

A CORPUS-BASED STUDY ON DISCOURSE MARKERS IN EXPOSITORY WRITING OF CHINESE COLLEGE STUDENTS

**Jiahuang Chen^{1*}, Cholthicha Sudmuk², Montarat Rungruangthum³,
Behrad Aghaei⁴, Suwaree Yordchim⁵**

^{1*,2,5} Graduate School, Suan Sunandha Rajabhat University, Bangkok, Thailand

³Faculty of Liberal Arts, Rajamangala University of Technology Phra Nakhon, Bangkok, Thailand

⁴Department of Middle East Studies, University of Michigan, Michigan, USA

E-Mail: s64584952004@ssru.ac.th^{1*}, cholthicha.su@ssru.ac.th², lect.montarat@gmail.com³,
aghaeib@umich.edu⁴, suwaree.yo@ssru.ac.th⁵

ABSTRACT

This study has applied a corpus-based approach with AntConc as a tool, following Fraser's (1999) Taxonomy of discourse markers to investigate the types and frequency of discourse markers in expository writings of 290 Chinese college students. There are four types of discourse markers: elaborative, inferential, contrastive, and causative, found in Chinese students' expository writings, however, there were none of the topic-relating markers. Among those four discourse marker types, elaborative markers were the most frequently used markers, followed by inferential, contrastive, and causative. Remarkably, coordinating conjunctions, such as *and*, *but*, *so*, were excessively used.

Keywords: Corpus-based approach, discourse markers, expository writing

INTRODUCTION

Expository writing is a style of composition that exposes or explains a topic using facts. The required skill to master expository writing is the ability to explain a topic logically and straightforwardly based on facts [1]. Expository writing ability is consistently considered necessary for English as a Second Language (ESL) or English as a Foreign Language (EFL) as it is tested in main language proficiency tests at the international level, such as the International English Language Testing System (IELTS) or the Test of English as a Foreign Language (TOEFL).

In China, expository writing is often tested in English exams for Chinese college students before graduation, such as the Chinese National Test for English Majors (TEM), Chinese College English Test Band 4 or Band 6 (CET4 or CET6). These tests emphasize that expository writing ability is one of the crucial skills of English proficiency in Chinese higher education.

In expository writing, discourse markers (DMs), a class of lexical expressions, such as *because*, *so*, *and*, *however*, *but*, etc., have been found frequently employed in expository writing. The efficient utilization of DMs is considered an essential element as they supply unity to the text and link the ideas of the topics [2].

However, the inefficient utilization of DMs in second-language writing is perceived as a rookie blunder [2]. In the case of Chinese ESL learners' expository writing, for instance, such discourse markers as *and*, *so*, *but*, are overused [3][4] whereas those markers such as *on the other hand*, *otherwise* [5], and conjunctions *and*, *but*, are misused[6].

The ineffective use of DMs, such as overuse, underuse, or misuse of DMs, by Chinese learners of English has been regarded as a persistent issue in their expository writing. In addition, the previous study on DMs by Chinese learners of English is rather few, and the methodology of the studies is open to question [7]. This paper aims to investigate the use of Discourse markers in expository writing by Chinese college students according to Fraser's (1999) taxonomy [8], which is the most comprehensive classification in written discourse [9]. The data in this study were analyzed by using a corpus-based approach because compared with manual or empirical data analysis; it is easier to conduct empirical studies on language variety and use, which leads to much more broadly applicable and valid research findings than would otherwise be possible [10].

RESEARCH OBJECTIVES

The study aims to 1) investigate types of DMs in Chinese College Students' English Expository Writing in Chinese Learner English Corpus (CLEC), using AntConc software; 2) evaluate the frequency of each type of DMs in Chinese College Students' English Expository Writing in Chinese Learner English Corpus (CLEC), using AntConc software.

RESEARCH SCOPE

This study focused on 290 expository writings with identical topics, written by first or second-year Chinese college students from various educational institutes in Chinese Learner English Corpus (CLEC) [11].

METHODOLOGY

This study has employed the corpus-based approach with AntConc as the corpus analysis tool to investigate types of discourse markers and their frequency in Chinese colleges' expository writings in CLEC.

The research process consists of the following steps: 1) Collect data from CLEC; 2) Prepare data by setting up criteria for choosing an appropriate expository topic, then clean data by removing the personal information, and compiling the cleaned data into Txt. Files; 3) Analyze data with Fraser's (1999) taxonomy criteria by AntConc software; and finally, 4) Conclude the findings.

DATA COLLECTION

The 290 expository writings with identical topics, written by first or second-year Chinese college students from various educational institutes were collected from CLEC, which is one of the written learner corpora in China. CLEC contains English writings collected from Chinese learners of English

with differing levels of proficiency, covering senior secondary school students, and English and non-English-major university students.

The criteria for choosing an appropriate topic were set up as follows: The topic should not be too broad or too technical and the compositions should be written by Chinese freshmen or sophomores. According to the criteria, the topic “Global Shortage of Fresh Water in Developing Countries” was chosen for the analysis. The data consists of two hundred and ninety expository compositions. Each of them ranged from 120-150 words. The overall data contained 47,447 words.

DATA PREPARATION

All data has been cleaned, that is, some personal information was removed. The cleaned data were compiled together into Txt. Files before data processing by using AntConc [12]. This program was utilized in this study as a standalone program, language compatibility, and user familiarity.

DATA ANALYSIS

The corpus-based approach with AntConc software as the corpus analysis tool was applied to the data following Fraser’s (1999) Taxonomy of DMs. This taxonomy delimits the types and variants of DMs in each type so specifically that data can be piloted easily in corpus-based tools like AntConc. According to Fraser (1999), a class of lexical expressions known as DMs is derived mostly from the syntactic classes of prepositional phrases, adverbs, and conjunctions. They indicate a link, with a few exceptions, between the interpretations of the previous section, S1, and the part they introduce, S2. Their more precise interpretation is “negotiated” by the linguistic and intellectual context, but their fundamental meaning is procedural rather than conceptual. There are two kinds: those that connect the topic of S2 to that of S1, and those that relate the explicit interpretation of S2 with some aspect related to the segment, S1. A summary of Fraser’s taxonomy of 97 DMs in 1999 is presented in Table 1.

Table 1: Fraser’s (1999) taxonomy of DMs

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

Since DMs are prevalent in speech and writing, no consensus on such issues as terminologies, definitions, classifications, and functions of DMs has been reached. Even though Fraser (1999) had specifically delimited the types and variants of DMs, those variants can never be counted as DMs under any circumstance. To scientifically count the frequency and types of DMs in AntConc, the data criteria according to Fraser's (1999) ideas were set up for this study as follows:

- 1) The canonical form is represented <S1. DM + S2>, so S1 and S2 are usually separated by a period or a comma otherwise DMs will not be counted. For instance, in the sentence "People take it for granted that the fresh water in the world is limitless, *but* we are short of fresh water very much" will not be counted as a DM.
- 2) A DM links a discourse segment, namely a complete sentence, rather than a word, a phrase, or an elliptical expression. For instance, *because of* in "There is a freshwater shortage in the developing countries, *because of* the pollution of the water" will not be counted as a DM.
- 3) Adverbs or prepositional phrases as DMs are usually detached from S2, namely, a DM is directly preceded or followed by a comma in S2. For instance, *then* in "We must protect the fresh water, so we can have much fresh water *then*." will not be counted as a DM.
- 4) Only five types of DMs are counted, namely, contrastive markers, elaborative markers, inferential markers, causative markers, and topic-relating markers. Other subclasses of markers will not be counted in the research.
- 5) A DM combined with other word(s) to form a new phrase that is not regarded as a DM will not be counted as a DM. For instance, either *and* or *so* in the phrase *and so on* will not be counted as a DM.

The data were analyzed in two main steps: the first step is counting the frequency of DMs following the taxonomy of DMs proposed by Fraser (1999), using a corpus program known as AntConc 4.2.4. and the second step is listing DMs counted in the types of DMs proposed by Fraser (1999).

RESULTS

By applying the corpus-based approach with AntConc as the corpus analysis tool following Fraser's Taxonomy of DMs, the findings can be summarized as follows:

1) Types of DMs in Chinese College Students' English Expository Writing in CLEC

Four types of DMs in Fraser's Taxonomy, including elaborative, inferential, contrastive, and causative, were found in Chinese college students' English expository writing. There were 32 DMs in the Chinese college expository compositions. However, the fifth type of DMs, topic-relating markers, did not appear in any Chinese college students' English expository writing. The four types of MDs are shown in Table 2.

Table 2: Types of DMs in Chinese College Students' English Expository Writing

Types	Discourse Markers
Elaborative Makers	and, in addition, furthermore, moreover, besides, also, above all, otherwise, what is more, for another thing, that is to say, or
Inferential Makers	so, then, of course, as a result, thus, therefore, hence, consequently, in this case
Contrastive Markers	but, on the other hand, however, though, on the contrary, although, in spite of, whereas
Causative Markers	because, since, after all

2) The frequency of each type of DMs in Chinese College Students' English Expository Writing in CLEC

Among the four types of DMs, elaborative markers are the most frequently used by Chinese college students (322 times), followed by inferential markers (314 times), contrastive markers (274 times), and causative markers (86 times). The frequency of each type of DMs is presented in Table 3.

Table 3: Frequency of the four types of DMs in 290 Chinese college students' expository writings

Types	Concrete Markers	Frequency	Types	Concrete Markers	Frequency
Causative Markers	because	68	Contrastive Markers	but	180
	since	15		on the other hand,	37
	after all,	3		however,	31
in total		86		though	14
Elaborative Markers	and	230		on the contrary	7
	in addition,	29	Inferential Markers	although	3
	furthermore	13		in spite of	1
	moreover	13		whereas	1
	besides	10		in total	274
	also	8		so	211
	above all	6		then	42
	otherwise	4		of course,	16
	what is more	3		as a result,	15
	for another thing	2		thus	13
	that is to say	2		therefore	10

or	2	hence	5
		consequently	1
in total	322	in this case	1
		in total	314

The elaborative makers have the largest number of variants (12), followed by inferential markers (9), contrastive markers (8), and causative markers (3). The frequency of variants of each type except topic relating markers varies greatly. For instance, in elaborative markers, the marker appears 230 times, but dozens of markers *likewise* appear zero times. A similar phenomenon exists in inferential markers and contrastive markers.

The most frequently used DM by Chinese college students in their expository compositions are conjunctions, three coordinating conjunctions *and*, *but* *so* in particular. In contrast, there was no usage of topic-relating markers in the data. The least frequently used markers involve 63 markers in the four types which appear zero times. It should also be noted that a lot of markers in Fraser's (1999) taxonomy have never or been restrictively used. For instance, 15 markers, such as *in spite of*, *consequently*, *for this/that case*, are restrictively used (less than 10 times), and 65 markers, including all topic relating markers, are not used by Chinese college students at all. The zero occurrences of 65 markers are shown in Table 4.

Table 4: Zero occurrences of 65 markers in Chinese college students' expository writings

Types	Concrete Discourse Markers
Elaborative Markers	better yet, more to the point, on top of it all, to cap it all off, I mean, in particular, well, in any event, namely, parenthetically, equally, analogously, by the same token, likewise, correspondingly, similarly, be that as it may
Inferential Markers	accordingly, as a consequence, as a logical conclusion, because this, because that, for this reason, for that reason, in any case, it can be concluded that, in conclusion, to conclude, under these conditions, under those conditions, all things considered
Contrastive Markers	in contrast to this, in contrast to that, nevertheless, still, in comparison to this, in comparison to that, in comparison with this, in comparison with that, conversely, contrast to this, contrast to that, rather than this, rather than that, instead of (doing) this, instead of (doing) that, nonetheless, despite this, despite that
Causative Markers	for this reason, for that reason
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, while I think of it, to return to my topic, with regards to, to return to my topic, with regards to

To summarize, there are five types of DMs in Fraser's (1999) Taxonomy, 97 DMs in total. However, the study showed that only four types of 32 DMs appeared in 290 Chinese college students' expository compositions, but 65 DMs were not used in the compositions at all. Among the 32 DMs, there were 12 variants of elaborative markers, followed by 9 variants of inferential markers, 8 variants of

contrastive markers, and 3 variants of causative markers. No variants of topic relating markers were employed by Chinese college students in their expository compositions. It seemed that frequency of variants of DMs was related to genres of writing and styles of discourse, for instance, elaborative markers often appeared frequently in expository writing due to their role in providing additional information, clarifying concepts, and connecting ideas[13]; however, topic relating markers were less frequent in written discourse compared with spoken discourse because written discourse tended to be more planned and carefully structured, with explicit organization and clearer topic development [14]. In addition, the frequency of DMs in Chinese college students' expository writing might also be related to the Chinese college students' unproficiency in English language, because highly proficient speakers or writers tend to use discourse markers more strategically and flexibly with a wider range of DMs and a better control over their frequency [15].

CONCLUSION AND FUTURE WORK

This study was conducted by a corpus-based approach with AntConc as the corpus analysis tool following Fraser's Taxonomy of DMs to investigate types and their frequency of DMs in Chinese colleges' expository writings in CLEC. Four types of MDs, elaborative, inferential, contrastive, and causative, were found. There were no topic-related markers in the data. The elaborative markers are the most frequently used ones, followed by inferential markers, contrastive markers, and causative markers.

The findings of the most frequently used DMs by Chinese college students in their expository writings are conjunctions accorded with He's [3] and Hu et al.'s [4] findings, namely, those markers like so, and, but are overused in Chinese EFL learners' expository writing. Moreover, the frequency of DMs in Chinese college students' expository writing reveals the Chinese college students' unproficiency in English language with a narrow range of DMs and awkward control over their frequency.

Discourse markers in English writing should receive more attention in teaching English as a foreign language due to the revealed findings of this study that Chinese college students have limited usage of DMs in their expository writings.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to Suan Sunandha Rajabhat University for invaluable help throughout this research. Special thanks also go to Dr. Cholticha Sudmuk, Dr. Montarat Rungruangthum, Dr. Behrad Aghaei, and Dr. Suwaree Yordchim for their professional guidance and revision.

REFERENCES

- Bruffee, K.A. (1993), *A Short Course in Writing: Composition, Collaborative Learning, and Constructive Reading*, New York: Harper Collins College Publishers.
- Al-khazraji, A. (2019), Analysis of Discourse Markers in Essays Writing in ESL Classroom, *International Journal of Instruction*, Vol.12, No.2, Pp.559-572.
- He, A.P. (2002), On the discourse markers so. In P. Peters, P. Collins & A. Smith (Eds.), *New Frontiers of Corpus Research: Papers from the Twenty First International Conference on English*

- Language Research on Computerized Corpora Sydney (Pp. 41-52), Amsterdam/New York: Rodopi.
- Hu, Z. et al. (1982), Some linguistic differences in the written English of Chinese and Australian students, *Language Learning and Communication*, Vol.1, No.1, Pp. 39-49.
- Dan-ni, S., & Zheng, C. (2010), Analyzing the micro coherence in English writing and implications for the teaching of English writing, *Sino-US English Teaching*, Vol.7, No.12, Pp.7-10.
- Zhang, M.S. (2000), Cohesive Features in the Expository Writing of Undergraduates in Two Chinese Universities, *RELJ Journal*, Vol. 31, No.1, Pp.61-95.
- Flowerdew, J.& Tauroza, S. (1995), The Effect of Discourse Markers on Second Language Lecture Comprehension, *Studies in Second Language Acquisition*, Vol.17, No.4, Pp.435–458.
- Fraser, B. (1999), What Are Discourse Markers? *Journal of Pragmatics*, Vol.31, No.7, Pp. 931–952.
- Rahimi, M. (2011), Discourse markers in argumentative and expository writing of Iranian EFL learners, *World Journal of English Language*, Vol.1, No.2, Pp.68-78.
- Biber, D. & Reppen.R. (2015), Introduction in Biber, D.&Reppen.R.(Eds.), *The Cambridge Handbook of English Corpus Linguistics*, (Pp.1-7). Cambridge: Cambridge University Press.
- Gui, S.C.&Yang, H.Z. (2003), *Chinese Learner English Corpus*. Shanghai: Shanghai Foreign Language Education Press.
- AntConc version 4.2.4.(2023), Retrieved November 20, 2023, from <http://www.laurenceanthony.net/software/antconc/>.
- Biber, D. & Gray, B. (2010), Challenging stereotypes about academic writing: Complexity, elaboration, explicitness, *Journal of English for Academic Purposes*, Vol.9, No.1, Pp.2-20.
- Schourup, L. C. (1999), Discourse markers, *Lingua*, Vol.107, No.3-4, Pp.227-265.
- Fraser, B. (1990), An Approach to Discourse Markers, *Journal of Pragmatics*, Vol.14, No.3, Pp. 383-395.