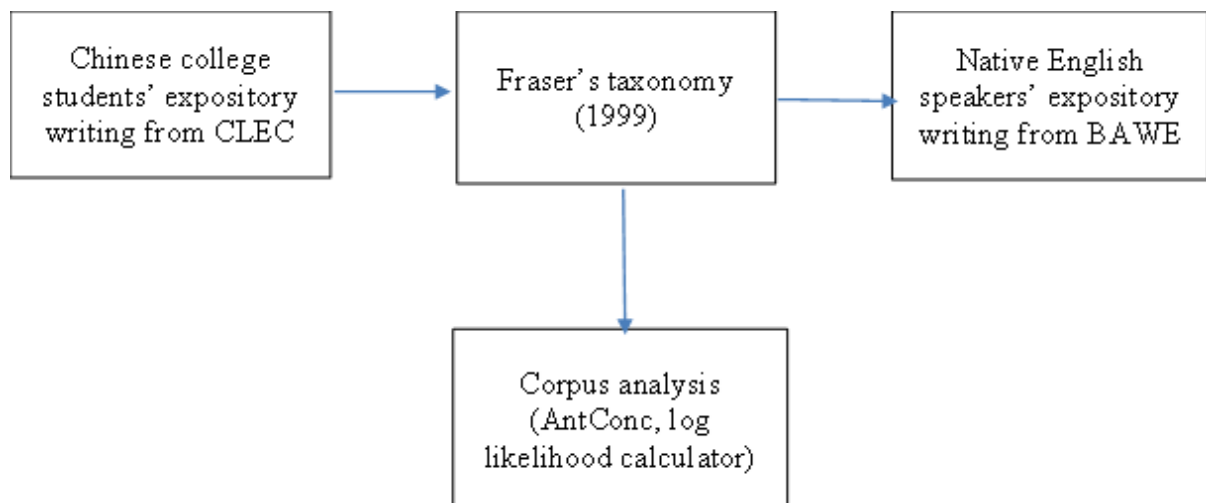


Figure 1 Conceptual Framework
(Source: Constructed by the researcher, 2024)

Research Findings

1. Types and Frequency of DMs

After expository compositions from two corpora had been selected and DMs delineated by Fraser (1999) had been clarified, the corpus analysis tool AntConc (Anthony, 2024) would be utilized to quantify the types and frequency of DMs within the selected expository writings from the CLEC and BAWE corpora, in accordance with Fraser's (1999) taxonomy. DMs had been identified in each sample of the two corpora, using a combination of AntConc and manual verification to ensure accuracy. In addition, the online log-likelihood calculator (<https://ucrel.lancs.ac.uk/llwizard.html>) had been adopted to determine the significance of the difference in frequencies between two datasets. The frequency and types of DMs in both corpora had been calculated and categorized respectively as shown in table 4.

Table 4 Frequency of DMs in the Sample from CLEC and BAWE

Types	Variants	Frequency		Log Likelihood
		CLEC	BAWE	
Inferential Markers	so	1,039	85	+955.84
	then	296	31	+248.28
	of course	72	8	+58.89
	as a result	61	24	+16.66
	because of this	4	0	+5.55
	in that case	3	0	+4.16
	as a consequence	1	0	+1.39
	as a logical conclusion; because of that; for that reason; it can be concluded that; under those conditions; all things considered	0	0	0.00
	accordingly	1	2	-0.34
	under these conditions	0	1	-1.39
	in any case	0	1	-1.39
	to conclude	0	2	-2.77
	for this reason	0	3	-4.16
	in this case	4	13	-5.02
	thus	77	121	-9.86
	in conclusion	15	39	-11.05
	therefore	77	136	-16.56
	consequently	4	26	-18.03
	hence	7	82	-74.35
Contrastiv e Markers	but	1,058	271	+498.01
	on the other hand	132	31	+67.37
	on the contrary	34	2	+34.46
	still	3	2	+0.20
	in spite of	2	2	0.00
	in contrast to this/that; in comparison to/with this/that; contrast to this/that; instead of (doing)this/that; despite that rather than this/that	0	0	0.00
	though	46	75	-7.02
	despite this	0	8	-11.09
	nevertheless	4	21	-12.67
	nonetheless	0	14	-19.41
	conversely	0	21	-29.11
	whereas	5	46	-37.98
Elaborativ e Markers	although	25	165	-115.43
	however	155	546	-231.1
	and	863	458	+126.18
	otherwise	62	0	+85.95

	besides	40	2	+42.14
	above all	26	1	+28.88
	that is to say	23	2	+20.72
	what is more	8	0	+11.09
	in addition	58	28	+10.69
	or	44	22	+7.48
	for another thing	3	0	+4.16
	moreover	44	33	+1.58
	better yet; on top of it all; to cap it all off; I mean; in any event; parenthetically; well; analogously; by the same token; be that as it may	0	0	0.00
	furthermore	43	50	-0.53
	more to the point	0	1	-1.39
	correspondingly	0	3	-4.16
	namely	2	9	-4.82
	equally	0	7	-5.06
	likewise	1	13	-12.20
	in particular	2	19	-15.90
	also	26	64	-16.56
	similarly	1	28	-31.50
Causative Markers	because	341	44	+260.07
	after all	15	11	+0.62
	for that reason	0	0	0.00
	since	37	53	-2.86
	for this reason	0	4	-5.55
	by the way	1	0	+1.39
Topic- Relating Markers	back to my point; before I forget; incidentally; speaking of; just to update you; on a different note; that reminds me; to change to topic; to return to my topic; while I think of it; with regards to	0	0	0.00

As depicted in Table 4, an analysis of the CLEC revealed the presence of all five categories of DMs within Fraser's (1999) taxonomy in the English expository writing of Chinese college students. Notably, topic-relating markers were observed with the lowest frequency, appearing only once in a sample comprising 243,163 words.

Among the remaining categories, inferential markers were utilized most frequently by the students, amassing a total of 1,661 occurrences. This was surpassed only by contrastive markers, which were used 1,464 times, followed by elaborative markers with 1,246 instances, and causative markers with 393 occurrences. However, the distribution of DM variants within each category was notably uneven. For instance, within the elaborative markers, the DM "and" was recorded 863 times, while others, such as "equally," and "correspondingly," were entirely absent from the sample. Similar disparities were noted across inferential, contrastive, and causative markers.

The analysis of the CLEC sample also highlighted a predilection for conjunctions, particularly "and," "but," and "so," among Chinese college students. This high frequency suggested a tendency towards simpler, more direct connective strategies, which might reflect ongoing development in their academic writing skills. Moreover, the underutilization or complete absence of several markers from Fraser's (1999) taxonomy in the students' writing was observed. Specifically, some markers, including "accordingly," "in spite of," "by the way," and "likewise," were infrequently used (occurring less than 10 times), and some markers, predominantly topic-relating ones, were not employed at all.

The infrequent or non-existent use of a substantial number of markers from Fraser's (1999) taxonomy by Chinese college students in their expository writing suggested a potential limitation in their repertoire of DMs. This finding underscored the necessity for increased exposure and practice with a broader range of DMs, which could significantly enhance their writing proficiency.

As indicated in Table 4, the BAWE corpus sample exhibited the presence of four categories of DMs according to Fraser's (1999) taxonomy: elaborative markers, inferential markers, contrastive markers, and causative markers. Notably, topic-relating markers were absent from the sample.

Within these four categories, contrastive markers were employed most frequently by British native English-speaking university students, with a total of 1,204 occurrences. Elaborative markers followed with 740 instances, inferential markers with 574, and causative markers with 112. The distribution of DM variants within each category was found to be highly uneven. For instance, within contrastive markers, "however" was used 546 times, while other markers such as "in contrast to this" were not observed. Similar patterns of variation were noted in the usage of inferential, contrastive, and causative markers.

The top five most frequently used DMs in the BAWE sample were the contrastive marker "however" (546 instances), the elaborative marker "and" (458 instances), the contrastive marker "but" (271 instances), the contrastive marker "although" (165 instances), and the inferential marker "therefore" (136 instances). The prevalence of contrastive markers in the expository writing of native English-speaking university students suggested a pronounced emphasis on the juxtaposition of ideas to bolster the clarity and depth of their arguments.

However, the usage of certain markers from Fraser's (1999) taxonomy was either infrequent or non-existent. Specifically, 18 markers, including "for this reason," "in any case," "above all," and "on the contrary," were used sparingly (less than 10 times), and 48 markers, encompassing all topic-relating markers, were entirely absent from the sample. This suggested that even within the writing of native English-speaking students, there may be a selective preference for certain DMs, potentially indicating areas for further linguistic development or a reflection of the rhetorical conventions prevalent in academic writing.

2. Overuse and Underuse of DMs

This research endeavored to examine the types and frequency of DMs, according to Fraser's (1999) taxonomy, in expository writings by Chinese college students. The analysis had been performed utilizing AntConc. Subsequently, a comparative assessment of the frequency of DMs extracted from the CLEC and the BAWE corpus was undertaken. The comparison was to ascertain the extent to which Chinese college students either overused or underused specific DMs in their expository compositions. To quantitatively determine these usage patterns, the study employed the online log-likelihood calculator (<https://ucrel.lancs.ac.uk/llwizard.html>)

designed to measure the statistical significance of observed frequency distributions in corpus linguistics research.

To ascertain the differential usage of DMs by Chinese college students, a two-pronged analytical approach was implemented. Initially, a comprehensive comparison of the DM frequency across the five taxonomical categories proposed by Fraser (1999) was executed for both corpora. Log-likelihood values online helped to determine the significance of the difference in frequencies between the two datasets. The interpretation of these values in terms of overuse or underuse generally followed these guidelines in Table 5.

Table 5 Log Likelihood Value and Significance of Difference in Frequency

Log Likelihood Value	Significance of Difference in Frequency	Note
> 15.13	Significantly Overused	This corresponds to a significance level of $p < 0.0001$, indicating a very strong statistical significance.
10.83~15.13	Moderately Overused	This range corresponds to a significance level of $p < 0.001$, indicating a moderate and statistically significant overuse.
6.63~10.83	Noticeably Overused	This range corresponds to a significance level of $p < 0.01$, indicating a noticeable and statistically significant overuse.
3.84~6.63	Slightly Overused	This range corresponds to a significance level of $p < 0.05$, indicating a slight but statistically significant overuse.
-3.84~3.84	Minimal Differences	Values within this range indicate that the difference is not statistically significant, suggesting minimal differences between the datasets.

Subsequently, a granular analysis focusing on the frequency of individual DM variants within each of Fraser's categories was conducted. The results of the overall frequency comparison for the five DM categories were encapsulated in Table 6, which served to illustrate the relative overuse or underuse of markers of the five taxonomical categories in the expository writing of Chinese college students.

Table 6 Overall Comparison of Frequency of Five Categories in Two Corpora

Types	Frequency CLEC	Size	Frequency BAW	Size	Log Likelihood
Inferential Markers	1,661	243,163	574	243,160	+551.76
Contrastive Markers	1,464	243,163	1,204	243,160	+25.27
Elaborative Makers	1,246	243,163	740	243,160	+130.35
Causative	393	243,163	112	243,160	+165.63
Topic-Relating Markers	1	243,163	0	243,160	+1.39

Note: (1) A plus (+) or minus (-) symbol before the log-likelihood value to indicate overuse or underuse respectively in corpus 1 relative to corpus. (2) It is commonly acknowledged that a log-likelihood value of 3.84 or higher, which corresponds to a p-value of 0.05, indicates that a difference in frequency is statistically significant.

The results of Table 6 indicated that five types of DMs in the sample from CLEC, within Fraser's (1999) taxonomy, were all overused relative to those in the sample from BAWE; in addition, there existed significant differences in the frequencies of various DMs between CLEC and BAWE corpus, as measured by the log-likelihood values:

(1) The log-likelihood value of +551.76 indicated an extremely significant overuse of inferential markers by Chinese college students compared to the BAWE sample.

(2) The log-likelihood value of +165.63 indicated a significant overuse of causative markers in CLEC compared to BAWE. While it was not as pronounced as inferential markers, the frequency was still considerably higher than expected.

(3) The log-likelihood value of +130.35 suggested a significant overuse of elaborative markers in the CLEC sample.

(4) The log-likelihood value of +25.27 showed a notable overuse of contrastive markers in CLEC.

(5) The log-likelihood value of +1.39 indicated a negligible difference in the use of topic-relating markers between the CLEC and BAWE samples. This value suggested that the frequency of topic-relating markers was not significantly different between the two corpora.

In summary, Chinese college students exhibited a significant overuse of inferential, causative, elaborative, and contrastive markers compared to their BAWE counterparts, while the use of topic-relating markers remained relatively similar across both corpora.

After comparing the overall DM frequency within Fraser's (1999) five categories between the CLEC and BAWE samples, a detailed comparison of each variant within these categories was performed. The detailed comparison of each variant between CLEC and BAWE samples was conducted in Table 4.

According to Table 4 and Table 5, we compared the frequency of each variant of five categories of DMs between the sample from CLEC and BAWE corpus, providing log-likelihood values to indicate the significance of differences in usage. The findings could be elucidated as follows.

(1) In the analysis of inferential discourse markers, CLEC exhibited a marked preference for "so," "then," "of course," and "as a result," with significant overuse relative to BAWE. Other markers, such as "because of this" and "in that case," are slightly overused, while "as a consequence," "accordingly," "under these conditions," "in any case," and "to conclude" show negligible differences in frequency. Conversely, "thus," "in conclusion," "therefore," "consequently," and "hence" are notably underused in CLEC.

(2) Regarding contrastive markers, CLEC demonstrated significant overuse of "but," "on the other hand," and "on the contrary," indicating a clear preference. The marker "still" showed no significant difference in usage. In contrast, "though" was noticeably underused, while "despite this" and "nevertheless" were moderately underused. Significant underuse was observed for "nonetheless," "conversely," "whereas," "although," and "however."

(3) Elaborative markers in CLEC were characterized by significant overuse of "and," "otherwise," "besides," "above all," and "that is to say," with "what is more" showing moderate overuse. "In addition" and "or" were noticeably overused, and "for another thing" was slightly overused. Minimal differences were seen for "moreover," "furthermore," and "more to the point." Slight underuse was noted for "correspondingly," "namely," and "equally," with "likewise" showing moderate underuse. "In particular," "also," and "similarly" were significantly underused in CLEC.

(4) Causative markers showed a significant preference in CLEC for "because," with minimal differences for "after all" and "since," and slight underuse for "for this reason."

(5) Topic-relating markers in CLEC were largely comparable to BAWE, with the exception of "by the way," which showed a minor difference in usage between the two corpora.

In conclusion, the comparative examination of DMs within the expository writings of Chinese college students with those of English native speakers had discerned a notable prevalence of certain markers that were either overused or underused. Specifically, the study had identified a subset of DMs that exhibited a significant, moderate, noticeable, slight, and minimal deviation from the normative usage patterns. However, the study was of particular interest to those significantly overused and underused markers due to their statistical prominence and scarcity respectively. These markers were meticulously detailed in Table 7, thereby offering a comprehensive overview of the usage discrepancies observed in the study.

Table 7 Significantly Overused and Underused Markers by Chinese College Students

Significantly Overused Markers			Significantly Underused Markers		
Types	Variants	Log Likelihood	Types	Variants	Log Likelihood log-likelihood
Inferential Markers	so	+955.84	Inferential Markers	hence	-74.35
	then	+248.28		consequently	-18.03
	of course	+58.89		therefore	-16.56
	as a result	+16.66		however	-231.1
Contrastive Markers	but	+498.01	Contrastive Markers	although	-115.43
	on the other hand	+67.37		whereas	-37.98
	on the contrary	+34.46		conversely	-29.11
				nonetheless	-19.41
Elaborative Markers	and	+126.18	Elaborative Markers	similarly	-31.50
	otherwise	+85.95		also	-16.56
	besides	+42.14		in particular	-15.90
	above all	+28.88			
Causative Markers	that is to say	+20.72	Causative Markers		
	because	+260.07			

Upon examination of Table7, it was evident that the markers "so" in the inferential category, "but" in the contrastive category, "and" in the elaborative category and "because" in the causative category were the most significantly overused marker in each category.

Discussion

The study aimed to compare the usage of discourse markers (DMs) in expository writing between Chinese EFL learners and native English speakers. It sought to identify types and frequencies of DMs, reveal overused and underused markers by Chinese EFL learners, and inform pedagogical approaches to teaching DMs in academic writing contexts.

The analysis revealed that Chinese EFL learners significantly overused certain DMs such as 'so,' 'but,' 'and,' and 'because,' while underusing others like 'thus,' 'in conclusion,' 'therefore,' and 'hence.' This overuse and underuse can lead to wordiness, disjointed ideas, and a lack of formal style, negatively impacting the clarity and coherence of academic writing. The study also found that the use of topic-relating markers was relatively similar between the two groups.

The findings reflect the challenges faced by Chinese EFL learners in mastering the nuanced use of DMs in academic writing. The overreliance on certain DMs may stem from a limited repertoire or a transfer from their first language. The underuse of certain DMs could indicate a lack of awareness of their functional roles in enhancing the structure and coherence of expository texts.

The proficient use of DMs is a critical component of academic writing proficiency for Chinese EFL learners. The study concludes that explicit instruction on the functional and formal aspects of DMs is necessary to enhance Chinese EFL learners' academic writing skills.

To summarize, the study reveals that Chinese EFL learners overuse certain DMs such as 'so,' 'but,' 'and,' and 'because,' while underusing others like 'thus,' 'in conclusion,' 'therefore,' and 'hence.' This impacts the clarity and coherence of their expository writing, suggesting a limited repertoire influenced by their first language and a perception of certain DMs as 'safe' choices. In contrast, native English speakers exhibit a more nuanced DM usage, enhancing textual coherence. These findings underscore the need for explicit DM instruction in EFL curricula, emphasizing functional and appropriate DM use.

Recommendations

Based on the findings of this study, the following recommendations are made to enhance discourse marker (DM) usage in Chinese EFL learners, improve their academic writing skills, and support their overall language proficiency development:

1. Theoretical Recommendations:

Further research should investigate the impact of targeted DM instruction on Chinese EFL learners' writing proficiency, focusing on specific DM types (e.g., inferential, contrastive) that pose challenges in academic contexts. This would clarify how instructional focus on different DMs can improve coherence and clarity in learners' writing.

Conduct cross-linguistic studies to examine DM usage patterns among learners from various language backgrounds. Such research can identify universal challenges and language-specific issues that affect DM acquisition, offering insights that are useful for creating adaptable teaching materials across different EFL populations.

2. Policy Recommendations:

Integrate explicit DM instruction into the EFL curriculum, with systematic guidance on the types, functions, and frequency control of DMs in academic writing. Such curricular inclusion would support students in understanding the appropriate and varied use of DMs to enhance logical flow and cohesion in writing.

Establish EFL writing standards that incorporate DM usage as part of assessment criteria, especially for advanced learners preparing for academic or professional environments. This could help create benchmarks for evaluating writing coherence, cohesion, and overall clarity.

3. Further Research Recommendations:

Explore effective teaching methods for DM instruction, such as providing students with examples of both correct and incorrect DM usage, developing exercises to encourage DM practice in varied contexts, and incorporating collaborative writing tasks with peer feedback. These strategies can support a deeper understanding of DMs' functional roles in academic discourse.

Conduct longitudinal studies to track the development of DM proficiency among EFL learners over time. Such studies can reveal how DM usage evolves with experience and identify key stages in DM acquisition, providing guidance for curriculum development tailored to learners' proficiency levels.

References

- Adewibowo, D. et al. (2018). A Study of Discourse Markers Used in the Theses Background Written by the Students of English Department of Bengkulu University. *Journal of English Education and Teaching*, 2 (3), 89-97.
- Alahmed, S., Mohammed, Y. & Kirmizi, ö. (2020). The Use of Discourse Markers in L2 English Writing by Iraqi Postgraduate Students at Karabuk University. *Eurasian Journal of English Language and Literature*, 2 (1), 107-115.
- Al-khazraji, A. (2019). Analysis of Discourse Markers in Essays Writing in ESL Classroom. *International Journal of Instruction*, 12 (2), 559-572.
- Anthony, L. (2024). AntConc (Version 4.3.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Badan, L. & Cenni, I. (2021). Discourse Markers in Italian Digital Context: Expressing Involvement and Reviewers' Attitude on TripAdvisor. In M. Johansson et al. (Eds.), *Analyzing Digital Discourses: Between Convergence and Controversy* (pp 77-103). Cham: Palgrave Macmillan.
- Blakemore, D. (2002). *Relevance and Linguistic Meaning: The Semantic and Pragmatics of Discourse Marker*. Cambridge University Press.
- Brinton, L. J. (1996). *Pragmatic Markers in English: Grammaticalization and Discourse Functions*. Mouton de Gruyter.
- Bruffee, K.A. (1993). *A Short Course in Writing: Composition, Collaborative Learning, and Constructive Reading*. Harper Collins College Publishers.
- Chen, C. W. Y. (2006). The use of conjunctive adverbials in the academic papers of advanced Taiwanese EFL learners. *International Journal of Corpus Linguistics*, 11 (1), 113-130.
- Dér, C. I. (2010). On the Status of Discourse Markers. *Acta Linguistica Hungarica*, 57 (1), 3-28.
- Dumlao, R. P. & Wilang, J. D. (2019). Variations in the use of Discourse Markers by L1 and L2 English users. *Indonesian Journal of Applied Linguistics*, 9, 202-209.
- Fludernik, M. (2003). Chronology, time, tense and experientiality in narrative. *Language and Literature*, 12 (2), 117-134.

- Frápolti, M. J., & Assimakopoulos, S. (2012). Redefining logical constants as inferential markers. *The Linguistic Review*, 29 (4), 625–641.
- Fraser, B. (1996). Discourse Markers. In J. Verschueren, J.-O. Östman, & J. Blommaert (Eds.), *Handbook of Pragmatics: Manual* (pp. 1-10). John Benjamins.
- Fraser, B. (1999). What Are Discourse markers? *Journal of Pragmatics*, 31, 931–952.
- Furkó, P.B. (2020). *Discourse markers and Beyond*. Palgrave Macmillan.
- Gui, S.C.&Yang, H.Z. (2003). *Chinese Learner English Corpus*. Shanghai: Shanghai Foreign Language Education Press.
- Jakupčević, E. (2019). Young Language Learners' Use of Discourse markers in L2 Narratives. *English Teaching & Learning*, 43,411–428.
- Jalilifar, A. (2008). Discourse Markers in Composition Writings: The Case of Iranian Learners of English as a Foreign Language. *English Language Teaching*, 1 (2), 115-122.
- Kizil, A. S. (2017). The Use of metadiscourse in spoken interlanguage of EFL learners: A Contrastive analysis. *Online*. Retrieved January 12, 2024, from <https://www.researchgate.net/publication/320001324>
- Liu, G.B. (2013). On the Use of Linking Adverbials by Chinese College English Learners. *Journal of Language Teaching and Research*, 4 (1), 149-155.
- Meyer, B. J., & Rice, G. E. (1982). The interaction of reader strategies and the organization of text. *Text*, 2 (1-3), 155-192.
- Nesi, H., Gardner, S., Thompson, P., and Wickens, P. (2004). British Academic Written English Corpus. Available from <https://www.coventry.ac.uk/research/research-directories/current-projects/2015/british-academic-written-english-corpus-bawe/>
- Rabab'ah, G., Ma'touq, A., Alghazo, S. (2022). Discourse Markers in Narrative Essays: A Case Study of Jordanian High School EFL Learners. *Jordan Journal of Modern Languages and Literatures*, 14 (1), 203-217.
- Rahimi, M. (2011). Discourse markers in argumentative and expository writing of Iranian EFL learners. *World Journal of English Language*, 1 (2), 68-78.
- Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, 143, 367-381.
- Schiffrin, D. (1987). *Discourse Markers*. Cambridge University Press.
- Schourup, L. (1999). Discourse Markers. *Lingua*, 107 (3), 227-265.
- Wang, C., & Modehiran, P. . (2023). Needs Analysis of Metacognitive Strategies Training in English Language Instruction Using Digital Technology and to Improve the Autonomous-Learning Ability and English Language Achievement of Chinese EFL Undergraduate Students. *International Journal of Sociologies and Anthropologies Science Reviews*, 3 (6), 83–94.
- Wei, M. (2011). Investigating the oral proficiency of English learners in China: A comparative study of the use of pragmatic markers. *Journal of Pragmatics*, 43 (14), 3455-3472.
- Zhao, H. (2013). A study on the pragmatic fossilization of discourse markers among Chinese English learners. *Journal of Language Teaching and Research*, 4 (4), 707-714.