

Lexical Cohesion of Sino-British College Students' EAP Writing

Chunhong Zhao

Huanghuai University, Henan, China

Abstract—English becomes an important tool of academic research and international communication with the rapid development of global economy. Producing a piece of EAP writing with completeness, clarity and coherence has become a pursuit of both college students and academic staff. The coherence and completeness of EAP writing is realized mainly by lexical cohesion. However, it has long been ignored by Chinese EFL learners and teachers. Using the corpora of BAWE and EAP writing by British and Chinese learners, this research compares and contrasts the indices of lexical cohesion generated by computational linguistics tool, Coh-Metrix, across different levels and between English native speakers and Chinese EFL learners. It is found that both native speakers and Chinese EFL learners tends to use less simple repetition and more complicated cohesion with the improving of English language proficiency while the average frequency of lexical cohesion employed by Chinese EFL learners is less than native speakers. Therefore lexical cohesion teaching and learning should be given more consideration for successful EAP writing for Chinese EFL learners.

Index Terms—EAP writing, Chinese EFL learners, lexical cohesion, corpus linguistics, Coh-Metrix

I. INTRODUCTION

EAP writing plays an important role not only in degree study but also pave way for future career success. More and more scholars' academic success depends on their English language proficiency and the connections between English language and specialty knowledge are increasingly closer (Hyland, 2005). Plenty of research and labour market investigation shows that college students can adapt themselves to the more competitive global and local economic development with proficient English for Academic Purpose (EAP) and English for Special Purpose (ESP) (Sun, 2010).

According to the latest UIS (UNESCO Institute of Statistics) data on the mobility of students (UIS, 2014), at least 4 million students in 2012 were enrolled in tertiary education abroad. The top destination universities are mainly in the countries with English as native language, such as USA, UK, Australia and Canada. International students not only have to acquire good communication skills and fluent English both in writing and speaking, but also absorb the western writing culture in style, structure, reference and the local policy towards academic integrity in academic writing. Even if international students acquire good score in English proficiency tests, they often find that academic papers in English-speaking universities are different from what they are used to.

As for international students from China, EAP writing is a great challenge for them due to the differences of language type and culture system. With more than 10 years of English study experience before Chinese EFL learners enter into tertiary education, many of them have acquired a certain amount of English vocabulary and grammar. However few of them could produce a piece clear, concise and coherent EAP writing. Besides the format of the EAP writing, the main reason for their difficulty in EAP writing comes from the unity of the idea, which is partly realized by the cohesion devices of the text.

In the hope of EFL learners could express themselves like a native speaker, traditional lexical-grammatical teaching method had been reproached and the modern EFL teaching focuses more on the communicative skills of the learners. It emphasizes the spoken communication and weakens the language input of lexis and grammar and thus hinders the development of language communicative skills. Writing is one of the key components of communicative capacity of EFL learners'. The level of writing proficiency is greatly influenced by the appropriate using of the words and phrases of the target language. Lexical cohesion, as an important device of discourse structure, has a great influence on discourse organization and cohesion.

According to Halliday and Hasan (1976), cohesion contributes to the wholeness and unity of the text and they divided it into grammatical cohesion and lexical cohesion. Hoey (1991) claimed that lexical cohesion is the only type of cohesion that regular forms multiple relationships and it is the dominant mode of creating texture. There have been many studies on lexical cohesion since the publication of the groundbreaking work by Halliday and Hasan in 1976. Although there are some researches carried out on the cohesion of academic writing (Witte and Faigley 1981), but only few studied especially lexical cohesion using in academic writing (Stotsky 1983, Chen 2008, Mirzapour 2011). Due to the limit resources to analysis, the relevant studies in China are about the works from students majored in English language or short argumentative writing on one topic.

With the development of corpus linguistics and computational linguistics, the authentic language resources provide first-hand research data and the computational tools make the study of large quantity information from corpora possible.

This study is going to compare the lexical cohesion using by both native speakers and Chinese EFL learners in their EAP writings with the indices generated by computational tool, Coh-Metrix. The purpose is to find the features of lexical cohesion in academic writing produced by Chinese EFL learners and thus give indication to Chinese EFL learning and teaching on EAP writing.

Four questions will be answered with this study.

1. Is lexical cohesion an effective index for distinguishing the levels of English language proficiency?
2. Can lexical cohesion be used to differentiate the writing from ENL speakers and EFL speakers?
3. Do Chinese EFL learners use more lexical cohesion than native speakers?
4. What are the main differences and similarities on lexical cohesions in EAP writing between native speakers and Chinese EFL learners?

II. LEXICAL COHESION: LANGUAGE PROFICIENCY

Lexical cohesion, as an important contributor to discourse organization, has been studied broadly in EFL and ESL on its relation to language proficiency. But due to different analyzing point, research method, data, the results are different. Mainly there are three different opinions about the relationship between the frequency of lexical devices and language proficiency: positive related, null related and negative related.

Through comparing and contrasting writing by 160 EFL learners from Arab, China, Japan and Spain, Ferris (1994) found the better EFL proficiency, the higher frequency of cohesive links. The idea is supported by the studies by Liu and Braine (2005), Clarence (2012). On the contrary, Crossley and McNarma (2012) found that users with higher EFL level tended to use less cohesions than those with lower levels.

As for the difference between EFL writers and ENL writers on cohesion links using, Mirzapour and Ahmadi (2011) found Persian EFL writers had excessive using of cohesion links; Spanish EFL writers used less cohesion links (Crossley & McNarma 2009); Johnson (1992) claimed that there was no significant difference on cohesion through comparing 20 Malaysian writers' and 20 native speakers' persuasive essays.

The different results of the previous researches are due to the different definition of cohesion, data collection and methodology. Many researchers have been trying to find the standard to direct the learning and classroom teaching of cohesion. So far there has been no satisfying result. There is even less research directly on lexical cohesion in EAP writing. This research is to find the features of lexical cohesion in EAP writings by Chinese college students. Hopefully the findings will help students and teachers to carry our more efficient work with considering the skillful using of lexical cohesion in EAP wiring.

III. COMPUTATIONAL TOOL FOR ANALYZING LEXICAL COHESION: COH-METRIX

The exactness and reliability of computational tool may help with clearing the above ambiguous studies. Coh-Metrix was developed by staff from Memphis University and its online system can be available at <http://Coh-Metrix.mephis.edu>. The initial development aim was to find the level of readability of text through analyzing lexical relations. Crossley and McNamara was the first to use it to analyze EFL writing. Its application and reliability have been tested by Granger, etc. for many times. Its indices and architecture is reported in Graesser, McNamara, Louwerse and Cai (2004) and McNamara (2011).

The five groups of indices generated by Coh-Metrix studies the lexical cohesion within one paragraph:

Index 1: Noun Overlap (NO) is the simple repetition of nous. Any other different forms (e.g. plural) are not included.

Index 2: Content Word Overlap (CWO) considers the proportion of explicit content words that overlap between pairs of sentences. Local and global indices, and their standard deviations are included in the indices. The mean of global indices is used in this research.

Index 3: Stem Overlap (SO) refers to a noun in one sentence matched with a content word (i.e., nouns, verbs, adjectives, adverbs) in a previous sentence that shares a common lemma (e.g., tree/treed; mouse/mousey; price/priced).

Index 4: Hypernymy of Nouns and Verbs (HNV) is reported in Coh-Metrix with the help of WordNet. In WordNet, each word is located on a hierarchical scale allowing for the measurement of the number of subordinate words below and superordinate words above the target word. Coh-Metrix provides estimates of hypernymy for nouns, verbs, and a combination of both nouns and verbs.

Index 5: Latent Semantic Analysis (LSA) is to measure the semantic repetition among sentences and paragraphs. Coh-Metrix offers 8 groups of indices about LSA, which is between 0 (low cohesion) and 10 (high cohesion). Examples are made in the introduction to Coh-Metrix version 3.0:

Text 1: The field was full of lush, green grass. The horses grazed peacefully. The young children played with kites. The women occasionally looked up, but only occasionally. A warm summer breeze blew and everyone, for once, was almost happy.

Text 2: The field was full of lush, green grass. An elephant is a large animal. No-one appreciates being lied to. What are we going to have for dinner tonight?

The indices of Text 1 are much higher than the indices in Text 2. Text 1 is all about a leisure scene in a park: green, grass, children, playing, summer, breeze, kites and happy. On the contrary, Text 2 has no focused semantic relation, which means the sentences are not overlapped semantically.

IV. CORPORA AND RESEARCH METHODOLOGY

The corpora are made up from two parts: BAWE and Huanghuai University (HHU) students' EAP writing. BAWE is part of the project of British Higher Institution Assessment Writing Genre Survey supported by Economic and Social Research Committee from 2004 to 2007, which including 6 million words and 13 different genres. The collection of the written assignments is from Year 1 to 3 of undergraduates with different discipline background. The HHU EAP writings are collected from the essays of EFL learners in Huanghuai University in China. From level 1 to Level 3, 30 essays are selected randomly from each British ENL learners and Chinese EFL learners. Altogether there are 270 essays are examined, including 180 essays from BAWE and 90 copies from HHU students' EAP writing.

It is relatively difficult to get EFL essays from college students in China because undergraduates and graduates in China are not required to write in English academically except for abstract. In order to study EAP writing of college students in China, the research have the samples of EAP writing from joint programs in HHU in China. HHU has cooperated with Bangor University, Aberystwyth University, Coventry University in UK and University of Mysore in China on different specialties, ranged from accounting, marketing, computer science, software engineering, art design, cartoon and animation, nursing and automotive engineering. Most joint program graduates from HHU will either seek for further education in the UK or India or get employed in international companies. Students receive intensive training of English for General Purpose (EGP). From Year 2 and then on students will be trained and assessed by both staff from partner universities and HHU. Some of the core modules assessment requires the submission of academic writings, which make it available to get EAP writing by college students in China.

One natural paragraph is selected from each essay in the corpora. The section avoids charts and tables and in-text references. Thus the name of the author and the year of publication in quotation were deleted on purpose to reflect the natural flow of the text. Then the text of each subtract is pasted into the online Coh-Metrix 3.0. After clicking the button of 'Submit', 108 indices are generated. This study saved each group of indices into an Excel file according to the levels and mother tongue of the writers. Only the five indices are examined in each group: noun overlapping, content words overlap, stem overlap, LSA and hypernymy (nouns and verbs).

V. INDICES OF LEXICAL COHESION IN EAP WRITINGS

The first group of indices is from the essays written by British ENL college students, which is shown in Table 1:

TABLE 1:
LEXICAL COHESION IN EAP WRITING ACROSS LEVEL 1-4 BY BRITISH ENL COLLEGE STUDENTS

Corpus	Level	NO	CWO	SO	LSA	HNV
BAWE-ENL	1	0.5465	0.1086	0.6416	0.3196	2.1715
BAWE-ENL	2	0.5540	0.1049	0.6492	0.3444	2.1120
BAWE-ENL	3	0.5772	0.1036	0.6653	0.3035	1.8685
Average		0.5592	0.1057	0.6520	0.3225	2.0507

Table 1 shows that noun overlap increase with the increasing of the language level. Occurrence of stem overlap from Level 1 to Level 3 keeps on going up. On the contrary, indices of both content word overlap and hypernymy of nouns and verbs drop down across different levels.

The second investigation is about the Chinese EFL learners who receive their education in British higher institutes. Their performance of lexical cohesion across different levels in EAP writing are shown in Table 2:

TABLE 2:
LEXICAL COHESION IN EAP WRITING ACROSS LEVEL 1-4 BY CHINESE EFL COLLEGE STUDENTS IN UK

Corpus	Level	NO	CWO	SO	LSA	HNV
BAWE-EFL	1	0.5000	0.1213	0.5500	0.3275	1.9843
BAWE-EFL	2	0.5348	0.1081	0.6804	0.3150	2.0160
BAWE-EFL	3	0.5755	0.0936	0.6951	0.3121	2.0249
Average		0.5368	0.1077	0.6418	0.3182	2.0084

Table 2 shows the indices of noun overlaps, stem overlap and hypernymy of nouns and verbs all increase by different degrees. The figure of content word overlap declines across level 1 to level 3. The indices of latent semantic analysis go down straight from level 1 to level 3.

Essays by EFL Chinese college students are divided into 3 levels from Year 2 to Year 4 in China as it is not required to produce academic writing in Year 1. The indices of their using lexical cohesion are shown in Table 3:

TABLE 3:
LEXICAL COHESION IN EAP WRITING ACROSS LEVEL 1-3 BY CHINESE EFL COLLEGE STUDENTS IN CHINA

Corpus	Level	NO	CWO	SO	LSA	HNV
HHU-EFL	1	0.5672	0.1396	0.6201	0.2565	1.7834
HHU-EFL	2	0.5895	0.1178	0.6226	0.2271	2.1637
HHU-EFL	3	0.6001	0.0998	0.6272	0.2036	2.1956
Average		0.5856	0.1191	0.6233	0.2291	2.0476

Table 3 shows the three groups of indices (noun overlap, stem overlap, hypernymy of nouns and verbs) of lexical cohesion by EFL college students in China go up while the other two groups (content word overlap and latent semantic analysis) go down across the three levels.

VI. DISCUSSION AND INTERPRETATION

The overall trend of noun overlap in EAP writing by three groups of college students increases across the different levels. According to Biber et al. (1999), the reason for the higher frequency of nominalizations in EAP writing is that the function of nominalizations within the context is “to treat actions and processes as abstract objects separated from human participants”, whereas fiction and speech “are more often concerned with people and use verbs and adjectives to describe how they are behaving”. And Ravelli (1996) claims that nominalizations are “prestigious, technical and formal, rather than coming from a more everyday realm”. This further suggests another important rhetorical effect produced by the use of nominalizations, which lower level of undergraduate students might notice and learn to use them with more exposure to academic study.

Using of LSA tends to be decreasing with the level improving. The possible reason for that is there might be more jargons within the discipline. Higher indices of LSA indicate the higher readability of the text and the decrease of LSA in ESP writing in the corpora means the articles tends to be more complicated and harder to understand for general readers. It is a bit unexpected to see that Chinese EAP writings have more frequency of using specific words than British ENL writing. What’s more, British native speakers tend to use less specific words with language proficiency level improving.

The decreasing of content word overlap and increasing of stem overlap indicate the improving of college students’ vocabulary. The reducing tendency of hypernymys shows that British ENL college students use more specific words with the improving of language proficiency.

From Table 1, 2 and 3, it is hard to conclude whether the overall using of lexical cohesion is increasing or not with the improving of language proficiency. However the lexical cohesion skills are obviously different across different level. From Table 4: The average indices of lexical cohesion by British ENL and Chinese EFL college students, we can see clearly that British ENL college students use less simple repetition (noun overlap and content word overlap) and more complex repetition (stem overlap, latent semantic analysis and hypernymy of nouns and verbs).

TABLE 4:
AVERAGE INDICES OF LEXICAL COHESION IN EAP WRITING BY BRITISH ENL AND CHINESE EFL COLLEGE STUDENTS

Learner	NO	CWO	SO	LSA	HNV
British ENL	0.5660	0.1061	0.6495	0.3220	2.0695
Chinese EFL	0.57845	0.1149	0.6442	0.2651	2.0515

With different degree of language environment exposure, lexical cohesion in EAP writing shows different figures. Table 5 shows the means of the above five indexes of lexical cohesion by Chinese EFL college students in the UK and China respectively.

TABLE 5:
MEANS OF INDICES OF LEXICAL COHESION IN EAP WRITING BY CHINESE EFL COLLEGE STUDENTS BETWEEN UK AND CHINA

Corpus	NO	CWO	SO	LSA	HNV
BAWE-EFL	0.5368	0.1077	0.6418	0.3182	2.0084
HHU-EFL	0.5856	0.1191	0.6233	0.2291	2.0476

From Table 5, the same results occur as what happens to Table 4, noun overlap, content word overlap, hypernymy of nouns and verbs by HHU-EFL are all overused than BAWE-EFL Chinese students. But the complex use of lexical cohesion, such as stem overlap and latent semantic analysis indices figures are lower than that of the Chinese learners in the UK.

With the above analysis, we can conclude that the overall frequency of lexical cohesion in EAP writing has no obvious increases or decreases across levels of language proficiency. Each individual index we researched does have growing or dropping tendency across different language levels. As for the difference in using lexical cohesion in EAP writing between Chinese EFL learners and British ENL learners, more simple lexical cohesion and less complex lexical cohesion are used by Chinese EFL learners. The better the writer's English proficiency is, the more complex lexical cohesion is used in their EAP writings.

VII. THE LIMITS AND THE IMPLICATION OF THE RESEARCH

One of the limits of the study can be the sample selection. The length of paragraph is generally shorter by Chinese EFL learners than by ENL speakers. The average length of paragraph by Chinese EAP writer is from 150 to 200 while the average paragraph length by native speakers is about 250. The number of words of the natural paragraph might influence the indices produced by Coh-Metrix. Another thing about the corpora is whether the essays written by joint program students in HHU can represent Chinese EAP writing. With the development of EAP teaching in China, there should be more options for assessed EAP writings in Chinese universities. Therefore extended text and general non-English major college students' EAP writing can be sampled. Besides the five indexes in this research, more aspects of lexical cohesion should be studied. The different computational tool for lexical cohesion analysis should be explored as well in the future research.

The focus on corpus linguistics is to highlight the communication function of language and the application of language: language is action and meaning is application (Mahlberg, 2006). In the natural use of language, the research tries to find some features of lexical cohesion in EAP writing across different EFL levels with comparing to British ENL speakers. Mother tongues positive transfer has quite limited influence on upper intermediate or advanced EFL learners and more rules of natural language development will be applied to them. As EFL learners, Chinese college students and classroom teaching should focus more on lexical cohesion in EAP writing, especially complex lexical cohesion. Through large amount of natural EAP writing, we can find the features of lexical cohesion by different language users across different levels. Independent learners and classroom instructor should both be aware of the significance of corpus linguistics in their learning and teaching. The true principles of language application are embedded in the vivid natural language.

REFERENCES

- [1] Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education, 61.
- [2] Chen, J. (2008). An investigation of EFL students' use of cohesive devices. *Asia Pacific Education Review*, 5(2), 215-225.
- [3] Clarence, G. (2012). A Computational Investigation of Cohesion and Lexical Network Density in L2 Writing. *English Language Teaching*, 5(8), 57-69.
- [4] Crossley, S., & McNamara, D. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, 18, 119-135.
- [5] Crossley, S., & McNamara, D. (2012). Predicting second language writing proficiency: the roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35(2), 115-135.
- [6] Ferris, D. (1994). Rhetorical strategies in student persuasive writing: differences between native and non-native English speakers. *Research in the Teaching of English*, 28(1), 45-65.
- [7] Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, 36, 193-202.
- [8] Johnson, P. (1992). Cohesion and coherence in compositions in Malay and English. *Journal of Language Teaching and Research*, 23(2), 1-17.
- [9] Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English (English Language)*. London: Longman.
- [10] Hoey, M. (1991). *Patterns of lexis in text*. Oxford: Oxford University.
- [11] Hyland, K. (2005). *Teaching and researching writing*. Beijing: Foreign Language Teaching and Research Press.
- [12] Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System*, 33(4), 623-636.

- [13] Mahlberg, M. (2006). Corpus linguistic theory and its application in English language teaching. *International Journal of Corpus Linguistics*, 11/3, 363-381.
- [14] McNamara, D. S., Ozuru, Y., & Floyd, R. G. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education*, 4, 229-257.
- [15] Mirzapour, F., & Ahmadi, M. (2011). Study on Lexical Cohesion in English and Persian Research Articles. *English Language Teaching*, 4(4), 243-255.
- [16] Ravelli, L. (1996). Making language accessible: successful text writing for museum visitors. *Linguistics & Education*, 8 (4), 367-388.
- [17] Stotsky, S. (1983). Types of Lexical Cohesion in Expository Writing: Implications for Developing the Vocabulary of Academic Discourse. *College Composition and Communication*, 34(4), 430-446. <http://www.jstor.org/stable/357899>.
- [18] Sun, Y.Z. (2010). Founded Speech. *ESP Teaching and Research in China*, 1, ii.
- [19] UIS. (2014). Global flow of tertiary-level students. <http://www.uis.unesco.org/Education/Pages/international-student-flow-viz.aspx>. (accessed 20/05/2014).
- [20] Witte S., & Faigley, L. (1981). Coherence, cohesion and writing quality. *College Composition and Communication*, 32(2), 189-204. <http://dx.doi.org/10.2307/356693>.

Chunhong Zhao received her M.A. degree in English Language and Literature. She works in International College in Huanghuai University as vice dean. Her mainly interests include English language teaching and internationalization of higher education.