Cohesion in Academic Writing: A Comparison of Essays in English Written by L1 and L2 University Students

Zhongqing He

School of Foreign Studies, University of Science and Technology Beijing, China

Abstract—Cohesion plays an important role in ensuring clarity, appropriateness, and comprehensibility in text. This study compares L1-Chinese and L1-English university students' use of cohesion in English essays and examines the cohesive features in academic writing of L2 students from three perspectives: density of cohesion, distribution of cohesive devices, and distance of cohesive ties. Data consist of 126 student-produced essays in two corpora: 63 by L1-Chinese students from the TECCL corpus and 63 by L1-English students from the BAWE corpus. Findings indicate that L2 students have a lower density of cohesion in their academic writing. Their essays are marked by underuse of lexical cohesion and demonstrative reference, initial positioning of conjunctions, and heavy use of temporal conjunction. In addition, L2 students use significantly fewer immediate and remote cohesive ties, and the length of the material between ties is found to be inadequate compared to L1 students' writing. The study has important implications for L2 writing pedagogy in China.

Index Terms—cohesion, L2, Chinese university students, academic writing

I. INTRODUCTION

Cohesion is a crucial language resource that contributes to the overall unity of a text (Thompson, 2004, p.179). Since Halliday and Hasan (1976), the study of cohesion has attracted much attention among scholars in such fields as text analysis and language teaching. In the past years, much of the study has focused on the relationship between cohesion and coherence (e.g., Carrell, 1982; Tanskanen, 2006), and the use of cohesion in texts of various registers or genres, such as conversations (González, 2011, 2013), business discourse (Johns, 1980), literature works (Moini & Kheirkhah, 2016), government documents (Trebits, 2009), and legal texts (Fakuade & Sharndama, 2012).

With the number of second language (L2) students increasing exponentially in universities in English-dominant countries in recent years (Staples & Reppen, 2016), cohesion usage in L2 students' writing has begun to draw increasing attention. A growing body of studies suggest that exploring how certain lexico-grammatical patterns are employed in a text to create unity provides information that will help L2 students improve cohesion in their writing (Aktas & Cortes, 2008). While many of these investigations have contributed much to our understanding on how L2 learners use cohesion in their writing, it is important to note that most published studies have been largely focused on cohesion and its effects on writing quality (Chiang, 2003; Crossley, 2012; Crossley, Kyle, & McNamara, 2016), cohesion errors (Liu & Braine, 2005; Liu & Qi, 2010; Ong, 2011), and cohesive features in students' writing (Hinkel, 2001; Rahman, 2013). These studies largely involve the counting of frequencies and the examination of realizations of certain cohesive devices, with very few studies focusing on the overall density of cohesion, and even fewer on distance of cohesion. While a few studies have examined all the three aspects of cohesion proposed by Halliday and Hasan (1976), namely distribution of cohesive devices, cohesion density, and distance of cohesive ties, they consist of data drawn from texts written by L2 student writers only. Little, however, is known about the three aspects of cohesion in English essays of L2 university students compared to L1 university students.

This study addresses these gaps by adopting Halliday and Hasan's (1976) framework and undertaking a comparative study of cohesion in English essays written by L1-Chinese and L1-English university students. Such an approach offers the opportunity to examine more fully the use of cohesion in academic writing of Chinese learners of English as a foreign language (EFL), and thus adds to a growing body of research on cohesive features in English essays of L1 and L2 student writers.

II. COHESION AND ACADEMIC WRITING

Cohesion refers to the linguistic devices by which bits of the text are connected to each other. It is one of the major resources in linguistic system that contributes to text construction (Halliday and Hasan, 1976, p.vii). Together with coherence, cohesion enables a text to have "texture", the property which distinguishes a text from a non-text. So far, Halliday and Hasan (1976) has given "the most comprehensive" treatment of the subject of cohesion (Akindele, 2011; Brown & Yule, 1983) and since then the study of cohesion has attracted considerable attention in areas of text analysis and language teaching.

In the past decades, studies of cohesion have largely pivoted around two issues, the interaction between cohesion and coherence, and cohesion usage in texts of various types. It is generally accepted that cohesion and coherence are not the same thing (Carrell, 1982; Eggins, 1994; Thompson, 2004). The distinction between cohesion and coherence is the distinction between something that exists within text and something that exists "in the mind of the writer and reader" (Thompson, 2004, p.180), between the internal properties and the contextual properties of a text (Eggins, 1994, p.87). Although they are conceptually interrelated (Witte & Faigley, 1981; Yang & Sun, 2012), there is no causal relationship between the two constructs (Carrell, 1982; Castro, 2004). Previous studies showed that a cohesive text may be minimally coherent (Witte & Faigley, 1981), and that at the same time, a coherent text will likely, but not necessarily, be cohesive (Norment, 1994).

Cohesion usage in texts of different registers or genres has also received much attention of scholarship in the past years. These studies demonstrate that the use of cohesive devices is somehow affected by the registerial and generic features of texts. Crowhurst (1987), for example, examined cohesion in argument and narration at three grade levels and found that narrative has higher use of cohesive ties than argument. Johns (1980) compared the use of cohesive items in letters, reports, and textbooks. The results showed that lexical cohesion is the most common cohesive device in all three genres and that the occurrence of subtypes of lexical cohesion varies across genres. González (2011, 2013) focused on lexical cohesion in conversational discourse. The results showed that repetition is the most frequent lexical cohesion device and that lexical cohesion plays an important role in establishing interpersonal relationships in conversations. In addition, studies also revealed that reference was used predominantly in legal texts (Fakuade & Sharndama, 2012), and that conjunction contributes to the forming of textual organization patterns in government documents (Trebits, 2009).

Recently, however, in response to the increasing number of international students in universities in English-dominant countries, such as the US, scholars in cohesion studies have gradually shifted their attention to cohesion usage in L2 students' writing. One of the key issues here is the correlation between cohesion and writing quality of L2 students. The results, however, are rather contradictory (Liu & Braine, 2005). Some researchers have found that there is no direct correlation between cohesion and the quality of writing (e.g., Castro, 2004; Zhang, 2000). Other researchers, however, have found the opposite. For example, Chiang (2003) argues, through his analysis of cohesive conditions and perception of writing quality in L2 learners' writing, that cohesion is "the best predictor of writing quality". Yang and Sun (2012) reached a similar conclusion and showed that the correct use of cohesive ties correlated significantly with the writing quality. Though it is still unclear whether there is a definite correlation between cohesion and writing quality, it is generally accepted that cohesion is an important aspect of L2 learners' writing quality.

Additionally, cohesion errors in L2 students' writing have received much attention, particularly among Chinese scholars. The errors that L1-Chinese learners make in using cohesion when they write in English range from overuse of conjunctions (Field & Yip, 1992; Zhang, 2000), misuse of collocations (Liu & Braine 2005), of references (Liu & Braine, 2005; Ong, 2011; Zhang, 2000), and of lexical devices (Zhang, 2000), to restricted use of lexical cohesion (Liu & Braine, 2005; Zhang, 2000). These studies suggest that L1-Chinese learners have difficulty in using cohesive devices correctly and appropriately in their writing (Yang & Sun, 2012; Wang & Sui, 2006), and that it is important to conduct a comprehensive analysis of L1-Chinese student writers' use of cohesion in their writing to improve this situation.

Although these studies have contributed much to our understanding on cohesive features in L2 learners' writing, there are still some underexplored areas which necessitate further research. First, these studies have mostly focused on certain types of cohesive devices, particularly conjunction (e.g., Bolton, Nelson, & Hung, 2002; Granger & Tyson, 1996; Lei, 2012; Shi, 2017) and lexical cohesion (Zhao, 2014), leaving the other types relatively unexplored. Second, these studies largely involve counting frequencies and examining realizations of cohesive devices. There is a dearth of literature investigating the overall density of cohesion and distance of cohesive ties in L2 students' English essays. While a few studies have examined all the three aspects of cohesion in L2 students' writing, they consist of data taken from L2 learners' writing only. Little, however, is known about the three aspects of cohesion in L2 learners' essays compared to those of L1 learners. The present study tries to fill the need by adopting a comparative study of English essays written by L1-Chinese and L1-English university students. Specifically, this study addresses the following research questions:

(1) What are the similarities and differences in the use of cohesion in English essays written by L1-Chinese and L1-English university students?

(2) What are L1-Chinese university students' general patterns in using cohesion in English essays in terms of density of cohesion, distribution of cohesive devices, and distance of ties?

III. METHOD

A. Data Collection

This study explores the cohesive features in L2 learners' academic writing by comparing English essays written by L1-Chinese and L1-English university students. To this end, the data for this study consist of essays written by L1-Chinese and L1-English students in two corpora. The L1-Chinese essays are derived from the essay subset of Ten-thousand English Compositions of Chinese Learners (TECCL) Corpus (Xue, 2015), which contains 9,864 texts written by Chinese EFL learners at different levels in 32 provinces in China. These texts are compositions produced in academic tasks within the English curriculum system, including assignments in and after class, and compositions in

mid-term and end-of-term exams. The L1-English essays come from the British Academic Written English (BAWE) corpus, which consists of 1,953 essays written by L1 speakers of English in UK.

The motivation for comparing essays written by L1-Chinese and L1-English university students is to use the latter as a reference point in order to shed light on similarities and differences in cohesion usage between L1 and L2 students. It is important to acknowledge that academic writing is also a "second language" to native-English-speaking students (Hyland, 2016), and a comparative analysis of essays written by L1 and L2 students, if not used properly, might unintentionally engender a false conception of good writing based on practices of L1-English students (Heng Hartse & Kubota, 2014). Therefore, in the present study, patterns in cohesion usage in essays written by L1-English students are not used as the "norms" or "benchmarks" for L1-Chinese students. Rather, they are used as one point of reference which might lead to a better understanding of L1-Chinese university students' use of cohesion in their essays and provide possible insights for L2 writing pedagogy (Bychkovska & Lee, 2017; Leedham, 2015).

To make the texts from the two corpora more comparable, three factors were taken into consideration: (1) background information; (2) topics; and (3) length. First, the texts we chose are all written by university students in the humanities and social sciences, particularly education, history, philosophy, politics, and sociology. Second, all the texts are essays produced in academic contexts, belonging to the category of "academic discourse" as defined in Hyland (2009, p.1), and covering similar topics, ranging from obesity, death penalty, college life, and population to air pollution. Third, preference was given to essays between 500 and 1200 words long, and no essay of more than 1300 words was included, in order to allow for a more in-depth analysis. Attention was also given to the wholeness of the texts, each text having a complete structure, with beginning and ending paragraphs, and body paragraphs.

With the help of the background information spreadsheets in the two corpora, the three factors above were utilized for text selection. After the initial screening, 96 texts from the TECCL Corpus and 74 texts from the BAWE corpus were obtained. Then on further inspection of the demographic information of the writers, 63 texts were chosen from these texts in the two corpora respectively to constitute the data for the present study. Therefore, there were altogether 126 texts, 63 of which were by L1-Chinese university students, labeled as CUS (Text 1–63), and 63 by L1-English university students, labeled as EUS (Text 1–63). Table 1 provides information with regard to the total number of words, the average number of words per text, the total number of T-units, and the average number of T-units per text in each corpus.

TABLE 1
GENERAL DESCRIPTION OF DATA

Data	No. of texts	Total words	Average words	Total T-units	Average T-units
CUS	63	50924	808.32	3452	54.79
EUS	63	51674	820.22	2921	46.37

B. Analytical Framework

The analytical framework was largely based on Halliday and Hasan (1976), which outlines a tripartite framework for the analysis of cohesion in text. The framework focuses on three dimensions simultaneously: (1) density of cohesion; (2) distribution of cohesive devices; and (3) distance of cohesive ties. The current study adopts this three-dimensional analytical framework.

Following Castro (2004), and Norment (1994), cohesion density is examined both per T-unit and per 100 words in this study. For distribution of cohesive devices, this study focuses on three types of cohesive devices, namely reference, conjunction, and lexical cohesion, on the grounds that the pilot analysis corroborated that substitution and ellipsis are very rarely used in the formal written language (Christiansen, 2011; Halliday & Hasan, 1976).

Distance of ties is defined in this study as the distance, or "text-span" (Witte & Faigley, 1981), between a cohesive tie and the presupposed element. There are four basic types of distances of ties: immediate, mediated, remote and cataphoric. When the presupposing element in a cohesive tie is immediately adjacent to the presupposed element in the previous T-unit, the distance is defined as immediate. If there are some intervening T-units between the presupposing element and the presupposed element, the distance might be mediated, if these T-units form a chain of presupposition, or remote, if they are not involved in a chain of presupposition. Finally, if a cohesive tie refers to the presupposed element in the following clause, the tie is defined as cataphoric.

In practice, however, types of distance of ties can be very complicated because cohesive ties tend to form "cohesive chains" (Yang, 1989), rather than occur in isolation. Therefore, a further distinction is usually made in mediated ties: immediate-mediated and remote-mediated (González, 2011). The distinction lies in whether a chain of mediated ties occurs in subsequent or non-subsequent T-units: if it occurs in subsequent T-units, it is immediate-mediated; if it occurs in non-subsequent T-units, then it is remote-mediated. In addition, mediated and remote cohesive ties can form a larger chain of cohesive ties, namely mediated-remote (Witte & Faigley, 1981).

C. Procedure

This study adopts the following procedures:
(1) locating and calculating words and T-units in each text;

(2) identifying cohesive devices in each text, and coding them with the help of an Excel sheet, using conventions largely based on Halliday & Hasan (1976, p.333-338);

(3) identifying the distance of cohesive ties in each text, and coding the information in an Excel sheet;

(4) counting the frequencies of cohesive ties in each text, and calculating the number of cohesive ties per T-unit and per 100 words;

(5) calculating the density of each type of cohesive device in the two types of data;

(6) conducting descriptive analyses and T-tests to see whether there is a significant difference in terms of density of cohesion, distribution of cohesive device, and distance of ties in the two categories of essays.

The coding of the data was conducted manually by the author and another researcher, who is also a teacher of English in the tertiary level. Both the inter-raters attended trainings in the coding of cohesive devices and of distance of cohesive ties. The inter-rater reliability result using Cronbach Alpha is 0.836, which suggests that there is very good agreement among the inter-raters. The Statistical Package for Social Sciences (SPSS) software (Version 20) was used to conduct T-tests in Procedure (6). A P value ≤ 0.05 indicates that there is a significant difference between the frequencies of two variants.

IV. RESULTS AND DISCUSSION

A. Density of Cohesion

To examine the density of cohesion in the two categories of essays, the number of cohesive ties in each text was calculated, and then cohesive density was computed by dividing the number of ties by the number of T-units and by the number of 100 words.

TABLE 2
COHESION DENSITY IN CUS AND EUS ESSAYS

Data	No. of ties	Density	
		per T-unit	per 100 words
CUS	6632	1.92	13.02
EUS	8341	2.86	16.46

As shown in Table 2, the density of cohesion in L1-Chinese essays, in terms of the number of cohesive ties both per T-unit and per 100 words, is lower than that in L1-English essays. The T-test result (t = -3.231, p = 0.002) shows that there is a significant difference, which indicates that L1-Chinese learners used significantly fewer cohesive ties essays than L1-English learners.

This finding is supported by Liu and Braine (2005), and Zhao (2014). Furthermore, the result indicates that the lower density of cohesion in L1-Chinese students' essays might be attributable to the tendency to use much shorter sentences in English writing compared to L1-English students. As Wang & Slater (2016) pointed out, the mean sentence length and the mean clause length in L1-Chinese students' essays were significantly lower compared to L1-English students' essays and L1-Chinese students also used more clauses, which, in all likelihood, result in a larger quantity of T-units in CUS essays. Given the already salient difference in the total number of cohesive ties in the data, the cohesive density, particularly in terms of the average number of cohesive ties per T-unit in CUS essays, is lower than that in EUS essays.

B. Distribution of Cohesive Devices

1. General distribution

Table 3 compares the distribution of the three major types of cohesive devices in CUS and EUS essays. The general distribution of the three types of cohesive devices is more or less the same in terms of frequency of occurrence. The most frequently used cohesive device in both the data is lexical cohesion, which accounts for 62.39% and 71.50% respectively in CUS and EUS writing, followed by reference (21.49% vs. 17.77%), and conjunction (16.12% vs. 10.73%). This finding corresponds with some previous studies (e.g., Field & Yip, 1992; Liu & Braine, 2005; Zhang, 2000), which have the same order of frequency of occurrence of cohesive devices. It also further corroborates Hoey's (1991, p.10) claim that lexical cohesion is "the dominant mode of creating texture".

TABLE 3
COHESIVE DEVICES IN CUS AND EUS ESSAYS

Type	CUS		EUS	
	N	%	N	%
Reference	1425	21.49	1482	17.77
Conjunction	1069	16.12	895	10.73
Lexical cohesion	4138	62.39	5964	71.50
Total	6632	100	8341	100

T-test result indicates that CUS writing has significantly fewer ties of lexical cohesion than EUS writing (t = -3.371, p = 0.003). This suggests that L1-English students use prominently more lexical cohesion ties in their essays than L1-Chinese students. The result is similar to some previous studies (e.g., Liu & Braine, 2005; Zhang, 2000; Zhao, 2014). T-tests also show that there is no significant difference in the distribution of reference and conjunction in CUS and EUS

writing.

2. Distribution of reference

Table 4 compares the distribution of reference in CUS and EUS essays. The most frequently used subtype of reference in CUS essays is personal reference (R1), followed by demonstrative reference (R2), and comparative reference (R3). In the use of reference ties, personal reference clearly dominates in CUS essays.

The situation in EUS essays, however, is quite different. The most frequently used subtype of reference is demonstrative reference, followed by personal reference, and then comparative reference. Furthermore, the three subtypes of reference are more evenly distributed in EUS essays than in CUS essays.

The result of T-tests shows that there is a significant difference in the use of demonstrative reference in CUS and EUS essays (t = -3.125; p = 0.001), which means that L1-English students use more demonstrative reference ties in their writing than L1-Chinese students.

TABLE 4
DISTRIBUTION OF REFERENCE IN CUS AND EUS ESSAYS

Subtype	CUS		EUS	
	N	*%*	*N*	*%*
R1	758	53.19	541	36.50
R2	369	25.90	667	45.00
R3	298	20.91	274	18.49
Total	1425	100	1482	100

In terms of the expressions of the three subtypes of reference, there are two important findings. First, in the expression of personal reference, both L1-Chinese and L1-English university students use more personal pronouns (e.g., *I*, *me*, *him*) than possessive determiners (e.g., *their*, *my*). Among the personal pronouns they employ, the three most frequently used in CUS data are *it*, *she*, and *they*, while in EUS data they are *it*, *he*, and *they*.

However, CUS and EUS essays differ from each other in the use of a pronoun which does not express a definite reference in the text. L1-Chinese students tend to be more flexible, employing not only the impersonal pronoun *one*, but also the second person plural pronoun *you* or *your*, which may have the danger of resulting in "the frequent shifting of pronominals" (Zhang, 2000). In contrast, L1-English students seem to be more conservative, usually sticking to the use of the first- person plural *we*, as illustrated in Examples (1) and (2).

(1) It is really common that there is someone plays an important role in a particular stage of your life and he or she may change **your** views of live and dig **your** potential fully. (CUS-1)

(2) Collectively, the upbeat rhythm combined with the repetition of "Forget not yet" seems to make the poem one of contrast, where **we** see a paradoxical irony between continuation and departure. (EUS-11)

Second, demonstrative reference in EUS essays is dominated by the demonstrative *this*, while CUS essays tend to employ more diverse expressions. L1-English students seem to be more proficient in the use of *this* to achieve cohesion in their writing. The demonstrative *this* is not only used as a nominal head, but, more importantly, it is often used in combination with general words or "shell nouns" (Aktas & Cortes, 2008) to create an "old-to-new flow of information" (Swales & Feak, 2012, p.43). In contrast, though also occupying the lion's share in the use of demonstrative reference ties, *this* in CUS essays is mostly used as a nominal head or as a deictic followed by the repetition of a lexical item or synonym. For example:

(3) Actually, **this** is a quarrel that have not been burst out, though I felt guilty for being impolite. (CUS-1)

(4) His early experiments involved coating of wounds with carbolic acid (phenol, a coal tar derivative), which would form

antiseptic crusts of coagulated blood on wounds. The beginning at **this** procedure were unsuccessful. (EUS-1)

3. Distribution of conjunction

Table 5 presents the distribution of conjunction in CUS and EUS essays. The first four subtypes of conjunction (additive, adversative, causal, and temporal) are more evenly distributed in CUS essays. In EUS essays, however, the use of conjunction ties is dominated by additive conjunction, which accounts for 48.94% in the total frequencies of conjunction ties used. The last subtype of conjunction, continuative, is hardly ever used in either of the two data.

T-tests show that except for temporal conjunction (t = 1.944, p = 0.041) there is no significant difference in the distribution of additive, adversative, causal, and continuative conjunction across CUS and EUS essays. This suggests that L1-Chinese university students use more temporal conjunction in their essays than L1-English students and that they do not differ much in the use of the other four subtypes of conjunction. This finding is corroborated by Zhang (2000), who found that overuse of temporals is one of the features of Chinese students' use of cohesion in writing.

TABLE 5
CONJUNCTION IN CUS AND EUS ESSAYS

Subtype	CUS		EUS	
	N	*%*	*N*	*%*
C1	386	36.11	428	47.82
C2	216	20.21	234	26.15
C3	188	17.59	119	13.30
C4	241	22.54	101	11.28
C5	38	3.55	13	1.45
Total	1069	100	895	100

Another interesting finding here is the initial positioning of conjunction in CUS essays. As observed by Ong (2011) and Zhang (2000), L1-Chinese students tend to put conjunctions such as *and*, *but*, *then*, and *so* at the beginning of a sentence. The four conjunctions in the initial position of sentences account for 45.37% of the total instances of these conjunctions used in CUS essays, while the percentage in EUS essays is only 17.02%. This may be due to the fact that these conjunctions are the most obvious expressions of conjunction and also the ones that were learnt by Chinese EFL learners at the early stage of their English learning. Therefore, when it comes to the use of conjunction in their writing, L1-Chinese students may instinctively turn to these conjunctions and put them in the initial position of a sentence in order to emphasize their role of connecting sentences.

With regard to expressions of conjunction, however, there are more dissimilarities than similarities. L1-Chinese students' use of additive conjunction was dominated by *and* while L1-English students tend to put equal emphasis on the usual expressions like *and* and *also*. Specifically, in CUS essays, more than half of the addictive conjunction are expressed by *and*; the other half are in the form of *again*, *additionally, besides*, *furthermore*, *in addition*, and *what's more*. In contrast, EUS essays mostly rely on *and* and *also* to express additive conjunction; together they account for 86.21% of the total frequencies of additive conjunction used.

Additionally, L1-Chinese learners seem to have difficulty in recognizing the difference between formal academic language and colloquial informal language in that they use many "informal items" (Leedham, 2015), such as *besides*, *what's worse*, and *what's more*, as shown in Example (5). Furthermore, as in (6), a large proportion of their use of *and* serve as an "empty filler", marking a link to previous discourse (Christiansen, 2011, p.163), but they are usually used without the cohesive purpose of either adding new information or exemplifying an idea (Ong, 2011).

(5) Her college charges her for 20 thousand Yuan a year for study, and a flight ticket to London cost her 8 thousand Yuan. **What's more**, she has to pay rent for her dormitory in London, which is super high in such a big city. (CUS-11)

(6) In China, most families have 4 grandparents, 2 parents and only one child. **And** this couple is responsible for their parents and the child. (CUS-10)

In expressing adversative conjunction, both CUS and EUS essays rely heavily on the use of *but* and *however*, but each group has its own preferences for other expressions. L1-Chinese students frequently use the adverbial *actually*, especially in the initial position of sentences, which may be another indication that they have difficulty in recognizing the features of academic language while they are writing. In contrast, in addition to *but* and *however*, L1-English students prefer the use of such expressions as *in effect*, *in fact*, and *yet*.

In terms of the expressions of causal conjunction, CUS and EUS essays generally resort to conjunctions that express a basic cause-effect relationship, such as *so*, *thus*, and *therefore*. As stated above, L1-Chinese students tend to put *so* at the beginning of a sentence in order to introduce the consequence of the preceding clause, as in Example (7).

(7) Before being under stress, we should realize that stress is a monster that does harm to ourselves, our friends and family, even our society. **So** it's of significance for us to learn more about the abominable effects of stress. (CUS-15)

In expressions of temporal conjunction, one distinctive feature is that L1-Chinese students rely heavily on the use the sequential and conclusive conjunctions such as *firstly*, *secondly*, *finally*, *in conclusion* and *in a nutshell*. These two specific types of temporal conjunctions account for 75.16% of the total instances. This may be due to the fact that they wish to demonstrate a clear and explicit connection between the arguments in their writing, as in Example (8). L1-English students, however, mainly resort to other resources, like lexical repetition, reference and additive conjunction to string their arguments, as in Example (9).

(8) **In conclusion**, I do agree with the argument that the advantages of studying abroad far outweigh its disadvantages. (CUS-45)

(9) I will also look at task processes as a group. **In addition**, I will mention some questions rising at each stage. (EUS-10)

As for continuative conjunction, since it is more commonly found in spoken text (Christiansen, 2011; Halliday & Hasan, 1976), a very limited number of instances are found in both the data. In addition to the most common continuative conjunctions such as *well*, *now*, and *of course*, L1-Chinese students also use quite a few adverbs, such as *honestly*, *surely*, and *seriously*, to serve as continuative in their writing, as evidenced in Example (10).

(10) **Seriously**, maybe it's too painful for you to endure melancholia and you may commit suicide. (CUS-55)

Here *seriously* may be the shortened form of *seriously speaking*, which in a way indicates that L1-Chinese students are conscious of using cohesive devices to make their writing cohesive and coherent, but due to their limited resources they sometimes try to create expressions of their own to serve as cohesive devices.

4. Distribution of lexical cohesion

As shown in Table 6, in the distribution of lexical cohesion, the order of frequency of occurrence of the five subtypes is the same for CUS and EUS writing. Both the data are dominated by repetition of the same words (L1), which accounts for 52.39% and 58.95% respectively, followed by collocation (L5), and then by synonyms or near synonyms (L2). General nouns (L4) and superordinates (L3) are used rarely.

Through T-tests, significant differences were found between CUS and EUS essays with regard to both repetition of the same words (t = -3.315, p = 0.002) and lexical collocation (t = -2.382, p = 0.031). This indicates that L1-English students (M = 55.81, SD = 10.84) use significantly more repetition of the same words in their essays than L1-Chinese students (M = 34.41, SD = 7.38), and that L1-English university students also use more collocations (M = 25.30, SD = 6.35 vs. M = 17.40, SD = 6.31).

TABLE 6
LEXICAL COHESION IN CUS AND EUS ESSAYS

Subtype	CUS		EUS	
	N	*%*	*N*	*%*
L1	2168	52.39	3516	58.95
L2	697	16.84	538	9.02
L3	73	1.76	115	1.93
L4	104	2.51	201	3.37
L5	1096	26.49	1594	26.73
Total	4138	100	5964	100

A similar result was found by Leo (2012), which identified lexical collocations as a sign of native-like fluency in the use of cohesive devices. However, unlike his study, the present study demonstrates that synonym or near synonym cannot be taken as an indicator of native-like competency of cohesion. As exemplified in (11), in academic writing, L1-English students tend to employ more repetition of the same lexical items in order to remain precise when they are referring to something in the preceding discourse, and when repetition of the same lexical items is not appropriate, they usually resort to collocation and synonyms or near synonyms.

(11) Scientists are constantly saying that levels of obesity are increasing. What is the reason for this increase? Obesity is increasing because the amount of energy individuals consume, from the food they eat, is higher than the amount they use up in the daily activities they perform. Body weight is maintained by a simple energy balance between the amount of energy consumed and the amount of energy expended. The national food survey has identified that over the last 25 years the amount of food we consume has decreased. This implies that the levels of obesity are increasing because the amount of physical activity performed is declining. There is also inconclusive evidence to suggest that a high fat, low carbohydrate diet favors the development of obesity. (EUS-04)
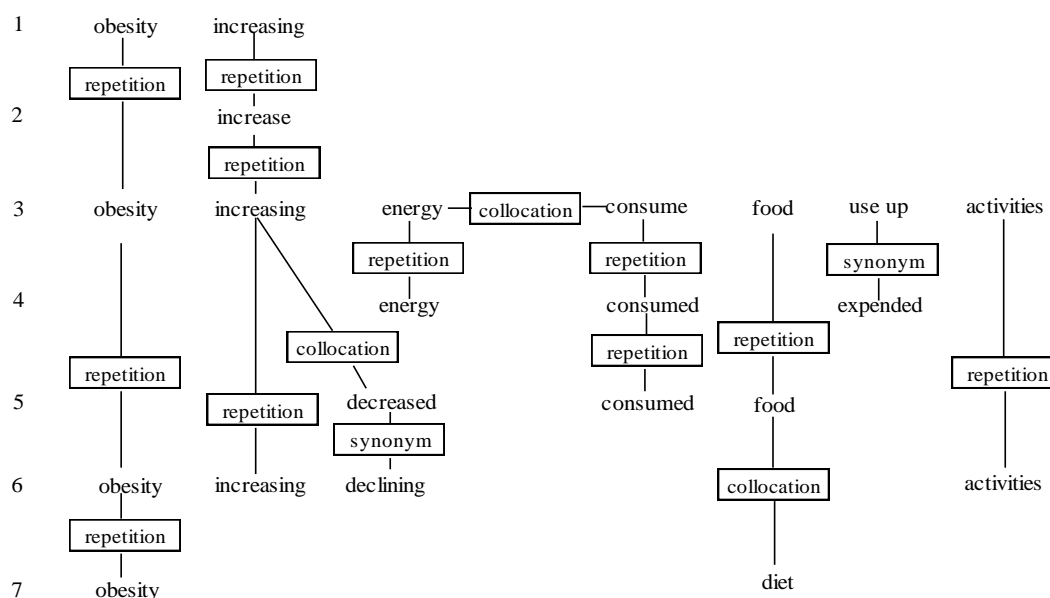

Figure 1 Lexical strings in Example (11)

Figure 1 presents the lexical strings in Example (11). In the seven T-units, the writer mostly employs repetition of the same lexical items to achieve cohesion: there are altogether eleven instances of repetition, three of collocation, and two of synonym. Therefore, simple repetition is the most important lexical cohesion employed by native speakers, followed by collocation and synonym or near synonym.

Lastly, no significant difference was found in the distribution of superordinates and general nouns between CUS and

EUS essays. Both subtypes were found to exist in very limited numbers in academic writing. In terms of the expressions of general nouns, L1-Chinese and L1-English students rely mainly on general nouns on human (e.g., *people*), place (e.g., *place*), things (e.g., *thing*), abstract entity (e.g., *change*, *difference*), and fact (e.g., *fact*, *idea*).

C. Distance of Ties

Table 7 displays in rank order the percentages of different types of distance of ties in the two data. As shown in the table, both CUS and EUS essays are dominated by immediate ties, which account for 53.38% and 51.60% respectively. These are followed by remote ties and the two subtypes of mediated ties, namely immediate-mediated and remote-mediated. Considerably less frequent are mediated-remote and cataphoric ties, which are hardly ever used in academic writing. This finding is largely consistent with Zhang's (2000) research, which identified the same order of frequency of occurrence with respect to the use of distance of ties by Chinese EFL learners in their expository writing.

TABLE 7
DISTANCE OF TIES IN CUS AND EUS ESSAYS

Subtype	CUS		EUS	
	N	*%*	*N*	*%*
Immediate	3502	53.38	4201	51.60
Remote	1801	27.45	2573	31.60
Immediate-mediated	611	9.31	649	7.97
Remote-mediated	335	5.11	378	4.64
Mediated-remote	293	4.47	313	3.84
Cataphoric	19	0.29	28	0.34
Total	6561	100	8142	100

T-tests further determined whether there was a significant difference in the distribution of distance of ties across the data. L1-English students (M = 66.68, SD = 8.92) use more immediate ties in academic writing than L1-Chinese students (M = 55.58, SD = 8.73) and this difference is statistically significant: t = -2.471, p = 0.024. In addition, there is also a significant difference in the use of remote ties between the two groups (t = -3.023, p = 0.004). L1-English students (M = 40.84, SD = 9.13) employ far more remote ties in their writing than L1-Chinese students (M = 28.59, SD = 6.02).

Another feature of the distance of ties is the average length of the intervening T-units. This feature applies to four types of distance of ties: remote, immediate-mediated, remote-mediated, and mediated-remote. The length of the intervening T-units in different types of distance of ties performs different functions in the text. In remote ties the length of intervening material usually reveals how sparsely certain themes are mentioned from time to time in the text. The greater the length, the more sparsely themes are brought up, and the more difficult it may be for readers to follow the connection between the themes. The average length of intervening material in remote ties in CUS essays is greater than that in EUS academic writing (5.74 vs. 4.28), which further confirms that L1-Chinese students use fewer remote ties in their writing, and at the same time indicates that remote ties in CUS essays are usually employed to refer to something far back in the preceding sentences.

With respect to immediate-mediated ties, however, the situation is different. The larger number of average length intervening T-units suggests that certain themes are more fully discussed in the text and that there are more T-units working together to contribute to the development of a theme. The average number of intervening T-units in immediate-mediated ties in EUS essays is larger than that in CUS essays (1.98 vs. 1.54), which may be an indication that L1-English students are more skilled in developing themes in their writing.

Similarly, remote-mediated ties are basically chains of remote ties referring to the same entity; they serve to string key themes within the text (Witte & Faigley, 1981). Like remote ties, the greater length of the intervening T-units in remote-mediated ties in CUS essays (13.21 vs. 11.18) may indicate that key themes are more sparsely scattered in CUS academic writing than in EUS writing. Last, mediated-remote ties are also employed to develop key themes in a text, and usually exist in small numbers. The larger average length of the intervening material in mediated-remote ties in EUS essays (8.67 vs. 6.26) suggests that key themes are more thoroughly developed in EUS academic writing than in CUS academic writing.

V. CONCLUDING REMARKS

This study has examined the use of cohesion in academic writing of L2 students by comparing the cohesive features in academic writing of L1-Chinese and L1-English university students. It is found that despite some similarities L1-Chinese learners displayed some unique features in their use of cohesion in terms of cohesion density, the distribution of cohesive devices, and distance of cohesive ties. Specifically, the cohesion density in L1-Chinese learners' academic writing is significantly lower than that in L1-English students' writing. L1-Chinse students' academic writing is also characterized by underuse of lexical cohesion and demonstrative reference, initial positioning of conjunctions, and heavy use of temporal conjunction. Furthermore, L1-Chinese learners used proportionally fewer immediate and remote ties of cohesion, with greater distance between ties.

These findings may have important pedagogical implications for L2 writing. This study indicates that some cohesive features in CUS essays, particularly the frequent shifting of pronouns in personal reference, informal expressions of additive and adversative conjunction, and the initial positioning of conjunctions, are largely due to L1-Chinese students' lack of knowledge of the features of formal academic writing. Therefore, it is important that when giving instructions in academic writing classes, teachers should provide students with explicit knowledge of the differences between informal colloquial language and formal academic language in English, and L2 students should have the opportunity to learn the features and styles of academic language. In addition, CUS essays are found to be marked by underuse of lexical cohesion, overuse of certain explicit cohesive devices, and inappropriate length of intervening material in ties. This suggests that L1-Chinese students might lack the knowledge and awareness of how to use cohesion in their writing appropriately (Yang & Sun, 2012; Wang & Sui, 2006). It is more constructive for teachers to teach students the features of different types of cohesive devices and provide models for analysis in L2 writing classes (Liu & Braine, 2005; Zhang, 2000). In this way L2 students might become more competent in using cohesion in their English essays.

REFERENCES

[1] Akindele, J. (2011). Cohesive devices in selected ESL academic papers. *African Nebula*, (3), 99-112.
[2] Aktas, R., & Cortes, V. (2008). Shell nouns as cohesive devices in published and ESL student writing. *Journal of English for Academic Purposes*, 7, 3–14.
[3] Bolton, K., Nelson, G., & Hung, J. (2002). A corpus-based study of connectors in student writing: Research from the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics*, 7(2), 165–182.
[4] Brown, G., & Yule, G. (1983). Discourse analysis. Cambridge: Cambridge University Press.
[5] Bychkovska, T., & Lee, J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, 30, 38–52.
[6] Carrell, P. (1982). Cohesion is not coherence. *TESOL Quarterly*, 16(4), 479–488.
[7] Castro, C. (2004). Cohesion and social construction of meaning in the essays of Filipino college students writing in L2 English. *Asia Pacific Education Review*, (5), 215–225.
[8] Chiang, S. (2003). The importance of cohesive conditions to perceptions of writing quality at the early stages of foreign language learning. *System*, 31, 471–484.
[9] Christiansen, T. (2011). Cohesion: A discourse perspective. Bern: Peter Lang.
[10] Crossley, S. (2012). Predicting second language writing proficiency: the roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35(2), 115–135.
[11] Crossley, S., Kyle, K., & McNamara, D. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, 32, 1–16.
[12] Crowhurst, M. (1987). Cohesion in argument and narration at three grade levels. *Research in the Teaching of English*, 21(2), 185–201.
[13] Eggins, S. (1994). An introduction to systemic functional linguistics. London: Pinter.
[14] Fakuade, G., & Sharndama, E. (2012). A comparative analysis of variations in cohesive devices in professional and popularized legal texts. *British Journal of Arts and Social Sciences*, 4(2), 300–318.
[15] Field, Y., & Yip, O. (1992). A comparison of internal conjunctive cohesion in English essay writing of Cantonese speakers and native speakers of English. *RELC Journal*, 23(1), 15–28.
[16] González, M. (2011). Lexical cohesion in multiparty conversations. *Language Sciences*, 33, 167–179.
[17] González, M. (2013). A reappraisal of lexical cohesion in conversational discourse. *Applied Linguistics*, 34(2), 128–150.
[18] Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 17(1), 17–27.
[19] Halliday, M. A. K., & Hasan, R. (1976). Cohesion in English. London: Longman.
[20] Heng Hartse, J., & Kubota, R. (2014). Pluralizing English? Variation in high-stakes academic texts and challenges of copyediting. *Journal of Second Language Writing*, 24, 71–82.
[21] Hinkel, E. (2001). Matters of cohesion in L2 academic texts. *Applied Language Learning*, 12(2), 111–132.
[22] Hoey, M. (1991). Patterns of lexis in text. Oxford: Oxford University Press.
[23] Hyland, K. (2009). Academic discourse: English in a global context. London & New York: Continuum.
[24] Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, 31, 58–69.
[25] Johns, A. (1980). Cohesion in written business discourse: Some contrasts. *The ESP Journal*, 1(1), 35–43.
[26] Leedham, M. (2015). Chinese students' writing in English: Implications from a corpus-driven study. London & New York: Routledge.
[27] Lei, L. (2012). Linking adverbials in academic writing on applied linguistics by Chinese doctoral students. *Journal of English for Academic Purposes*, 11, 267–275.
[28] Leo, K. (2012). Investigating cohesion and coherence discourse strategies of Chinese students with varied lengths of residence in Canada. *TESL Canadian Journal*, 29(6), 157–179.

[29] Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System*, 33, 623–636.
[30] Liu, L., & Qi, X. (2010). A contrastive study of textual cohesion and coherence errors in Chinese EFL abstract writing in engineering discourse. *Intercultural Communication Studies*, 19(3), 176–187.
[31] Moini, M., & Kheirkhah, F. (2016). Use of cohesive devices in children and regular literature: Conjunctions and lexical cohesion. *International Journal of Comparative Literature & Translation Studies*, 4(4), 12–23.
[32] Norment, N. Jr. (1994). Contrastive analysis of cohesive devices in Chinese and Chinese ESL in narrative and expository written texts. *Journal of the Chinese Language Teacher Association*, 29, 49–81.
[33] Ong, J. (2011). Investigating the use of cohesive devices by Chinese EFL learners. *Asian EFL Journal*, 13(3), 42–65.
[34] Rahman, Z. (2013). The use of cohesive devices in descriptive writing by Omani student-teachers. *SAGE Open*, 2013, October-December, 1–10.
[35] Shi, J. (2017). A corpus-based study of contrastive/concessive linking adverbials in spoken English of Chinese EFL learners. *Studies in Literature and Language*, 14(2), 17–25.
[36] Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing*, 32, 17–35.
[37] Swales, J., & Feak, C. (2012). Academic writing for graduate students, 3rd ed. Ann Arbor: The University of Michigan Press.
[38] Tanskanen, S. (2006). Collaborating towards coherence: Lexical cohesion in English discourse. Amsterdam & Philadelphia: John Benjamins Publishing.
[39] Thompson, G. (2004). Introducing functional grammar, 2nd ed. London: Edward Arnold.
[40] Trebits, A. (2009). Conjunctive cohesion in English language EU documents: A corpus-based analysis and its implications. *English for Specific Purposes*, 28, 199–210.
[41] Wang, S., & Slater, T. (2016). Syntactic complexity of EFL Chinese students' writing. *English Language and Literature Studies*, 6(1), 81–86.
[42] Wang, H., & Sui, D. (2006). Measuring Coherence in Chinese EFL Majors' Writing through LSA (Latent Semantic Analysis). *Asian EFL Journal*, 11(6), 1–24.
[43] Witte, S., & Faigley, L. (1981). Coherence, cohesion and writing quality. *College Composition and Communication*, (3), 189–204.
[44] Xue, X. (2015). Ten-thousand English compositions of Chinese learners (The TECCL Corpus) (Version 1.1). The National Research Centre for Foreign Language Education, Beijing Foreign Studies University.
[45] Yang, A. (1989). Cohesive chains and writing quality. *Word*, 40, 235–254.
[46] Yang, W., & Sun, Y. (2012). The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency level. *Linguistics and Education*, 23, 31–48.
[47] Zhang, M. (2000). Cohesive features in the expository writing of undergraduates in two Chinese universities. *RELC Journal*, 31, 61–95.
[48] Zhao, C. (2014). Lexical cohesion of Sino-British college students' EAP writing. *Theory and Practice in Language Studies*, 4(10), 2123–2128.

Zhongqing He is an associate professor of English at University of Science and Technology Beijing (USTB), China. He obtained his PhD at Beijing Normal University and was a visiting scholar at the University of Michigan (2016-2017). His current research focuses on systemic functional linguistics (SFL) and SFL-informed EAP studies.