Contrastive Discourse Marker Usage in Native and Non-Native Undergraduate Academic Writing: A Corpus-Based Study

Sattawat Chaiyasit[1], Wanwisa Changkwian[2], Piyapohn Pohnsuriya[3*]

[1]Faculty of Science and Arts, Burapha University, Chanthaburi Campus

[2]International College, Burapha University

[3*]Language Institute, Burapha University

E-mail: piyaporn.po@go.buu.ac.th*

*Corresponding Author

Abstract

This preliminary corpus-based study investigated how native (NS) and non-native (NNS) undergraduate students use contrastive discourse markers (CDMs) in academic expository writing. Following Fraser's (1999) taxonomy, 120 essays (60 per group) from the British Academic Written English (BAWE) corpus, totaling 299,279 tokens, were analyzed using AntConc and log-likelihood statistical tests. Nine CDMs were examined: but, however, although, though, nevertheless, whereas, on the other hand, on the contrary, and even though.

Results revealed both similarities and differences in CDM usage patterns. While both groups demonstrated comparable use of basic markers such as but, however, and although, NNS writers significantly overused formal connectors, particularly on the other hand (LL = -40.88) and whereas (LL = -9.97). Positional analysis indicated that NNS writers showed a strong preference for sentence-initial placement (e.g., however at 70.2%), whereas NS writers employed more flexible positioning strategies throughout their texts.

These findings suggested that NNS writers rely heavily on formulaic and explicit contrastive structures, likely reflecting instructional influences and limited rhetorical flexibility. The study provides alternative insights for enhancing academic writing pedagogy in EFL contexts by highlighting the need for more contextually appropriate and varied CDM instruction.

Keywords: Contrastive Discourse Markers, Corpus Linguistics, Academic Writing, Native-Non-native Comparison, EFL Writing

Introduction

Higher education students majoring in English are usually expected to write academic reports in English for their senior projects before graduation or in the final year so as to universities' requirements. These tasks are intended to show both academic understanding and fluency in the English language. In order to gauge their level of English proficiency, they also need to take the English proficiency as exit exam for graduation. English proficiency scores are also required for recent graduates in order to apply for jobs or pursue higher education. However, when it comes to writing, Thai students face challenges to write their tasks in English For example, they must learn how to write in the correct grammar (Jaroenkitboworn, 2014), The investigation by Boon Yarattanasoontorn (2017) has also reported that the significant problems made by Thai university students in English writing concerns using correct grammar in all types of writing, the difficulty of writing tasks (e.g., letters, reports, and projects), and selecting appropriate vocabulary for each writing context respectively. Considering these above issues, students also need to take standardized tests like TOEIC, TOEFL, and IELTS to test their English proficiency level before or after graduation to apply for their jobs (Wongsothorn, Hiranburana, & Chinnawongs, 2002).

Considering these above issues, one of the key features that students must learn is to be able to produce clear and well-organized writing texts. A Key element of this writing text is to correctly use the discourse markers. The ability to use Discourse Markers (DMs) is one of the marking schemes. Many EFL students, however, have difficulties writing academic texts that are logically structured and coherent, according to earlier research (Hyland, 2005; Liu, 2013). The inability to clearly construct reasoning or argumentation is a common weakness in EFL writing, frequently brought on by the inefficient use of discourse markers (DMs) (Granger and Tyson, 1996; Leedham & Cai, 2013). Discourse markers are textual devices that help the reader follow the logical flow of a writer's ideas by signaling the relationship between discourse segments (Fraser, 2015; Hyland, 2005). DMs are particularly crucial for expressing contrast, cause, result, or elaboration in academic writing.

One aspect of research studies in discourse markers is the investigation on the use of contrastive discourse markers or CDMs. CDMs are especially important for constructing arguments by using words or phrases such as but, however, on the other hand to make a contrast sense in sentences. Prior corpus-based research has demonstrated how EFL writers frequently overuse fundamental CDMs e.g. however, often used in formulaic sentence-initial positions, sometimes without the proper contrast (Altenberg & Tapper, 1998; Huneety et al. 2023), and advanced ones are

underused. Developing students' ability to write logical and convincing academic texts is a primary objective for English teachers. Students' writing can become clearer and more logically structured if they are taught how to use discourse markers appropriately, especially contrastive discourse markers because it might enable students to use contrastive discourse markers more confidently in their academic texts.

This pilot study focused on Contrastive Discourse Markers (CDMs) because they are especially difficult for many EFL learners to use correctly in academic writing. CDMs, such as but and however, show contrast in meaning and require a good understanding of logic and argument structure (Fraser, 2015; Hyland, 2005). Discourse markers, unlike other markers, often confuse learners who do not fully understand the language because it requires proficiency in writing as well. Previous studies have shown that learners tend to use some CDMs too often and others too little, leading to problems with text flow and clarity (Altenberg & Tapper, 1998; Leedham & Cai, 2013).

Although various research studies have explored discourse markers used in L1 and L2 writing, most have focused on general marker types. Few studies have looked closely at contrastive discourse markers (CDMs) in academic writing at the undergraduate level, particularly in expository essays. Therefore, this study aims to compare the use of CDMs in academic texts written by native and non-native English-speaking undergraduate students. The findings might offer insights into usage patterns and pedagogical implications for improving academic writing instruction in EFL contexts and can help improve how they are taught, especially when students use them in fixed or unnatural ways (Huneety et al., 2023). By focusing on nine common CDMs, this study gives detailed insights that could guide teaching and future research on discourse markers.

Related Studies on CDMs Used by Native and Non-Native Speakers

Recent studies have shown increasing attention to how contrastive discourse markers (CDMs) are used in academic writing among native and non-native writers across different proficiency levels. Based on Fraser's (2009) framework, researchers have examined the frequency, variety, and functional accuracy of CDMs such as but, although, however, whereas, and nevertheless. These markers play an important role in expressing contrastive relations between discourse segments, and their use often reflects a writer's level of proficiency. For example, Ağçam (2017) conducted a study to investigate the use of CDMs among students in three groups with different mother tongues. Similar to Ali's work (2023) found that the discourse markers usage in the conversations of 14 Jordanian Spoken Arabic. Salem et. al (2024) conducted a research study to analyze the use of discourse markers used in academic writing in EFL learners. Together, these studies consistently highlight three key differences in the use of CDMs

between native and non-native writers: (1) non-native writers tend to use a limited range of CDMs, (2) they often overuse basic markers like but and however, and (3) they show limited flexibility in placement. Such patterns often result in weaker cohesion and less effective rhetorical organization. Therefore, explicit instruction on the functions and positions of contrastive discourse markers is necessary to help EFL learners improve their academic writing quality.

Research Objectives

1. To investigate the Contrastive Discourse Markers (CDMs) used by native and non- native writers at tertiary level in academic writing.

2. To analyze the similarities and differences of Contrastive Discourse Markers (CDMs) used by native and non- native writers at tertiary level in academic writing.

Research Methodology
Selection of Corpus Analysis Tool

There are various types of corpus analysis tools, each with unique features, benefits, and drawbacks. Additionally, there are two primary categories to choose from: web-based and standalone tools. An overview of toolkits was provided as shown in Table 1.

Table 1 A summary of corpus analysis toolkits

Tool	Type	Log-Likelihood	POS Tagging	Upload Corpus	Free / Paid	Limitations
AntConc	Standalone	✓	-	✓	Free	POS tagging can be used Tagant.
Sketch Engine	Web-based	✓	✓	✓	Paid (Free Trial/Student Plan	Subscription, complicated to use
LancsBox	Standalone	✓ (basic)	✓	✓	Free	Subscription, learning curve
Wmatrix	Web-based	✓	✓	✓	Free	May crash with large corpora

Tool	Type	Log-Likelihood	POS Tagging	Upload Corpus	Free / Paid	Limitations
Voyant Tools	Web-based	-	-	✓	Free	Not for deep corpus work; web-based only
WordSmith Tools	Standalone	✓	-	✓	Paid	Windows-only, outdated UI
BYU Corpora (COCA, BNC)	Web-based	✓ Limited	-	-	Free	Cannot upload corpora
#LancsBox X	Hybrid	✓	✓	✓	Free	Beta features may be unstable
KH Coder	Standalone	✓	✓	✓	Free	Complex setup
TXM	Standalone	✓	✓	✓	Free	Java & technical skill required
TextStat / Textalyser	Web-based				Free	Only basic statistics, no corpus tools
UAM CorpusTool	Standalone		✓ (manually)	✓	Free	Complicated to use

Web-based tools may have limitations on the size of the data when compared to stand-alone tools for analysis. The study, therefore, used Antconc's concordance, part-of-speech (POS) tagging, and log-likelihood functions to evaluate the frequency and analyze overuse and underuse of contrastive discourse markers (CDMs). These tools made it possible to examine CDM usage patterns in academic writing corpora that were both native and non-native, both qualitatively and quantitatively. TagAnt is another sub-tool that Antconc provides for POS tagging that serves the same purpose. Consequently, Antconc and TagAnt were used as corpus analysis tools in this investigation.

Criteria of CDMs based on Fraser 1999 Taxonomy

CDMs can function differently based on part of speech and the following clause. Table 2 presents criteria of CDMs selection methods.

CDMs	Inclusion Criteria and Example	Exclusion Criteria and Example
but	Coordinating conjunction linking two clauses: "The course outline	Within noun phrases: "She wore a blue **but** beautiful dress."

CDMs	Inclusion Criteria and Example	Exclusion Criteria and Example
	was well-designed, **but** the evaluation was difficult."	
however	Sentence-initial contrastive marker: "The model was useful. **However**, it failed to account for new data."	Stance softener (mid-sentence): "He is, **however**, a talented player."
although	Subordinating conjunction joining clauses: "**Although** the evidence was weak, the claim was accepted."	Sentence fragment without clause connection: "It's a good result, **although**."
though	Contrastive conjunction: "The results were promising; **though**, not entirely conclusive."	Final conversational tag: "Nice try, **though**."
nevertheless	Concessive marker linking sentences: "The hypothesis was flawed. **Nevertheless**, it sparked discussion."	
whereas	Formal contrastive conjunction: "NNS essays relied on formulaic markers, **whereas** NS essays varied more."	
on the other hand	Formal contrastive phrase: "The plan is expensive. **On the other hand**, it could save lives."	
On the contrary	Strong opposition marker: "Many assume the task is simple. **On the contrary**, it requires significant expertise."	
while	Contrastive conjunction showing opposition: "Thai favors indirectness, **while** English tends toward directness."	Temporal meaning: "He read while she cooked."
yet	Coordinating conjunction showing contrast: "The theory is popular, yet it lacks empirical support."	Adverb meaning "still": "He has yet to submit the report."

CDMs	Inclusion Criteria and Example	Exclusion Criteria and Example
even though	Concessive conjunction (strengthened form of "although"): "Even though he disagreed, he supported the decision."	

In this study, the inclusion of contrastive discourse markers (CDMs) follows Fraser's (1999) framework, which defines CDMs as lexical items that signal a contrast between two discourse clauses (see Table 2).

Only those markers that clearly function by linking clauses or sentences were included. Examples are but, however, although, though, nevertheless, still (when meaning "nevertheless"), yet (when used contrastively), even though, whereas, while (when showing contrast), on the other hand, on the contrary, and instead (in sentence-initial position). These were counted only when they contrasted two full ideas or arguments. On the other hand, markers that appear inside phrases or function as prepositions—such as in spite of, in contrast to, and in comparison to—were excluded, as they do not connect clauses but instead modify noun phrases. Some uses were also excluded due to ambiguity. For example, but in "a blue but beautiful dress" is not contrastive at the discourse level; still in "he is still working" is temporal, not contrastive; and though in "nice try, though" acts as a conversational tag, not a true connector.

In conclusion, these criteria help ensure that only functional, discourse-level markers of contrast were analyzed, avoiding overcounting and maintaining consistency with Fraser's taxonomy.

Data Preparation and Data Analysis
Corpus Selection Rationale

This study selected the British Academic Written English (BAWE) corpus because it provides authentic student writing from real academic settings across several UK universities. The corpus includes a wide range of disciplines and levels, making it suitable for examining how undergraduate students use discourse markers in academic texts. BAWE is widely used in previous studies on discourse markers. For example, Gardezi and Nesi (2009) analyzed conjunctive ties in economics student writing. The corpus also includes clear metadata on student background and task type, which allows for reliable comparison between native and non-native writers (Alsop & Nesi,

2009). Its structure and content make it appropriate for corpus-based studies focusing on contrastive discourse marker use in academic writing.

Essay Writing Selection

The selection criteria were based on both essay topic relevance and writer proficiency. Essays were filtered by genre (expository) and language background—distinguishing between native English speakers and non-native speakers for comparative purposes. In this study, native speakers are defined as those whose first language is English, while non-native writers were from Asian contexts where English is taught as a foreign language, specifically China, Korea, Japan, and Thailand.

A total of 120 explanatory essays were purposively selected from the British Academic Written English (BAWE) corpus to prepare the dataset for analysis. The corpus was divided evenly between 60 essays written by native English speakers (NS), comprising 139,367 tokens, and 60 essays written by non-native English speakers (NNS), totaling 159,912 tokens. All essays were written by the undergraduate students at tertiary level at the undergraduate level across a range of academic disciplines, including economics, business, politics, biology, engineering, and archaeology. The essay prompts covered general academic topics, such as Why do women have children? Why do some not do so?, The effect of group processes on individuals and work teams, and How powerful is the President of the United States?

After organizing data, each essay was converted into plain text (txt) format. After that, they were divided into two sub-corpora: folders that were texts written by native and non-native students. Tables, references, writers' names, and topics were removed to maintain anonymous and to reduce errors before data analysis. Sentence structure was maintained by checking punctuation and spacing, particularly around discourse markers (e.g., comma use after however). Some words, like but, have multiple meanings and were tagged as part of speech to make the analysis process clear and prevent ambiguity. TagAnt was used to complete the tagging process.

First, the British Academic Written English (BAWE) corpus was used to download the source material. Following the data preparation procedure, the data was categorized into two corpora: native and non-native. For each Contrastive Discourse Marker (CDM) in Fraser's (2009, 2015) frameworks—but, however, although, though, nevertheless, nonetheless, in contrast, and on the other hand—the native and non-native corpora were examined using Antconc (4.3.1). According to Granger (2002) and Biber et al., (2002) the frequencies of contrastive discourse markers (CDMs) were computed for normalization per 1,000 tokens and presented in Excel. (Biber et al.,

2002; Hunston, 2000). The frequencies for each side were computed using log-likelihood to compare the similarities and differences between the two groups after all frequencies and normalization of CDMs for both native and non-native essays were completed. Alongside the frequency analysis, a qualitative approach was used to examine how contrastive discourse markers (CDMs) were positioned within sentences. This positional analysis focused on whether markers appeared at the beginning, middle, or end of a sentence, offering insight into structural and rhetorical patterns. While frequency shows how often markers are used, positional analysis reveals how effectively they are deployed in context which is an important aspect for understanding coherence and fluency in academic writing, especially among EFL learners. The results were then reported.

Data Interpretation

To identify whether a contrastive discourse marker was overused or underused, this study used log-likelihood (LL) values generated from a web-based calculation from the University of Lancaster. Given Corpus A as NS and Corpus as NNS, direction of − means NNS overuse. However, if the direction is +, it refers to NNS underuse the CDMs. The critical value is LL qual to +/-3.84 (p<0.05), which indicates statistical differences (see Table 3).

Table 3 refers to interpretation of LL

Condition	Interpretation	Explanation
LL ≥ 3.84 and Direction = "−"	Overuse by NNS	NNS use the marker significantly more than NS.
LL ≥ 3.84 and Direction = "+"	Underuse by NNS	NNS use the marker significantly less than NS.
LL < 3.84 (any direction or "=")	No significant difference	The usage difference is not statistically significant (p ≥ 0.05)
LL ≥ 6.63	Strong difference	p < 0.01 means stronger statistical significance.
LL ≥ 10.83	Very strong difference	p < 0.001 means very strong statistical evidence.

Table 3 reports the LL value interpretation. The LL value compares the frequency of each marker between the native and non-native corpora. A result was

considered statistically significant if the LL value was equal to or greater than 3.84 (p < 0.05). The direction of difference was shown by a plus sign (+) or minus sign (−). A positive sign (+) indicates that the marker was used more frequently in the non-native corpus (i.e., overused), while a negative sign (−) shows that the marker was used less frequently in the non-native corpus (i.e., underused). In addition to LL values, raw frequency counts were also examined to support the interpretation.

Results

To answer Research Objective 1, this study investigates the use of contrastive discourse markers (CDMs) in expository essays written by native and non-native undergraduate writers.

CDMs Usage by Native and Non-Native writers

120 essays written by native and non-speakers for 60 essays each based on Fraser's (1999) framework were analyzed for its frequency use and revealed patterns in the use of Contrastive Discourse Markers (CDMs).

Table 4 CDMs Use in Academic writing in Native and Non-Native Undergrads

CDM	Frequencies		Normalized		LL	Interpretation
	NS	NNS	NS	NNS		
but	50	52	0.359	0.325	+0.22	No significant difference (LL < 3.84)
however	283	279	2.031	1.745	+3.05	No significant difference (LL < 3.84)
on the other hand	12	73	0.086	0.457	-40.88*	Significantly Overused by NNS (LL < 3.84)
though	46	46	0.330	0.288	+0.41	No significant difference (LL < 3.84)
on the contrary	2	13	0.014	0.081	-7.62*	Significantly Overused by NNS (LL < 3.84)
although	81	115	0.581	0.719	-2.27	No significant difference (LL < 3.84)
whereas	15	42	0.108	0.263	-9.97*	Significantly Overused by NNS (LL < 3.84)
nevertheless	13	32	0.093	0.200	-5.59*	Significantly Overused by NNS (LL < 3.84)

CDM	Frequencies		Normalized		LL	Interpretation
	NS	NNS	NS	NNS		
even though	8	3	0.06	0.019	+3.10	No significant difference (LL < 3.84)

Native speakers (NS) and non-native speakers (NNS) used Contrastive Discourse Markers (CDMs) similarly and differently. There are two main key results of not significantly different and significantly overused by NNS.

No Significant Difference between NS and NNS

Five CDMs did not show statistically significant differences in usage between native and non-native speakers. The coordinating conjunction **"but"** was used at almost the same frequency by both groups (NS = 50, NNS = 52, LL = +0.22). Likewise, **"however,"** which was the most frequently used CDM overall, showed a comparable distribution (NS = 283, NNS = 279, LL = +3.05). The subordinators **"though"** (NS = 46, NNS = 46, LL = +0.41) , **"even though"** (NS = 8, NNS = 3, LL = +3.10), and **"although"** (NS = 81, NNS = 115, LL = -2.27) also revealed no significant difference between the two groups.

Significantly Overused by NNS

Four CDMs were found to be significantly overused by non-native speakers, with varying levels of statistical strength. The most notable difference was observed in the use of **"on the other hand"** (NS = 12, NNS = 73, LL = -40.88), which marked the highest log-likelihood value in the dataset, indicating extreme overuse by NNS. Additionally, three other markers exhibited moderate to strong overuse: **"whereas"** (NS = 15, NNS = 42, LL = -9.97), **"on the contrary"** (NS = 2, NNS = 13, LL = -7.62), and **"nevertheless"** (NS = 13, NNS = 32, LL = -5.59).

The overuse of "On the other hand"

NS	NNS
Standardized testing ensures consistency. **On the other hand,** it can disadvantage students from diverse backgrounds.	Some people like dogs. **On the other hand,** others like cats.

This represents the strongest statistical significance in the dataset. The markedly higher frequency of use by NNS writers (73 vs. 12 occurrences) suggests that NNS possibly indicated a preference for a more formal or cautious writing style than NS.

The sample usage by NS indicated that NS seem to use more complicated sentences to support the previous sentence.

The overuse of "On the contrary"

NS	NNS
Many assume the task is simple. **On the contrary,** it requires significant expertise and patience.	He said it's boring. **On the contrary,** I don't like it.

The marker "on the contrary" was also significantly overused by NNS (NS = 2, NNS = 13, LL = -7.62). Although both less frequency use, the use by NNS may reflect an attempt to emphasize strong argument or to appear more formal tone. This pattern could indicate that NNS writers tend to prefer using markers of contrast, even when native speakers might opt for less formal.

The overuse of "whereas"

NS	NNS
Urban populations are growing rapidly, **whereas** rural areas continue to shrink.	I like red. **Whereas,** my friend likes blue.

The overuse of "Nevertheless"

NS	NNS
The method had limitations; **nevertheless,** the findings were consistent with previous studies.	I was tired. **Nevertheless,** I go shopping.

The marker "whereas" was notably overused by non-native speakers (NS = 15, NNS = 42, LL = -9.97). This possibly suggests that NNS writers may favor this formal contrastive conjunction to structure opposing ideas clearly, potentially influenced by academic writing instruction that emphasizes explicit logical connectors.

Similarly, "nevertheless" appeared significantly more in NNS texts (NS = 13, NNS = 32, LL = -5.59). This usage may reflect a tendency among NNS writers to mark concession. Possibly as a strategy to enhance coherence or to compensate for limited lexical variation. In contrast, native writers may achieve similar rhetorical effects through more varied or implicit cohesive devices. Native writers, on the other hand, tend to employ a wider range of cohesive strategies, often relying on more implicit or subtle

means to convey contrast. To answer research question 2, the similarities and differences of CDMs usage by native and non-native writers at tertiary level in academic writing were presented. Based on Table 2, the qualitative analysis of the use of CDMs by Native and Non-Native writers, the significant usage in differences were examined to see the similar and different patterns of CDMs placed in sentences (see Table 5).

CDMs	NS			NNS		
	Initial	Medial	Final	Initial	Medial	Final
but	36	364	0	55	452	0
however	152	131	0	196	83	0
on the other hand	5	7	0	54	21	0
though	9	180	0	16	166	0
on the contrary	1	1	0	9	4	0
although	34	47	0	67	48	0
whereas	3	12	0	2	40	0
nevertheless	5	8	0	17	15	0
even though	8	0	0	3	0	0

The analysis of CDMs shows both similarities and differences between native and non-native speakers in academic writing. For similarities, both groups used basic markers such as but, though, and although at similar rates, with no significant differences. Both also prefer to use high-frequency markers like but in the middle of the sentence (NS = 91%, NNS = 89%). No CDMs were found at the end of sentences in either group, suggesting both follow similar patterns in marker placement.

However, some key differences were also found. Non-native speakers overused formal markers such as on the other hand (LL = -40.88), whereas (LL = -9.97), on the contrary (LL = -7.62), and nevertheless (LL = -5.59). This shows that non-native writers may rely more on clear contrast signals to make their ideas easy to follow. Native speakers, in contrast, used fewer markers and showed more balanced use. There were also differences in position. Non-native speakers used many markers at the beginning of sentences, especially however (70.2% initial) and on the other hand (72.0% initial). Native speakers preferred to place markers in the middle of sentences to keep the writing smooth and natural.

Overall, native speakers used CDMs in a more flexible and natural way, depending on the context. Non-native speakers used them in a more fixed and formal pattern to make their writing clearer. These differences suggested that non-native

writers may rely more on obvious signals to organize their ideas, while native writers use contrast more indirectly.

Discussion

This study found that while both native and non-native writers used core contrastive discourse markers (e.g., but, however, although) while non-native writers significantly overused formal CDMs such as on the other hand, nevertheless, and on the contrary, especially at the beginning of the sentences. In contrast, native speakers used a wider range of CDMs more flexibly and more appropriately by using them in various sentence positions, for example in the middle of the sentence. Both groups showed similar use of basic markers and shared an understanding of some discourse conventions, such as avoiding CDMs at the end of sentences. These findings reflect Fraser's (2009) theoretical framework and align with previous research studies (e.g., Lee & Subtirelu, 2021; Al-Rubaye, 2022; Mahmoud & Al-Fraihat, 2023), which highlight key differences in CDM usage based on students' proficiency level and their writing experience. Interestingly, non native English writers often use formal contrastive words like on the other hand, nevertheless too much because they may learn these words from their English class without practicing enough to use them naturally. As a result, they may think that using the former markers will make their writing sound more academic or professional. This is called hypercorrection, the state when learners try too hard to sound correct and ending up sounding unnatural. Also, non native writers usually use contrastive markers at the beginning of the sentences. For example, they start their sentences with however, or on the other hand without the context. This shows that they may follow the grammatical rules from their first language. For native speakers, they are more flexible to use the discourse markers. They can put the contrastive words in different positions in the sentence, not just at the beginning. This shows that their writing texts have a better flow or rhythm and are more natural. Native speakers not only use the contrastive markers to connect ideas, but also use them to highlight the key points of their writing. These findings are consistent with previous studies mentioned earlier. For example, Ağçam (2017) compared the use of CDMs among three learner groups and found that Turkish speakers tended to overuse them, while Spanish speakers underused them. Ali (2023) also observed that both sophomore and senior students often overused simple markers, suggesting that increased academic level does not always lead to better DM usage. While these studies focus on general learner groups, the present study also shows how Asian learners, particularly Thai students, use CDMs differently. For instance, earlier studies reported that Asian learners

often misused basic contrastive markers such as so and but, or substituted them incorrectly. In contrast, Thai learners in this study tended to overuse formal contrastive markers like on the other hand and nevertheless, even when the context did not require strong opposition. This shows that although overuse and misuse are common among non-native writers, the way markers are selected and applied can vary across L1 backgrounds. Therefore, the present study contributes further by offering detailed, corpus-based evidence of these patterns in the Thai academic context and highlights specific pedagogical needs in relation to discourse marker instruction.

Implications for Classroom Instructions

The findings of this study shed light on the CDMs are taught in EFL writing instruction. Rather than focusing only on memorizing the list of the markers, explicit instruction and more practice should be applied in the classroom as well. Teachers and instructors should integrate the tasks that require students to use different CDMs, and practice to put CDMs in different positions in the sentences. Another way is to make the students more exposed to the native speakers' academic writing texts in order to make them more familiar with the natural usage. In addition, explicit instruction should address the differences between similar markers, and clarify when a marker is necessary or unnecessary in a sentence. These pedagogical changes can support learners to move from the same writing patterns to a more advanced writing of how to create well-structured arguments in their writing tasks.

Overuse Patterns and Pedagogical Influence

One key finding was that NNS use formal contrastive discourse markers much more often than NS, especially on the other hand (LL = -40.88), which showed the biggest difference. This suggested that many EFL learners were taught to use this type of marker as a safe and formal way to show contrast. Other formal markers like whereas, nevertheless, and on the contrary were also overused. This may show that non-native writers possibly rely on fixed, familiar phrases instead of choosing the most suitable marker for the context. This supported Granger and Tyson's (1996) idea of "hypercorrection," where learners overuse features they believe sound academic. Their writing may reflect classroom instruction that focuses on using clear linking words but does not fully teach when and how to use them in a natural and flexible way.

Positional Preferences and Discourse Flow

There are also clear differences in where native and non-native writers place CDMs in sentences. NNS often use them at the beginning of sentences—for example,

however (70.2%) and *on the other hand* (72.0%). This may show a more mechanical or rule-based use, likely influenced by how they were taught. In contrast, NS used CDMs more flexibly, placing them in different parts of the sentence. This showed better control of sentence rhythm and emphasis. It also supported Hyland's (2005) idea that skilled writers use discourse markers in a more subtle and natural way.

Similarities and Shared Patterns

Both native and NNS used basic contrastive markers like but, however, though, and although at similar frequencies. This suggested that most learners have a good understanding of basic contrast in academic writing. In addition, neither group used CDMs at the end of sentences, which showed a shared understanding of English grammar and typical patterns.

Limitations

This study has some limitations that should be reported. First, the BAWE corpus includes essays from different academic fields, but the distribution may not be balanced between native and non-native writers. Differences in discipline may influence how CDMs are used. Second, although all participants were undergraduate students, we did not control for their English proficiency levels. Some patterns may be due to individual language ability rather than L1 background. Third, the non-native group likely includes writers from various first-language backgrounds, which may affect their use of discourse markers. This study did not explore those differences in detail. The analysis focused on individual CDMs rather than their role in full arguments. Lastly, the data came from expository essays in a British academic setting, so the results, while primarily, may not apply to other genres, institutions, or current writing practices. However, it still provides an alternative guidance in terms of setting criteria, program selection and other factors that might result in the study.

Suggestions for Future Research

Future research can further investigate the students' first language that affects their use of contrastive discourse markers, especially how their native language use contrastive markers in writing. For researchers who examine individual learner profiles including proficiency levels, educational background, and exposure to English could offer deeper insights into CDM acquisition. This could give a better understanding of how students learn to use CDMs. Another way is to do the longitudinal study of how students use CDMs over time and across genres. Comparative studies between learners

in EFL and ESL environments could also shed light on the influence of instructional context. Finally, expanding research to genres beyond expository essays, for example, argumentative writing, reports, and personal reflections, would provide a more comprehensive picture of CDM usage across academic contexts.

Recommendations

Based on the findings that non-native speakers significantly overuse formal contrastive discourse markers, several key recommendations emerge for improving academic writing instruction. Pedagogically, educators should develop targeted teaching materials that explicitly address the contextually inappropriate use of markers like "on the other hand" and "nevertheless," helping students understand when these formal connectors enhance versus hinder their arguments. This instruction should move beyond simply teaching discourse marker types to foster rhetorical awareness of how marker choice affects reader perception and text effectiveness. For curriculum development, academic writing courses should integrate corpus-based examples that demonstrate authentic usage patterns, allowing students to see how successful writers actually employ these devices rather than relying on formulaic templates that lead to overuse. Research-wise, future studies should expand beyond frequency analysis to examine the actual communicative impact of discourse marker choices through reader-response studies, while also investigating how different cultural and linguistic backgrounds influence marker selection patterns. Institutionally, writing centers and assessment policies should incorporate discourse marker appropriateness as a criterion for effective academic writing, training tutors and instructors to recognize and address formulaic language use that may signal deeper issues with rhetorical flexibility. These recommendations collectively aim to help non-native writers develop more sophisticated, contextually appropriate discourse marker usage that enhances rather than diminishes their academic arguments.

Acknowledgment

References

Ağçam, R. (2017). Contrastive Discourse Markers in Native and Non-native English Academic Writing. *Journal of Linguistics and Language Teaching, 8*(2), 29–42.

Ali, E. A. (2023). The use of discourse markers in Jordanian Spoken Arabic. International Journal of Linguistics, 15(2), 82–(end not specified). https://doi.org/10.5296/ijl.v15i2.20805

Alsop, S., & Nesi, H. (2009). Issues in the Development of the British Academic Written English (BAWE) Corpus. *Corpora, 4*(1), 71-83. https://doi.org/10.3366/E1749503209000227

Altenberg, B. & M. Tapper. (1998). *The Use of Adverbial Connectors in Advanced Swedish Learners' Written English*. In Learner English on computer, ed. S. Granger. London: Longman.

Anthony, L. (2023). *AntConc* (Version 4.3.1) [Computer software]. Waseda University. https://www.laurenceanthony.net/software

Biber, D., Conrad, S., & Reppen, R. (2002). *Corpus Linguistics: Investigating Language Structure and Use*. UK: Cambridge University Press.

Boonyarattanasoontorn, P. (2017). An Investigation of Thai Students' English Language Writing Difficulties and Their Use of Writing Strategies. *Journal of Advanced Research in Social Sciences and Humanities*, *2*(2), 111-118.

Fraser, B. (1999). What are discourse markers?. *Journal of Pragmatics, 31*(7), 931–952.

Fraser, B. (2009). An account of discourse markers. *International Review of Pragmatics, 1*(2), 293–320.

Fraser, B. (2015). The combining of discourse markers: A beginning. *Journal of Pragmatics, 86*, 48–53.

Gardezi A. & H. Nesi (2009). *Variation in the Writing of Economics Students in Britain and Pakistan: the Case of Conjunctive Ties*. In M. Charles, D. Pecorari & S. Hunston (eds), Academic Writing: At the Interface of Corpus and Discourse. London: Continuum.

Granger, S. (2002). A bird's-eye view of learner corpus research. *ResearchGate*, July 2013, 1-27. https://www.researchgate.net/publication/239327978_A_Bird's-eye_view_of_learner_corpus_research

Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers. *World Englishes, 15*(1), 17–27.

Huneety, A., Alkhawaldeh, A., Mashaqba, B., Zaidan, Z., & Alshdaifat, A. (2023). The Use of Discourse Markers in Argumentative Compositions by Jordanian EFL learners. *Humanities & Social Sciences Communications, 10*(1), 41. https://doi.org/10.1057/s41599-023-01525-0

Hunston, S. & Francis, G. (2000) *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English.* Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/scl.4

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.

Jaroenkitboworn, K. (2014). A Study of English Acknowledgements Written by EFL Thai Learners, *PASAA, 47*(1), 97-128. DOI: 10.58837/CHULA.PASAA.47.1.4

Leedham, M., & Cai, G. (2013). Besides on the other hand: Using a corpus approach to explore the influence of teaching materials on Chinese students' use of linking adverbials. *Journal of Second Language Writing, 22*(4), 374–389.

Liu, M. (2013). A study of the use of discourse markers by Chinese EFL learners. *International Journal of English Studies, 13*(2), 29–48.

Salem, K. & Al-Thunayyan, A. & Alhuqbani, M. (2024). Analysis of Discourse Markers and Their Combinations in Nonnative Academic Writing. *Journal of Language Teaching and Research*. 15. 2010-2019. 10.17507/jltr.1506.26.

University of Lancaster. (n.d.). Log-likelihood calculator. http://ucrel.lancs.ac.uk/llwizard.html

Wongsothorn, A., Hiranburana, K., & Chinnawongs, S. (2002). English language teaching in Thailand today. *Asia Pacific Journal of Education, 22*(2), 107–116. https://doi.org/10.1080/0218879020220210