



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Research Commons

<http://researchcommons.waikato.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

**Different or similar: patterns of lexical cohesion in Chinese
non-narrative text and their applications in the study of Teaching
Chinese as a Foreign Language**

– A corpus-based contrastive study

A thesis

submitted in fulfillment

of the requirements for the degree

of

Doctor of Philosophy

at

The University of Waikato

by

SHOUJI LI



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2012

Abstract

This study identifies and investigates patterns of lexical, cohesive devices in Chinese argument texts by two groups of writers: native speakers and learners of *Chinese as a Foreign Language* (hereafter CFL). The purposes of the research are twofold: first to inform instruction in the writing of extended Chinese texts to CFL learners and secondly to provide a basis for the further investigation of the linguistic competences that relate to the writing of such Chinese texts. To achieve such purposes, the study employs two corpora: a corpus of 50 native-speaker Chinese texts written for the *National College Entrance Examination* (hereafter the NCEE) and a corpus of 50 CFL texts written for the *HSK* (the Chinese language proficiency test). The overall framework adopted for the data analysis was *Contrastive Analysis Study* and the texts were analyzed using corpus-based methods. The first stage of the research involved using the NCEE Corpus of native-speaker texts as the basis for a critical review of Halliday and Hasan (1976) and Hoey (2000)'s approaches to the classification of cohesive devices in English, and through this review, adapting and extending the categories of these two theoretical approaches in order to account for salient linguistic elements used to achieve cohesion in Chinese texts. Using this modified system of classification of lexical cohesive devices that was developed for Chinese texts, the NCEE Corpus was then reanalyzed and the patterns of the use of lexical devices were identified. The same analysis was then performed on the texts from the HSK Corpus. The findings of the analyses of the two corpora were then compared to identify the different patterns of use of lexical cohesive devices by native-and non-native speaker writers of Chinese.

The findings derived from the comparison of the analyses of the NCEE texts and the HSK texts show that there are a number of important differences in the use of lexical devices between the two corpora. The findings seem to suggest that learners of CFL are less competent in the use of a range of lexical devices, such as simple and

complex paraphrase, constituent repetition, superordinate and hyponymic repetition. The comparison also reveals that learners of CFL have relatively low capacity to use lexical devices to create adequate and effective cohesive links across texts between non-adjacent sentences, and that they have difficulty in employing various types of lexical items to form repetitions appropriate to the development of argument in this type of text.

On the basis of the findings, discussion of the pedagogical implications for the teaching of Chinese writing is provided. A further outcome of the research is that it appears to indicate the need for more in-depth investigations to inform both the teachers of Chinese writing and the developers of teaching and learning materials.

Acknowledgements

I would like to gratefully and sincerely thank my supervisor Dr. Ian Bruce for guidance, encouragement, understanding, and most importantly, his mentorship throughout the process of my academic journey. I have especially valued his guidance in leading me into the research field of Applied Linguistics and helping me develop my own academic individuality and independence. I would also like to thank the members of my supervision panel, especially Dr. Xiaoning Wang and Dr. James McLellan, for providing valuable academic guidance and suggestions. Their contributions of time, ideas, have made my Ph.D. experience productive and stimulating. My thanks also go to Dr Roger Barnard for his encouragement and concern about my academic activities.

I am most grateful to the administrative staff, especially Professor Dejin Sun, from Beijing Language and Culture University for providing me with CFL corpus texts. Without this valuable resource, this corpus-based research would not have been successful.

I would like to acknowledge the financial, academic and technical support of the University Waikato and its staff, particularly in the award of a doctoral scholarship that provided the necessary financial support for this research.

Finally, I would like to thank my wife Judy Xiao, my son Eden Li, and my daughter Sophie Li, who have made this journey with me and who are inspirations in my life.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Introduction to the study.....	1
1.2 Aims of the study	6
1.3 Significance of the study	7
1.4 Overview of the chapters.....	9

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction	11
2.2 Contrastive Rhetoric as a field of study	11
2.2.1 The establishment of contrastive rhetoric	12
2.2.2 The development and expansion of contrastive rhetoric.....	13
2.2.3 The critiques of contrastive rhetoric.....	14
2.2.4 New directions proposed for contrastive rhetoric	16
2.2.5 Contrastive rhetoric research in China	17
2.2.6 Summary of contrastive rhetoric and implications for the study	20
2.3 Textual cohesion: theory and research	22
2.3.1 The definition of cohesion in text.....	22
2.3.2 Different types of cohesion	24
2.3.3 Lexical cohesion.....	26
2.3.4 The organizational role of lexical cohesion	28
2.3.5 Systems of cohesion analysis	29
2.3.6 Implications of the study of cohesion study for language teaching	30

2.3.7 The study of cohesion in Chinese.....	31
2.3.8 Summary of the research on cohesion and its implications for the study	35
2.3.9 Implications of the studies on cohesion for the proposed research	36
2.4 Corpus linguistics	37
2.4.1 The definition of Corpus Linguistics.....	37
2.4.2 Various types of corpora	38
2.4.3 The advantages of the Corpus Linguistics methodology	39
2.4.4 Methodologies for corpus-based studies	41
2.4.5 Methodologies for learner corpus based studies	43
2.4.6 Corpus Linguistics and Language teaching.....	44
2.4.7 The development of Corpus Linguistics in China.....	46
2.4.8 Summary of Corpus Linguistics and its implications for study	47
2.5 Summary	49

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction	50
3.2 Research questions	50
3.3 The theoretical frameworks and research design	52
3.4 The data collection and sampling procedure for the two corpora	54
3.4.1 The nature of the data	55
3.4.2 Selection of the samples.....	60
3.5 The corpus design.....	64
3.5.1 The design of reference corpus of native Chinese writing	64
3.5.2 The design of the Chinese interlanguage writing corpus	65

3.5.3 Attributes of the corpora of the study	67
3.6 Methods of data analysis	67
3.6.1 Step 1: The manual analysis	68
3.6.2 Step 2: The computational analysis	73
3.6.3 Step 3: The statistical analysis and interpretation.....	79
3.7 Summary	80

CHAPTER 4: STUDY 1 FINDINGS

4.1. Introduction	81
4.2. Non-lexical cohesion.....	81
4.2.1 Thematic chunk.....	82
4.2.2 Parallel-Structure	85
4.2.3 The use of idioms and proverbs	89
4.2.4 The use of quotations from classics	92
4.2.5 The use of metaphor.....	93
4.2.6 The use of summary words	94
4.2.7 Summary and brief discussion	96
4.3. Lexical Cohesion.....	97
4.3.1 Simple lexical repetition	97
4.3.2 Complex lexical repetition.....	102
4.3.3 Constituent repetition.....	105
4.3.4 Simple paraphrase.....	111
4.3.5 Complex paraphrase.....	112
4.3.6 Superordinate, hyponymic, and co-reference repetition.....	114
4.4 The overall view of the cohesive devices found in the NCEE Corpus	118

4.5 The corpus statistical analysis	119
4.5.1 Statistics for the use of various types of lexical cohesion in the NCEE Corpus	120
4.5.2 Statistics for the number of bonds formed in the NCEE texts	123
4.5.3 The numbers of bonds formed between non-adjacent sentences to that between adjacent sentences.....	126
4.5.4 The ratio of sentences that form bonds with other sentences to those form no bonds with others	129
4.5.5 The average number of words that form repetition with other words in one sentence	142
4.6 Summary	144

CHAPTER 5: STUDY 2 FINDINGS

5.1 Introduction	145
5.2 Non-lexical cohesion	145
5.2.1 Thematic chunk	147
5.2.2 The parallel structure	149
5.2.3 The use of idioms and proverbs	149
5.2.4 The use of quotations from classics	151
5.2.5 The use of summary words	153
5.2.6 The use of metaphor.....	155
5.3 Lexical cohesion.....	158
5.3.1 Simple lexical repetition	159
5.3.2 Complex lexical repetition	160
5.3.3 Simple lexical paraphrase	162
5.3.4 Complex paraphrase.....	163

5.3.5 Co-reference repetition	165
5.3.6 Summary of the findings of the use of lexical cohesion in the HSK Corpus	166
5.4 The statistical corpus analysis	166
5.4.1 The use of various lexical repetitions in texts of the HSK Corpus	166
5.4.2 The number of bonds formed in the texts of the HSK Corpus	168
5.4.3 The number of bonds formed between non-adjacent sentences to that between adjacent sentences in the HSK Corpus	171
5.4.4 The number of sentences that form bonds with other sentences to those form no bonds with others in the HSK Corpus.....	174
5.4.5 The statistics of the average number of words that form lexical cohesive links with other words in one sentence in the HSK Corpus	176
5.5 Summary	179

CHAPTER 6: COMPARISON OF THE DATA

6.1 Introduction	180
6.2 The comparison of the use of non-lexical devices in the two corpora	180
6.3 The comparison between the use of lexical devices in the two corpora	182
6.3.1 Comparison of the A2 data in the two corpora	183
6.3. 2 Comparison of the B2 data in the two corpora	184
6.3. 3 Comparison of the C2 data in the two corpora	184
6.3. 4 Comparison of the D2 data in the two corpora	185
6.3. 5 Comparison of the E2 data in the two corpora	185
6.4 Summary	186

CHAPTER 7: DISCUSSION OF THE USE OF NON-LEXICAL COHESIVE

DEVICES

7.1 Introduction	187
7.2 The discussion of the patterns of the use of non-lexical devices in the two samples of ten texts.....	187
7.2.1 Use of the thematic chunk in the two corpora.....	188
7.2.2 Use of the parallel structure in the two corpora	198
7.2.3 Use of idioms and proverbs in the two corpora	201
7.2.4 Use of metaphors in the two corpora	206
7.2.5 The overall distribution of non-lexical devices in the two samples of ten texts	210
7.2.6 Implications for the study of cohesion and for TCFL pedagogy	213
7.3 Summary	219

CHAPTER 8: DISCUSSION OF THE USE OF LEXICAL COHESIVE

DEVICES

8.1 Introduction	220
8.2 The patterns of the use of each type of lexical devices	220
8.2.1 The patterns of the use of constituent and superordinate-hyponymic repetition.....	221
8.2.2 The patterns of use of simple-paraphrase and complex paraphrase	227
8.3 The patterns of bond density in the two corpora	233
8.4 The patterns of the average ratio of the bonds formed between non-adjacent sentences to that between adjacent sentences in the two corpora	246

8.5 The patterns of the average ratio of sentences that form bonds with other sentences to those form no bonds with others in the two corpora.....	250
8.6 The patterns of the average number of words that form repetition with other words per sentence in the two corpora	257
8.7 Implications for the study of cohesion and the TCFL pedagogy	263
8.8 Limitations of the study.....	268
8.9 Summary	268
References.....	269
Appendix I: Sample of Chinese composition test in the NCEE of 2007.....	289
AppendixII: Sample of Chinese composition test in the HSK.	290
Appendix III: Letter requesting permission to access data texts.....	291
Appendix: IV: Chinese version of Letter requesting permission to access data texts.....	293
Appendix V: Application for Ethical Approval	295

LIST OF TABLES

Table 2.1	Significant collocates of COMMIT in BROWN Corpus	42
Table 3.1	The number of characters in each of the 50 HSK texts	62
Table 3.2	Text attributes of the NCEE and the HSK Corpora	67
Table 3.3	The tagging system for the study	71
Table 3.4	A summary of lexical repetition in the NCEE Text 4	75
Table 3.5	Repetition matrix of the NCEE Text 4	76
Table 4.1	Various types of statistical data to be extracted from the NCEE Corpus	119
Table 4.2	The various types of lexical cohesion and their frequency per 100 characters in the NCEE Text 1	121
Table 4.3	The average frequencies of the seven types of cohesion in the 50 NCEE texts	123
Table 4.4	The bond density and the average bond density in the NCEE Corpus	125
Table 4.5	The ratio of bonds formed between non-adjacent sentences (BNA) to that between adjacent sentences (BA) in the NCEE Corpus	128
Table 4.6	The ratio of the number of the sentences that form bonds with other sentences (SFB) to those form no bonds (SFNB) in the NCEE Corpus	140
Table 4.7	The items that form connections (IFC) with other items in the NCEE Text 4 情感的权力	142
Table 4.8	The average number of the items that form connections (IFC) with other items in each sentence in the NCEE Corpus	143
Table 5.1	The results of the analysis of the use of idioms and proverbs in the ten CFL texts	151
Table 5.2	The results of the analysis of the use of quotation from classics in ten CFL learner texts	153

Table 5.3	The results of the analysis of the use of summary words in the ten CFL learner texts	155
Table 5.4	The results of the analysis of the use of metaphor in the ten CFL learner texts	157
Table 5.5	The number of the various types of lexical cohesive devices and their frequency per 100 characters	167
Table 5.6	The average frequencies of the seven types of repetition in the HSK Corpus	168
Table 5.7	The individual bond density and the average bond density in texts in the HSK Corpus	169
Table 5.8	The ratio of bonds formed between non-adjacent sentences (BNA) to that between adjacent sentences (BA) in texts in the HSK Corpus	171
Table 5.9	The ratio of the number of the sentences that form bonds with other sentences (SFB) to those form no bonds (SFNB) in texts in the HSK Corpus	174
Table 5.10	The average number of the items that form connections (IFC) with other items in one sentence in texts in the HSK Corpus	176
Table 6.1	The number of the non-lexical devices used in each text and their average value	181
Table 6.2	The comparison of the use of non-lexical devices and their average values	181
Table 6.3	Comparison of the average frequency (per 100 characters) of each type of lexical cohesive device used in 50 texts	183
Table 6.4	Comparison of the average bond density in 50 texts in the two corpora	184
Table 6.5	Comparison of the average ratio of the bonds formed between non-adjacent sentences to that between adjacent sentences	185
Table 6.6	Comparison of the average ratio of sentences that form bonds with other sentences to those form no bonds with others in the two corpora	185
Table 6.7	Comparison of the average number of words that form repetition with	

	other words per sentence in 50 sample texts	186
Table 8.1	Sentence coordinates of the NCEE Text 4	251
Table 8.2	Sentence coordinates for the CFL Text 1	251
Table 8.3	The percentages of marginal and central sentences in both the NCEE Text 4 and the CFL Text 1	253
Table 8.4	NCEE Text 4 words that form repetition with others	258
Table 8.5	CFL Text 1 words that form repetition with others	258
Table 8.6	Lexical cohesive links in the above outline	262

LIST OF FIGURES AND DIAGRAMS

Figure 2.1	Hu's (1994) multi-level model of cohesion and coherence	33
Figure 3.1	Theoretical frameworks for the study	54
Figure 3.2	The lexical repetition entrance flow chart	70
Figure 3.3	The interconnection net of bonded sentences derived from Table 3.5	78
Figure 4.1	The heuristic flow chart for simple repetition	100
Figure 4.2	The heuristic flow chart for complex repetition	104
Figure 4.3	The heuristic flow chart for constituent repetition	108
Figure 4.4	The argumentative structure of the NCEE Text 14 好奇心 [Curiosity]	109
Figure 4.5	The heuristic flow chart for paraphrase	113
Figure 4.6	Cohesive devices found in the sample of ten texts taken from the NCEE Corpus	118
Figure 8.1	The semantic network of the constituents of the word 好奇心 [curiosity]	223
Figure 8.2	An example of the contrast between the average bond densities in the two corpora	240
Figure 8.3	An example of the number of bonds formed between adjacent and non-adjacent sentences in NCEE Text 4	247
Figure 8.4	A typological pattern of the argument development in the NCEE Corpus	249
Diagram 8.1	The correlation between the sentence length and number of repetitions	260
Diagram 8.2	The relation between lexical repetition, semantic network, and textual organization	261

CHAPTER 1: INTRODUCTION

1.1 Introduction to the study

According to Halliday and Hasan (1976), cohesion refers to “relations of meaning that exist within the text and that define it as a text” (p. 4). Cohesion can be achieved either by grammatical (reference, substitution, ellipsis, conjunction) or lexical links within a text or between sentences, links often referred to as *cohesive ties*. However, Hoey (2000) claims that Halliday and Hasan (1976) do not recognize that lexical cohesion is “the single most important form of cohesive tie” (p. 9), and, therefore, further explores the role lexical cohesion plays in text-organizing in English non-narrative texts. On the basis of his analysis, Hoey argues that lexical repetition is more important than any other type of cohesion, especially when forming cohesive ties over large spans of text. The study of the role that lexical repetition plays in textual organization, therefore, is very valuable, not only to theoretical and descriptive linguistics, but also to language teaching (especially second or foreign language teaching) and Applied Linguistics. Drawing on Halliday and Hasan’s (1976) and Hoey’s (2000) analytical framework, a large number of studies have been conducted to explore the role of lexical cohesion in the organization of text to inform the teaching of English as Second Language (ESL) (see, for example, Connor, 1984; Chiu, 1993; Jin, 2008; Kafes, 2012). These explorations have laid a solid foundation for the further investigation of how lexis contributes to the creation of coherence within a text.

Despite this body of research, there is still scope for further investigation of cohesion within other research frameworks and methodologies, such as contrastive rhetoric, second language acquisition, and corpus linguistics and in relation to the teaching of languages other than English, such as Chinese. Most recent studies of cohesive patterns, lexical errors and teaching strategies of writing have investigated English text, with ESL learners and ESL teaching as their research objects. By contrast, few research studies have explored the patterns of lexical cohesion in other languages, such as in Chinese non-narrative text. According to Peng’s (2004) survey, among 540 articles (published in academic

journals in China from 1994-2005, which discuss the teaching of the creation and processing of text, 503 studies focus on the teaching of foreign language text (the vast majority are concerned with English text), only 37 discuss relate to the structure and teaching of Chinese text. This contrast reflects the fact that the study of cohesion in the Chinese language may be still at a preliminary stage, and consequently, as yet unable to reveal the role that the cohesive devices play in Chinese texts. Similarly, in the latest literature in the field of *Teaching Chinese as a Foreign Language* (henceforth TCFL), the majority of the studies that discuss cohesion focus on grammatical cohesion (Chen, 2005; Huang, 2005; Liu, 2005; Xiao, 2001; Yang, 2004; Zhao, 2005a & 2005 b), and only a few focus directly or indirectly on the use of lexical cohesion in CFL learners' writing (Cao, 2002; Sun, 2007; Tian, 2006). Although these studies contribute much to the understanding of the differences between the use of cohesive devices in English and Chinese, and between their use in native Chinese speakers' and CFL learners' writing, they are far from being at the stage of providing an overall picture of the use of cohesion in Chinese language and the CFL inter-language context in relation to their text-organizing functions.

Furthermore, those studies that have attempted investigations in this area have not, so far, provided detailed and systematic descriptions of lexical patterns because of the methodologies adopted and the amount and the types of data employed. For example, in his book *Textual Cohesion and Coherence*, Hu (1994) examines various types of lexical cohesion in Chinese text. Hu's work is probably the first systematic description of lexical devices that contribute to the organization of Chinese text. However, there are aspects related to this description may need to be further explored. Firstly, since the text that Hu used to analyse the lexical cohesion, is a novel called 子夜 [*Midnight*] written in 1932, it may not be able to represent the linguistic features of contemporary Chinese. Secondly, the examples that Hu presents in his book are from a variety of textual sources, some are conversational, some descriptive, and some argumentative. Although these examples help to explain how lexical devices work in creation of cohesion, they may not be helpful to describe systematically some of the predominant lexical features that are specific to certain types of texts. In addition to these issues, the research is more qualitative than quantitative. In other words, although he

describes the use of some of the lexical devices in the text that he used, he does not present an overall picture of the how lexical cohesion contributes to the creation of cohesion in Chinese text by quantifying various types of lexical devices. Similarly, there are a few studies of lexical cohesion conducted in the TCFL field, such as those of Sun (2007) and Tian (2006). For example, Sun (2007), drawing upon Hoey's (2000) analytical framework, compares the use of lexical cohesion in narrative Chinese texts written by a Vietnamese learner and a native Chinese speaker. Although he examines different types of lexical cohesion in both texts and compares the patterns of the use of these devices between the two writer, the data set for this research does not seem to be sufficiently large to claim that the gap between the writers represents the collective features of the use of the lexical cohesion in Vietnamese learners' texts, as the article's title suggests. Furthermore, the text that Sun (2007) used to analyse is a narrative, which may not be suitable to use Hoey (2000)'s analytical framework for lexical analysis, as Hoey (2000) clearly states that his lexical analysis "is not applicable to narrative texts" (p. 188). The above examples show that more in-depth studies that examine various types of Chinese text using large-scale data sets and appropriate methodologies are needed to further explore the text-organizing functions of lexical cohesive devices in Chinese text.

In addition, the teaching of writing in Chinese as a Foreign Language has long been among the most challenging issues facing teachers and learners of this language. With the development of the TCFL, a large number of frontline CFL teachers have conducted a number of research studies with the aim of finding more effective and scientific ways of teaching writing, such as those of Cao (2002) and Wu (2011). Although these studies are valuable and inspirational, there is still a lack of systematic and holistic research that focuses on exploring the patterns of the use of lexical cohesion in both Chinese text and CFL learner text. Studies are needed which provide an overall picture of patterns of the lexical cohesion in Chinese text, and which identify the difficulties that CFL learners may have in their learning of writing. For example, Cao (2002) points out that CFL learners' performance in Chinese writing is far from ideal, one of the reasons for this being that the teaching of lexis is still limited to the traditional approach, namely presenting new vocabulary plus example sentences. She argues that although it is

better than teaching student lexical items only, it does not encourage the learners to acquire vocabulary items in the right contexts in which they are used. She claims that the key to the success of acquiring vocabulary is to teach vocabulary at the text level right from the beginning of their learning, and that, only in this way learners, not only integrate the meaning of vocabulary explained in dictionaries with their contextual connotations, but also are able to express their ideas at paragraph level by grasping the cohesive functions of the lexis taught. She, therefore, calls on more studies in this area to improve on the teaching of lexis in CFL classrooms.

Taking a similar perspective, Peng (2004) recognises that the problem of the teaching of writing lies in the lack of an overall teaching framework and the specification of content. She suggests that some aspects of Chinese text, such as reference, conjunctive words and phrases, time expressions, lexical cohesion, and interaction between the selection of sentence patterns and coherence, should be explored further and integrated into the teaching of writing (p. 51). In relation to writing instruction, Wu (2011) finds that the one of the major challenges that CFL learners face is a lack of pre-writing activities to prepare them at both lexical and textual level. By comparing the results from a control group and an experimental group, he claims that learners that are encouraged to engage in activities, such as discussing given topics in groups, brainstorming key words and their lexical and semantic associations, organizing and outlining and their ideas, writing up their essays either in groups or on their own, produce better organized and more coherent compositions than those who do not participate in such activities. These discussions, experiments, suggestions, and findings, therefore, motivated me to conduct this research in order to address some of these problems faced by CFL learners in their learning of writing, the findings of which may potentially inform the teaching of CFL in ways that are both theoretical and practical.

Over the last four decades, *contrastive rhetoric* (henceforth CR), as an interdisciplinary field, has contributed to research knowledge in applied linguistics and informed the teaching of *ESL* and *EFL* as well as *EAP* (Kaplan, 1966; Hints, 1984; Connor, 1996). Although some CR studies have come under sharp criticism for their overgeneralization and oversimplification of complex

issues, and for their methodological weakness, CR scholars agree that the need for L2 writers to learn how to organize L2 written text still exists. Consequently, CR as a field has great potential and has been employed by L2 scholars (see, for example, Connor 2004). With the development of the CR theory and the improvement of research methodologies, many CR studies have investigated lexical cohesion in ESL writing (Al-Jubouri, 1984; Connor, 1984; Hinkel, 2002; Reynolds, 1995, 1996; Scarella, 1984; Tierney & Mosenthal, 1983; Witte & Faigley, 1981; Zhu, 2001). These studies, have contributed to understandings of various aspect of writing, such as the composing process, its organization, and teaching methodologies. Consequently, their findings provide some insights into the exploration of the linguistic features of both Chinese text and CFL text as well as for the improvement of methodologies for the teaching and learning of CFL writing. For example, from the contrastive rhetoric perspective, Hinkel (2002) carried out a large scale corpus-based L2 text analysis and reports that “even advanced and trained L2 writers have a severely limited lexical and syntactic repertoire” (p. 622) and because of this, he “calls for changes in the methodologies for teaching L2 writing that are based on the pedagogy intended for teaching compositions to NNs” (p. 622). To address some of the acquisition issues regarding lexis, such as L2 learners’ limited knowledge of productive vocabulary, the ignorance of the means of repetition of lexical items in adjacent sentences, some scholars, such as Kony (2005) suggest that it is effective for native English teachers to adopt a contrastive rhetoric approach to assist Chinese learners of English to increase their variety of vocabulary and choice of lexis by training them to gain the ability to “find a word that closely parallels their intended meaning” in its context (p. 1). Compared with the large number of studies of a range of aspects of English use in the CR literature and the insightful knowledge contributed by researchers, little research appears to have been undertaken to investigate lexical cohesion in CFL writing from the perspective of CR. Therefore, this study aims to use a CR approach to expand the scope of research in this area and make a significant contribution to the existing literature.

In recent years, *corpus linguistics*, as a new computer-based research methodology, has been widely recognized and embraced by large number of researchers in the field of Applied Linguistics. Because of their heuristic power

and potential for quantitative analysis, corpus based techniques have helped to bring out distinctive patterns of actual language use (Granger, 2002; Stubbs, 1996). Since its introduction to China in the 1980s, the concept of lexical cohesion has received considerable attention in ESL corpus based studies. In the newly-developed discipline of TCFL, although the corpus-based approach have been adopted by many researchers to undertake studies on various linguistic features of CFL writings (Chang, 2011; Gui, 2011; Liu, 2011; Wu, 2010; Xiao & Wang, 2009; Yang, 2011), it is difficult to find a study that adopts a corpus-based methodology to explore the patterns of the use of lexical cohesion in CFL learners' writing. The present study, therefore, adopts a corpus linguistics methodology, to compare the similarities and differences in the use of lexical cohesion in both native Chinese speaker and CFL learners' writings. The two corpora used in this study are *National College Entrance Examination* (henceforth the NCEE Corpus) and *Hanyu Shuiping Kaoshi* [Chinese Proficiency Test] (henceforth the HSK Corpus). Both are specifically designed for this study according to a list of corpus-establishing standards, detailed description of which is provided in Chapter 3 Methodologies.

1.2 Aims of the study

This study is primarily a corpus-based contrastive study on the use of lexical cohesion in the two corpora, one of texts written by native Chinese speakers and one of texts written by learners of Chinese as a foreign language. In particular, this study aims to provide a description of the extent to which the two language groups differ in the use of lexical cohesive devices in their writing. On the basis of this description, the study then goes further to identify the difficulties that non-native Chinese speakers may have in assigned writing tasks. It is hoped that the results of the statistical data and their interpretation will provide insights into the teaching of CFL writing as well as the development of teaching and learning materials for CFL writing instruction.

1.2 Significance of the study

This study firstly explores the text-organizing function of both grammatical and lexical cohesive devices in non-narrative Chinese text, which has not been studied previously using large-scale corpus methodology in the TCFL field. It aims, thereby, to provide a more comprehensive understanding of the ways in which text is organized by various types of cohesive device; to deepen the concept of cohesion in textual organization; and to expand the scope of the cohesion research. By using the findings of the NCEE Corpus analysis as a basis for comparison, the study then examines patterns of the use of cohesive devices in the HSK Corpus, and by doing this, highlights one of the most challenging areas faced by CFL learners in their Chinese writing. These findings may help CFL teachers to recognize some of the factors that influence the effectiveness of their teaching of writing, and the recommendations provided in the discussion section may be informative to both teachers and writing course material developers to strengthen and improve their work in the future. The sections below provide a more detailed description.

Firstly, in the current literature of TCFL field and especially teaching the writing of extended Chinese texts, most studies that consider the issue of cohesion adopt the taxonomy of Halliday and Hasan (1976), which firstly makes a distinction between grammatical and lexical cohesive ties, classifying the former in terms of four types: reference, conjunction, substitution, and ellipsis. However, this approach to the classification of cohesive devices in Chinese may not fully reflect the ways in which Chinese text is organized because it is mainly based on the analysis of English texts. The findings of this study show that, apart from reference, conjunction, substitution, and ellipsis, there are a considerable number of other cohesive device used in native Chinese texts. Therefore, the findings give rise to a proposal for a number of additional grammatical categories of cohesion found in Chinese texts. The analysis of the lexical cohesive devices used in the NCEE Corpus also indicates that there are more categories in Chinese texts than that proposed by Hoey (2000). The proposal of the taxonomy of cohesive devices in the Chinese texts in this study, therefore, offers an overall framework of the cohesive devices in Chinese non-narrative text, which may better inform the study

of cohesion in TCFL (and especially the writing of Chinese), and consequently may contribute to the understanding of the salient, cohesive elements of Chinese texts.

In addition, the study compares the patterns of the use of cohesive devices, especially the lexical devices, between NCEE Corpus and HSK Corpus. The HSK texts show a remarkable low frequency of the use of some of the grammatical and lexical devices. This gap in the use of cohesive devices between native and non-native writers of Chinese helps firstly to identify the challenges that CFL learners face in their Chinese writing, and secondly, to identify the areas that the researchers and teachers in the TCFL field may need to study and develop further. The interpretation and discussion of the gap in the last two chapters has a two-fold significance. It first recommends the need for a wider range of categories to describe cohesion in Chinese non-narrative text, and consequently to provide teachers with a sound theoretical knowledge so that they can be better-informed in this area. In relation to TCFL pedagogy, the discussion also identifies some of the aspects of teaching methodology that teachers may need to improve. By providing such insights, it is hoped that CFL teachers may be better-equipped in terms of both subject-matter content knowledge and pedagogical understandings of teaching CFL learner cohesion in writing. Such provision may help to raise CFL teachers' competence in this area, and consequently may result in more effective instructional processes and outcomes. CFL learners may also benefit from some of the learning strategies discussed in the last two chapters. For example, the suggestion that learners be aware of the importance of their lexical competence by engaging in more pre-writing preparation activities could help them to adopt more active, effective, and efficient ways to acquire their vocabulary as well as increase their organizational ability to writing non-narrative texts.

Since this study is corpus-based, interlanguage research, the discussion of some of the design principles, as well as some of the crucial factors that may have impacted on the result of the research may provide some practical reference for future cohesion studies in Chinese texts. Although there are publications that have discussed the HSK Corpus design and its relevant parameters, including those related to HSK Composition Corpus (Zhang 2003, Shen & Huang 2010a, 2010b), the discussion in this study, together with the methodologies for annotation and

concondancing, focusing in particular on research in lexical cohesion in Chinese non-narrative texts. They, therefore, may be especially useful for future studies that examine various aspects of the use of lexical cohesion in Chinese texts, CFL learner texts, as well the comparison between the two corpora.

Several theoretical implications also arise from the findings of this study. The taxonomy of cohesion proposed by Halliday and Hasan (1976) is originally the result of the analysis of English texts. Although it has been widely adopted by scholars to study cohesion in other languages, such as Chinese, the taxonomy may not be able to fully reflect the salient linguistic features of other languages. Despite the fact that cohesion is a cross-language, linguistic phenomenon, the findings of this study, and especially the new categories that this study proposes, raise questions about whether or not cohesive devices have cross-cultural and cross-genre variations. If these new categories of cohesion can be found to be employed in other languages, the commonly adopted taxonomy of cohesion may be enriched and expanded. Further investigation, therefore, may need to explore whether or not these devices contribute integrally to the organization of argumentative texts written in other languages, the results of which may offer more depth in understanding the properties and functions of cohesion in texts, and therefore may help further contribute to overall theories of cohesion.

1.3 Overview of the chapters

Chapter 2 reviews the literature that is related to the current study, and then goes further to explain how the study is linked to existing research on cohesion as well as the gap in research knowledge that the study attempts to fulfil. This chapter covers four main areas: contrastive rhetoric, research relating to cohesion, corpus linguistics, and the research in the TCFL field.

Chapter 3 outlines the overall research methodology employed in the present study. It first presents the research questions that the study intends to address, followed by the research design, where the major theoretical frameworks are

illustrated and explained. The last part of the chapter discusses corpus design related issues, such as data collection, corpus annotation, and concordancing.

Chapters 4 and 5 are the two findings chapters, where the patterns of the use of cohesion, especially lexical cohesion, in both the NCEE and the HSK corpora are analysed and presented in details separately.

Chapter 6 is a short chapter, which mainly presents a summary of the comparison between the two sets of data reported in Chapter 4 and 5. This chapter prepares for the discussion in Chapters 7 and 8.

Chapter 7 discusses the gap in the patterns of the use of non-lexical cohesion between the NCEE Corpus and the HSK Corpus. Although non-lexical cohesion is not the central focus of the present study, this discussion helps to understand the overall framework of cohesion in Chinese non-narrative texts and, how cohesion contributes to the organization of text.

Chapter 8 interprets the gap in the patterns of the use of lexical cohesion between the NCEE Corpus and the HSK Corpus. It discusses the possible reasons for the existence of the gap, identifies the areas that may have contributed to the gap. The chapter also discusses implications for future studies in the TCFL field, as well as implications for CFL pedagogy and development of teaching and learning materials. The limitations of the present research are also discussed at the end of the chapter.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

In the current TCFL literature, few studies have investigated lexical cohesion in Chinese non-narrative text by using corpora of both native Chinese speakers and CFL learners. Drawing on contrastive rhetoric research, lexical cohesion theory, and corpus linguistics, this research is an attempt to address this lack. To provide a theoretical framework for the study, and to highlight the gaps that the study aims to fill, this chapter reviews research and theory from each of these three research strands. Section 2.1 reviews contrastive rhetoric, Section 2.2 lexical cohesion, and Section 2.3 the application of corpus linguistics to the area of language teaching. Following the review of each research strand, its implications for the current study are discussed. It is important to note at this point that the critical reviews of the literature of each of these three research strands are regarded as more complementary than contradictory in terms of their analytical applications to this study.

2.2 Contrastive Rhetoric as a field of study

Beginning in the 1960s, CR has become established as a discipline combining a wide, interdisciplinary body of theory and research, drawn from composition and rhetoric, text linguistics, translation studies, genre analysis and cultural anthropology, for the purpose of comparing discourse structures across cultures and genres. CR aims to address the writing problems of ESL learners and to help teachers of writing to design content and instructional materials that will help students to write according to expected conventions, especially at the paragraph and text levels. In this section, aspects of the research in the field of CR are discussed, such as the establishment of CR as a new branch of Applied Linguistics, the development of CR, the critics it received and the new directions of CR. The implications for the present study are summarized at the end of this chapter.

2.2.1 The establishment of contrastive rhetoric

Contrastive rhetoric research was pioneered by Kaplan (1966), an American applied linguist, in his study of the organization of paragraphs in ESL student essays. On the basis of his study, and drawing on the Sapir-Whorf hypothesis of linguistic relativity, Kaplan suggests that language and writing are cultural phenomena. He describes the structure of English exposition as linear. For example, he proposes that a paragraph in English typically begins with a topic sentence supported by examples that are related to the central theme, while the paragraph development in other languages involves different organization patterns, which he characterized in the forms of circles, parallels or zigzags. This assumption implies that ESL learners will encounter difficulties caused by differences between the discourse-level organization of their first language and English. He, therefore, recommends that ESL writers should learn how to organize their writing in ways that are accepted by native English speakers.

On the basis of Kaplan's (1966) analytic framework, a considerable number of research studies have been undertaken comparing writing patterns and styles in many languages and cultures, such as Arabic, Chinese, Japanese, Korean, German, Finnish, Spanish, and Czech. Among these studies in the field of CR, those by Hinds (1984, 1987, 1990) are worthy of note. For example, by comparing texts from a particular newspaper column written in Japanese with their English translation, Hinds (1987) concludes that Japanese is a reader-responsible language while English is a writer-responsible one. Hinds (1990), therefore, suggests that Japanese ESL learners might need to be taught that they should not assume that a reader's task is to "think for themselves, to consider observations made, and to draw their own conclusions", and that a writer's task is "to stimulate the reader into contemplating an issue or issues that might not have been previously considered" (p. 99-100). In his further exploration of the differences between some oriental languages, such as Chinese, Thai, and Korean with English with respect to organizational patterns and coherence, Hinds claimed that the writing style of these languages is what he terms a "quasi-inductive style" or "delayed introduction of purpose", while English text has a more explicit inductive style.

2.2.2 The development and expansion of contrastive rhetoric

Over the last two decades, the field of CR has expanded from focusing only on research into paragraph-level features, cultural thought patterns and cognitive factors to the exploration of different genres, social contexts, subcultures as well as composing processes. For example, in her important monograph on contrastive rhetoric research, Connor (1996) presents a comprehensive overview of the field of CR, followed by an exploration of how it interfaces with other relevant disciplines. Using examples from a wide variety of languages, Connor examines the influences of the learner's L1 on the acquisition and the development of literacy skills in a second language and how this can facilitate the teaching and learning of writing in ESL classrooms. In response to criticisms, such as the lack of a unified methodology, lack of a single research programme and a preoccupation with studies on surface structures, Connor argues that the very multidimensional nature of CR actually allows researchers to analyze a single issue from multiple perspectives. Thus, as a result, Connor's redefining of CR has helped "expand the field of CR to include different languages, text types, genres, and research methodologies" (Casanave, 2004, p. 40). CR, therefore, should not be considered as limited solely to study the effects of linguistic transfer from L1 to L2. The emphasis should also be on writing processes, as well as on cross-cultural variations and on writing for specific genres.

The new trends which have emerged in the study of CR can be summarized as follows: an expansion into genre-specific studies and a shift of emphasis from the written product of L2 writers to the social context of writing. The first trend of genre-specific CR research is exemplified by studies which involve various types of text. For instance, Connor analyses texts such as interview data, European Union grant proposals prepared by English and Finnish scholars, case reports in business management, and argumentative essays written by students from four countries who participated in the IEA (*International Association for the Evaluation of Education*) project (Connor et al., 1995; Purves, 1988). The second trend in CR involving a shift in emphasis from product to process is illustrated by the argument that the activity of writing should be considered as being socially situated; each writing activity has its own specific contextual situation within

which researchers may be required to examine its audience, purpose, interaction, collaboration, and composing strategies (Connor 2002). Some scholars, such as Atkinson (2004), claim that much larger discursal and ideological factors as well as individual ideology should also be taken into account in the analysis of ESL writing.

2.2.3 The critiques of contrastive rhetoric

Although Kaplan's (1966) famous diagrams summarizing five types of paragraph formation and Hinds' (1984, 1990) conclusion that English is a writer-responsible language have been widely reproduced and cited by many language writing books and publications, they are now criticized by many scholars and researchers for simplifying and overgeneralizing a complex idea and providing simple treatments. The critiques of CR focus on the following aspects: the assumptions of CR, concepts of CR and their political and ideological connotations, the methodologies, and the pedagogical implications.

In her critique of Kaplan's assumptions, Péry-Woodley (1990) claims that he failed to pay attention to a number of important variables that could impact on students' writing other than L1 rhetorical conventions. These variables can be "projected audience, writer-reader interaction, communication situation, text-type...or sociolinguistic factors" (p. 148). Atkinson (2002) argues that because there are so many genres, styles, text-types, and registers in different sub-cultures, any generalization that "writing in English is linear" must be obviously wrong. Atkinson (2002, 2004) also takes a critical perspective on the definition of "culture", and claims that CR needs a better conceptualization of culture in the field of L2 writing for the 21st century. Instead of viewing culture as "received", "big", "national", and "product", he proposed "new definitions of culture that allow writing researchers to view culture as fluid and dynamic following current thinking in cultural and post-modern studies. Kubota (2004) claims that "CR has tended to construct static, homogeneous, and apolitical images of the rhetorical patterns of various written language" and has "implicitly reinforced an image of the superiority of English rhetoric"(p. 9). The binary image of rhetoric parallels

colonial dichotomies between the colonizer and the colonized and suggests “the hidden political and ideological nature of conventional knowledge created by CR (p. 24).

Among many problems in CR research, comparability is perhaps the most serious. In her book, Connor (1996) acknowledges some methodological problems of CR which have been identified in the critiques: “small sample size, a mix of genres, and generalizing from L2 data to L1 behaviour ... lack of adequate levels of comparability in the studies” (p.162). Swales (1990) points out that it is important to consider the variability of genres in contrastive rhetoric studies. For example, he critiques Ostler’s 1987 study that compares Arabic students' English writing to English paragraphs taken from books. Swales asserts that researchers need not only to describe text types (e.g. narrative, descriptive, argumentative) or situations (writing in particular discourse communities), but also to consider who is doing the writing or what their purpose is in writing. Matsuda (1997) criticizes CR that only looks at the writer’s and reader’s linguistic, educational, and cultural background. The writer’s agency is denied. Casanave (2004) notes that Connor’s interest in text analysis leads her “to neglect attention to important sociolinguistic, developmental, or ideological factors that influence how, why, and what people write” (p. 40).

Only focusing on differences in the writing conventions between L1 and L2 is another weakness claimed by many CR critics. Leki (1997) suggests that CR’s primary focus on differences rather than similarities “has a distancing and exoticizing effect”(p. 242), which in reality does not help us much to understand the complex issues of L1 and L2 writing. Atkinson (2004) calls for more attention to what he terms plurality, continuity, universality, hybridity, and similarity within and across languages and cultures.

On the basis of theory and research, a considerable number of pedagogical recommendations have been made by contrastive rhetoricians with the aim of facilitating the teaching and learning of L2 writing. These recommendations include: reorganizing scrambled paragraphs and filling out an outline following

given topic sentences (Kaplan 1966, 1972); composing by imitating models or following outlines (Kaplan 1967); making students aware of that writing is a social activity that requires more than attention to syntactic and lexical items (Kaplan 1988); identifying topic sentences in texts, sorting supporting information from its generalization, and focusing on a particular topic to match the writer's world knowledge (Grabe & Kaplan, 1989); making ESL students aware that English-speaking readers are convinced by facts and statistics, expect explicit links between main topics and subtopics, and value originality (Leki, 1991, 1992).

These recommendations and suggestions, however, have not been considered to be “effectively translated into the practice of teaching organizational structures” (Matsuda, 1997). Leki (1997) points out that because scholars have paid insufficient attention to the ideological implications of their work and highlights the dangers of cultural essentialism (p. 244). Consequently, the tendency of stereotyping students according to their ethnic or cultural group membership may result in ignoring students' self-concepts, experiences, beliefs and their sense of identity and overlooking the underlying ideological presuppositions, biases in particular written text. Matsuda therefore suggests that in pedagogical applications of CR, the writer's and reader's background needs to be expanded. He proposed a “dynamic theory of L2 writing” for overcoming the limitations of traditional CR and its implications for the pedagogy of L2 writing, which should at least include expanding the conception of backgrounds of L2 writing and considering the discourse community as a space shared by both the writer and the reader, and placed at the intersection of the backgrounds of the writer and the reader.

2.2.4 New directions proposed for contrastive rhetoric

The controversies in CR research not have only made scholars ask thought-provoking questions about CR, but also inspire researchers to look for new directions for exploration of CR study and application. The following are some of the major suggestions made for CR research in the future.

Attempting to address some recent critiques of CR and to suggest new directions for a viable CR, Connor (2004) proposes the term *intercultural rhetoric* to reflect what has been advanced to broaden cross-cultural studies of writing. Corresponding to this change, Connor proposes new methodologies for CR research, which adopts research-situated reflexivity and “is becoming more sensitive to the social context and the local situatedness and particularity of writing activity” (Connor, 2002, p. 506).

In terms of helping with classroom decision-making process, Casanave (2004) recommends that writing teachers take an *investigative pedagogical approach* in their classroom practice. To implement this approach in the classroom, teachers need to bear several points in mind: firstly, the CR issues are perhaps not appropriate to be introduced to beginners; secondly, not all students need to learn how to write academic essays or articles, decisions have to be made on the basis of students’ real need in L2 writing courses; thirdly, if there is a need, teachers and students can investigate texts in students’ L1 and L2, including consideration of the kind of writing instruction that students have received, and the differing expectations of readers in the L1 and L2. Lastly, students can also be involved with investigations about how teachers of writing in different cultures respond to writing or how they themselves respond.

Critical contrastive rhetoric, as an alternative of CR, is proposed by Kubota (2004) and suggests a re-conceptualization of cultural difference in rhetoric by incorporating post-structuralist, postcolonial, and post-modern perspectives of language and culture, which expands the focus of CR from on culture in written communication of to concepts such as power, discourse, and subjectivities. As to the pedagogy of critical contrastive rhetoric, Kubota (2004) maintains that it “affirms multiplicity of languages, rhetorical forms, and students’ identities, while problematizing the discursive construction of rhetorical and identities, and thus allowing writing teachers to recognize the complex web of rhetoric, culture, power, and discourse in responding to student writing” (p. 7).

2.2.5 Contrastive rhetoric research in China

Over the last two decades, a great number of scholars and researchers in China have focused on the challenges and difficulties of Chinese students learning English as a foreign language specifically for the purpose of facilitating and improving the teaching and learning of English writing. The research has focused on the following three main aspects of writing:

- the cultural differences associated with writing practices in Chinese and English;
- the influence of learners' L1 on their L2 writing; and,
- the study of the process of writing and the possible process model of L2 writing.

In China, many studies have focused on exploring and summarizing the different writing patterns influenced by particular Chinese philosophies and mode of thinking. For instance, Ji (2006) discusses some of the methods of reasoning that appear in Chinese academic writing, which are directly influenced by Chinese thought patterns. Liu and Zhou (2004) investigate how Chinese EFL students are affected by the differences between Chinese and English language and culture. Specifically, they propose that:

- the major difference between Western and Chinese cultures can be described as linearity versus circularity. Western linearity places stress on the two oppositions and division: man to nature, subject to object, whereas Chinese circularity stresses their interdependence and integration;
- in scientific study methods, Westerners place strong importance on evidence and analysis and there is a strong focus on logical thinking and abstract reasoning. By contrast, in Chinese tradition, people are more accustomed to observing and judging things as an enclosed whole. Experiment and analysis are not rejected in the studies, but intuition and instinctive experience and entire synthesis predominate;
- differences can also be reflected in the patterns of writing. In Western writing

or speech, one can know the content of an article by looking at its title, subtitles, and topic sentences of each paragraph. The main points are often signaled by cohesive devices, such as “first,” “second,” “next,” “finally”, “because,” “however,” and “actually,” alerting the reader to sequence, time, and logical relationships. On the other hand, in Chinese writing or speech, great importance is attached to the unity and harmony of the whole piece. Attention is normally given to the correspondence between introduction and conclusion and the natural transition from one point to another, rather than the clear-cut division between different sections;

- when quoting other people’s opinions in an article, Western writers like to acknowledge sources for their ideas. In contrast, the traditional practice of Chinese writing involves drawing on a wide range of sources without specific and obvious citation; and,
- the cultural differences between English and Chinese can also be found in various aspects of the language system. For instance, many researchers discuss the widely accepted notion of the importance of hypotaxis in English syntax and parataxis in Chinese. This contrast can be reflected at every level of the language system, from phonetic forms to frequency of the use of noun, verb and preposition, from sentence structure to tense and aspect, from cohesive devices to text and discourse structure.

Over the last two decades, a large number of studies in China have been undertaken to investigate the transfer between the learners’ L1 and their L2. Adopting methodologies similar to Cumming (1989), Kobayashi and Rinnert (1992), and Lay (1982), Wen (1998, 2004) conducted a survey of the writing of 180 Chinese students majoring in English at two Chinese universities. Her findings suggest that students’ L1 writing ability, command of vocabulary, and capacity for text construction may greatly facilitate their English writing. However, other scholars propose that students’ L1 writing practices may actually exert a negative influence on their learning of English writing. By analyzing lexical, grammatical and textual errors made by Chinese EFL students, Liu and Wang (2003) find that many students have difficulties in selecting appropriate vocabulary, subject-verb agreements, use of tense and aspect and voice,

organization and construction of text due to their lack of the knowledge of the target language and culture. Pan and Wu (2006) point out that the misunderstanding of pragmatic knowledge and rhetorical devices of the English language could also lead to unsuccessful writing performance.

Some researchers have focused on investigation into the process of L2 writing. Guo and Liu (1997), Wang and Wen (2002) find that the L2 writing process is a bilingual event, in other words, L2 writers use both their L1 and L2 when they are composing in L2. Adopting a combination of research methodologies, such as think aloud, observation, retrospection, and interviews, Wang and Wen (2002) investigate the writing processes of 16 English majors in Chinese tertiary institutions. Their qualitative and quantitative study shows that the influence of the L1 decreases with the development of the L2, and the tendency to use the L1 varies with individual composing activities. They divide the process of writing into five different stages: analyzing tasks, planning content, constructing structure, text output, and process management. Among these stages, analyzing task and text output involve the least amount of the use of L1, while the process management involves the highest levels of L1 use.

2.2.6 Summary of contrastive rhetoric and implications for the proposed study

CR is an interdisciplinary field which identifies rhetorical problems in L2 writing and attempts to explain them on the basis of the assumption that language and writing are cultural phenomena. Its pedagogical goal is to provide a knowledge base to help L2 teachers and students and to raise the awareness about differences in writing patterns across cultures. However, early CR studies have faced sharp criticism for their overgeneralization, over-interpretation and over-simplification of complex issues. Therefore, CR researchers have found it necessary to broaden their interests to include not only the written texts as finished products, but also the processes that students go through as they compose, and the social and cultural contexts within which those processes are situated. There also has been an increase in the study of different types of written texts and genres and in the

employment of more qualitative and quantitative methods. To advance the development of CR, Connor (2004) proposes a new term intercultural rhetoric to reflect the new trends of study in CR, and Kubota (2004) calls for establishing a critical contrastive rhetoric which may be a solution to tackling the complex web of culture, rhetoric, power and discourse. Casanave (2004) also suggests that an investigative pedagogical approach will help teachers of L2 writing to make their decisions in classrooms. Despite the problems that remain unresolved in the field of L2 writing, researchers agree that the need for L2 writers to learn the organization of L2 written text still exists. Consequently, CR as a field still appears to have considerable potential and deserves the attention of L2 scholars.

CR research provides in-depth insights and useful references to the proposed study. As one of the main text organizational devices, lexical cohesion has been among the research topics of many ESL writing studies (Al-Jubouri, 1984; Connor, 1984; Guo, 2007; Reynolds, 1995 & 1996; Scarcella, 1984; Tierney & Mosenthal, 1983; Witte & Faigley, 1981; Zhu, 2001). However, compared to these studies, the research on CFL writing is still in its infancy stage. For example, little is known about the use of lexical cohesion in the CFL writing. Therefore, further and deeper explorations are needed to inform pedagogical knowledge and professional practice in this area. The consensus that CR researchers have reached on that L2 writers need to learn the organization of L2 written texts, therefore, provides the researcher motivation to compare the organizational patterns between the texts written by native Chinese speakers and CFL learners. The CR study of both differences and similarities in organizational patterns of writing will justify the purpose of the proposed study. Despite the controversies and debates in CR, many scholars agree that the differences in organizational patterns of writing across and within disciplines, languages and cultures, similar characteristics in rhetorical patterns also need to be explored (Leki, 1997; Atkinson, 2002; Casanave, 2004; Kubota, 2004; Moreno, 2004). These studies therefore provide a firm basis for the proposed exploration of differences and similarities between English and Chinese non-narrative text in terms of lexical cohesion.

In terms of research methodology, the existing CR studies and debates also shed

light on how to deal with sampling issues. Comparability has been one of the major methodological issues in CR research. Over the years of study and exploration, researchers have come to realize the importance of defining genres and discourse communities in contrastive rhetoric analysis. Swales (1990) points out that contrastive analyses need to compare and contrast texts written for similar purposes in similar situations. This criterion of sampling in contrastive study – namely genre-specific research- is crucial to the proposed topic, which requires collecting data from academic writing and L2 writings.

The implications of some CR studies will serve as a guide for the proposed study to explore possible pedagogical solutions to CFL learners' writing problems. The findings of CR have helped to raise awareness about linguistic and cultural differences among teachers and students of ESL writing, which in turn is valuable for teachers and students for psychological and practical reasons. Similarly, the application of the findings of this study will be discussed for the purpose of providing suggestions on how to help teachers and students to investigate the lexical features in Chinese text and choose appropriate teaching and learning strategies accordingly.

2.3 Textual cohesion: theory and research

This section reviews the following aspects of the current literature on cohesion related research: various definitions of cohesion, types of cohesion, lexical cohesion and its organizational role in written text and different approaches to cohesion analysis. The implications of this research for language study and the study of cohesion in Chinese are also evaluated. The last part of this section is a summary and discussion of the implications of this literature for the present study.

2.3.1 The definition of cohesion in text

Cohesion is a very important concept in text linguistics. After the publication of Halliday and Hasan's work *Cohesion in English* (1976), their particular approach to defining and describing cohesion became widely known and widely used

within the field of text linguistics. However, it is important to note, that despite the considerable influence this approach, there is not a general consensus as to the definition of the concept of cohesion. For example, van Dijk (1972), Gutwinski (1976), de Beaugrande and Dressler (1981), and Hoey (2000) all offer differing definitions of cohesion. Nevertheless, in this section, the account of cohesion provided by Halliday and Hasan (1976) will be briefly outlined since “this is by far the most comprehensive treatment of the subject and has become the standard text in this area” (Brown & Yule, 1983, p. 190). In presenting some of their key concepts relating to cohesion, mention will be also made of the challenges and concerns that other theorists raise in relation to some of these concepts.

Halliday and Hasan (1976) take the view that “the concept of cohesion accounts for the essential semantic relations whereby any passage of speech or writing is enabled to function as text” (p. 13). Cohesion connects a string of sentences to form a text rather than a disconnected sequence of sentences. They hold that the primary determining factor of whether a set of sentences can or cannot constitute a text depends on *cohesive relations* within and between the sentences, which create *texture*. “A text has texture and this is what distinguishes it from something this is not a text” (p. 4). The texture is provided by the cohesive relation that exists between a referring item and an item that it refers to. In other words, cohesion occurs where the interpretation of certain element within a text is dependent on that of another.

To refer to an occurrence of a pair of cohesively related items, Halliday and Hasan use the term *cohesive tie*. They state that “[t]he concept of a tie makes it possible to analyze a text in terms of its cohesive properties, and give a systematic account of its patterns of texture” (p. 4). An example of what a cohesive relation is and how it is set up can be seen in the following example:

Example 1 (taken from Halliday & Hasan, 1976, p. 3):

- (1) Wash and core six cooking apples. Put them into a fireproof dish.

In Example (1), the pronoun “them” in the second sentence refers back to the “six cooking apples” in the first sentence, and it is the cohesive tie between “them”

and the “six cooking apples” that gives cohesion to the two sentences, so that these two sentences can be interpreted as a text.

Halliday and Hasan (1976) also make a distinction between cohesion and linguistic structure. They maintain that “a text...is not a structural unit; and cohesion...is not a structural relation” (p. 6). “Cohesive relations have in principal nothing to do with sentence boundaries” (p. 8) and, therefore, are not restricted by sentences. They also stress that cohesion is not a synonymy of discourse structure such as paragraph, episode, or topic unit, which is normally “used to refer to the postulated unit higher than the sentence, for example the paragraph or some larger entity such as episode or topic unit” (p. 10).

2.3.2 Different types of cohesion

Halliday and Hasan (1976) believe that cohesion is expressed through the different organizational strata of language. Just like meaning is expressed through grammar and vocabulary, and grammar and vocabulary are in turn realized through sounds and writing, cohesion is also “expressed partly through the grammar and partly through the vocabulary” (p. 5). On the basis of this multiple coding system of language, Halliday and Hasan divide cohesion into two general categories: *grammatical cohesion* (*reference, substitution, ellipsis, conjunction,*) and *lexical cohesion*. Since lexical cohesion is the focus of this study, it is discussed in next section. The following is a brief review of their approach to grammatical cohesion.

Reference is a well-researched area in text linguistics (see, for example, Chastain, 1975; Geluykens, 1994; Lyons, 1979; Tyler, 1978; Webber, 1978). Halliday and Hasan (1976) define reference as a case where the semantic interpretation of a certain item depends on referring to another item. This is exemplified the following example:

Example 2 (taken from Halliday & Hasan, 1976, p. 31):

(2) Three blind mice, three blind mice.

See how they run! See how they run!

In Example (2), the word “they” refers to “three blind” mice. In “See how they run!”, “they” means not only three blind mice, “but the ‘same three blind mice that we have just been talking about” (p. 31). According to the distinction made between items being interpreted by something that is in the context of the situation or that within the text, Halliday and Hasan (1976) divides reference into two classes:

- *exophoric reference*, reference to some aspect of the context of situation, i.e. a reference to the outside world;
- *endophoric reference*: reference to some element within the text of the discourse itself; it is either *anaphoric*, referring to preceding text; or *cataphoric*, referring to text that follows.

Halliday and Hasan (1976) also identify the following types of reference: *personal*, *demonstrative* and *comparative* reference. Consider the following examples:

Examples 3, 4, and 5 (taken from Halliday & Hasan, 1976, pp. 50, 60, 79):

(3) My husband and I are leaving. We have seen quite enough of this unpleasantness. (4) We went to the opera last night. That was our first outing for months. (5) It’s the same cat as the one we saw yesterday.

In Example (3), “we” is a personal reference, which refers back to “my husband and I”; in (4), “that” is a demonstrative reference, which refers to “we went to the opera last night”; in (5), “same” is a comparative reference, which refers to “the one we saw yesterday”.

In defining substitution, Halliday and Hasan (1976) describe it as a relation within a text and a “substitute is a sort of counter which is used in place of the repetition of a particular item” (p. 89). Ellipsis can be defined as “SOMETHING UNDERSTOOD, where understood is used in the special sense of ‘going without saying” (p. 142). According to Halliday and Hasan (1976), substitution and

ellipsis essentially are the same process: “substitution as the replacement of one item by another”, while “ellipsis can be interpreted as that form of substitution in which the item is replaced by nothing” (p. 88). There are three types of substitution: *nominal*, *verbal* and *clausal*. Consider the following examples:

Examples 6 and 7 (taken from Halliday & Hasan, 1976, pp. 89, 143):

(6) My axe is too blunt. I must get a sharper one.

(7) Joan brought some carnations and Catherine some sweet peas.

Sentence (6) is an example of substitution, where “one” is substituted for “axe”. Sentence (7) is an example of ellipsis, where the verb “brought” in the second clause is omitted in the context.

Halliday and Hasan (1976) notes that conjunction, as a kind of cohesive relation, differs in nature from reference, substitution and ellipsis, because “they express certain meanings which presuppose the presence of other components in the discourse” (p. 226), instead of being devices which connect with into the preceding (or following) text. In other words, conjunction is a relationship indicating how the subsequent sentence or clause relates to the preceding or the following sentence or parts of sentence. The most frequent conjunction relationships are: *causal* (for, because, so), *temporal* (before, after, then), *adversative* (but, however, yet), *additive* (and, furthermore, or) or *discourse markers* (now, well, after all) (p. 242). However, taking a different approach, de Beaugrande & Dressler (1981) define conjunction as a *default junction*, a term they use to refer to “a clear device for signaling the relationships among events or situations” (p. 71). They identify four major types of junction: *conjunction*, *disjunction*, *contrajunction* and *subordination*.

2.3.3 Lexical cohesion

According to Halliday and Hasan (1976), lexical cohesion is “the cohesive effect achieved by the selection of vocabulary” (p. 274), but this does not mean that a lexical item always has a cohesive function. A lexical item may enter into a

cohesive relation, “but by itself it carries no indication whether it is functioning cohesively or not”, and “that can be established only by reference to the text” (p. 288). There are two types of lexical cohesion: *reiteration* and *collocation*. Reiteration may be a repetition — simply repeating the same word in a context, or in the form of synonymy, hyponymy, superordinate, or general word. Consider the following examples of each type of reiteration:

Examples 8, 9, 10, and 11 (taken from Halliday & Hasan, 1976, pp. 278, 275):

Repetition

(8) There is a large mushroom growing near her, about the same height as herself ... She stretched herself up on tiptoe, and peeped over the edge of the mushroom...

Synonymy

(9) Accordingly... I took leave, and turned to the ascent of the peak. The climb is perfectly easy...

Hyponymy

(10) Henry's bought himself a new Jaguar. He practically lives in the car.

General word

(11) Can you tell me where to stay in Geneva? I've never been to the place.

In Example (8), there is a repetition: “mushroom” refers back to “mushroom”. In (9), “climb” refers back to “ascent”, and “climb” is a synonym of “ascent”. In (10), “car” refers back to “Jaguar”, and the former is a superordinate of the latter. In (11), the “place” refers back to Geneva, and “place” is a general noun functioning cohesively.

According to Halliday and Hasan (1976), collocation is defined as “a systematic relationship between a pair of words...that stand to each other in some recognizable lexico-semantic (word meaning) relation (p. 285)”. The pair of words can not only be synonyms and near-synonyms such as climb...ascent, beam...rafter, and superordinates such as elm...tree, boy...child, but also pairs of opposites of various kinds, complementaries such as boy...girl, stand up...sit down, or antonyms such as like...hate, wet...dry (p. 285). The concept of collocation has become controversial in the field of text linguistics (Halliday & Hasan, 1976; Mel’cuk & Pertsov, 1987; McCarthy, 1991; Hoey, 2000). Halliday and Hasan (1976) themselves acknowledge that collocation is “the most problematic part of lexical cohesion” (p. 284). McCarthy (1991) maintains that “it is debatable whether collocation properly belongs to the notion of lexical cohesion” (p. 65), because collocation only implies the probability that lexical items will co-occur, but it does not mean there is a semantic relation between words. Hoey (2000) states that Halliday and Hasan can be criticized for labeling relations such as boy-girl, wet-dry, order-obey, sunshine-sky as collocations, because collocation has long been used for referring to “the relationship a lexical item has with items that appear with greater than random probability in its (textual) context” (p. 7).

2.3.4 The organizational role of lexical cohesion

In his (2000) work, Hoey points out that his aim of his reported study is to “describe a new system of analysis based on the study of cohesion, particularly on lexical cohesion” (p. 3). He further states that what distinguishes his study from previous research on cohesion is that “the attention is not primarily on itemizing cohesive features but on observing how they combine to organize text” (p. 3). In other words, his study focuses more on the organizational role of lexical cohesion. This differs greatly from Halliday and Hasan’s (1976) study. According to Hoey (2000), although “Halliday and Hasan (1976) have a chapter entitled ‘The meaning of Cohesion’, it is disappointingly fragmentary and has little to say about the issues we have raised.” The questions raised by Hoey are:

1. How does the presence of cohesion contribute to the coherence of a text?

2. How does the presence of cohesion affect the ways in which sentences are perceived to be related to each other as complete propositions?
3. Does cohesion contribute to creating the larger organization of a text (if such exists)? (p. 11)

In his discussion about the above questions, Hoey summarizes these questions as the following questions or statements: “Question 1 becomes a question about how the presence of a cohesive tie predisposes a reader to find a text coherent; Question 2 is concerned with the statement that sentences have a meaning together that is more than the sum of their separate parts; Question 3 is about the belief that text is patterned and that “cohesion, particularly lexical cohesion, does contribute importantly to a text’s organization” (p. 14). After analyzing Halliday and Hasan’s (1976) own sample of cohesion analysis, Hoey shows that the lexical cohesion accounts for over forty percent of the ties in Halliday and Hasan’s own analyses, which leads Hoey to believe that “lexical cohesion is the only type of cohesion that regularly forms multiple relationships” and that lexical cohesion is the “dominant mode of creating texture” (p. 10).

In his discussion of the above questions, Hoey acknowledges previous research studies in lexical cohesion he draws on, such as Hasan (1984), Winter (1974, 1979), and Phillips (1985). After reviewing these works, Hoey concludes that “the implication of the work of Hasan, Winter, and Phillips is that cohesion does contribute to coherence” (p. 25).

2.3.5 Systems of cohesion analysis

Since Halliday and Hasan’s (1976) publication of ‘Cohesion in English’, the methodologies of cohesion analysis have undergone considerable changes. In the last chapter of their work, Halliday and Hasan (1976) suggest a method for the analysis of cohesion in text, which demonstrates how a text may be analyzed using their coding and describing system.

Dissatisfied with Halliday and Hasan’s system of analysis, Hoey (2000) develops a new system of analysis of cohesion for the purpose of proving that lexical

cohesion plays the most important part in the organization of a text. In his analytic system for lexical cohesion, he first defines his sample data as a set of non-narrative texts. He then classifies instances of lexical repetition into simple repetition, complex repetition, simple mutual paraphrase, simple partial paraphrase, substitution, co-reference, ellipsis, deixis. Because his intention is to explore the ways in which repetition links sentences, he constructs a repetition matrix which could accurately reflect the non-linear complexity of links and, therefore, help to interpret how they serve the text-organizing function. Hoey's system of cohesion analysis offers a new perspective on how the lexical cohesion can be examined and how to interpret them.

With the development of computational implications in the field of Applied Linguistics, a variety of analytic systems of cohesion have been proposed recently. For example, the Princeton *WordNet* is a large lexical database of English within which nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. This structure is useful in the area of computational linguistics and natural language processing. Since its establishment, a great number of WordNet- based studies have been undertaken to explore and analyze lexical patterns in English texts (Makuta, 1997; Harabagiu, 1999; Teich & Fankhauser, 2003; Vechtomova, Karamuftuoglu, & Robertson, 2006). These studies, to a great extent, have enriched and supplemented the study of lexical cohesion and will continue to advance the exploration in the field.

2.3.6 Implications of the study of cohesion study for language teaching

Halliday and Hasan's (1976) study on cohesion has not only inspired numerous research studies in the field of text analysis, but also encouraged scholars to consider how the findings can be applied to classroom instruction, especially in the teaching of reading and writing (see, for example, Hoey, 2000; McCarthy, 1991). For instance, in English language teaching, as Hoey (2000) comments, "it has become commonplace for materials to have units devoted (implicitly or explicitly) to encouraging the recognition and use of cohesive devices" (p. 4).

On the basis of his own findings, Hoey (2000) discusses the implications of cohesive devices for the teaching of reading and writing. Firstly, he recognizes that the role of the reader is a positively creative one rather than a passive one. He points out that when reading quickly, the reader normally look for signals of large-scale patterns of organization. When the purpose of the reading is looking for information or reference, however, reading is actually a matter of recognizing connections— making links between earlier and later parts of a text. This is also crucial in the EFL classroom, where identification of bonded pairs needs to be part of what is taking place. Moreover, contrary to what is most commonly practiced in second language classrooms, Hoey (2000) argues that “what should be taught, instead of unfamiliar and infrequent words in the text, are the bond-making items” (p. 241). For example, *disinvestment*, *disinvest*, *disinvesting*, *investment*, *investors* are probably the more frequently-encountered forms of *invest*, the teacher should teach students these forms instead of just teaching them the base form *invest* in a certain context.

In his discussion about developing the writing skill, Hoey (2000) suggests that the key point of learning to write is to learn how to make connections across the text with regard to topic sentences. He argues that although “the teaching and practice of cohesion has become a regular part of many teaching programmes” (p. 242), it has not been proved that this is either necessary or sufficient to produce good writers. He maintains that students who are learning to write, whether in their own or another language, should be encouraged to think of writing as non-linear, which is “to make connections between what they are currently saying and what they have previously said and later intend to say” (p. 243). To help learners of writing to develop the ability to use more morphologically complex words, he suggests that materials that could teach students how to bond back to earlier sentences need to be developed.

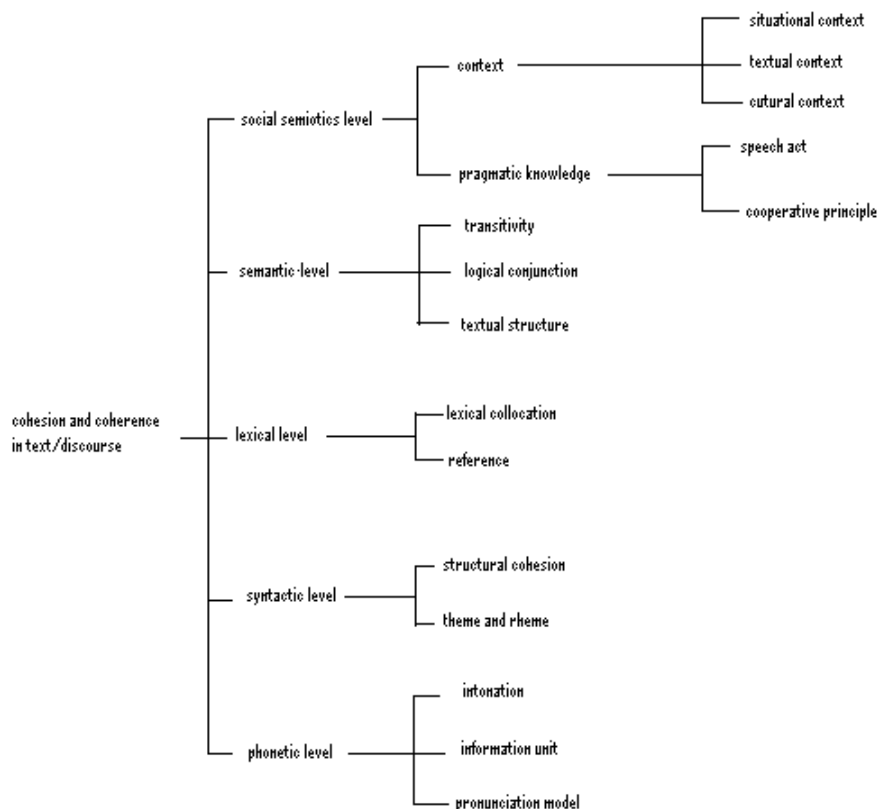
2.3.7 The study of cohesion in Chinese

On the basis of the theoretical framework of cohesion originated and developed by many Western scholars, a number of researchers in China have made their own

contribution to the advancement of the theory of cohesion and its relationship with coherence (Hu, 1994; Zhang, 2001; Zhu, 2001). Some of their points of views are unique and worth noting because they have attempted to incorporate some relevant linguistic features of Chinese language to the theory of cohesion. For example, drawing on theories from the fields of pragmatics, semantics, and phonetics, Hu (1994) expands Halliday and Hasan's (1976) framework of cohesion by integrating multi-level factors, namely symbolic-level factors, semantic-level factors, lexical-level factors, syntactic-level factors, and phonetic-level factors, of language into the system. His multi-level model of cohesion and coherence is shown in Figure 2 .1 (translated from the model in Chinese, p. 225).

In the preface of his book *Cohesion and Coherence in Text*, Hu (1994) states that apart from retaining Halliday and Hasan's (1976) basic concept of cohesion, his work attempts to incorporate different theories, such as transitivity, theme and rheme, pragmatics, tone group, intonation, and textual structure into the study of cohesive devices. He believes that, since language is always used in specific contexts and situations, it is closely associated with cultural context and situational context, and even more constrained by common knowledge and cooperative principles held by the language user in the context. He argues that cohesive devices are not only expressed through grammatical and lexical systems, but can be found at other levels, such as phonological, semantic, and social

Figure 2. 1 *Hu's (1994) multi-level model of cohesion and coherence*



symbolic levels. For example, he notes that pronunciation models, such as alliteration, rhyme, consonance, may help organize a text by rhyming sounds with each other. He holds that a pronunciation model has a cohesive function as long as it goes beyond the sentence boundary. Compared with Halliday and Hasan's (1976) discussion of the cohesive function of intonation, Hu's (1994) exploration appears somewhat more comprehensive.

Hu's (1994) study is conducted on the basis of the analysis of both English and Chinese text. This helped him to identify differences in cohesion between these two languages. For example, he points out that, there is no definite article equivalent to the English '*the*', which is under the category of demonstrative reference in Halliday and Hasan's (1976) classification of reference. In places where *the* is used in English text, Chinese text uses *zhe* (this), *na* (that) or puts words in different positions to achieve the same cohesive function. Another instance is the use of zero-anaphora in both English and Chinese texts. While it can be found in both languages, Chinese text tends to have more zero-anaphora than English (p. 65).

To summarize, by contrasting and comparing the cohesive devices in both English

and Chinese, Hu's (1994) study on cohesion expands the scope of existing approaches to cohesion, and enriches the categories of cohesive device and deepens our understanding of cohesion.

However, it is worth noting that some of the sample texts that he uses in his study may not reflect linguistic features of contemporary Chinese, as many of his examples are from works of literature such as *Honglou Meng* [A Dream of the Red Chambers], and *Ziye* [The Midnight]. The former was written in the middle period of the Qing Dynasty, and the latter was written in 1930s. In addition, since his study is not based on a large Chinese corpus, there may be other cohesive devices commonly employed in Chinese texts not identified in his study. Therefore, it would seem that there is a need to explore further the phenomenon of cohesion in Chinese text.

Apart from Hu (1994), Zhu (1995, 1997) also attempts to develop the theory of cohesion. One of his research interests lies in the study on cohesive devices, especially the complex issue of collocation. He points out that in Halliday and Hasan's (1976) classification, the semantic relations between words that have cohesive functions, such as *ill* and *doctor*, *garden* and *dig*, *bee* and *honey* are difficult to describe and explain. This is actually a question of how broad the scope of collocation should be. He suggests that these words can be classified according to their semantic function in the transitivity system. For instance, the relationship between *bee* and *honey* can be categorized as *actor + goal* (bee + produces + honey); while *garden* and *dig* can be *process + goal* (dig + garden).

Besides these theoretical contributions to the study of cohesion, a number of research studies have been devoted to the exploration of differences in the use of cohesive devices between certain types of English and Chinese texts. For instance, Zhang (2005) compares 52 English and Chinese news texts on political and economic issues (26 in each language) to find out whether the frequency of use of certain cohesive devices is different in English and Chinese news. Her findings suggest that the devices used in the news in two languages are basically the same, but the frequency of the use of these devices differs in these respects: Firstly, English uses more referential devices than that in Chinese. Secondly, the two

languages differ in the use of *zero anaphora* and ellipsis. By *zero anaphora*, we mean some omitted referential words which normally should appear in the appropriate position. The following are examples of zero-anaphora.

Examples 12 and 13 (taken from Hu, 1994, pp.64-65):

(12) But, (0) admitting that we were all of English descent, what does it amount to?

(13) Wojiarugeming, (0) ye zaoguoduoshao ci de fan.

In Examples (12) and (13), both (0) are zero anaphora. Zhang's (2005) statistics suggests that in Chinese, zero anaphora accounts for 10.4% in her total number of cohesive devices, while in English only 2%. Similarly, the use of ellipsis in Chinese news is much more frequent than that in English. Lastly, the difference also lies in the use of simple lexical repetition and collocation. As mentioned previously, Chinese texts use more simple lexical repetition and fewer referential devices than English does. Zhang (2005) also suggests that in her survey more lexical collocations are used in English news than in Chinese news.

2.3.8 Summary of the research on cohesion and its implications for the proposed study

The concept of cohesion was firstly advocated by Halliday and Hasan (1976) in their representative work *Cohesion in English*, in which they provide a detailed description of what a text is and what cohesion is in a text. They also present taxonomy of the types of cohesive relationship that may occur within a text, which includes grammatical cohesion (substitution, ellipsis, conjunction, reference) and lexical cohesion (including collocation). To describe and analyse cohesion, they use the term *tie* to refer to the occurrence of a pair of cohesively related items. Among the later discussions about the role of cohesion in the organization of text, Hoey's 2000 study is clearly significant. By re-analysing Halliday and Hasan's (1976) own sample analyses at the end of the 1976 book, Hoey claims that lexical cohesion is the single most important form of cohesive

tie, which convinced him that lexical cohesion is the dominant mode of creating texture and, therefore, of the greatest importance in text organization.

Thus, over the 30 years of study of cohesion in text, various systems of cohesion analysis have been developed and applied to research in cohesive ties and relations in text. Some of the most influential research approaches are Halliday and Hasan's (1976) coding system that focuses on cohesive ties, and Hoey (2000) repetition matrix which is used to investigate lexical cohesion in terms of repetition bonds between sentences. Also a variety of computational models and computer-based or on-line lexical database have been developed to explore the lexical patterns in English text. This research on cohesive devices in text has been widely applied to the field of second language teaching. As a result, it has become a common practice in language classrooms to encourage the recognition and use of cohesive devices. Studies on lexical cohesion have been found useful to answer difficult questions related to various aspect of language use, such as reference, word order, tense and aspect choices (McCarthy, 1991, p. 39). Hoey's (2000) study also provides significant insight into how to teach lexis, how to improve students' reading and writing skills with the help of repetitive patterns in text.

2.3.9 Implications of the studies on cohesion for the proposed research

The theoretical framework of cohesion provided by Halliday and Hasan (1976) provides a firm foundation on which the proposed research will be undertaken. Their comprehensive description of cohesion and issues surrounding the concept presents an overall view of how cohesive relations are realized through the stratal organization of the language. Their taxonomy of cohesive relations, namely grammatical and lexical cohesion, enables a deepened understanding of how text is organized and, therefore, makes it possible for us to focus on one particular aspects of the textual property, for instance, lexical cohesion in the case of proposed study.

Halliday and Hasan's (1976) position on separating the situational factors of a text from the linguistic features — cohesive devices in this case — helps to justify the

methodological choices made in the proposed research. Since the present study aims to investigate the patterns of the use of lexical cohesive devices in the NCEE and the HSK corpora, it mainly focuses on the linguistic features of text organization. This may raise objections that the study does not consider the situational factors of a text, for example, the social and ideological context in which the writer is situated. Halliday and Hasan (1976) maintain that, although the reader or listener does not separate the internal and the external aspects of “texture”, “when the linguists seek to make explicit the basis on which these judgements are formed, they are bound to make observations of two rather different kinds” (p. 20). In other words, it may be necessary to separate the linguistic features from the situational factors in the analysis of text. The separation of the internal from the external aspects, therefore, does not mean that the study fails to consider those situations related to texts, and results in a lack of merit. Instead, it actually demonstrates the researcher’s recognition of the importance of the situational properties of texts. However, since it lies outside of the research scope, it is not discussed in the present study.

Research on lexical cohesion, especially Hoey’s (2000) system of lexical analysis, is particularly relevant to the proposed study. The analytic approach adopted by Hoey also provides guidance on how to analyse genre-specific Chinese texts. However, Hoey’s study is mainly based on English material, and his conclusions are more English-centric. In his notes at the end of his 2000 book, Hoey himself acknowledges that there is no evidence that the patterns of lexical cohesion in English that he describes could be applied to all languages. He reports that languages such as German, French, and Portuguese share the similar characteristics of lexical cohesion, but his approach does not seem to apply to Finnish. Since Chinese, as an East Asian language, differs from English in many different aspects, it is meaningful to explore whether the two languages share the same patterns of lexical cohesion in text.

2.4 Corpus linguistics

In this section, the following aspects of the corpus linguistic are reviewed: the definition of corpus linguistics, the various types of corpora and the advantages of adopting a corpus-based research methodology. Since the present study investigates the HSK Corpus, which is a learner corpus, the methodologies used for learner corpus-based studies, and the development of corpora in China and their implications for teaching in Chinese are also evaluated. The section concludes with a summary of the review and its implications for the present study.

2.4.1 The definition of Corpus Linguistics

According to Sinclair (1991), a corpus is “a collection of naturally occurring language text, chosen to characterize a state or variety of a language” (p. 171). Since the establishment of the first computer readable corpus, the Brown Corpus, in the early 1960s, *Corpus linguistics* (henceforth CL) has become a major research area in Applied Linguistics. With the popularization of corpus study and the rapid progress in computer technology, corpus linguistics has had a great impact on many areas of linguistic research, such as register and genre studies, translation studies, second language acquisition, and language teaching. In the current literature of CL research, despite differing views on the nature and processes of CL, it has been widely recognized as a method for obtaining and analyzing linguistic data both quantitatively and qualitatively rather than a theory of language or even a separate branch of linguistic, such as sociolinguistics or Applied Linguistics. As Granger (2002) states, “corpus linguistics can best be defined as a linguistic methodology which is founded on the use of electronic collections of naturally occurring texts” (p. 4). As a relatively newly-emerged research methodology, the main focus of corpus linguistics is the discovery of patterns of authentic language use through analysis of actual usage. In CL there have emerged basic principles and techniques for the investigation of language facts using corpora, and as a methodology, it has contributed to the discovery of new facts about language use, and enabled the verification of some theories and the facilitation of language teaching.

2.4.2 Various types of corpora

Depending on the purpose of specific research or project, the compilations of corpora vary from one context to another. The following section lists some of the most common types of corpora, including:

- general corpora, such as the *British National Corpus* or the *Bank of English*, contain a large variety of both written and spoken language, as well as different text types, by speakers of different ages, from different regions and from different social classes;
- synchronic corpora, such as *Frown and Lancaster-Oslo/Bergen (F-LOB) Corpus*, record language data collected for one specific point in time, e.g. written British and American English of the early 1990s;
- historical corpora, such as *ARCHER* and *the Helsinki Corpus*, consist of corpus texts from earlier periods of time. They usually span several decades or centuries, thus providing diachronic coverage of earlier stages of language development;
- learner corpora, such as the *International Corpus of Learner English* and the *Cambridge Learner Corpus*, are collections of data texts produced by foreign language learners, such as essays or written exams; and,
- corpora for the study of varieties, such as the *International Corpus of English*, represent different regional varieties of a language.

Since there are many corpus compilation tools, techniques, and digitalized texts available, researchers can now also choose to create their own special purpose corpora purely for their own use. The basic principles, the annotation methods, the procedure of building a corpus are also available both from publications and relevant Internet web sites. For instance, John Sinclair provides a detailed description of how to develop linguistic corpora on the web site: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/index.htm>. These tools, techniques, texts and information provide resources for researchers who want to build corpora that suit their own needs.

2.4.3 The advantages of the Corpus Linguistics methodology

Since its development as a methodology, CL has been used in numerous studies to describe language features and to test hypotheses formulated in various linguistic frameworks. More and more researchers and scholars have realized that corpus-based study has advantages in relation to examining large-scale, authentic texts, which was not possible in the pre-computational era. The following points summarize some of the principal advantages of corpus linguistic methodology (Biber, Conrad, & Reppen, 1994, 1996; Volk, 2002; Wang, 2005):

- a corpus-based approach is empirical. Researchers utilize large and principled collections of natural texts to look for actual patterns of language use in natural situations;
- researchers do not have to rely on their own or other native speakers' intuition or on made-up examples. Rather, they draw on a large amount of authentic, naturally-occurring, language data produced by a variety of speakers or writers;
- a corpus-based approach makes extensive use of computers for analysis, using both automatic and interactive techniques; computers make it possible to identify and analyze complex patterns of language use, allowing the storage and analysis of a larger database of natural language than could be dealt with by hand;
- the analyses are consistent, reliable, and also can use computers interactively, allowing researchers to make difficult linguistic judgments while the computer takes care of record-keeping; and,
- corpus linguistic research can involve both quantitative and qualitative analytical techniques. Both qualitative measurements of certain language patterns and statistical data can help researchers to propose functional interpretations for the existence of certain patterns;

Despite these advantages, there are methodological issues with corpus compilation and conducting corpus-based research, as pointed out by some corpus linguists (e.g. Granger 2002). Below are some of the points that researchers need to be aware of in corpus-based studies.

- corpora should be compiled according to strict design criteria. A random collection of heterogeneous data or a poorly designed corpus does not qualify as a corpus and can lead to invalid research results.
- corpus-based analysis should be regarded as an approach that is complementary with more traditional approaches, rather than as the single correct approach. In fact, research questions for corpus-based studies often grow out of other kinds of investigation.
- corpus-based research result should not be over-generalized. In particular, if the study is based on a very small corpus of certain type of text, say written American English, claim should not be made that the result is valid for American English as a whole or even for English in general. Research results need to be qualified by saying that the result is only valid as far as written academic American English is concerned, and that further research into other types of texts needs to be conducted for more general conclusions about the features.

2.4.4 Methodologies for corpus-based studies

CL involves a number of research methods. According to Wallis and Nelson (2001), CL has what can be called the *3A perspectives: annotation, abstraction and analysis*.

Annotation refers to the application of a scheme to texts, which may include structural markup, tagging, and a number of other representations. According to Leech (2003), annotation is “the practice of adding interpretative linguistic information to a corpus” (p. 25). Depending on the needs of researchers, the types of annotation vary from one corpus to another. The most common type of annotation is *Part of Speech* tagging (POS tagging). The following example of POS tagging is given by Leech (2003):

present_NN1 (singular common noun)

present_VVB (base form of a lexical verb)

present_JJ (general adjective)

Apart from POS tagging, there are other types of annotation, such as phonetic annotation, semantic annotation, pragmatic annotation, discourse annotation and lexical annotation.

Abstraction refers to the translation of terms in the scheme to terms in a theoretically motivated model or dataset. Abstraction mainly involves linguist-directed searching. With the help of concordancing programmes like LEXA, Concapp, TACT, WordSmith, WORDCRUNCH, and WINMAX, a researcher can make various queries according to predefined research questions or hypotheses to obtain, sort and store frequency, probability, occurrence, and collocation information of language use or patterns.

Analysis consists of statistical investigation, manipulation and generalization from the dataset. Analysis may include statistical evaluations, optimization of rule-bases or knowledge discovery methods. The information produced by concordancing programmes are normally lists of values or numbers, which need to be analyzed and investigated both quantitatively and qualitatively following certain statistical procedures. Below are two examples taken from Zhang (2010, p. 453) to demonstrate what quantitative and qualitative analyses are:

Table 2.1 shows that the significant collocates of the word *commit* in the Brown Corpus. The first column of the table lists the collocates of *commit*; in the second, F(c) refers to the frequency of these collocates in Brown and in the third column F (n, c) refers to the frequency of the co-occurrence of *commit* and its collocates. When its MI score reaches three or above and the Z-score is two or above, the word is considered to be a significant collocate.

Table 2.1 *Significant collocates of COMMIT in BROWN Corpus*

Collocates	FC	F(n,c)	MI	Z-score
suicide	17	6	12.777	91.711
adultery	3	1	12.694	36.384
murder	76	5	10.353	36.034
crimes	14	2	11.472	33.651
murderer	21	2	10.887	27.452
sin	54	3	10.109	25.631
theft	10	1	10.957	19.893

perfection	11	1	10.82	18.963
conscience	40	1	8.9577	9.872
assumption	41	1	8.922	9.7482
errors	44	1	8.82	9.403
assault	95	1	7.709	6.3162
fraud	112	1	7.472	5.792
robbery	123	1	7.34	5.511
effort	145	1	7.1	5.047
death	286	1	6.12	3.461

To summarize the features of the collocation of *commit*, the qualitative analysis is firstly conducted. Table 2.1 shows that 11 out of 16 lines have the colligation of *commit* + N (with negative meaning), accounting for 69%, and among these, the MI scores and Z-scores of *suicide*, *adultery*, *murder*, *crimes*, *sin*, *murderer*, *theft* are both significantly higher than 10. Following this, the qualitative analysis is then carried out: the above figures indicate that *commit* + N (with negative meaning) is a typical collocation, and it may be concluded that the word *commit* tends to attract negative words and so exhibits an obvious negative semantic prosody.

Quantitative and qualitative analysis are very useful as it allows researcher to describe the linguistic features of the language use or patterns investigated, or to make generalizations about data, or to verify hypotheses which have been proposed.

2.4.5 Methodologies for learner corpus based studies

Besides the methodologies discussed above, learner corpus-based study also has its unique research methodologies because of the nature of learner corpora. According to Granger et al. (2002), “linguistic exploitation of learner corpora usually involves one of the following two methodological approaches: *Contrastive Interlanguage Analysis* (CIA) and *Computer-aided Error Analysis* (CEA)” (pp. 11-12).

The Contrastive Interlanguage Analysis (CIA) involves two kinds of comparison:

Native Speaker (NS) / Non-native Speaker (NNS) comparison and NNS/NNS comparison. NS/NNS comparison is intended to “shed light on non-native features of learner writings and speech through detailed comparisons of linguistic features in native and non-native corpora” (p. 12). Through comparing the two corpora, a range of features of non-nativeness in learners’ writing and speech can be highlighted, such as errors, underrepresentation and overrepresentation of words, phrases and structures. The CIA approach is central to learner corpus-based study, because it can help researchers to assess the extent of the deviation of learners’ language use, and facilitate foreign language teaching. NNS/NNS comparison aims to help researchers to improve their knowledge of interlanguage. By comparing different learner populations, researchers can distinguish “between features which are shared by several learner populations and are therefore more likely to be developmental and those which are peculiar to one national group and therefore possibly L1 dependent” (p. 13).

Computer-aided Error Analysis (CEA) involves two types of methods. One is to retrieve all instances of the misuse of a particular item by scanning a corpus with the help of standard text retrieval software tools. The other involves devising a standard system of error tags and tagging all the errors in a learner corpus. After these have been done, the analyst can retrieve the instances of misuse of certain items in the learner corpus.

2.4.6 Corpus Linguistics and Language teaching

Implications for language teaching

In recent years, interest has grown dramatically in the use of language corpora and computer tools for language teaching. It is seen that corpus-based study can assist the language teacher and can lead to more principled classroom materials and activities. According to Barlow (2002), CL can be applied to at least three areas of language education, namely: syllabus design, materials development, and classroom activities.

A syllabus decides what should be taught in classroom according to students' needs. Frequency and register information can, therefore, be incorporated into course planning and design. By conducting an analysis of a corpus which is relevant to the purpose a particular class, the teacher can retrieve reliable information about frequency and register to determine the content of the course.

When developing teaching materials, teachers often rely on their intuitive sense of what students should be taught. With the help of a corpus, teachers could design exercises based on real examples which reflect the actual use of language features. In this case, a corpus serves as a reference guide. In the area of foreign language teaching, a learner corpus could provide teachers with learner errors information which is helpful for further improvement of the learner-centered teaching activity.

By using concordancing programs and some chosen corpora, students can conduct their own language analyses and interact with these corpora at various levels. For instance, students can investigate a language pattern individually or in a group and report what the findings are. The teacher can assist this process and illustrate possible activities. This encourages learners to problematize language, to explore texts and to draw their own conclusions about language use. In this scenario, the teacher's role becomes a research facilitator rather than the more traditional imparter of knowledge. This student-centered discovery learning gives students the access to real examples of language use, which come from real contexts rather than being constructed for pedagogical purposes (Gavioli & Aston, 2001).

Controversies in the relationship between CL and language teaching

Although CL has been generally welcomed as a powerful tool for facilitating language teaching, the relationship between what linguists describe and what language teachers need to prescribe as appropriate for learning remains controversial. For example, Widdowson (1991) raises questions about direct application of corpus data to language teaching. He argues that language teaching cannot simply be based on descriptive facts, and the result of corpus analysis can be only one point of reference because these corpus descriptions do not necessarily meet the conditions of pedagogic effectiveness. In response to his concern, Stubbs (2001) regards it as "much ado about nothing" and "a non-issue"

(p. 170). He argues that corpus linguists do not put “the attested” in conflict with “the possible”, but give priority to “describing the commonest uses of the commonest words” (p. 151). Sinclair (1991) states that “corpus linguistics has no direct bearing on the way languages may be presented in pedagogical context” (p. 489). This statement, however, is inconsistent with other statements that Sinclair has made elsewhere. For instance, when offering guiding principles in books for teachers, Sinclair (1997) says explicitly that one of the precepts for language teachers is “present real examples only” (p. 30).

There are also further exchanges and wide-ranging discussion of the roles of corpora in language teaching. For instance, Prodromou (1996) and McCarthy and Carter (1996) discuss the transferability of research in general to the pedagogical domain before a process of mediation takes place in order to ascertain and meet learners’ and teachers’ needs and wants. These discussions clearly indicate that there is considerable scope for negotiation and dialogue between corpus linguistics and language teaching. These exchanges will also deepen researchers’ and teachers’ understanding of the relationship between the two sides of this fundamental issue, language and teaching, and makes linguists and language teachers fully aware of the complexity of the issue.

2.4.7 The development of Corpus Linguistics in China

Since CL was introduced to China in the 1980s, a large number of scholars have been involved in the compilation of various types of corpus and corpus-based studies, and have developed a considerable body of research knowledge in these areas.

The compilation of corpora in China

Over the last two decades, a number of English corpora and Chinese Learner English corpora have been built up in China. The first corpus compiled in China is the *Jiao Da English for Science and Technology Corpus* (JDEST), which was developed by Shanghai Jiaotong University in 1985. This corpus consists of over one million words and covers ten subject areas in science and technology. In 2003, *Chinese Learner English Corpus* (CLEC) was completed and published

internationally. Apart from these written text corpora, some spoken English corpora were also compiled in a number of Chinese universities, such as *Spoken English Corpus of Chinese Learners* (SECCL) developed by Nanjing University.

In recent years there has been a rapid development of Chinese corpora in China. For example, Beijing University has developed three corpora: *The Modern Chinese Corpus*, *The Ancient Chinese Corpus* and *The Chinese-English Parallel Corpus*. The Beijing Language and Culture University also compiled a *Modern Chinese Patterns Corpus* in 1995. In 2005, the National Chinese Language Committee also completed a fully annotated general Chinese language corpus entitled *The Modern Chinese Language Corpus* (MCLC) with over 70 million characters.

Compared with the compilation of other Chinese corpora, the development of corpora of the second language learners' output has been relatively neglected. Up to now, there are only two Chinese learner corpora that have been published and fully annotated for the purpose of study on TCFL. One is *The Chinese Interlanguage Corpus* which was completed in 1995; the other is the *HSK Dynamic Composition Corpus*, which was completed in 2006. Both corpora were developed by Beijing Language and Culture University.

Corpus based studies in the area of TCFL

As TCFL in China is still at its infancy, the corpus-based approach has not been used widely in research in this field. Since the establishment of the Chinese Interlanguage Corpus in 1995, a number of studies have been devoted to the investigation of the learner errors. For instance, Chen (1996) investigates the misuses of adverb *ye* in CFL learners' compositions. Cui (2003) explores the misuse of propositional structure in the writings of Japanese, Korean CFL learners. These studies introduce a corpus-based approach to the exploration of the errors made by CFL learners', and promote its use for the further investigation of CFL interlanguage features. However, in the current research literature, most studies concentrate on error analysis based on frequencies and simple qualitative discussions. More advanced employment of corpus statistical tools and measurements are still rare in the field of CFL research. Furthermore, many

studies are limited to the exploration of certain single words or particular patterns. Systematic and detailed descriptions of the grammatical, lexical features of the CFL interlanguage have not been provided. For example, to date, few research studies have been conducted to investigate the actual use of lexical repetition in CFL learners' composition and made comparisons with good writing by Chinese L1 writers.

2.4.8 Summary of Corpus Linguistics and its implications for the proposed study

Corpus linguistics has emerged mainly as a research methodology rather than a separate branch of Applied Linguistics. Depending on the purpose of the compilation, the types of corpora that are developed may vary considerably, such as from large general corpora to smaller, discipline- and genre-specific corpora. The process of implementing a corpus-based study involves three aspects: annotation, abstraction and analysis. The use of computer technology and software in CL affords a number of advantages: the analysis is fast and reliable, and the analysis can involve both qualitative and quantitative research methodologies. For research that uses data from learner-language corpora, two analytical approaches are widely used: contrastive interlanguage analysis and computer-aided error analysis.

CL can be applied to language teaching in the area of syllabus design, materials development, and classroom activities. However, there are controversies in the terms of relationship between the linguistic description the CL affords and its suitability for pedagogical application. In China, two corpora have been compiled to facilitate TCFL, and a number of the studies carried out have focused on the investigation of errors made by CFL learners. In future studies of this area, there is potential for the scope of the research to be expanded beyond the area of error analysis. Research methodologies need to be more diversified and research results need to be more convincing. For example, in the current TCFL literature, there are a large number of studies on error analysis, such as those of Chen (1996), Sun

(2008), and Zhao (2005), however, few studies have used CIA approach to investigate the differences of the use of language between native Chinese speakers and CFL learners. In terms of the research findings, the claim made by some studies such as Sun (2007) may need more sufficient data analysis result to support.

Implications for the proposed study

As a newly emergent research methodology, the corpus-based approach integrates information technology into the linguistic research of both naturally-occurring language and interlanguage systems. In the field of second language teaching, research can use CIA and CEA to compare and explore the differences in use of language between native speakers and language learners. This kind of investigation can be used to assist teachers and researchers to develop a clearer understanding of the characteristics of learners' interlanguage systems, for the purpose of further developing and improving second language teaching and learning. The present study will make use of the CIA approach to explore the differences in the use of lexical repetition in both Chinese native speakers' and CFL learners' compositions. Drawing on the design criteria, principles, and frameworks proposed by corpus linguists, such as Sinclair (1991) and Granger (1998), this study will build two corpora to undertake these research purposes. The aim is that the findings of this study will reveal new aspects of the lexical features of the Chinese interlanguage system, and contribute to the literature of the TCFL studies in China. .

In China, many scholars have attempted to apply corpus-based approaches to their research. These compilations of corpora and their exploration have laid a solid foundation for further CIA and CEA studies. For example, the HSK Dynamic Composition Corpus provides sample texts for the present study. Also, the findings from the existing TCFL literature inform the conceptualization and design of the present study. The problems and challenges in existing, corpus-based CFL research, such as the small amount of corpus-based research exploring the grammatical and lexical features in Chinese interlanguage, and the fact that most corpus-based studies of CFL focus on error analysis are also taken into account.

Therefore, this review of the current literature of corpus-based studies in the field of TESOL and TCFL undertaken in this chapter has direct significance for the present study.

2.5 Summary

This chapter has provided a review of previous studies in the field of CR, cohesion study, applications of corpus linguistics, and outlined how these are relevant to the present study. In this review of the literature, the constraints of research scope and methodological limitations of previous studies, together with the gaps that the present study attempts to fill have also been discussed. Informed by the theoretical frameworks and empirical studies reviewed in this chapter, the next chapter will describe the research methodology employed, including the principles of the research design, data collection, data annotation, and the data analysis procedures used in the present study.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

Chapter 3 describes the research design and methodology employed in the analysis of texts written by both Chinese native speakers and CFL learners. Section 3.1 introduces the aim of the study, the framework that the study adopts, and the main research questions. Section 3.2 briefly revisits the theoretical framework for the study in order to frame the presentation of research questions. Section 3.3 then relates this theoretical framework to the nature of the data and the data collection procedures that are employed. This section also provides a justification for the use of the selected data. In the subsequent Sections 3.4 and 3.5, the design of both the NCEE and HSK Corpora are described, and the methods of data analysis of the text are also discussed. This includes the demonstration of examples of text analysis, and the step-by-step procedures employed in the annotation of data. An overall summary of the methodological procedures follows these sections.

3.2 Research questions

This study will use cross-linguistic comparison of two corpora, one of texts written by learners of Chinese as a foreign language and one of texts written by native speakers, in order to explore the role of lexical cohesion in written Chinese text. Specifically, the aim of this study is to describe the extent to which the two writer groups differ in the use of lexical cohesive devices in their writing. The description will then be used to identify the difficulties that non-native Chinese speakers may have in assigned writing tasks. It is hoped that the research findings may inform the development of teaching materials and guide classroom practice that addresses the writing challenges facing non-native Chinese speakers.

The overall theoretical approach of this research is a corpus-based *Contrastive Interlanguage Analysis* (CIA) approach. As Granger (2002) stresses, detailed

comparison of linguistic features in native and non-native corpora can reveal non-native features of learner writing, such as deviance, avoidance, overuse, underuse of certain lexical and grammatical items. Thus, native and non-native language corpora may provide relevant information about what is typical in a particular language and what might be challenging for language learners. Following the framework of CIA (Granger *et al.*, 2001), the study will establish two corpora, namely, a reference corpus of native-speaker Chinese writing and a Chinese interlanguage writing corpus for the purpose of carrying out the proposed exploration of lexical cohesion in the texts of both language groups.

Thus, the analytical framework of this study draws together contrastive rhetoric (Chapter 2, Section 2.2), textual analysis (Chapter 2, Section 2.3) and corpus linguistics (Chapter 2, Section 2.4). It involves a detailed text analysis of two sets of texts produced by two language groups in similar test situations. While the predominant focus in this study is on quantitative data, it also has qualitative aspects, such as consideration of the data in relation to possible contextual and pedagogic issues that may relate to the types of challenge that CFL learners encounter when writing Chinese text.

The overall research question of this study is:

How do Chinese native speakers and CFL learners differ in their use of lexical cohesive devices in non-narrative text?

To address this overall question, a series of subsidiary questions will be asked at different stages of the analysis so as to explore the patterns of lexical cohesion in both language groups' compositions.

At the reviewing stage

- Does lexical cohesion account for the largest proportion of cohesive ties?
- How many types of repetition are employed in the native speaker texts?
What are they?
- What are the frequencies of each type of repetition in the native speaker texts?

- Are the categories of repetition in Chinese texts same as those in English?

At the comparison stage

- How many types of cohesive devices are employed in CFL learners' texts?
- What are the frequencies of each type of repetition in CFL learners' texts?
- What are the frequencies of each type of repetition in the texts of both groups?
- What are the characteristics of the use of each repetition device in the texts of both groups?
- What are the densities of repetition in both groups' texts?
- Are there any differences in the connectivity of each sentence in the texts of both groups?

At the problem discussion stage

- Are there any features of the use of repetition devices in CFL learners' texts which are significantly different from that in the native speaker texts?
- If so, do these significantly different uses of repetition in CFL learners' texts prevent them from producing native style fluency in composition?
- What are the possible causes for their failure to produce natural Chinese?

The questions presented here relate to the whole research project. This chapter, however, will only focus on how the study addresses the first group of questions.

3.3 The theoretical frameworks and research design

This study falls within three research paradigms: contrastive rhetoric, the study of cohesion, and corpus linguistics (especially in the area of learner corpus based contrastive interlanguage analysis). Despite their different theoretical orientations, the three paradigms share a commonality of exploring the organizational patterns in genre-specific texts across and within disciplines, languages, and cultures.

Originally contrastive rhetoric sought to address rhetorical problems in L2 writing, and attempted to explain them on the basis of the hypothesis that writing patterns and culture are synonymous. However, as the theory has developed and expanded in terms of the scope of CR investigations, present approaches to contrastive rhetoric are working towards a discipline- and genre-specific study of the relationship between rhetorical patterns of texts and their cultural contexts. Contrastive rhetoric, then, provides an overall theoretical framework for the study. In the study, some basic principles of contrastive rhetoric are employed to explore the differences and similarities between the texts written by native speakers and L2 learners of Chinese in terms of lexical repetitions. For example, Swales (1990) emphasizes that to achieve comparability in contrastive analyses, it is necessary to compare and contrast texts written for similar purposes in similar situations. Following this principle, the present study considered carefully about various aspects of the compilation of the NCEE and HSK Corpora to ensure comparability of the data.

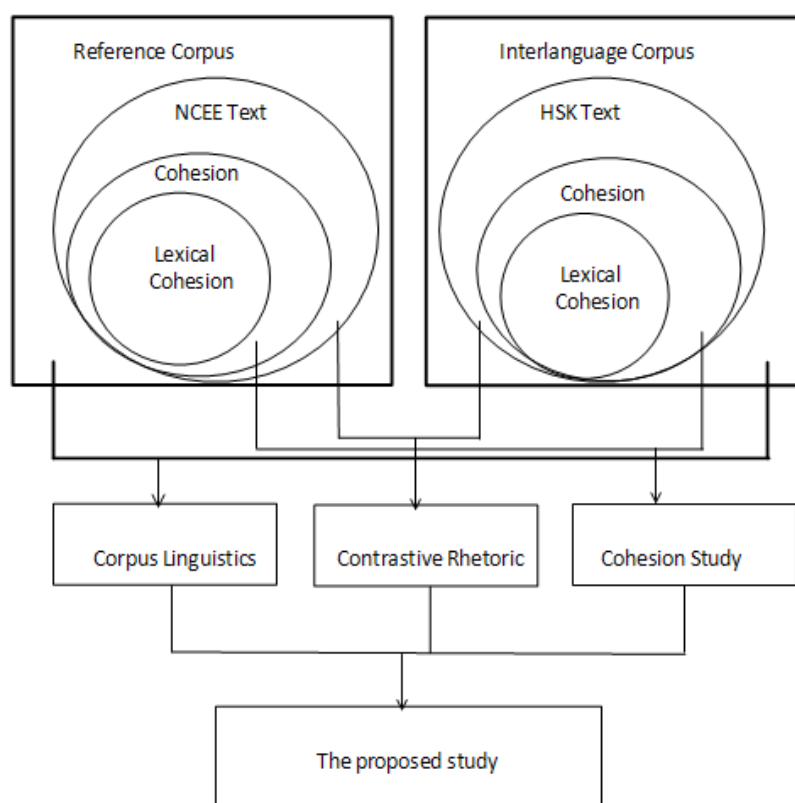
The present study also draws on the theories and research findings relating to cohesion in text, especially on Halliday and Hasan's (1976) theoretical framework and Hoey's (2000) analytical approach in the study of lexical cohesion in non-narrative texts. For the study, Halliday and Hasan's (1976) taxonomy of types of cohesive relationships within a text, which include grammatical cohesion (substitution, ellipsis, conjunction, reference) and lexical cohesion, provides the major starting point for the research. Hoey's (2000) framework for analyzing lexical repetitions is especially relevant to the exploring the use of lexical devices in the texts of the NCEE and HSK Corpora. Three analytical approaches and techniques adopted in Hoey's (2000) analysis, namely the design of heuristic chart to distinguish different types of repetition, the establishment of the repetition matrix to reveal the density of repetition among sentences, and the interpretation of the significance of how the lexical repetition works in the organization of a text, are employed to address the research questions of the study.

As discussed in Section 2.4.1, corpus linguistics is more regarded as a research methodology rather than a particular branch of Applied Linguistics. The strength of corpus linguistics lies in its capacity for processing large scale data and

revealing descriptive linguistic features of the collected data. In particular, the learner corpus based Contrastive Interlanguage Analysis (CIA) approach, proposed by scholars in the field (e.g. Granger, 2002), has been used to assess the extent of the deviation of learners' language use, in order to improve researchers' knowledge of interlanguage, and to facilitate foreign language teaching. For the purpose of the present study, the CIA approach is, therefore, employed to provide practical guidance in relation to corpus design, collection of data, annotation of the data, abstraction of the data and the interpretation of the statistical data.

On the basis of these three theoretical paradigms, the framework used for the study can be illustrated as in Figure 3.1.

Figure 3.1 *Theoretical frameworks for the study*



3.4 The data collection and sampling procedure for the two corpora

In this section, the nature of the data used for compilation of the two corpora and various aspects of the sampling procedure are discussed.

3.4.1 The nature of the data

As Hoey (2000) states, his analysis of lexical patterns only applies to non-narrative texts, and is not applicable to narrative texts. Using his analytical framework, the study will focus on non-narrative text in Chinese. In particular, the non-narrative texts in this study are compositions written by Chinese students during the *National College Entrance Examination* (NCEE) and CFL learners during the HSK (Chinese Language Proficiency Test). The former is used as text samples to establish a reference corpus of native Chinese writing and the latter an interlanguage corpus of CFL writing.

The justification for the use of the NCEE and HSK texts

In the field of contrastive rhetoric, it has been emphasized that corpus-based text studies should be conducted with strict adherence to the principle of establishing a *tertium comparationis* (common platform of comparison or shared similarity) between texts and approaches to their analysis (Connor, 2004; Connor & Moreno, 2005). For Connor and Moreno (2005), the concept of *tertium comparationis*, or common platform of comparison, provides a standard that is important at all levels of contrastive analysis, namely: choosing the primary data, establishing comparable textual concepts, and comparable analyses of the linguistic and textual data.

Taking this principle into account, the present study will choose compositions written by native-speaker Chinese students for the NCEE and CFL learners for the HSK as its primary sample texts. The justification for the use of these two sets of texts and their comparability is explained in detail in the following two sections.

The native-speaker Chinese writing in the NCEE

As a prerequisite for admission to tertiary education at the undergraduate level, the National College Entrance Examination (NCEE) is an academic examination held annually in China, and is the only criterion for higher education entrance. With nearly 9.5 million high school students taking the examination in 2008, it is

the largest examination of its kind in the world. It is a high-stakes assessment in the lives of young Chinese people, the outcome of which can greatly affect their future in an intensely competitive society. In the examination, three subjects are compulsory: Chinese, English and Mathematics. Other subjects, such as history, geography and politics, or physics, chemistry and biology can be chosen by students as a combination, depending on what kind of tertiary education major they want to study. The overall mark system of the examination is generally a weighted sum of their subject marks.

According to the *Outline of the 2008 National College Entrance Examination—Chinese*, the Chinese examination comprises five testing areas, namely Chinese language knowledge and expression, common knowledge of Chinese literature and famous works, reading comprehension of ancient Chinese works, reading comprehension of modern Chinese works, and composition writing. Although it is only one of the five areas tested, the composition test is the key part of the Chinese exam, usually accounting for 60 to 75 of the total 150 marks. The outline clearly states that the overall purpose of the composition test is to examine students' ability to write some of most commonly used types of essay, such as narration, argumentation, and description. The emphasis in the test is on such aspects of writing as: the use of vocabulary, the selection of the supportive material, the rationality of the argument and the organization of the structure.

The current composition test is task-based writing. Students taking the test are required to write an essay of not less than 800 Chinese characters on a given topic or an eliciting prompt. The time allowance for the whole Chinese exam is two and half hours. Although there is no suggested timing for each testing areas given by the NCEE Examination Panel, it is generally agreed that students should spend at least 50 - 60 minutes on the composition test. Students can choose whatever style they like except poetry (a sample of a composition test is attached in Appendix I). To maintain fairness and objectivity in grading, each composition written in the test will be marked by at least two senior examiners. If a discrepancy between the marks given by the two examiners is more than six marks, then the writing will be examined by a third examiner. The final mark will be the mean score between the two marks which are closest to each other.

Despite deficiencies noted by some experts in China (e.g. Ke, 2010), the composition test is still regarded as the most authoritative and standard assessment of students' writing ability. The grade that a composition receives is, therefore, considered to be a comprehensive indicator of a student's writing skills. In particular, those compositions that receive full marks are widely known as outstanding pieces of writing, demonstrating excellent literacy skills and a profound background knowledge of literature. These compositions, however, are extremely rare. For instance, there is only one composition which received full marks among nearly 100,000 compositions in Shanghai city in 2008 NCEE Chinese test, and 13 among 614,000 in Guangdong Province in the same year.

The present study will choose compositions that have received full marks in the NCEE examination as its sample texts. The time span of when the compositions were written is from 1999 to 2010, and the writers are from all over the country. The primary text type is argumentative. According to a statistic reported by the leader of the Chinese examination paper marking team in Guangdong Province, although the exam does not restrict candidates to this type of writing, 96% of the compositions are argumentative (some argumentative texts may include segments of text that involve narration or description in support of an argument). For this reason, the study will focus on argumentative texts which have received full marks in the NCEE, as it is the most common and representative type of text written in the test.

The learner of Chinese writing from the HSK

HSK, the romanized abbreviation of 汉语水平考试, is the national, standardized proficiency test designed and developed by the *HSK Center of Beijing Language and Culture University* to evaluate the Chinese language proficiency of non-native Chinese speakers (including foreigners, overseas Chinese and people from Chinese national minorities). There are three categories in the HSK: beginning level (HSK Basic), elementary to intermediate level (HSK Elementary-Intermediate), and advanced level (HSK Advanced). The HSK is held regularly in both China and in other countries each year, and HSK certificates are issued to those who meet the required scores. The HSK test and the issuing of HSK

certificates are supervised by the *State Commission of the Chinese Proficiency Test* under the authority of the Ministry of Education. The HSK is jointly administered by the Office of the State Commission of the Chinese Proficiency Test, under the supervision of the State Commission of the Chinese Proficiency Test, and by the HSK Centre of Beijing Language and Culture University.

The HSK has become increasingly popular both in China and overseas with the rapid growth of the Chinese economy. HSK Certificates can now be used to certify that the holder has acquired the required level of Chinese proficiency to enter a college or university as an undergraduate or graduate student. The certificates may also allow their holders to be exempt from Chinese courses at the corresponding level in some educational institutions. With respect to employment in China, the certificates may serve as a basis for employers to evaluate the Chinese language proficiency of job applicants. The Certificate of HSK Advanced can certify that the holder is qualified to undertake ordinary work by using Chinese as a communicative tool.

Both the HSK Basic and HSK Elementary-Intermediate are standardized tests. The former test is divided into three sections: listening, sentence structure and reading comprehension, while the latter comprises four parts: listening, sentence structure, reading comprehension and a cloze test.

The HSK Advanced consists of three separately-timed subtests: a 120-item standardized written test, a writing composition section and an oral test. The 120-item written test is further divided into three subsections: listening, reading and cloze. A variety of objective-type questions are used including multiple-choice, matching, filling in the blanks and writing Chinese characters. The oral test takes the form of an interview.

Similar to the NCEE composition test, the HSK composition test involves task-based writing. Candidates have to draft a composition of 400 to 600 characters in 30 minutes, according to the requirements of a task. The type of writing varies from test to test: candidates may be required to write their opinions on a given topic, such as *my opinion on euthanasia*, or to write a job application letter (a sample is attached in Appendix II). To maintain the quality and the objectivity and

fairness of the grading system, the *Methods and Detailed Rules of HSK Composition Marking* has been formulated by the HSK Examination Centre. This document stipulates that each composition should be marked by three examiners individually. The final score will be the mean value of the three marks if they are very close to one another. If there is an apparent discrepancy between the three marks, the writing will be reviewed by the three examiners together until an agreed score is given by the three examiners unanimously.

To establish a *tertium comparationis* between the two language groups, the study will choose HSK Advanced compositions of CFL learners which are argumentative texts as sample texts, that is only those giving opinions on certain topics will be selected. These compositions will have received the grade of A, the highest level of the HSK.

The above descriptions of both the NCEE and HSK aim to establish comparability between the two sets of tests. Firstly, since both are standardized national tests, they are the most valid tests of writing in Chinese available. Secondly, the grading systems are relatively objective and fair, which establishes the reliability of the test. Thirdly, the NCEE is the only examination that serves as the admission to higher education. The HSK certificate is also a proficiency test that allows the holder (at the advanced level) to enter a college or university as an undergraduate or graduate student. Furthermore, both the NCEE and the HSK compositions employ argumentation as one of the possible text types to test candidates' writing skills.

However, while every attempt should be made to provide comparable data sets, it is not possible to achieve an exact match between the two data sets. For example, one potential weakness of the comparison between the two sets of texts is the length of the respective essay texts. Although the HSK requires the candidates to write 400-600 characters, the length of the L2 learners' texts may be commonly shorter than that of required because of the nature of their developmental language proficiency. Consequently, it is not valid to compare simply the total number of lexical cohesive devices used in the texts of both language groups without considering this factor. To compensate for this possible problem, the study adopts

the following approach: Firstly, the total number of lexical cohesive ties used in each sample text is recorded; secondly, the total number of characters in each text is also counted using the NotePad programme. The third step is to calculate the frequency of the use of cohesive ties in each sample. For the convenience of comparison, the study compares the number of repetition ties used per 100 characters, which can be worked out by using the following formula:

$$\text{the frequency} = \frac{\text{the total number of cohesive ties} \times 100}{\text{the total number of characters}}$$

Using this formula, the comparison of the frequency scores is carried out to explore the difference in the use of the lexical repetition ties between the texts of the both samples.

3.4.2 Selection of the samples

As discussed, the study involves the analysis of two samples of Chinese compositions that present an argument: compositions written by native Chinese students in the NCEE examination and compositions by CFL learners in the HSK Chinese language proficiency test.

The collection of the compositions written by native Chinese students in the NCEE began with an exploration of the NCEE preparation resources available in China. As the NCEE is the predominant higher education entrance examination in China, there are numerous books published every year on how to prepare for the composition test, and how to learn from examples of excellent writing. Similarly, there are countless websites providing the same kind of learning resources. For the convenience of data collection and process, the compositions that receive full marks were obtained from relevant Internet resource web sites:

<http://houdongm.banzhu.net/article/houdongm-43-101919.html>;

<http://www.eol.cn/article/20060714/3199975.shtml>;

<http://www.zuowen.org/html/07/10356.html>;

http://www.0817china.com/gz/syzyx/onevs_yxzw.asp?id=67;

<http://www.jsgaokao.net/web/2008-07-03/7238.html>;

<http://baike.baidu.com/view/18288.htm?fr=topic>

<http://gaokao.eol.cn/html/g/manfen10n.shtml>

http://post.sooedu.cn/topic_1289.html

<http://www.51test.net/show/1274463.html>

As the texts available on these web sites fall within the public domain, and are not protected by copyright, the required data can be openly accessed and obtained. However, following common practice in academic research, all information retrieved from the internet will be acknowledged through appropriate in-text and reference list citation practices.

On these websites, the compositions that receive full marks are often followed by comments given by authoritative Chinese teachers or professors. The comments normally cover aspects of writing, such as the type of text, the structure or the organization of the text, the argument, the use of supporting material and the use of lexis. These comments, therefore, provide strong support when distinguishing the requirements of typical argumentative texts from other types of writing. During the collection of the data, these comments, as well as the compositions themselves, were examined carefully to ensure that the compositions obtained are valid argumentative texts. Altogether 50 argumentative texts written from 1999 to 2010 were found from the above mentioned websites.

In developing the corpus of Chinese second language learner writing, compositions written by CFL learners were obtained from the HSK Dynamic Composition Corpus at the Beijing Language and Culture University (BLCU) in China. The HSK Dynamic Composition Corpus is an academic project developed and administered by the Centre for Studies of Chinese as a Second Language at BLCU. This 4,000,000 Chinese-character-corpus consists of compositions from 1992 to 2005 of the high level HSK test, written by non-native speakers taking the test. The purpose of the corpus building is to provide a review and research platform for TCFL.

To obtain the required data, a letter of request was sent to the administrators of the

HSK Composition Corpus. (Appendices III and IV include copies of the letters requesting permission to access texts written in both English and Chinese). An agreement on terms and conditions regarding the use of the data was requested from the corpus administrator to ensure the procedure meets the requirements of the University of Waikato Human Research Ethics Regulations (2008) (see Appendix V). After the researcher and the corpus administrator agreed upon the terms and conditions of the use of the data, the researcher was provided with 650 sample texts. As it was necessary to establish a *tertium comparationis* between the texts written by native Chinese speaker and by the learners of Chinese, a selection of the latter was made. On the basis of the criteria for establishing comparability discussed previously, and according to the methodology of the corpus design, 50 sample texts were selected from the HSK Corpus to form the learner corpus. In order to establish a balance between the two sets of data, and to ensure that the data were comparable, the first selection criterion was to choose HSK texts that fall within the category of an argumentative text. As the 650 texts were all written in one particular examination, they all have the same topic, which required candidates to give their opinions on an issue. The selection, therefore, was straightforward and met the criterion well. The second task was to select texts that received grade A. Among the 650 sample texts obtained, there are 63 texts that received grade A. The last task is to select longer texts among the 63 A-graded texts. As discussed in Section 3.3.1, the length of L2 learners' writing is usually shorter than that of required because of the nature of their developmental language proficiency. For this reason, every effort was made to make up this deficiency by choosing sample texts as long as possible without compromising on the grade received. With the help of the NotePad programme, each sample text was counted and 50 texts were chosen according to their character counts (see Table 3.1).

Table 3.1 *The number of characters in each of the 50 HSK texts*

Text Title	Total number of characters
自己创业与进公司 1	593
自己创业与进公司 2	534
自己创业与进公司 3	531

自己创业与进公司 4	522
自己创业与进公司 5	517
自己创业与进公司 6	515
自己创业与进公司 7	514
自己创业与进公司 8	507
自己创业与进公司 9	486
自己创业与进公司 10	475
自己创业与进公司 11	456
自己创业与进公司 12	442
自己创业与进公司 13	424
自己创业与进公司 14	580
自己创业与进公司 15	576
自己创业与进公司 16	575
自己创业与进公司 17	571
自己创业与进公司 18	568
自己创业与进公司 19	566
自己创业与进公司 20	562
自己创业与进公司 21	552
自己创业与进公司 22	551
自己创业与进公司 23	550
自己创业与进公司 24	549
自己创业与进公司 25	548
自己创业与进公司 26	508
自己创业与进公司 27	505
自己创业与进公司 28	500
自己创业与进公司 29	499
自己创业与进公司 30	498
自己创业与进公司 31	495
自己创业与进公司 32	492
自己创业与进公司 33	479
自己创业与进公司 34	478
自己创业与进公司 35	467
自己创业与进公司 36	445
自己创业与进公司 37	493
自己创业与进公司 38	491
自己创业与进公司 39	490
自己创业与进公司 40	481
自己创业与进公司 41	474
自己创业与进公司 42	473
自己创业与进公司 43	472
自己创业与进公司 44	470
自己创业与进公司 45	466
自己创业与进公司 46	465
自己创业与进公司 47	464
自己创业与进公司 48	463
自己创业与进公司 49	462
自己创业与进公司 50	460

3.5 The corpus design

The methodology of corpus linguistics, as in any other field of applied linguistics, requires explicit design criteria to control all the possible variables that may affect the reliability and validation of the research results and the possible replication of relevant research (Atkins & Clear, 1992; Engwall, 1994). The following sections discuss the principles and justification of the corpora design for this study.

3.5.1 The design of reference corpus of native Chinese writing (The NCEE Corpus)

As discussed in Section 2.4.5, Granger (2002) identifies two types of Contrastive Interlanguage Analysis (CIA): native speaker (NS)/non-native speaker (NNS) comparison and non-native speaker (NNS)/non-native speaker (NNS) comparison. For the purpose of this study, the first approach NS/NNS is employed. The intention of the NS/NNS comparison in this research is to reveal the possible non-native speaker features of the use of lexical cohesive devices in learner writing through detailed comparison of native and non-native speaker compositions. In relation to the establishment of a native-speaker corpus, a crucial issue, as Granger (2002, p. 12) points out, is the selection of the particular dialectal variant of the texts of that corpus. It is well-known that there are many dialects in China, and it is vital to clarify that in this study the Chinese language means Mandarin, which is the official standard language used by the People's Republic of China. As a common language of communication for speakers of all Chinese speech varieties, Mandarin is used in government agencies, in the media, and as a language of instruction at all levels of schools and tertiary institutions. In terms of the written language, despite the fact that dialectal languages vary in their spoken form throughout the country, all Chinese use the same writing system, namely Chinese characters, so all publications and writings can be understood by anyone throughout the country. In this study, the use of compositions written for the NCEE, therefore, can be justified in relation to the issue of sampling of dialectal variant. Another issue to consider in the design of the native-speaker corpus is the level of proficiency of the native speakers (Granger, 2002; Lorenz, 1999). Lorenz

(1999) argues that it is both unfair and descriptively inadequate if learner texts are compared with texts produced by professional native-speaker writers. Therefore, native-speaker student writing arguably provides a better type of data for the purpose of comparison with CFL learner writing if the aim of the comparison is to describe and evaluate the interlanguage as fairly as possible. Because this study has the objective of describing and highlighting features of the use of lexical cohesive devices in learner texts, it was decided to choose to use NCEE compositions produced by native students as its reference corpus.

3.5.2 The design of the Chinese interlanguage writing corpus (The HSK Corpus)

Because of the heterogeneous nature of the learner data as well as the learning situations, any learner based corpus also needs explicit design criteria. When discussing the compilation of learner-based corpora, Granger (2002, p. 9) lists some of the major variables which need careful consideration. The first is the authenticity of the data. In the area of learner corpora research, the term ‘authentic data’ covers various degrees of authenticity, ranging from data collected from genuine communications where people going about their normal businesses to data gathered from authentic classroom activities. For Granger, “learner corpora of essay writing can be considered to be authentic written data” (p. 8). As the HSK Corpus designed for this study uses compositions written by learners of Chinese as its raw data, it therefore meets this criterion well. The second issue Granger (2002) points out is that a learner corpus has to use textual data. By textual data, he means that “the language sample must consist of continuous stretches of text, not isolated sentences or words” (pp. 8-9). This criterion is directly relevant to this study, as its purpose is to explore the use of lexical devices in non-narrative written text. The texts used in the HSK Corpus are compositions written for the HSK test, specifically compositions that are required to give the writer’s point of view. Consequently, they can be safely considered as textual data and meet Granger’s second criterion for learner corpus data. Other criteria relevant to the design of this corpus are the learners’ levels of proficiency, the task settings such as time limit for a task, whether the learners are allowed to

use a reference book, and whether the task is carried out in an examination situation. In relation to these criteria, the attributes of the selected sample data for the HSK Corpus is to a great extent homogeneous. With regard to the level of proficiency, at the time of collecting sample data, the prerequisite for taking HSK composition test was that their Chinese proficiency was supposed to have reached advanced level. It is, therefore, fair to say that collected texts were written by candidates whose Chinese were at the same level. As described in Section 3.3.1, the HSK is a closed test and no reference book is allowed to be used at the time of the examination. The time allowance for the composition is 30 minutes. As a result, the conditions under which every HSK candidate writes their response to the composition task are exactly the same.

One point worth noting is that the variable of the learner's mother tongue is not considered in the design of the HSK Corpus despite the study employing the NS/NNS comparison approach. The reason is that the corpus comprises anonymous texts from the HSK test corpus and the personal details of each writer, such as their nationality and L1 background, were not available in order to develop a profile of the writers of the sample. Often, a study adopting the NNS/NNS comparison involves consideration of the learner's first language background in order to differentiate between L2 linguistic features that are common across different L1 learner populations, which can be considered to be developmental, from L2 linguistic features common to one national group and, therefore, likely to be L1-influenced. This study, however, employs the NS/NNS approach with the aim that the non-nativeness of the use of lexical cohesive ties in Chinese learners' writing will be examined across the whole sample. Had personal data about each writer been available, the non-native linguistic features of the learner's writing may have been further explored by differentiating them into L1-specific. However, this type of exploration falls outside of the scope of the current research.

3.5.3 Attributes of the corpora of the study

As discussed above, all the relevant criteria were taken into account in the design of the two corpora in relation to their comparability. The attribute of the finalized corpora and the procedure for compiling the corpora are summarised in Table 3. 2.

Table 3.2 *Text attributes of the NCEE and the HSK Corpora*

Name of corpus	The NCEE Corpus	The HSK Corpus
Text attributes		
Text length:	about 800 characters	400-600 characters
Text type:	argumentative text	argumentative text
Grade:	full marks	A
Number of texts:	50	50
Text-written situation	examination	examination
Writers' first language	Chinese	Various of languages
Time allowed for writing	about 50-60 minutes	30 minutes

According to these of pre-defined attributes, the two sets of data were obtained. The two corpora were then compiled and stored separately. On the basis of the revised lexical cohesion classifications, the sample texts were closely examined and tagged in Microsoft Word, and then converted into plain text format for further computational analysis and statistical interpretation.

3.6 Methods of data analysis

After the compilation of the two corpora was completed, three steps of data analysis were carried out to uncover the similarities and differences in the use of lexical repetition between the two writer groups. These were:

Step 1: Manual analysis

- defining and categorizing various types of repetition under the name of non-lexical and lexical repetition in the NCEE texts

- developing heuristic charts to distinguish various types of repetition and facilitate the process of analysis
- establish a tagging system for the annotation of the corpora and completing the annotation

Step 2: Computational analysis

- abstracting the data using the NotePad programme
- establishing of the repetition matrix for both the NCEE and the HSK corpora

Step 3: Statistical analysis and interpretation

- summarizing the statistical data
- reporting the statistical data
- discussing the findings

These three steps are described in detail in the following three sub-sections.

3.6.1 Step 1: The manual analysis

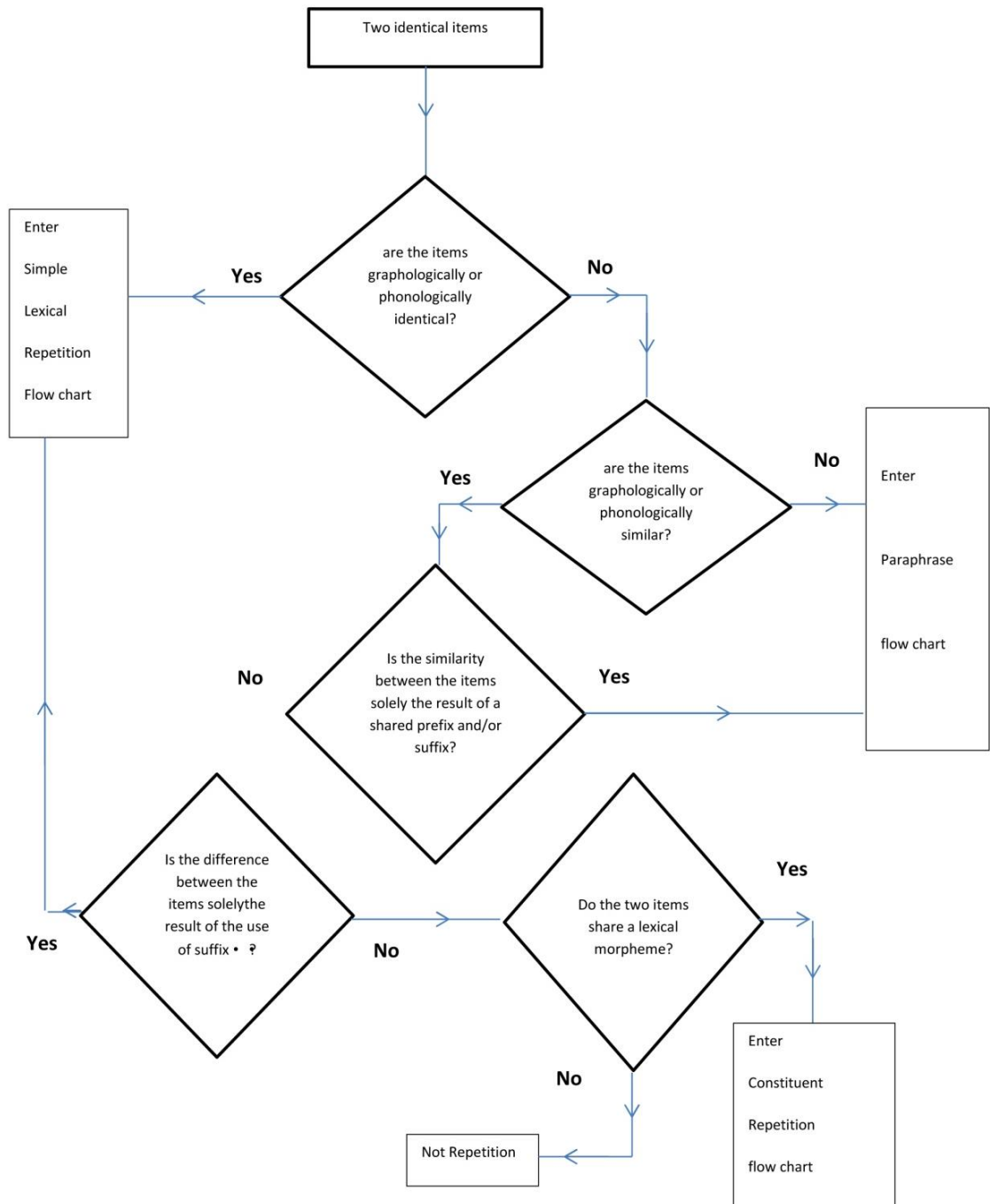
Drawing on Halliday and Hasan's (1976) framework and taxonomy, the study begins by testing whether or not the taxonomy applies to Chinese non-narrative texts. For the purpose of evaluation, only ten NCEE texts were manually analyzed, and the evaluation was conducted by the researcher as the sole rater. In Halliday and Hasan's (1976) taxonomy of cohesion, there are two broad categories of cohesive devices, grouped under grammatical cohesion and lexical cohesion. The NCEE texts were then systemically examined for each of the devices from these two categories by marking up paper copies of each text. The result of the manual analysis suggested that this taxonomy does not have the capacity to categorize some of the types of cohesion found in the ten manually analyzed NCEE Sample Texts 1-10. For this reason, the study developed a modified taxonomy to classify various types of cohesion found in Chinese texts (see the overview of the categories in Section 4.4).

After defining and categorizing the various types of cohesion found in the Chinese texts from the NCEE Corpus, following Halliday and Hasan's (1976) approach, a further analysis of the NCEE Texts 1-10 was conducted by the researcher to see whether or not the lexical cohesive devices were the main means of creating cohesive ties in Chinese non-narrative text as it is in English text, with an aim of ascertaining whether or not Hoey's (2000) claim of the predominant role of lexical cohesion in English non-narrative text and his classifications of lexical cohesion also applied to the analysis of Chinese text. On the basis of the results of the analysis, the repetition categories were reviewed and revised accordingly.

After defining and categorizing the various cohesive ties, the following step was to develop heuristic charts to classify various types of repetition and, thereby, to facilitate the later process of computer-mediated analysis of the whole corpus. For the purpose of the analysis in this study, the heuristic flow charts that were developed are based mainly on those of Hoey (2000) although they were revised to account for the differences between Chinese and English texts in terms of the linguistic features used to create cohesive ties (see Hoey, 2000, pp. 58, 59, 60, 68). For instance, the criterion that Hoey (2000) uses to distinguish simple lexical repetition from complex lexical repetition is to examine whether two items are different solely in terms of different choices from a grammatical paradigm, such as singular or plural paradigm (as in **bear** and **bears**), and verb-form paradigm (as in **try-tried**). This criterion, however, can not apply to the analysis of lexical repetition in Chinese text, as Chinese language is not overtly inflected. In English, many nouns are inflected for **number** with the inflectional plural affix **-s** (as in the above example **bear-bears**), and most verbs are inflected for **tense** with the inflectional past tense affix **-ed** (as in the above example **try-tried**). On the other hand, these types of inflection don't exist in Chinese. For this reason, changes have been made to the heuristic charts that Hoey (2000) designed to achieve the purpose of this study. As Hoey (2000) notes, the heuristic charts designed for the analysis of lexical repetition are not supposed to be exact definitions of various lexical repetitions. Rather, these charts are mainly used to distinguish various kinds of repetition from each other in contexts. In other words, the lexical relations between the items, such as synonymous and antonymous are mainly

contextual and only identified by assuming close approximation in context. An example of the chart has been provided as Figure 3.2, more detailed charts can be found in Chapter 4.

Figure 3. 2 *The lexical repetition entrance flow chart*



The last step of the manual analysis was to establish a tagging system and use it as a tool to annotate the all the 50 sample texts. Consequently, this annotation helped

the researcher to obtain, sort, and store statistical information with the help of the concordancing programme NotePad. In corpus linguistics research, there are a number of ways to annotate text depending on the needs of researcher. This study uses XML as its markup language, which is a modern system for annotating a text in a way that is syntactically distinguishable from that text. On the basis of the revised categories of lexical cohesive ties in Chinese text, seven types of tags are used to annotate the corpora data of this study (see Table 3.3). An example of a tagged text is given below to explain how it used to annotate the data.

Table 3.3 *The tagging system for the study*

Types of lexical repetition	Types of tags
Simple lexical repetition	<L type1="sr" sn="sn" lk="" ></L>
Complex lexical repetition	<L type2="cr"sn="sn" lk=""></L>
Constituent repetition	<L type 3=" constituent- re" sn="sn" lk=""></L>
Simple paraphrase	<L type4="sp"sn="sn" lk=""></L>
Complex paraphrase	<L type5="cp"sn="sn" lk=""></L>
Superordiante and hyponyms	<L type6="super-hypo" sn="sn" lk=""></L>
Co-reference	<L type7==" co-reference " sn="sn" lk=""></L>

L=lexical repetition, lk=linking item

The following example is a tagged sentence (numbered 11) in the NCEE Text 4 情感的权力 [the power of emotion].

Example 12:

11 而在 <L type5="cp"sn="sn11"lk2=" 历史 "> 现今 </L><L type1="sr"sn="sn11"lk9="社会"">社会</L>, 小到班级学校, 教师们不因为个人与某些学生的<L type1="sr"sn="sn11"lk9="亲疏"lk12="亲疏"><L type5="cp"sn="sn11"lk8="近"><L type5="cp"sn="sn11"lk8="疏">亲疏</L>而<L type2="cr"sn="sn11"lk1="错">错</L>判正误;大到一个国家不因与某些国家的邦交厚薄而任意进行军事政治支援纵容 <L type5="cp"sn="sn11"lk7="正义" lk8="正义" lk12="正义">非正义</L>战争

的肆意扩大，伤及无辜;这些，都是<L type5="cp"sn="sn11"lk1="错">正确</L>处理<L type1="sr"sn="sn11"lk2="情感"lk4="情感" lk5="情感"lk7="情感" lk8="情感"lk1="情感" lk12="情感"><L type5="cp"sn="sn11"lk12="真相 ">情感 </L> 与 <L type1="sr"sn="sn11"lk8=" 真理 "><L type5="cp"sn="sn11"lk1="情感"><L type5="cp"sn="sn11"lk7="情感"><L type5="cp"sn="sn11"lk12="情感">真理</L>的极佳明证，不论在什么时间，什么地点， <L type4="sp"sn="sn11"lk1=" 真相 "><L type4="sp"sn="sn11"lk12=" 真相 ">真理 </L> 永远是 <L type5="cp"sn="sn11"lk2="情感"><L type5="cp"sn="sn11"lk4="情感">真理</L>，我们不能掩盖或<L type5="cp"sn="sn11"lk8="颠倒黑白">歪曲</L>它，所谓"真金不怕火炼"么！

In the first line of the example, the word 现今[nowadays or these days] is marked with a Type 5 tag, which reads like <L type5="cp" sn="sn11"lk2="历史">现今</L>. The **L** within the angle bracket refers to lexical repetition, **type5="cp"** and means that this repetition is a type of complex paraphrase repetition, **sn="sn11"** indicates that the word 现今[nowadays or these days] is located in Sentence 11 of this text, while **lk2="历史"** can be interpreted as the word 现今[nowadays or these days] forms a repetition link with the word 历史 [history] located in Sentence 2 in the same text.

To support this tagging system, the annotation work began with numbering each sentence in the order in which they occurred in each text. For the purpose of the analysis of this study and for the convenience of referring to items that form links with other items in the same text, this study uses the sentence as its basic unit of analysis.

Once all the texts were numbered, each type of repetition was then identified and tagged with the help of the heuristic chart and the tagging system described above. Care was taken to ensure that every item was checked against every other to ensure that all lexical connections were checked and identified. It should be pointed out that the categorization of each link was not always clear-cut, rather, on

some occasions it involved a somewhat subjective decision. Consequently, among the links identified there were some cases which are either arguable or doubtful within the immediate context. In this case, a + symbol will be marked following the name of certain type of repetition within the quotation marks. For instance, if a link is an arguable simple repetition, it will be marked as `<L type1="sr+" sn="sn" lk="" ></L>`. Although it is important to acknowledge the existence of these doubtful cases, it does not, however, invalidate the general procedure and are of relatively little practical importance. As Hoey (2000) argues, the significance of the analysis of the lexical repetition used in non-narrative text does not lie in the presence of isolated links, but the clustering of links. Consequently, once the clustering of links was identified, the doubtful cases ceased to matter.

3.6.2 Step 2: The computational analysis

After the manual analysis stage, the tagged texts were converted to plain text format and made ready for sorting and running queries of the data. On the basis of the predefined research questions, the researcher used the concordancing programme NotePad to run various queries with the aim of obtaining such statistical information as: the number of types of lexical repetition used in a text; the number of connections a sentence forms with other sentences; the frequency of each type of lexical repetition which appears in a text. For instance, to obtain such statistical information from the sample text 情感的权力 [the power of emotion], the researcher run various queries with different search terms, such as: *Type1*, *Type2* (the name of certain types of lexical repetition) and *lk1*, *lk2* (the numbered linkage formed between certain items). The following is an example of the result of the query with the search term *type1*:

Example 13:

Total number of tokens of "type1" found in efo~3.txt = 36

情感的权力 1 在我所听闻过的<L type1="sr"sn="sn1"lk5=" 悲剧 "><L type5="cp"s

?>错</L>是占大多数的，因为<L type1="sr"sn="sn1"lk2=" 情感 "lk4="

情感 " lk

"sn="sn1"lk11=" 真理 ">情感</L>的<L type1="sr"sn="sn1"lk4=" 权力
"lk7=" 权力 " lk

人企及 ">太大</L>, 大得等不及<L type1="sr"sn="sn1"lk12=" 真相 "><L
type4="sp"

, 就将一切画上了句号。 2 翻开<L type1="sr"sn="sn2"lk9=" 历史 "><L
type5="cp"s

子</L>一折接一折, 这大概就是<L type1="sr"sn="sn2"lk1=" 情感 "lk4=" 情感 " lk

舌, 三言两句套套近乎, 便成了<L type1="sr"sn="sn3"lk6=" 亲信 ">亲信
</L>,

">情感</L>被这些居心回测之心<L type1="sr"sn="sn4"lk8=" 操纵>操纵
</L>, <

"sn4"lk8=" 当权者 ">君主们</L>的<L type1="sr"sn="sn4"lk2=" 情感
"lk1=" 情感 " lk

...

As can be seen, the total number in the example indicates that there are 36 simple lexical repetition items found in the sample text. By running similar queries to uncover other types of repetition, all the necessary information in relation to the use of various lexical repetition were then gathered for further exploration and discussion. Table 3.4 provides an example of the analysis of the types of repetition used in the text 情感的权力 [the power of emotion], including the numbers of instances of each type, and the frequency of occurrence of each type per 100 characters of text.

Given that the study intends to explore the ways in which lexical repetition serves a text-organizing function, the data in Table 3.4 show that there is considerable variation in the types of repetition used, the number of occurrences of each type, and their frequency of occurrence in the sample text. However, the significance of this type of statistical information does not lie in what emerges from one particular text, but in the possible commonalities or trends summarized on the basis of the analysis of the distribution of such features in a large number of texts.

As discussed in Section 3.3, the number of texts in the NCEEE analyzed for this purpose is 50.

Table 3.4 A summary of lexical repetition in the NCEE Text 4 情感的权力[the power of emotion]

The title of sample text: 情感的权力[the power of emotion]		
Total number of characters: 885		
Name of repetition type used	Number of the repetition items used	Frequency (per 100 characters)
Simple lexical repetition (type1 sr)	36	$36 \times 100 / 885 = 4.07$
Complex lexical repetition (type2 cr)	11	$11 \times 100 / 885 = 1.24$
Constituent repetition (type3 constituent re)	0	0
Simple paraphrase repetition (type4 sp)	34	$34 \times 100 / 885 = 3.84$
Complex paraphrase repetition (type5 cp)	74	$74 \times 100 / 885 = 8.36$
Superordinate and hyponym (type6 super-hypo)	6	$6 \times 100 / 885 = 0.68$
Co-reference (type7 co-re)	0	0

On the basis of the query results, the construction of repetition matrix was carried out. A repetition matrix, according to Hoey (2000), is a form of representation which allows a record to be kept of the number and the nature of the links identified between lexical items in one or more texts. An example of repetition matrix is given as Table 3.5, from which the following information is obtained:

Table 3.5 *Repetition matrix of the NCEE Text 4情感的权力 [the power of emotion]*

(1)												
(2)	3	(2)										
(3)	0	2	(3)									
(4)	6	3	0	(4)								
(5)	3	1	0	3	(5)							
(6)	0	0	2	0	2	(6)						
(7)	1	0	0	1	0	0	(7)					
(8)	16	5	0	7	2	0	4	(8)				
(9)	3	9	0	3	2	0	0	5	(9)			
(10)	0	1	0	1	0	0	2	4	0	(10)		
(11)	12	4	0	2	0	0	4	9	2	0	(11)	
(12)	6	1	0	1	1	0	6	9	3	2	9	

Total cells:	66	100%
Empty cells	26	39%
One repetition	8	12%
Two repetitions	9	14%
3 or more repetitions	23	35%

In Table 3.5, the horizontal columns indicate a sentence's lexical connections with any previously occurring sentences in the text, while the vertical columns show its connections with later sentences. Thus, for instance, Sentence 4 is shown by its horizontal column to have no link with Sentence 3, which links with Sentence 2, and 6 links with sentence 1. Its vertical column shows it to have three links with Sentences 5 and 9, one link each with Sentences 7, 10, and 12, seven links with Sentence 8, two links with Sentence 11, and no link with Sentence 6.

When discussing and interpreting a repetition matrix like this, Hoey (2000) also

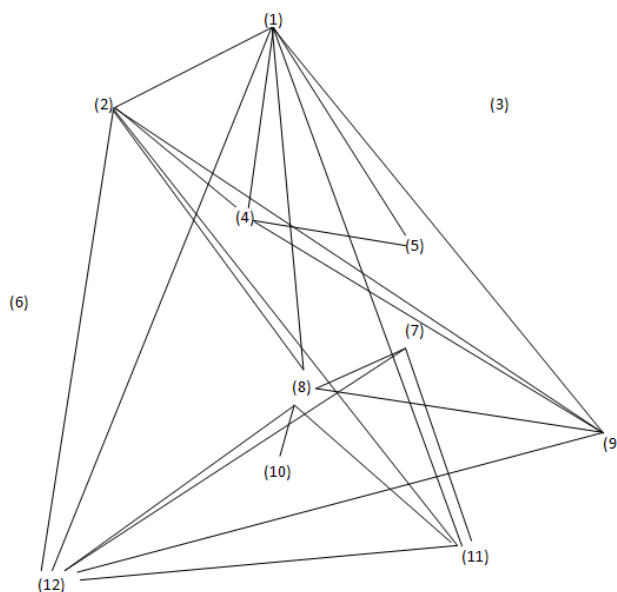
uses the term *bonding* to describe relations between sentences. According to Hoey (2000), a bond is established whenever there is an above-average degree of linkage between two sentences, and a bond can be defined as "a connection between any two sentences by virtue of there being a sufficient number of links between them"(p. 91). Hoey argues that the number of links which constitute a bond is dependent upon the type of text and on the average number of links within the text (p. 91), but the least number of links is three because of "the greater likelihood of two repetitions occurring in a pair of sentences by chance"(p. 190). As can be seen from the figures below the matrix, out of 66 cells in the matrix, 26 (39%) demonstrate an absence of links; of the remainder, a further 17 show one and two repetitions. This means that nearly two-thirds (65%) of sentence pairs are not significantly linked by lexical repetition. The rest of the sentence pairs (35%), however, are connected by an appreciable number of links, and out of 23 repetitions, only five of them are adjacent pairs (two adjacent sentences). These are the ones that are worthy of careful investigation. For this study, the cut-off point for claiming a bond is, therefore, not less than three links between a pair of sentences. In the above extract of the repetition matrix, the highlighted numbers are bonds.

Repetition matrices, such as in Table 3.5 are of great importance to the current study on lexical patterns in text. To begin with, they allow the researcher to demonstrate accurately the *non-linear complexity* of the links between sentence pairs. Non-linear here means that the connections in a text do not always occur between adjacent sentences, but are more often separated by intervening sentences. For instance, in Table 3.5, out of 66 cells, there are 28 connections which are formed between adjacent pairs, the other 38 are between sentences that are not adjoining to each other. In the study of the use of lexical cohesive repetition, the concept of non-linear complexity is clearly of great significance. As Hoey (2000) suggests, "one of the characteristics of mature native speaker writing is that the writer's sentences will relate to each other in non-linear ways" (p. 242); they tend to use repetition devices to connect over a considerable distance in a text, whereas some second language speakers, even though they are quite skillful in the use of various cohesive devices, tend to "repeat at shorter distances, typically within the paragraph boundaries" (p. 242). In the light of the discussion of non-linear

complexity, this study examined the features of this aspect of the use of lexical cohesive devices in texts written in the NCEE. Detailed description and discussion can be found in Chapter 4.

Another purpose that the matrix serves is to help research to generate a topological diagram, which Hoey (2000) terms a *net*. A net is a representation which demonstrates the web-like interconnection of bonded sentences. The principle of generating such as diagram is that each pair of sentences that form a bond is represented by a line. It is worth noting that such a diagram is only regarded as a typological representation, the position, length, or angle of line are therefore of little significance. Figure 3.3 illustrates the net derived directly from Table 3.5.

Figure 3.3 *The interconnection net of bonded sentences derived from Table 3.5*



A close look at the net can visually reveal some organizing patterns of text. Firstly, Sentences 1 and 2 appear to be the topic sentences for the whole text, introducing the theme, opening the discussion, and letting the argument flows through, while Sentences 3, 6 and 10 do not seem to connect with other sentences to any great extent and, therefore, contribute less to the main theme. It also appears as though Sentences 7 and 8 function as sub-topic sentences, as they connect more subsequent sentences than preceding sentences. Sentences 9, 11 and 12 clearly play a topic closing role in the text. From a text-organizing perspective,

Sentences 1, 2, 7, 8, 9, 11, and 12 can, therefore, be said to be *central sentences*, whereas the rest can be regarded as *marginal sentences*. It is hoped that the examination of features of such typological diagrams could help to deepen our understanding of how lexical cohesive devices contribute to the organization of the texts from the NCEE. This, however, deserves a more detailed discussion, which can be found in Chapter 4.

3.6.3 Step 3: The statistical analysis and interpretation

The last step of data analysis involves interpreting the statistical information obtained from the various query results and reporting the findings. According to Jensen (2008), “statistics can be used in two ways: to describe a data set or to draw inferences outside of the data set (descriptive and inferential statistics, respectively)” (p. 2). For the purpose of this study, the query results were first listed to present an overall picture of the use of lexical cohesive ties in the NCEE text. The data were organized in ways similar to Tables 3.4, 3.5, and Figure 3.3. This was a rather descriptive step without any assumptions made in the process. The statistical information mainly covered the following aspects: the number of various type of lexical repetition used; the frequency of the use of each type of repetition; the bonds formed in each text and the frequency of the formation of bonds; the degree of non-linear complexity of each text, and the number of central sentences and marginal sentences in each text. Once the numbers and the figures were ready, the next step was to summarize the statistical information with an aim of examining whether there are any patterns demonstrated in the use of the lexical devices. This step was more inferential, and involved an intensive process of observation, making logical judgments and reasoning. The following step was to report the inferences and discuss any issues that are worth further investigation in the future. The results of the above analysis are systematically classified and reported in detail in Chapters 7 and 8. The findings mainly fall under the following categories: the frequency of the use of each type of lexical repetition, the discrepancy between the various type of lexical repetitions used in text, the characteristics of the use of various repetition devices, and the degree of linear and non-linear complexity. Based on the above analysis, some problems facing

CFL learners in relation to the use of repetition devices are then identified and discussed.

3.7 Summary

This chapter has provided an outline of the research methodology employed in the study. It outlines the data collection procedures, corpus design, the collection and analysis of the corpus data, and the reporting of the findings. Justifications and examples were also provided to help clarify some research relevant issues. The main focus is concerned with the compilation of the NCEE and HSK Corpora, and on the basis of which to carry out the analysis of the use of cohesive devices in both corpora. Chapter 4 presents the analysis of the use of cohesive devices in the NCEE Corpus and reports the findings accordingly.

CHAPTER 4: STUDY 1 FINDINGS

4.1. Introduction

Chapter 3 discusses the overall research design for the study and outlines the analytical framework employed to address the research question. This chapter reports on the results of the analysis of cohesive ties used in texts from the NCEE Corpus (see Section 3.6.2). Two different types of analysis were carried out, one is a sole-rater manual analysis of the NCEE Texts 1-10, the other is a computational analysis of 50 texts in the NCEE Corpus. The purpose of the former is to examine whether or not Halliday and Hasan's (1976) taxonomy of cohesion can be applied to categorize the cohesive devices used in the NCEE texts. On the basis of the result of the sole rater analysis, the computational analysis firstly attempts to identify the various types of non-lexical repetitions used in each sample text. This includes the definition of each type of repetition, and explanation of what their function is in the organization of text. On the basis of Hoey's (2000) analytical approach to the study of lexical repetition, the chapter then explores the various categories of lexical repetition used in each text. Following the description of each type of repetition, examples in Chinese characters with English translations in brackets are also given to illustrate and to help understand the points being made. Finally, using the concordancing programme, this chapter examines some possible lexical patterns in Chinese texts by carrying out a corpus statistical analysis, which reports the total number of types of lexical repetition, the number of each type, and the frequency of each type used in the sample texts respectively.

4.2. Non-lexical cohesion

Halliday and Hasan (1976) classify the cohesive ties that occur in English text into the following categories: reference, conjunction, ellipsis, substitution and lexical cohesion. These devices, as many studies on the cohesion in Chinese text have shown, also play the similar role in the organization of Chinese text. However, this does not imply that Chinese text is without its own particular cohesive devices. On the basis of a rater analysis of NCEE Texts 1-10, the types

of cohesion described in Sub-sections 4.2.1 to 4.2.7 following are found to constitute a high proportion of cohesive ties in Chinese non-narrative text. These are termed: thematic chunk, parallel structure, the use of idioms and proverbs, the use of quotations from classics, the use of metaphor, and the use of summary words. In these sub-sections, these types of non-lexical repetition are defined and illustrated. It should be noted that that when giving examples of these cohesive devices, the corresponding English translation is also presented within square brackets to help reader to understand the meaning of the Chinese text. To highlight the structural and lexical features of the Chinese text, the translation approach adopted in this study is more literal. As a result, in some cases the structure of certain sentences, or the vocabulary used in some particular circumstances may seem not conform to the regular usage of the English language.

4.2.1 Thematic chunk

These chunks usually serve as the title of the essay and are repeatedly used throughout the text. They may appear in the introduction, helping to open up the discussion, or appear in the beginning or the end of each paragraph in the main body part acting as the topic sentence. They can also be found in the conclusion part of the text severing as the concluding word or final touch.

Although clearly playing an important role in the organization of the text, these chunks are difficult to classify under one of the categories which have been well researched in TCFL –namely reference, conjunction, substitution, and ellipsis.

It is worth noting that these chunks are different from idioms and sayings, which are often widely appreciated and used by native speakers. These chunks, however, are often uniquely conceived and constructed by the candidates themselves — often closely related to the given topic of the composition in the NCEE — and used by the writer of the text to achieve their specific argumentative purpose. Consider the following examples:

Example 14 (taken from the NCEE Text 27):

(1) 勿以好恶论断之 [Never let your judgment be colored by your likes and dislikes]

This chunk appears six times in Text 27: one as the title, one as the introducing point of the text, three as the leading sentences in the discussion part of the main body, and one as the concluding point in the end of the essay. Clearly the writer uses it as a chunk to serve as a thematic thread.

Example 15 (taken from the NCEE Text 50):

(2) 怎一个“情”字了得? [How can the word *emotion* express it completely?]

This chunk also occurs six times in Text 50 怎一个“情”字了得? [How can the word *emotion* express it completely?]. The writer borrowed the widely quoted expression 怎一个“愁”字了得? [how can the word *sorrow* express it completely?] and replaced the word *chou* (sorrow) with *qing* (emotion) to serve his argumentation. The chunk appears six times in the essay, and is used in almost the same way as the chunk in Example (14).

From our observation, the use of the Thematic Chunk in the NCEE Corpus is very common. Below is a list of examples found in the NCEE Corpus:

Example 16:

缤纷成就和谐 [profusion (or variety) makes harmony]

没有付出哪会有累累硕果 [no pain, no gain]

诗意地生活 [live a poetic life]

不要轻易说“不” [do not say no hastily]

公则生明 [impartiality leads to honesty]

切莫走近 [a hedge between keeps friendship green]

不要让眼光带上感情 [don't let your emotions affect your judgment]

切莫感情用事 [never act impetuously]

能思考的苇草 [the reed that can think]

情感的权力 [the power of the emotion]

调准感情焦距 [adjust the emotional focus]

板桥体 [the Banqiao calligraphic style]

理性的天平 [the rational scale]

不以亲疏好恶来定是非[don't judge right and wrong by one's relationship and taste]

勿以亲疏定善恶 [do not judge good and evil by one's relationships]

From the examples above, we may summarize the commonly shared features of the Thematic Chunk as following, which:

- 1) may be a phrase, or a sentence;
- 2) mostly serves as the title of the text;
- 3) is repeated three or more times;
- 4) may be used independently or as part of the sentence; and,
- 5) may have variations occasionally, but the meaning always remains the same.

4.2.2 Parallel-Structure

Hu (1994) discusses this structure in his analysis of the structural cohesion in Chinese text. He points out that this type of cohesion actually exists in both English and Chinese text. Consider the following English example:

Example 17 (taken from Hu 1994, p. 84):

(17) Cliff: Eek! Eek! I am a mouse.

Jimmy: A randy little mouse.

Cliff: (dancing round the table, and squeaking up). I'm a mouse. I'm a mouse. I'm a randy little mouse.

Osborne: *Look Back in Anger*

Hu (1994) argues that examples like (17) can hardly be found in Halliday and Hasan's (1974) examples of ellipsis and substitution. He maintains that it would be too simplistic to categorize the three sentences "I'm a mouse. I'm a mouse. I'm a randy little mouse" as mere lexical cohesion, as although there are indeed repetitions of words like *I*, *'m*, and *mouse*, the structure of the three sentences are parallel as well. And the textual meaning of *randy little* would be completely lost if the sentences are only regarded as lexical cohesion. It is the parallel structure that makes the last sentence highlighted and helps to lead the paragraph to the climax.

He then suggests that parallel structures do not involve only similarities of surface structure or of meanings, but the unity of structure and meaning. This kind of parallel structure is frequently found in the analyzed sample texts of Chinese writing. Consider the following two examples:

Example 18 (taken from the NCEE Text 13 诗意地生活 [Live a poetic life])

(10) 内心和谐，充溢着对万物的爱的人才可以诗意地生活。 [People can only live a poetic life when they have inner harmony and overflow with love towards all things.]

(11) 特蕾莎修女给每一个穷人带去关爱，爱让她的生活充满诗意； [Sister *Teresa* brought her love to every poor person, this love then made her life full of poetry;]

皮埃尔神父为无家可归者送去温暖，关怀让他的生活充满诗意； [Father *Pierre* sent warmth to the homeless, this concern then made his life full of poetry;]

袁隆平院士为饥饿中的人们送去希望，爱让他的生活充满诗意。 [The academician *Longping Yuan* sent hope to the hungry, the love then made his life full of poetry.]

(12) 当人们内心和谐，带着对万物的爱去生活，人们才可以学会如何诗意地生活。 [People can only learn how to live a poetic life when they have inner harmony with the love towards to all the things.]

(13) 诗意地生活，人们需要从阅读中汲取养料。 [To live a poetic life, people need to absorb knowledge from reading.]

(14) 让阅读成为生活的一部分，生活才可以变得诗意。 [Life can only become poetic if reading has become part of life.]

(15) 当人们从儒、道、墨法的典籍中探究为人之道， [When people seek the principal of living from *Confucian, Taoist, and Mohist* classics,]

当人们从司马迁的竹简中开启历史的明镜， [When people unlock the mirror of the history from writings of *Maqian Si*,]

当人们从唐宋八大家的作品里顿悟文章之法， [When people explore the way to write from the Eight Masters in *Tang* and *Tsong* dynasties,]

智慧便开始滋润人们干涸的灵魂，给灵魂注入诗意。 [Wisdom then begins to quench the thirsty soul of people, and to infuse poetry to their soul.]

(16) 当人们从陶渊明的菊花中思索隐逸之士的情怀, [When people reflect on the sentiments of recluse from *Yuanming Tao's The Chrysanthemum* ,]

当人们从史铁生的地坛中感念生与死的变迁, [When people sense the change between life and death from *Tiesheng Shi's Ditan*,]

当人们从卢梭的瓦尔登湖畔拾获人生的宁静, [When people appreciate the peace of life from *Rousseau's The Walden Pond*,]

思维便开始给心灵插上翅膀, 让心灵翱翔于诗意的天空。 [Thinking then begins to free people's hearts and let their thoughts soar in the poetic sky.]

(17) 从阅读中收获智慧, 从阅读中学会思考, 人们才能懂得诗意的生活。 [People can only appreciate the poetic life when they harvest wisdom from reading, learn to think from reading.]

Example 19 (taken from the NCEE Text 27 勿以好恶论断之 [Never make judgments by likes and dislikes])

(10) 感情的亲疏可以蒙蔽你的双眼, 让你看不清事物的真面目, 阻碍你对真理的认识, 成为你求知路上的绊脚石。 [Emotional closeness may blind your eyes, causing you hardly to know the truth, prevent you from perceiving the truth, and become the obstacle on your way to seeking knowledge.]

(11) 勿以好恶论断之, 楚王不听信郑袖之辞而疏远屈原, 自不会为秦国所吞并。 [Never make judgments by likes and dislikes. Had he not fallen for the word of Zheng Xiu and therefore alienated *Qu Yuan*, the *Chu* Emperor would not have been defeated by the *Qin* Kingdom.]

(12) 勿以好恶论断之, 李存勖不宠幸伶人, 重用伶官, 不谈国事, 自不会因一人之乱而七庙隳, 为天下苍生所笑。 [Never make judgments by likes and dislikes. Had he not favoured actors and put them in important positions, *Li Cunxu* would not have been laughed at by the whole nation, and his reign would not have been overturned by one people's rebellion.]

(13) 勿以好恶论断之, 诸葛亮不听信马谡的信誓旦旦, 自不会痛失街亭,

一败千里，死不瞑目。[Never make judgments by likes and dislikes. Had he not believed Ma Su's overstatement, Zhuge Liang would not have been utterly defeated and died with great regret because of the fall of Jieting.]

(14) 举目四望，古今中外，有多少英雄豪杰因内心之好恶而惨遭滑铁卢？[Looking back the history of China and worldwide, how many heroes sustained the failure such as that of Waterloo due to the judgments colored by their personal likes and dislikes?]

In the above two examples, parallelisms can be identified both within sentences and between sentences. As the purpose of this study is focused on cohesive ties between sentences, the following discussion will only take the parallelism that occurs between sentences into account. In Example 4, there are two groups of parallel structures, one is between Sentences 10, 12 and 17, sharing the same structure and key words ____才____诗意____生活 [People can only ____poetic life when_____]; the other is between Sentences 15 and 16, both has the same structure and key words 当人们从_____, 当人们从_____, 当人们从_____, _____便开始_____ [When people _____ from_____, When people _____ from_____, When people _____ from_____, _____ then_____]. In example 5, the parallel structure lies between Sentences 11, 12 and 13, as the three share the structure and key words 勿以好恶论断之, _____自不会_____ [Never make judgments by likes and dislikes. Had he not_____, _____would not have _____].

The characteristics of the parallel structure can be summarized in that they:

- 1) have the same grammatical structure;
- 2) share key words and structure markers;
- 3) are repeated two or more times; and,
- 4) may be consecutive or between distant paragraphs.

4.2.3 The use of idioms and proverbs

The use of idioms

While every language has its own idioms, it is commonly known that in many languages the definition of idiom itself remains controversial. In China, scholars, such as Liu (1982), regard idioms as words, phrases, or expressions that cannot be taken literally (p. 104). In other words, when used in everyday language, they have a meaning that differs from their dictionary meaning. Others, such as Zhou (2001), argue that idioms are fixed expressions that are sanctioned by long-standing custom and usage through history. For the purpose of this study, the more traditional definition of idiom is preferred. Similar to Liu's (1982) definition, idiom is defined here as "a word or group of words whose meaning in a particular context does not match the normally-accepted, literal meaning". This definition assists in differentiating an idiom from a fixed expression, which does not have a literal meaning and an idiomatic meaning.

It is worth noting that in the analyzed sample texts, the use of idiom constitutes a considerable percentage of cohesion ties. These can be traditional ones in Chinese literature, or borrowed from other languages such as English. Consider the following example:

Example 20 (taken from the NCEE Text 37):

(12) 如今，不也有领导干部任人唯亲，从而祸起萧墙的事例吗？

[Nowadays, aren't there many leaders, who appoint people by favouritism, which therefore give rise to some trouble?]

(13) 可见，认知事物，为人处世，若戴上感情的有色眼镜，危害大矣！

[It is obvious that the consequence is serious if one is affected by his own personal feelings when perceiving the world or conducting ourselves in the life.]

In this example, there is an idiom 祸起萧墙. Its literal meaning is *trouble arises from one's own door*, while here it means *consequence is caused by one's*

favouritism. It is therefore used to constitute the cohesive tie: 祸起萧墙 [trouble arises from one's own door] -危害 [consequence].

Example 21 (taken from the NCEE Text 43 任人唯亲, 任人唯贤 [Appointing people by favoritism, appointing people on their merit]):

(7) 在现代, 有没有韩非子笔下的"富人"呢? [In the current age, are there many people like the rich man created by Feizi Han?]

(8) 有! [Yes, there are.]

(9) 譬如在干部的任用问题上就出现了"任人唯亲"的弊端, 一些领导在任用下属时, 不是视其是否有才能, 而是看其与自己远近、使一些无德无能的人混入干部队伍, 正所谓"一人得道, 鸡犬升天"。[For instance, on the subject of the use of cadres, there are some problems of appointing people by favoritism. When appointing their subordinates, some cadres let some incompetent people and people without virtue sneak into the contingent of cadres, not depending on their ability, but on their relationship with them. This is just like what the idiom says—one turns immortal, his chicken and dogs go to the heaven as well.]

而真正有志有才之士被拒之门外。[While those competent and capable people are turned away.]

In this example, 一人得道, 鸡犬升天 is an idiom. The literal meaning of it is *when one turns immortal, his chicken and dogs goes to the heaven as well*, while its meaning in this context is *when someone get to a higher official position, his relatives and followers get promotion after him*. So in this context, 鸡犬[chicken and dogs] is used to describe those 无德无能的人[incompetent people and people without virtue], which is opposite to the phrase 有志有才之士 [people with high aspirations and with ability] in the following sentence. The following is an example that uses idiomatic expression borrowed from other languages such as English.

Example 22 (taken from the NCEE Text 27):

(13) 勿以好恶论断之，诸葛亮不听信马谡的信誓旦旦，自不会痛失街亭，一败千里，死不瞑目。 [Never make judgments by likes and dislikes. Had he not believed Ma Su's overstatement, Zhuge Liang would not have been utterly defeated and died with great regret because of the fall of Jieting.]

(14) 举目四望，古今中外，有多少英雄豪杰因内心之好恶而惨遭滑铁卢? [Looking back the history of China and worldwide, how many heroes sustained the failure such as that of Waterloo due to the judgments colored by their personal likes and dislikes?]

In the above example, Sentence 13 is an example of a historical figure in Chinese history who had sustained failure because of being affected by his personal feelings when appointing people. The underlined part in Sentence 13 is a description of the failure he sustained, and the underlined part in Sentence 14 is used to form cohesive ties with the preceding underlined part in Sentence 13. In Sentence 14, the idiom “the failure sustained in Waterloo” is translated from English literature, which has now been widely accepted in Chinese literature.

The use of proverbs

Like idioms, proverbs are also among those very important devices to make cohesive ties in Chinese texts. Most proverbs used in the analyzed text are preceded by the phrase 俗话说 [the sayings/the proverb says] or 所谓 [so-called], which make them very easy to recognize. Consider the following two examples:

Example 23 (taken from the NCEE Text 7):

(18) 俗话说：“防人之心不可无。” [The saying says, one should always guard against the harm others might do to him]

(19) 即使是心腹，也应防他那未露的尖刀。 [Even if someone is your confidant, you should be vigilant of the sharp knife he is hiding.]

Example 24 (taken from the NCEE Text 37):

(5) 自古以来，人们就颇为重视感情因素对事件成功与否的影响。[People have long been putting emphasis on the role the personal feeling plays in the way to success.]

(6) "打虎亲兄弟，上阵父子兵。"[One has to be dependent on his blood brother and father if he goes to hunt tiger or fight in a war.]

(7) 诚然，感情上的亲密对事业的成功也许会起到一定作用，然而，在认知事物本质上，感情上的亲疏有时也会给人带来错觉以及误导。[Indeed, close relationship may contribute to one's success in career; however, it sometimes may mislead people in the course of seeking the truth]

In Example 23, not only the word 防 [to guard against, to be vigilant of] in Sentences 18 and 19 forms a pair of repetition, and 不可 [should not] in Sentence 18 and 应 [should] in Sentence 19 are coherent with each other, the two sentences actually also form a relation of repetition, as Sentence 19 reinforces the idea that is expressed in Sentence 18. Sentence 18, therefore, can be regarded as using a non-lexical cohesive device used for achieving a cohesive tie.

4.2.4 The use of quotations from classics

In addition to the use of idioms and proverbs, it was also found that the repetition of quotations from classics can also function as a cohesive device. Consider the following example:

Example 25 (taken from the NCEE Text 14 好奇心[curiosity]):

(6) “好”，是一种欲望，求知欲学的愿望。[Hao is a kind of desire, a wish that shows the desire of seeking knowledge and learning.]

(7) 孔夫子早就回答过子贡“敏而好学，不耻下问”。 [Long before Confucius had taught Zigong to be sensitive and studious and not to be shamed for asking and learning from people who are at lower status.]

(8) “好”是一种态度，一种实事求是的精神。[Hao is an attitude, a spirit of

seeking the truth from facts.]

(9) 陆游曾告诫后人“纸上得来终觉浅，绝知此事要躬行”。[Lu You once warned his descendants that one must learn the truth through doing, and it is really shallow to learn only by reading books.]

In this example, two quotations are used in Sentences 7 and 9. They are both taken from ancient Chinese literature, and are used to repeat what has been explained in Sentences 6 and 8. In Sentence 7, the quotation is 敏而好学，不耻下问 [being sensitive and studious and not to be shamed for asking and learning from people who are at lower status], which is used to form a link by repetition with Sentence 6. Similarly, in Sentence 9, the quoted part is 纸上得来终觉浅，绝知此事要躬行 [one must learn the truth through doing, and it is really shallow to learn only by reading books], which is used to explain what is being said in Sentence 8.

4.2.5 The use of metaphor

In some cases, the figurative use of words, phrases or sentences can also be used to form cohesive ties. Consider Sentences 3, 4 and 19 in the following example:

Example 26 (taken from the NCEE Text 27):

(3) 勿以好恶论断之，这是一种怎样的人生境界，这是一种何等的明智选择！[Never make any judgments by one's likes and dislikes, what a spiritual realm it is! What a wise choice it is!]

(4) 当一个人可以将对事物的判断游离于感情的亲疏之外，那他必是明哲之人了！[A man must be a wise and philosophical one if he does not let his personal feelings affect his judgment.]

(19) 勿以好恶论断之，也许你会发现，狗尾草也会开出美丽的花。[Never make judgments by likes and dislikes. Maybe you will find that even the fox tail can bloom beautiful flowers (with a pursuing heart, even the mostly common people will become wise when making judgments).]

The expression 狗尾草也会开出美丽的花 [even the foxtail grass can bloom beautiful flowers] here is obviously a figurative use, meaning that even average people can become very wise and insightful if they do not let their personal feelings color their judgment. So the word 狗尾草[foxtail grass] means *average people*, while the phrase 美丽的花 [beautiful flowers] means *being wise and insightful*.

Here it is worth noting that this type of metaphor is one that is not necessarily frequently used or widely-known, a metaphor of this type (used as a cohesive device) can be invented by the writers to achieve their argumentative purposes. The notion of an invented metaphor is acceptable in the NCEE writing as long as the figurative use of the words or phrases is appropriate, and can convey the message the writer intends to express.

4.2.6 The use of summary words

Hu (1994) points out that using a summary word is a cohesive device that can be found in both English and Chinese. Below is an example taken from his work:

Example 27 (taken from Hu, 1994, p. 117):

University of California's former President David Gardner has approved a \$100 million extension of a program that offers 100 interest mortgages for faculty members and executives, a decision that comes in the face of possible student fee increases and staff pay cuts...*The action* was disclosed in the Oakland Tribune on Wednesday.

—Los Angeles Times 92.7.4

Here the summary word **action** refers back to the first half of the sentence in the first paragraph. McCarthy (1991) also discusses this type of word, such as *issue*, *problem*, *assessment*, and he terms it *discourse-organizing* word. The following are examples of this general word used to form cohesive ties.

Example 28 (taken from the NCEE Text 34):

(10) 在人类社会，自“大道既隐，各亲其亲各子其子”以来，亲情就成为了维系这个社会的稳固的牢不可破的精神纽带，人们往往认为亲人们总是对自己好，因而他们总是对的，没有错的，而对于旁人，难免存着此疆彼界之心，所谓“人心隔肚皮”，难免对别人有着成见，故事中的富人即此类典型人物，而有两位人物的行为却值得我们的赞叹。 [In our society, since people have increasingly become caring more for their own families and children, the love and affection between blood-relations seem to have become the solid and unbreakable spiritual tie that sustains the society. As a result, people often believe that their loved ones are the people who always treat them well and are always right. Towards other people, however, they are often vigilant and suspicious, often and have prejudice against them. The rich person in this story is a typical example of this. Contradictory to this, however, there are two people whose actions deserve our admiration.]

(11) 祁黄羊的“内举不避子，外举不避仇”，想必大家是耳熟能详了，不以亲疏好恶来定是非，而是惟贤是举，诚可贵矣。 [Maybe everyone is already familiar with the story of Qi Huangyang, who did not avoid recommending his own son for the right position among his own families, and did not avoid recommending his enemy among his connections. It is really valuable that he recommended people only according to their capabilities, not acting upon his own likes and dislikes.]

(12) 另一位是中国革命的先驱者孙中山先生，他在广大的人民群众中享有崇高的威望，在担任了南京临时政府大总统之后，权力可谓达到政治生涯的顶峰。 [The other one is Sun Yat-sen, the pioneer of the Chinese Revolution, who enjoyed high prestige among the masses at his time. After taking office of the president of the provisional government at Nanjing, he reached the apex of power in his political career.]

(13) 各界人士一致推举孙中山的哥哥孙眉担任广东的都督，当此之时，孙中山先生只需微微首肯，便定能保证自己的哥哥飞黄腾达，然而这位

伟人没有这样做，他只是婉言地谢绝了，说：“我哥哥才力有限，不堪担此重任。”[At that time, people from all walks of life unanimously put forward his older brother Sun Mei for the appointment of Guangdong Governor. Under that circumstance, with only a slight nod of his head, he would have made his older brother a great success in his life. Instead of doing so, the great man politely rejected the recommendation. My older brother is not competent for this position and cannot bear such great responsibility, he said.]

(14) 如此高风亮节怎能不让人仰慕不已。 [Because of such noble character and sterling integrity, what can stop people admiring him unceasingly?]

In the above example, there are two words that act as summary words, the first one is 行为 [actions] in Sentence 10, and the other is 高风亮节 [noble character and sterling integrity] in Sentence 14. The word 行为 [actions] in Sentence 10 summarizes the two actions of the two person 祁黄羊 [Qi Huangyang] and 孙中山 [Sun Yat-sen], and the word 高风亮节 [noble character and sterling integrity] summarizes the nature of the two actions. Clearly, these two words play an organizational role in the text and are therefore worth further investigation.

4.2.7 Summary and brief discussion

The rater analysis of the NCEE Texts 1-10 shows that apart from non-lexical categories proposed by Halliday and Hasan (1976), namely reference, conjunction, substitution and ellipsis, there are a considerable number of other cohesive devices used in native Chinese texts, such as thematic chunk, the parallel structure, the use of idioms, the use of proverbs, the use of classical quotation, the use of metaphor and the use of summary words. These findings, therefore, suggest that these additional categories of cohesion found in Chinese texts need to be considered by both TCFL teachers of writing and scholars in the TCFL area. On the other hand, in Chapter 6, it is seen that the analysis of the HSK texts shows a remarkable low frequency of the use of these devices. This gap in the use of

cohesive devices between native and non-native writers of Chinese is, and the need for a wider range of categories to describe cohesion is, therefore, worthy of further exploration and discussion in order to inform both the teaching of Chinese writing and the further investigation of the linguistic competences that relate to the writing of Chinese texts.

4.3. Lexical Cohesion

As discussed in Chapter 3, Section 3.6.1, a sole rater analysis of the NCEE Texts 1-10 was manually conducted to examine whether Halliday and Hasan's (1976) taxonomy of cohesion has the capacity to categorize the types of cohesion found in the NCEE texts. The following section describes approaches to categorizing the various types of repetition in non-narrative Chinese text. In the process of identifying different types repetition, a group of heuristic flow charts drawn from those of Hoey (2000) were used to represent the decision processes that being made when a lexical item is identified as being a simple or complex repetition of a previous item. These charts were presented after the definition of each category in the following sections. In Chapter 3, Figure 3.2 is presented as an 'entrance' chart, which establishes whether or not a lexical item is worth investigating and then provides a heuristic route to ascertain which category it belongs to.

4.3.1 Simple lexical repetition

As the simplest form of repetition, simple lexical repetition occurs when a lexical item that has already occurred in a text is repeated. The heuristic chart shown in Figure 4.1 is designed to help to decide whether a lexical item can be identified as being a simple repetition of a previous item. Consider the following example:

Example 29 (taken from the NCEE Text 5):

(17) 当他发现大厦设计存在漏洞，大厦可能在狂风中被折断时，他毅然向市政当局提交了调查报告。 [When he found out that there were building design flaws, which could pose great danger to the skyscraper if blown by

fierce winds, he requested the city council decisively to investigate it.]

(18) 一切的荣誉、金钱、掌声都似乎离他远了，但他却在广播中发出了真挚的话语：“我不乞求原谅，只求大厦能尽快加固。” [Because of this, the honor, money, the applause seemed not to be around him anymore. However, his speech that broadcasted on the radio was very sincere: "I do not seek for forgiveness, what I am really concerned about is to strengthen the building as soon as possible" .]

In this example, 大厦 [the skyscraper] in Sentence 18 is a simple lexical repetition of 大厦 [the skyscraper] in Sentence 17, since they have the same referent, both referring to the same building designed by 他 [he].

As Hoey (2000) explains, only open-set lexical items, such as nouns, adjectives can enter into such a repetition link. Other repetition connections between grammatical items, such as determiners, prepositions, auxiliaries, negatives, coordinators, subordinators, sentence conjunctions (or conjuncts), sub-modifiers or particles will not be treated as simple repetition. For example, the occurrences of the words 在 [in] in Sentences 17 and 18 are both prepositions, however, they will not be regarded as lexical repetition as they are members of close-set items, not open-set ones. Likewise, the structural particle 了, which appears in Sentences 17 and 18 will not be treated as an example of lexical repetition.

In Hoey's (2000) example of simple lexical repetition, **bears** and **bear** are treated as simple lexical repetition because the alternation of the two items is within a closed grammatical paradigm. The only graphological or phonological differences between these two items are the former is in a plural form, while the latter the singular. Some other examples Hoey (2000) gives are pairs like: **author-authors**, **work-works**, and **try-tried**. This type of example, however, is rare in Chinese, which shows very little overt inflection. In English, many nouns are inflected for number with the inflectional plural affix -s (as in the above example **bear-bear-s**), and most verbs are inflected for tense with the inflectional past tense affix -ed (as in the above example **try-tried**). On the other hand, in Chinese, there is very limited declension of nouns, with most nouns having no inflection at all, such as

in the following examples:

Example 30:

一张桌子 (one desk)--三张桌子 (three desks)

一个苹果 (one apple)--十个苹果 (ten apples)

一个工人 (one worker)--五个工人 (five workers)

The underlined words are nouns and their singular and plural forms are exactly the same. It is worth noting that there is one word 们 in Chinese which can function as a plural affix in some particular circumstances. However, the grammatical function and the context in which it appears are still very different from that of affix in English. For example, in the analyzed text there are examples like the following:

Example 31:

战士 (soldier or soldiers) - 战士们 (soldiers)

君主 (emperor or emperors) - 君主们 (emperors)

科学家 (scientist or scientists) - 科学家们 (scientists)

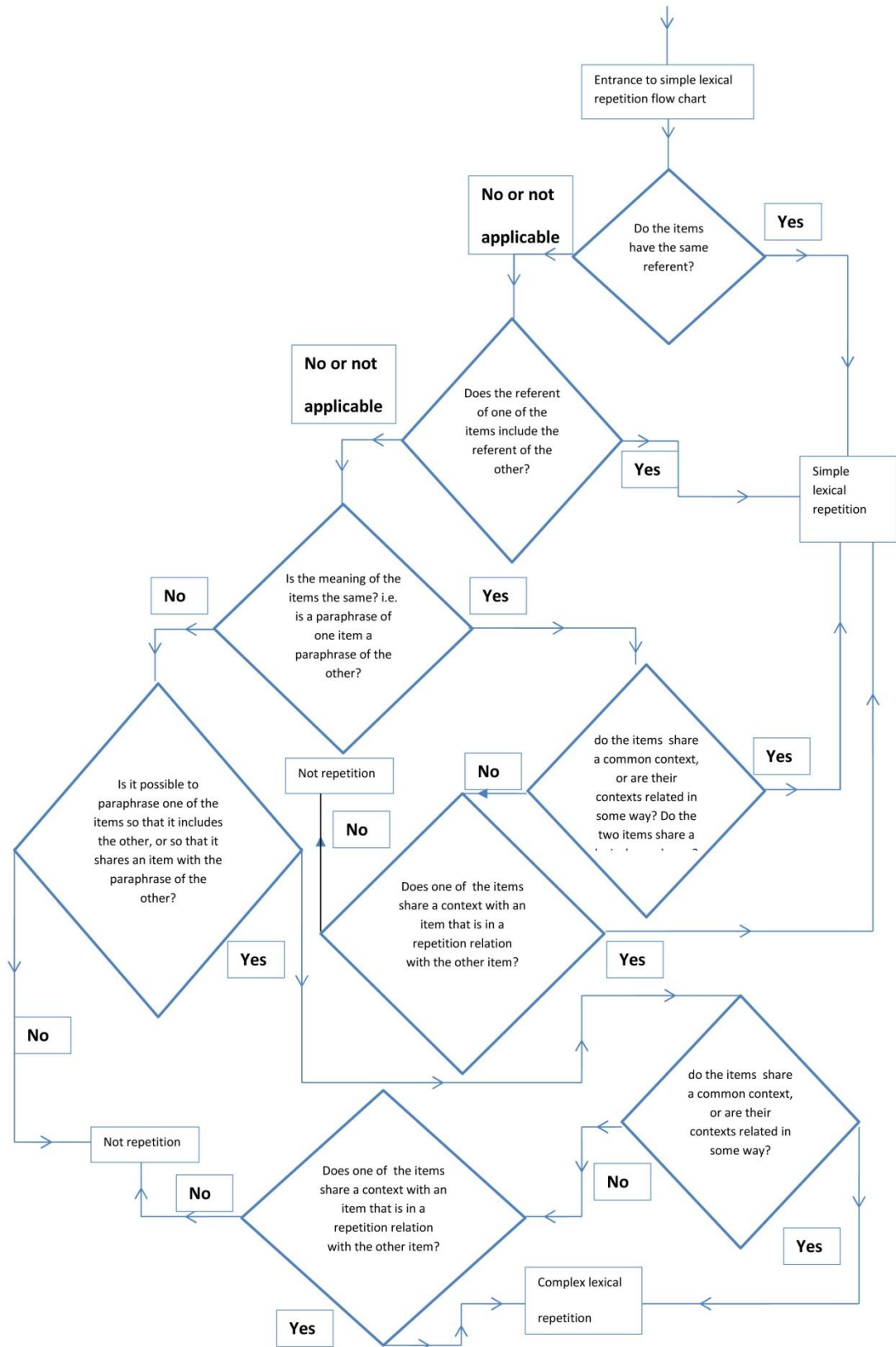
The underlined words can indicate one particular object or a group of objects, depending on the context. However, once they have 们 as their affix, they can only mean a group of objects. In this sense, 们 can be regarded as an inflectional plural affix. Nevertheless, the grammatical context 们 appears in Chinese is not as free as a common affix does in English. For instance, in the following example:

Example 32:

The soldiers are waiting for equipment.

战士们在等待武器的到来。

Figure 4.1 *The heuristic flow chart for simple repetition*



The two sentences have the same meaning, and both are grammatically correct. While in the following two sentences, they have the same meaning as well, but the Chinese sentence is grammatically wrong, because 们 cannot occur in this type of structure.

Example 33:

These two soldiers differ from each other.

*这两个战士们各不相同。

Below is an example where the suffix 们 is used to express the plural concept.

Example 34 (taken from the NCEE Text 46):

(3)自以为聪明的**猴子们**哪里知道，早上多出的那颗栗子，是从晚上那四颗中“挪”出来的，它们连半颗栗子都没多得到。[However, these self-conceited monkeys did not know that the extra chestnut they got in the morning was actually taken from the four chestnuts they would normally get in the evening, which meant that they did not actually get any extra chestnuts at all.]

(4)坐在考场中，我不由想到了减负中的我们，多少和这群**猴子**有点相像吧？[Sitting in the examination hall, I could not help thinking that the situation of us, who were supposed to be taken some of their burden off, is a bit similar to those monkeys.]

In the above example, the word 猴子 [monkeys] in Sentence 4 forms a simple repetition with the word 猴子们 [monkeys] in Sentence 3. Although the function of the suffix 们 is slightly different from the plural form -s or -es in English, it expresses the concept of plural in this circumstance. In this study, therefore, the suffix 们 in Chinese is regarded as one, and the only one grammatical paradigm of nouns.

Like most nouns, verbs in Chinese are also not inflected, either. Regardless of who is talking, when they are talking, or to whom they are talking to, the form of

verb remains the same. Verbs in Chinese do not have inflectional tense affix either. Rather, temporal nouns are used to indicate whether something will happen in the future, is happening now, or happened in the past. The Chinese language relies heavily on the use of temporal nouns and adverbs to communicate what English and many other languages do with different verb tenses. For example, in Chinese, the verb **to go** is 去; it does not change its form if it is used in a sentence indicating a event happened in the past.

Example 35:

I went to Auckland yesterday.

昨天我去奥克兰了。

The inflectional form of past tense of **to go** is **went**, while 去 does not change at all, the concept of the past tense is expressed by the temporal noun 昨天 (yesterday).

For the reasons discussed above, there will, therefore, be far fewer pairs of items which are entirely explicable in terms of a closed grammatical paradigm in Chinese text than that in English text. As a result, the vast majority of simple lexical repetition is in the form of exactly identical lexical items.

4.3.2 Complex lexical repetition

Complex lexical repetition occurs either when two lexical items share a lexical morpheme or a word, but are not formally identical or when they are formally identical, they have different grammatical functions. Figure 4.2 shows how to decide whether the relation of two lexical items is complex repetition. Consider the following example:

Example 36 (taken from the NCEE Text 14):

(23) 虽有夸张和虽有夸张和唯心主义的色彩，但却能看出人们对于“心”的**重要性**的认识。 [Although this claim has an exaggerating and idealistic flavor, it certainly shows the importance of the mentality that people have recognized]

(24) 古人说过，“心如止水”。 [Since people in the past said that, mentality is like tranquil water,]

(25) 由此可见，一份安静的心情是多么**重要**。 [It can be seen that how important it is to have a calm mood.]

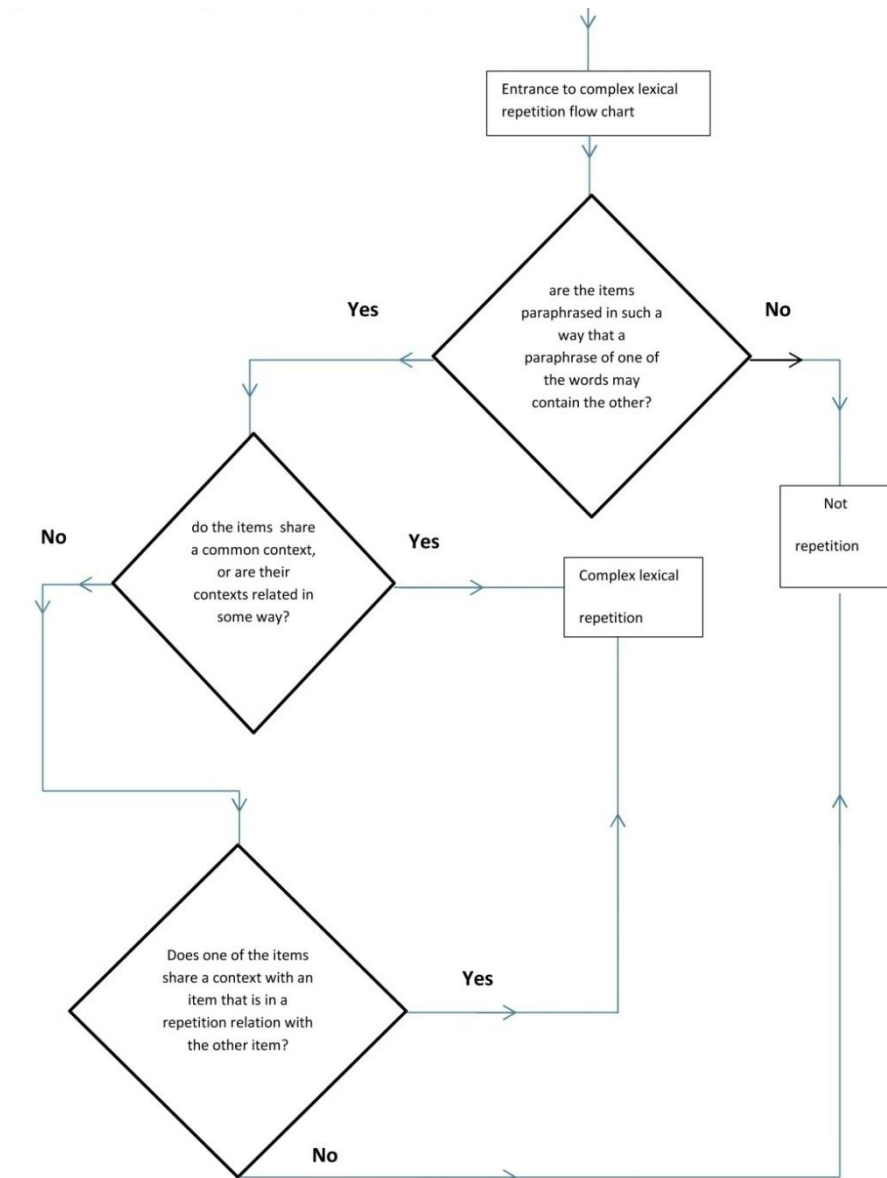
In this example, the word 重要性 [importance] and 重要 [important] form a complex lexical repetition, because they can be paraphrased in the context of the text in which they appear that the paraphrase of the former includes the latter: 重要性 [importance] can be explained as 重要的性质 [the nature of being important]. Likewise, words like 科学 [science] and 科学家 [scientist] can also be regarded as one example of complex lexical repetition. In the context of the text, 科学家 [scientist] may be roughly explained as 精通科学(尤指自然科学)的人 [people who have professional knowledge in the field of science, especially in natural science field]. Clearly, 科学 [science] and 科学家 [scientist] share the same word 科学 [science], and the paraphrase of 科学家 [scientist] includes 科学 [science]. The following figure is concerned with the complex repetition.

The following example demonstrates that if two lexical items are identical, they can still form the relation of complex repetition if they have different grammatical functions.

Example 37 (taken from the NCEE Text 32):

(2) 如果你是一名律师，你的亲人犯了罪，在法庭上你是依法办事，一切按法理讲，还是运用你的口才，抓一些漏洞，为亲人尽量摆脱罪名，这实在很难选择：一边是法理，一边是人情，该顾哪边？ [If you are a barrister and you

4.2 The heuristic flow chart for complex repetition



are pleading for your loved one in court who has committed a crime, will you act righteously in accordance with the law, or will you use your eloquence to find a loophole so that your love one can be freed of charges? It is really difficult for you to choose, as you are in the middle of the two conflicting sides, one is your relation and the other is the law, which side would you take?]

(4) 选择正义，有罪就该罚，这才是你的正确选择，切莫感情用事。
[Choose the righteousness and let the one who commits the crime be punished,

this is the right choice that you should make. Do not act only upon your emotions.]

In this example, the word 选择 [to choose] in Sentence 2 is a verb, while the word 选择 [choice] in bold in Sentence 4 is a noun, although the two are formally identical, they have different grammatical function, hence they are regarded as an example of complex lexical repetition.

4.3.3 Constituent repetition

In categorizing another type of device, it was noted that, when applying the criterion that the paraphrase of one item needs to be synonymous with or include the other in the context of the text, one type of repetition appear to be somewhat different from the previously discussed categories of simple and complex lexical repetition. In the sample texts, it was found that there are links between words and their constituents, or between idiomatic chunks and their constituents. Although they satisfy the criteria to be classified as complex repetition---the paraphrase of a word or chunk includes that of a constituent, this type of repetition seems involve more than complex repetition in terms of its role in helping the writer organize a text. In other words, the constituent of a certain word not only forms a link with the word, but more importantly, it serves as a thematic thread in the organization of a text. It is therefore worthwhile discuss this type of repetition under a different heading---constituent repetition. The flowchart in Figure 4.3 shows how instances of constituent repetition were identified.

Constituent repetition occurs when a word or a fixed expression forms a connection with one of its constituents, which is used self-sufficiently somewhere else within the text and still retains the same meaning. To achieve their argumentative purpose, writers often break down a word into constituents and use these components as their argument points. See the following example:

Example 38 (taken from the NCEE Text 14):

(4) 何为好奇心? [What is *haoqixin* (curiosity)?]

(5) 于我来说, 不过三字而已: 好, 奇, 心。 [To me, it just three words: *hao* (desire), *qi* (novelty), and *xin* (mind).]

(6) “好”, 是一种欲望, 求知欲学的愿望。 [*Hao* is a kind of desire, a wish that shows the desire of seeking knowledge and learning.]

(8) “好”是一种态度, 一种实事求是的精神。 [*Hao* is an attitude, a spirit of seeking the truth from facts.]

(12) 何谓“奇”? [What is *qi* (novelty)?]

(13) 不过是遇到“奇”发现“奇”使之不“奇”。 [It is just to come across *qi*, to discover *qi*, and to make it not *qi*.]

(21) “心即理也。”[The mind is principal.]

(22) 明代的王阳明是这样解释心与理的关系。 [This explanation for the relation between the mind and the principal is provided by Wang Yangming in Ming dynasty.]

(30) 好而不奇是庸才, 奇而无心是蠢才, 只有用心好之奇之, 才会有所发现, 有所收获。 [People who have desire to explore yet with no action to discover new things is a mediocrity, people who take action to discover new things yet without using their mind are fools. Only people who have the desire to explore, take action to discover new things, and always use their mind, will make discoveries and accomplishments.]

(31) 人生凭“好奇心”而活, 人类凭“好奇心”兴旺发达, 永续发展。 [It is the curiosity that makes us live in the world, and it is the curiosity that makes human kind grow and flourish, and therefore maintain its existence and development in the world.]

In the above example, the word 好奇心[curiosity] in Sentence 4 is broken down

into three parts: 好 [desire], 奇 [novelty], and 心 [mind]. The writer then uses these three constituents to explain what he thinks 好奇心 [curiosity] is. Clearly in this context, the word 好奇心 [curiosity] forms links with the three constituents respectively.

Constituent repetition may be regarded as a special device which is often used by writers to organize their argumentative text. Figure 4.4, which illustrates the semantic structure of the NCEE Text 14 好奇心 [Curiosity], further illustrates how this device works.

There are also idiomatic expressions in the sample texts which often form lexical links with their constituents. Although these links do not play a similar role of organizing the text to the one discussed above, they do differ greatly from complex lexical repetition.

They are therefore categorized under constituent repetition. Consider the following example:

Example 39 (taken from the NCEE Text 3):

(5) 包拯，包青天是妇孺皆知的清官，其为官之清不在于其断案是否详尽周密，明查秋毫，而在于其得到真相后敢杀恶人，六亲不认，敢有大义灭亲之举，这也是我最欣赏的地方。 [Bao Zheng is a well-known upright and honest official. His uprightness and honesty do not lie in his ability to carry out highly detailed and careful investigations, but in having the courage to execute evil ones, to turn his back on his flesh and blood, and to sacrifice his blood ties to uphold the uprightness, which is also the virtue that I admire the most.]

(11) 勿以亲疏定善恶，它还包括不要认为是亲人，就可以姑息养奸，亲人的手一样可以将亲人引上断头台，更要让亲人分清善恶，惩恶扬善，不时之时更要大义灭亲。 [Never judge the good or evil by the relationship. It also means that one should not nurse a viper in his bosom just because the offender is one's family members, and one need to realize that even his family

members could lead him to the gallows. Therefore, one should teach their beloved ones to distinguish the good and the evil and to promote justice. Sometimes one even has to sacrifice his own blood ties to uphold the justice.]

Figure 4.3 *The heuristic flow chart for constituent repetition*

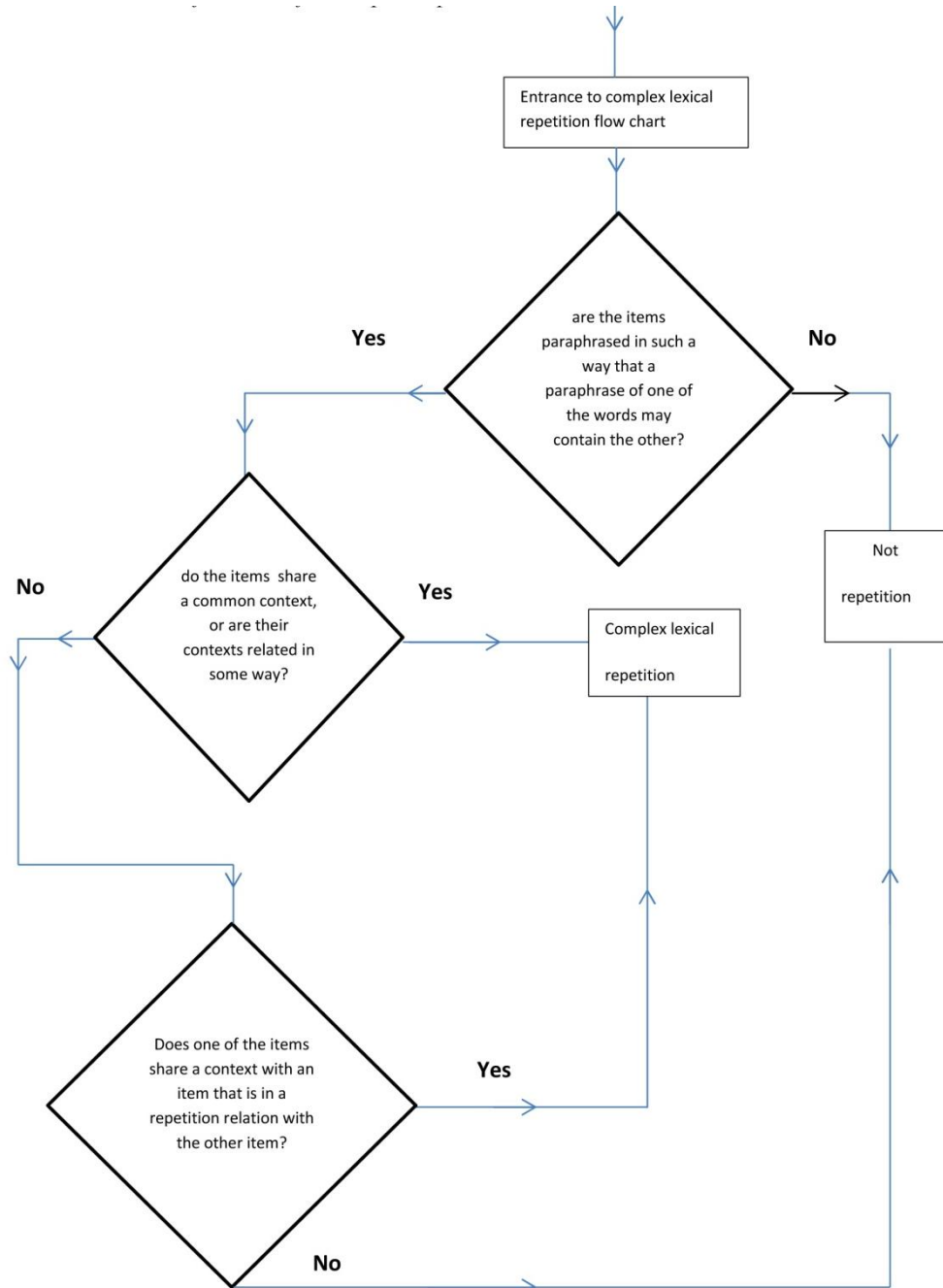
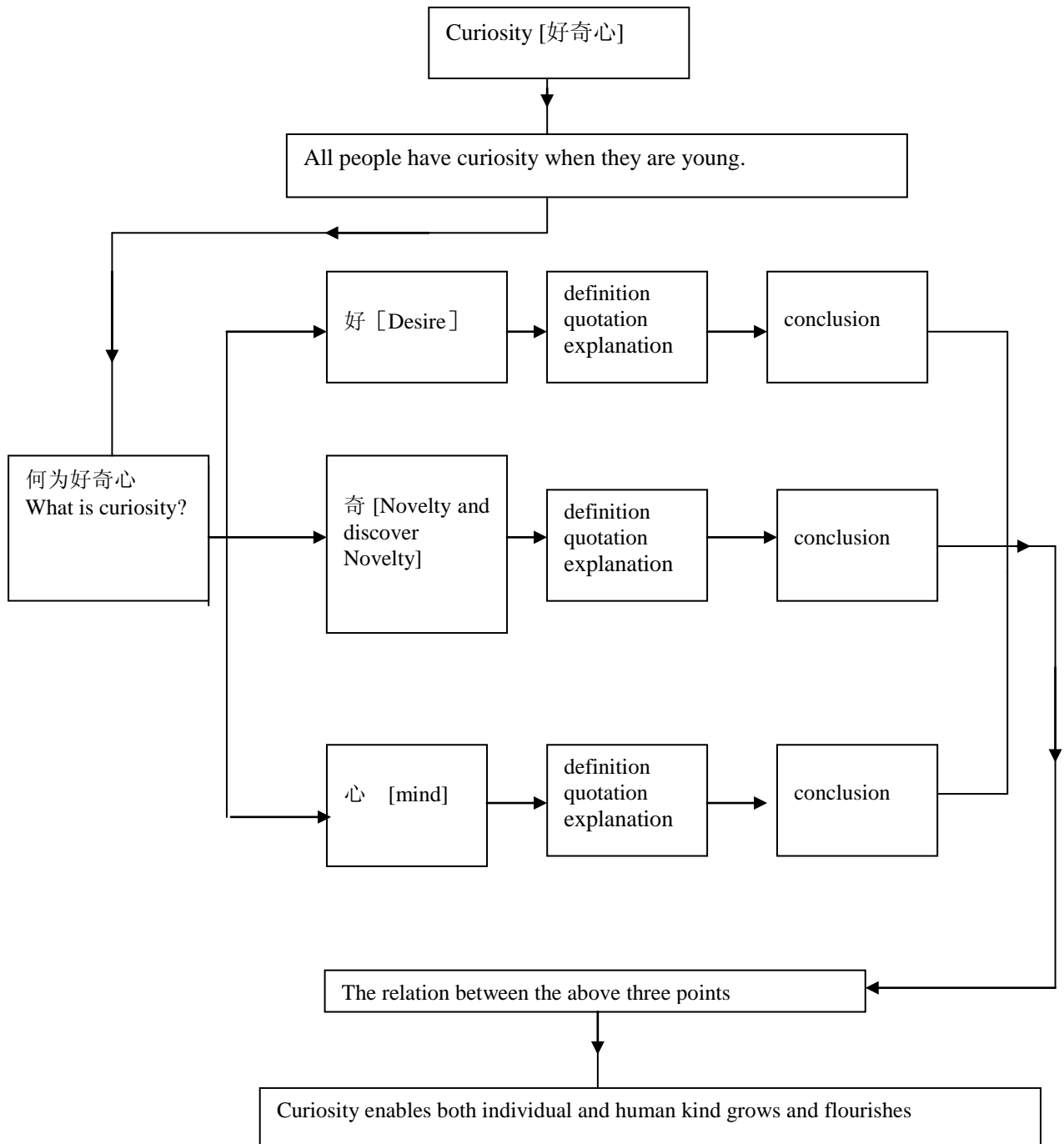


Figure 4.4 The argumentative structure of the NCEE Text 14 好奇心 [Curiosity]



In Example 39, there are two idiomatic chunks 六亲不认 [turn one's back on his flesh and blood] and 大义灭亲 [sacrifice one's blood ties to uphold the uprightness]. As a constituent morpheme of the two fixed expressions, 亲 literally means one's relatives, family members, and is equivalent to 亲人 [one's flesh and blood] in Modern Chinese. Although in some contexts, it can be used in a general sense to refer to one's friends and connections, it refers to one's flesh and blood in the context of this particular text. This constituent morpheme 亲 therefore forms a tie with the item 亲人 [relatives, family members, one's flesh and blood etc] in the example of Sentence 11. Therein lies a problem. At the surface level, it is the constituent morpheme 亲 [one's flesh and blood] that forms a tie with the word 亲人 [one's flesh and blood]. However, as defined in the methodology chapter, the basic analytical unit in this study is established at the level of words and not morphemes, therefore the connection between 亲 [one's flesh and blood] and 亲人 [one's flesh and blood] cannot be regarded as an example of complex lexical repetition. As idiomatic chunks like 六亲不认 [turn one's back on his flesh and blood] and 大义灭亲 [sacrifice one's blood ties to uphold the uprightness], are fixed expressions, and have grammatical functions similar to word in sentences, they therefore can form links with 亲人 [one's flesh and blood], because their paraphrases includes the latter.

It is clear that the difference between 科学 [science] and 科学家 [scientist] differs somewhat from that between 六亲不认 [turn one's back on his flesh and blood] and 亲人 [one's flesh and blood]. 科学 [science] and 科学家 [scientist] differ from the suffix-like element 家 [-ist] being added to the word 科学 [science], which makes the latter change its meaning. Other examples like: 妒忌 [jealous]-妒忌心 [jealousy], 同情 [sympathize]-同情心 [sympathy]. The difference between 大义灭亲 [sacrifice one's blood ties to uphold the uprightness] and 亲人 [one's flesh and blood] occurs because the latter is equivalent to the morpheme 亲 in the idiom, and which is only one of the constituents of the fixed expression.

4.3.4 Simple paraphrase

Simple paraphrase occurs when a lexical item can be substituted by another in context without loss or gain in specificity, and with no discernible change in meaning. The chart in Figure 4.5 shows how to identify whether two items are in the relation of simple paraphrase and complex paraphrase. In the following example 胜利 [victory] in Sentence 11 and 成功 [success] in Sentence 13 can be substituted with each other, and the meaning of the sentences they are in will remain the same.

Example 40 (taken from the NCEE Text 17):

(11) 狄更斯和爱迪生就是靠坚持而取得最后的胜利的。 [It is the perseverance that made Dickens and Edison obtain their eventual victory.]

(12) 坚持, 使狄更斯为人们留下许多优秀著作, 也为世界文学宝库增添了许多精品; 坚持, 使爱迪生攻克了许许多多的难关, 为人类的进步作出了不可磨灭的贡献。

[Perseverance had let Dickens leave many excellent works for people and added many master pieces to the world literature treasure. Perseverance, had helped Edison overcome numerous barriers and difficulties and make indelible contribution to the progress of man.]

(13) 可见, 坚持能够使人取得事业和学业上的成功。 [It is clear that, perseverance can help people succeed in their career and study.]

Hoey (2000) proposes that a simple paraphrase may be partial or mutual. A partial paraphrase means the substitution of item works in one direction only, while mutual paraphrase means the two items are interchangeable. The above example can be regarded as mutual paraphrase, while in the following example the words 克服 [to overcome] in Sentence 5 and 摒除 [to discard or dismiss] in Sentence 16 are in a partial paraphrase relation. In the following context, it seems natural to replace 摒除 by 克服, but it seems less acceptable to substitute 克服 for 摒除.

Example 41 (taken from the NCEE Text 26):

(5) 人们首先得克服感情上的亲疏远近才能客观地认识事物。 [People first have to overcome emotional closeness then they can objectively perceive things.]

(16) 让我们摒除感情的干扰，理智公正地在世为人吧。 [Let us dismiss emotional disturbance and live in the world rationally and impartially]

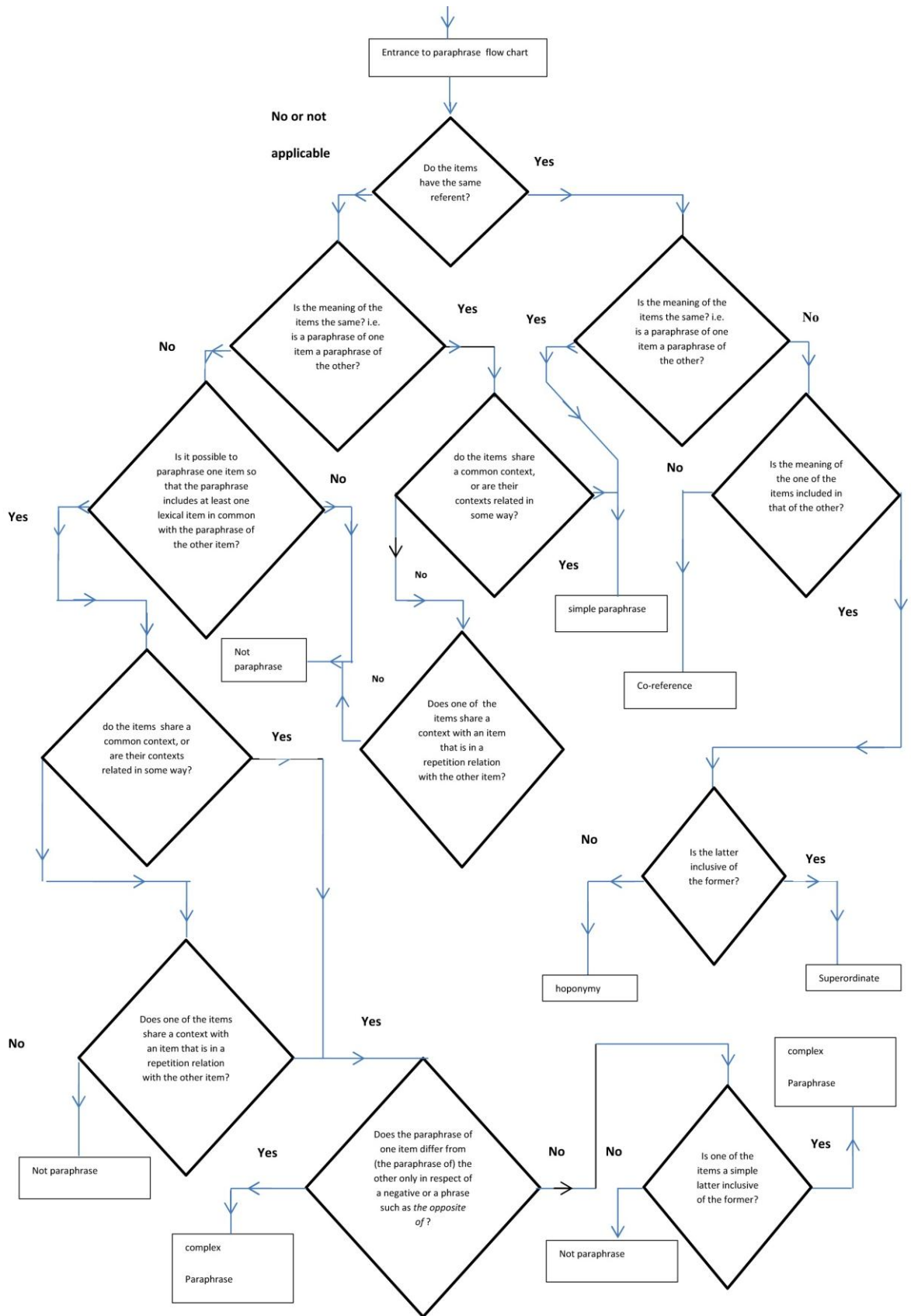
4.3.5 Complex paraphrase

According to Hoey (2000), complex paraphrase, in a broad sense, occurs “when two items are definable such that one of the items includes the other, although they share no lexical morpheme” (p. 64). This definition, however, as Hoey (2000) claimed, will relate a large number of lexical items in ways that are very difficult to control. For example in English, *sickness* and *doctor*, *carol* and *Christmas*, may be said that have a relation of complex paraphrase, as *doctor* may be crudely defined as *people who treat people’s sickness*, which includes the word *sickness*; *carol* may be defined as *a Christmas song or hymn*, which includes the word *Christmas*. For this reason, Hoey therefore restricted the application of complex paraphrase to three situations only.

The first type of complex paraphrase is antonymy. Examples of antonymy are like 浮躁 [impulsive] and 安静 [peaceful] in text 好奇心 [curiosity], 感情 [emotion] and 理智 [reason] in the NCEE Text 25 莫把情云遮慧眼 [don’t let your emotion cover your eyes].

It is worth noting that in his discussion of antonymy, Hoey (2000) points out that in English, the majority of antonyms, for example, *happy*, *unhappy*, *audible* and *inaudible*, can be categorized under complex repetition as they usually share a morpheme. However, this is not the case in Chinese. According to a rough statistic, it seems that the vast majority of antonyms in Chinese should be classified as complex paraphrases, as they do not usually share morphemes.

Figure 4.5 The heuristic flow chart for paraphrase



The second situation in which a complex paraphrase may be identified occurs when there is a linking triangle existing among three lexical items. For example, in the NCEE Text 27 勿以好恶论断之[don't judge by one's likes and dislikes], 掺入 [to mix) is a complex repetition of 掺杂 [to mingle], and also an antonym of 游离 [to separate -from], in such circumstances, it is reasonable to acknowledge a complex paraphrase link between 掺杂 and 游离。 Other examples are like:

认识 (verb)[to understand] – 论断 [to judge] –探求 [to explore]; 感情 [emotion]-情感 [feeling] – 理智[reason] in the sample text 莫把情云遮慧眼 [don't let your emotion cover your eyes].

4.3.6 Superordinate, hyponymic, and co-reference repetition

Superordinate and hyponymic repetition refers to a situation in which a more general word follows an earlier item in the context and both of them contain the same information within the context of a text. Consider the following example:

Example 42 (taken from the NCEE Text 32):

(6) 物理史上，人们在理解光的本质时，由于著名物理学家牛顿坚持微粒学说，光波学说其时已经得到验证，但牛顿在科学界的影响很大，人们宁愿相信牛顿，也不愿相信真理，使光波学说埋没了一百多年，直到在实验室里面重新被演示以后，人们才慢慢相信，接受光波学说。 [In the history of physics, when it comes to understanding the essence of the light, while the Wave theory of the light had been justified, people would rather believe the well-known physicist Newton than believe the truth because Newton preferred the Particle theory of the light. As a result, the Wave theory of the light was neglected for over 100 years. It was not until the experiments which demonstrated the theory anew in the lab that people slowly began to believe and accept the theory.]

(7) 在科学史上，人们宁愿相信著名科学家，而不去相信真理，这一切都是因为人们感情用事，认为名人不会错，孰不知名人也会有失误乃至错误的时候。[In the history of science, the reason why people would rather believe well-known scientist than truth is that people often act impetuously. They believe that people with great reputation will never go wrong. What they don't know is that even well-known people may make mistakes sometimes.]

In the above example, Sentence 7 is obviously used to summarize Sentence 6, and is a further development of the writer's argument-that people often act impetuously. Therefore, the item 科学史 (the history of science) in Sentence 7 supplies no more information than the word 物理史 (the history of physics) contains within this particular context. Likewise, there is no increase in meaning in the shift from the word 物理学家 (physicist) in Sentence 6 to 科学家 (scientist) in Sentence 7. Here the superordinates are 科学史 (the history of science) and 科学家 (scientist), and the related hyponyms are 物理史 (the history of physics) and 物理学家 (physicist).

Below is another example of this type of repetition.

Example 43 (taken from the NCEE Text 20):

(11) 真正的好奇心需要观察。[The real curiosity lies in people's being observant.]

(13) 用心观察，即使最寻常的事物也会化为“神奇”。[If one observes attentively, he will find that even the most usually things will become marvelous.]

(14) 在远古时代，先人们仰望星空，探斗转星移的奥秘，品云淡云聚的风采；到现在，我们通过望远镜与另一个星球招手，通过显微镜感知微观宇宙。[In the old days, our ancestors looked up the sky to explore the myth of the changing seasons and to enjoy the magnificent clouds; nowadays, people explore other planets in the space using telescope and inquire the micro world with microscope.]

(15) 可以说，没有观察与帮助我们更好观察的工具，好奇心也会闭上心房上的窗，落满灰尘。 [We can say that, without observation and the instruments which enable us to observe things better, curiosity will hide itself and have no use.]

The central point of this paragraph is in Sentence 11. Sentence 13 is the argument that supports the point and the Sentence 14 provides an example of the point. Sentence 15 on the other hand contrasts with Sentences 13 and 14, reinforcing the point from the opposite side. Clearly, Sentence 15 is based upon Sentences 13 and 14, and closely related to Sentence 14. Within this context, the superordinate word 工具 (instrument) in Sentence 15 forms two repetition links with the hyponymic words 望远镜 (telescope) and 显微镜 (microscope) respectively.

In relation to the identification of this type of repetition, it is important to note the appearance order of the superordinate and hyponym. To form a repetition of superordinate and hyponym, a hyponym must come first in the text and the superordinate follows. In this way, within the immediate context, the later item contains the same information as the earlier. However, if they are in reverse order, then there is an increase in the specificity of the meaning in a shift from a superordinate to a hyponym. The later item therefore cannot serve the function of repeating. To illustrate the principle, Hoey (2000) gives two examples. The first is that to imagine a text with the words *man* and *woman* in the first sentence, he argues that if hyponyms were regarded as repetitions, then each and every appearance of human beings that followed, no matter by name, relation, or profession, would be instances of superordinate-hyponymic relation as long as they share the immediate context. Likewise, if the word *thing* occurs early in a text, then all the objects that subsequently appear in the text would be regarded as repetitions as long as they share the same context (p. 69). These two examples exemplified that the controlling requirement for the identification of the superordinate-hyponymic relation is the later item must contain the same information as the earlier, and why instances of a hyponym following a superordinate within a shared context cannot treated as repetition.

Co-reference

In a shared context, if two items are interpreted as having identical referents, they then will be treated as co-reference repetition. In the NCEE Text 37 勿以亲疏定是非, 诸葛亮 (Zhuge Liang) and 孔明 (Kongming), 诸葛亮 (Zhuge Liang) and 仇人 (enemy) refer to the same person, they are therefore classified as co-reference repetition. Consider the following example

Example 44 (taken from the NCEE Text 50):

(28) 晋代历史学家陈寿的父亲因犯错误, 受到诸葛亮的严惩而郁闷致死。
[In the Jin dynasty, the father of the historian Chen Shou died of depression because of being severely punished by Zhuge Liang for his mistakes.]

(29) 然而陈寿并未因为孔明是仇人, 就徇“情”忘“理”, 歪曲历史,
[However, Chen Shou was not swayed by his own emotion and did not distort history because that Zhuge Liang was his enemy.]

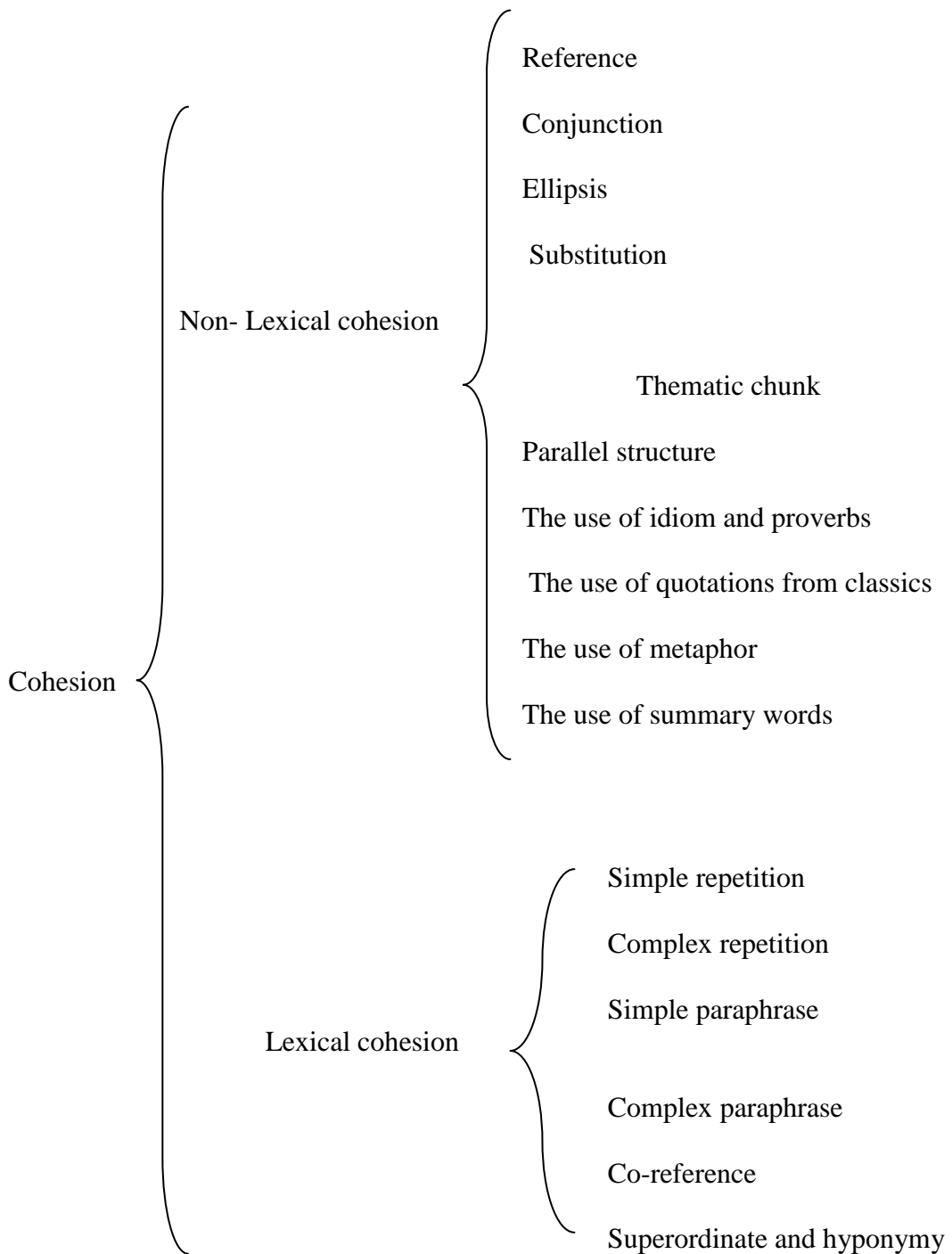
(32) 可他竟能抛开个人的爱恶, 如此“歌颂仇人”, 大讲实事求是, 也绝非“认知”出现了毛病。[However, that he could unexpectedly lay aside his own personal feelings, praise highly his enemy, and seek the truth from the historical facts, is definitely not something wrong with his ability to understand the world.]

Here in Sentence 28, the writer mentioned a historic name 诸葛亮, and then in Sentence 29 the word is used 孔明 instead to refer to the same person. This is because in historical China, people use to have two names, one is called 名 (ming)--how people call them, one is called 字 (zi)-- name given by birth. So in this case, 诸葛 (Zhuge) is a surname, 亮 (Liang) is a 名 (ming), while 孔明 Kongming is his 字 (zi). So the two items are used to refer to the same person. They are therefore co-referential. The other pair of co-referential items is 诸葛亮 (Zhuge Liang) in Sentence 28 and 仇人 (enemy) in Sentence 32, as they refer to the same people as well.

4.4 The overall view of the cohesive devices found in the NCEE Corpus

The following Figure 4.6 summarizes the different types of cohesive devices found in the sample of ten texts taken from the NCEE Corpus:

Figure 4.6 *Cohesive devices found in the sample of ten texts taken from the NCEE Corpus*



4.5 The corpus statistical analysis

The previous two sections are devoted to categorizing the various types of cohesion, so that each type of cohesion in texts can be identified more precisely. As the purpose of this study is to explore the role that lexical cohesion plays in the organization of non-narrative text, this section, therefore, only focuses on the statistical analysis of the use of lexical cohesive devices in the texts written by Chinese native speakers. Although the non-lexical devices discussed in Section 4.2 clearly show their contribution to creating the organization of a text, and need further investigation in order to complete the picture of textual organization, this work falls out of the scope of this research, and requires separate research.

On the basis of the categories of lexical cohesion defined in Section 4.3, the 50 texts of the NCEE Corpus were tagged using the tagging system described in Chapter 3. In order to explore the way lexical cohesion organizes a text, the following statistics of the use of the lexical cohesion were investigated using the concordancing programme NoteTab.

Table 4.1 *Various types of statistical data to be extracted from the NCEE Corpus*

Code	In each individual text	Code	In the 50 texts
A1	the frequency of each type of repetition used (per 100 Chinese characters)	A2	the average frequency of each type of repetition used (per 100 Chinese characters)
B1	bond density (the number of bonds formed per 100 Chinese characters in one text)	B2	the average bond density (the number bonds formed per 100 Chinese characters in 50 text)
C1	the ratio of the bonds formed between non-adjacent sentences to that between adjacent sentences	C2	the average ratio of the bonds formed between non-adjacent sentences to that between adjacent sentences

D1	the ratio of sentences that form bonds with other sentences to those form no bonds with others	D2	the average ratio of sentences that form bonds with other sentences to those form no bonds with others
E1	the average number of words that form repetition with other words in one sentence	E2	the average number of words that form repetition with other words in one sentence

4.5.1 Statistics for the use of various types of lexical cohesion in the NCEE Corpus

The purpose for carrying out such an investigation and the interpretation of each group of information are presented immediately after the presentation of the statistical figures. Table 4.1 lists the various types of statistical data to be extracted from the NCEE Corpus

In his discussion of the use of repetition in different languages, MacCarthy (1991) quoted from Hinds' (1979) observation that Japanese tend to use direct repetition in their conversations (p. 71). Zhang's (2005) comparison of the use of different types of cohesion in English and Chinese news articles also indicates that Chinese seems to use more simple repetition than that English. These discussions imply that there may be a variation of the preferences of the use of different types of lexical cohesion among different language user groups. By using the information provided by the A1 and A2 statistics (Table 4.1), this section aims to describe the distribution of various types of lexical cohesion in sample texts written by Chinese native speakers, and on the basis of which, to establish a comparing ground for the comparison of the distribution of the use of these lexical repetition in texts written by Chinese native speakers and by learners of Chinese.

The first step of the analysis is to obtain the required information from the NCEE Corpus and to present it (as demonstrated in Table 3.4 in Chapter 3 showing a sample repetition matrix). For the convenience of discussion and reference,

another similar example of the data obtained from the sample text 能思考的苇草 (The reed that can think) is given as Table 4.2:

Table 4.2 *The various types of lexical cohesion and their frequency per 100 characters in the NCEE Text 1*

The title of sample text: 能思考的苇草 [The reed that can think]		
Total number of characters: 688		
Name of lexical cohesive device	Number of the repetition items used	Frequency (per 100 characters)
Simple lexical repetition (type1 sr)	50	$50 \times 100 / 688 = 7.27$
Complex lexical repetition (type2 cr)	8	$8 \times 100 / 688 = 1.16$
Constituent repetition (type3 constituent re)	0	0
Simple paraphrase (type4 sp)	24	$24 \times 100 / 688 = 3.49$
Complex paraphrase (type5 cp)	36	$36 \times 100 / 688 = 5.23$
Superordinate and hyponym (type6 super-hypo)	12	$12 \times 100 / 688 = 1.74$
Co-reference (type7 co-re)	4	$4 \times 100 / 688 = 0.58$

The figures in Table 4.2 clearly show the actual use of the different types of lexical cohesion in this particular text. This makes it possible to identify the variety of means the writer used to connect sentences and his or her preferences for using different types of cohesion. Since we are interested in exploring the possible lexical patterns that are largely common in all the sample texts written by Chinese native speakers, such individual information obtained from a single piece of writing obviously does not serve any functional purpose of the study. Hence, it is important to go further to collect similar type of information from the 50 sample texts and, on the basis of the findings from the analysis of this sample, to obtain the average value of the frequency of each type of repetition used in 50 sample texts by using the following formula:

$$AF_x = \frac{FT1x + FT2x + \dots + FT49x + FT50x}{50}$$

In the formula, AF means the average frequency, x stands for a certain type of repetition, F stands for frequency, T means sample text, and the number 50 in the denominator position indicates the total number of the sample text used in this corpus. For example, the following formula is used to calculate the average value of the frequency of the simple repetition in 50 sample texts:

$$AF_{sr} = \frac{FT1sr + FT2sr + \dots + FT49sr + FT50sr}{50}$$

In the formula, the AF_{sr} stands for the average value of the frequency of the simple repetition, FT1_{sr} means the frequency of the use of simple repetition in Sample Text 1, and FT2_{sr} means the frequency of the use of simple repetition in Sample Text 2, and so on. By using this formula, it is possible to work out the average frequency of the simple repetition in 50 sample texts. Likewise, the following formula is used to work out the average frequency of the complex repetition in 50 sample texts:

$$AF_{cr} = \frac{FT1cr + FT2cr + \dots + FT49cr + FT50cr}{50}$$

By following the similar formula, the average frequencies of the seven lexical cohesive devices were obtained and listed in Table 4.3 respectively.

From Table 4.3, we can see that the average frequency of simple repetition is the highest among the seven categories, which is almost double the number for complex paraphrase. The frequency of simple paraphrase is 2.75, which is slightly lower than that of complex paraphrase. In comparison, the frequency of superordinate-hyponymic, complex repetition, constituent repetition, and co-reference are markedly low, with co-reference is only at 0.11, the lowest among all the categories.

As outlined at the beginning of this section, this information establishes a basis for comparison of the use of various types of lexical cohesion in texts written by Chinese native speakers and by learners of Chinese. In Chapter 6, similar type of information is listed to show the average frequencies of the seven types of cohesion in texts written by learners of Chinese. A detailed comparison and discussion were then presented in Chapters 7 and 8.

Table 4.3 *The average frequencies of the seven types of cohesion in the 50 NCEE texts*

Category	Average Frequency	Example
simple repetition	6.22	真理 (truth)- 真理 (truth)
complex repetition	0.47	科学(science) - 科学家 (scientist)
constituent repetition	0.30	好奇心(curiosity)- 奇(novelty)
simple paraphrase	2.75	成功(success) -胜利(victory)
complex paraphrase	3.19	成功(success)-失败(failure)
superordinate-hyponymic	0.90	工具 (instrument)-显微镜 (microscope)
co-reference	0.11	诸葛亮(Zhuge Liang)- 孔明 (kongming)

4.5.2 Statistics for the number of bonds formed in the NCEE texts

The next group of statistics under investigation is those referred to as B1 and B2 in Table 4.1. Statistics B1 indicates the closeness of connection between bonded sentences and B2 shows the density of repetitions in a text. In Chapter 3, the notion of bond is briefly discussed to introduce the research methodology for this study. It is, however, necessary to make a few more points about the nature and the properties of the bond formed between sentences, so that the significance of carrying out such statistics analysis is justified.

Firstly, as Hoey (2000) argues, bonded pairs are always semantically related and make sense together (p. 162). Consequently, sentences connected by bonds are normally coherent (or virtually coherent) when placed together (p. 133). Using Winter's (1974) concept of systematic repetition and significant replacement, Hoey further explains this coherence effect created by the lexical bonds formed between sentences. When a pair of sentences is connected by three repetitions (the number three is relative, varying from text to text), it can be said that the pair of

sentences is bonded. The function of these repetitions is to provide a context or framework for the writer to offer something new in the later sentence, and the new information, which replaces something from the earlier sentences, is to maintain the lexical uniqueness of the later sentence. The analysis of a pair of bonded sentences is, therefore, a process of recognizing the relationship between repetition (old information) and replacement (new information). It is in this way that lexical repetition provides sufficient information for readers to recognize bonded pairs as intelligible and supplies a context or framework for writers to create parallel sentences to be interpretable.

Because of this nature of the bond, the number of bonds formed in a text can to a great extent reflect the degree of closeness of connection between sentences. The more a sentence bonds with other sentences, the more it builds lexically upon what has gone before and provides the lexis for subsequent statements. This degree of closeness of connection can be quantified and measured by means of the notion of the *bond density*. This can be calculated by using a formula similar to the one used to work out the frequency of the use of certain type of repetition.

$$\text{bond density} = \frac{\text{the number of bonds} \times 100}{\text{the total number of characters in the text}}$$

By comparing the number of bonds formed per 100 characters, it is hoped to measure the extent to which the sentences in a text are cohesive and coherent. A smaller bond density may indicate that sentences in a text are less closely related, while a greater density may reflect that sentences in a text are more closely connected to each other. One may argue that the number of bonds is not necessarily proportional to the degree of closeness of connection between sentences in some texts. For example, this may be possible if one invents a text with a great number of cohesive bonds yet the number does not produce a high degree of coherence in the text. In response to this argument, Hoey (2000) acknowledges that “it would not be difficult to invent texts where the presence of three shared items guaranteed no such closeness”. However, he argues that “in authentic text we find that even the pairs which are not coherent are normally close related” (p. 134). Table 4.4 provides an overall picture of the distribution of

the bond density in the 50 texts of the NCEE Corpus.

Table 4.4 *The bond density and the average bond density in the NCEE Corpus*

	Text Title	Total number of bonds	Total number of characters	B1	B2
1	能思考的苇草	36	688	5.23	3.13
2	情感理性认知	47	968	4.86	
3	勿以亲疏定善恶	19	725	2.62	
4	情感的权力	23	885	2.60	
5	诚信抉择	18	827	2.18	
6	偶然	15	725	2.07	
7	珍惜拥有和懂得 舍弃拥有的方式	15	1029	1.46	
8	让理性主持宣判	12	850	1.41	
9	兼听则明	8	699	1.144	
10	知之思之行之	7	917	0.76	
11	缤纷成就和谐	25	805	3.10	
12	谈意气	21	991	2.10	
13	让纪念闪耀理性光芒	33	846	3.90	
14	好奇心 1	29	776	3.73	
15	没有付出哪有累累硕果	35	778	4.498	
16	诗意地生活	24	905	2.65	
17	坚持就是胜利	29	775	3.74	
18	铭记八荣八耻	35	1096	3.20	
19	好奇心 2	39	940	4.15	
20	好奇心 3	16	797	2.00	
21	好奇心 4	29	740	3.91	
22	清欢	42	853	4.92	
23	不要轻易说“不”1	13	799	1.62	
24	不要轻易说“不”2	25	825	3.03	
25	莫把情云遮慧眼	36	835	4.31	
26	公则生明	25	814	3.07	
27	勿以好恶论断之	14	710	1.97	
28	感情不是保险绳	46	830	5.54	
29	天平和七弦琴	12	860	1.40	
30	切莫走近	38	711	5.34	
31	不要让眼光带上感情	29	762	3.81	
32	切莫感情用事	27	750	3.60	
33	呼唤理智	36	735	4.90	
34	亲戚与明理	19	808	2.35	
35	调准感情集距	21	702	2.99	

	拥有清晰世界			
36	做一回理性自己	13	840	1.55
37	勿以亲疏定是非	22	777	2.83
38	莫让浮云遮望眼	37	933	3.97
39	独一无二	39	884	4.41
40	生命的意义	31	919	3.37
41	小杯子大道理	10	958	1.04
42	杯中窥人生	29	811	3.58
43	沉默的大多数	17	818	2.08
44	生命的痕迹	28	786	3.56
45	遭遇挫折，笑对痛苦	12	835	1.44
46	减负中的“朝三暮四”	43	982	4.38
47	变味的善良	36	949	3.79
48	角色转换之间	41	825	4.97
49	站对人生的舞台	19	847	2.24
50	怎一个“情”字了得	38	1211	3.14

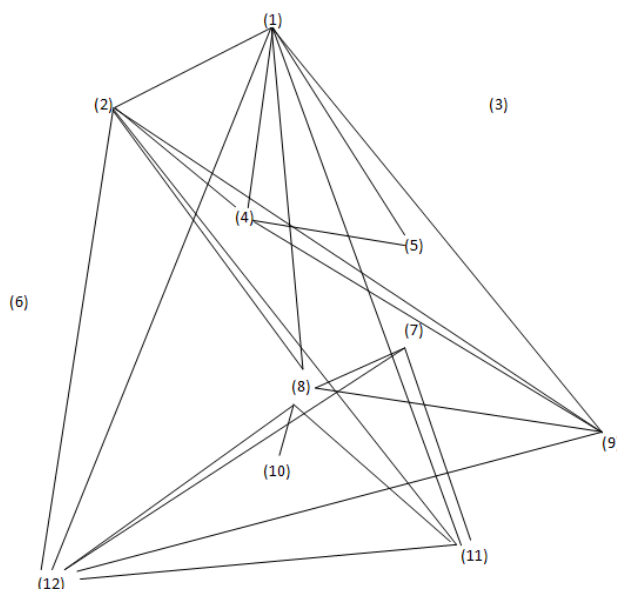
4.5.3 The numbers of bonds formed between non-adjacent sentences to that between adjacent sentences

The ratio of the bonds formed between non-adjacent sentences to adjacent sentences in a text can also reflect the degree of cohesion and coherence. In the light of the notion of repetition and replacement pioneered by Winter (1974) and the approach to the analysis of the lexical repetition adopted by Hoey (2000), it can be said that repetition in a text is an indicator of mutual relevance between sentences. By relevance, here it means that the extent to which the sentences are semantically and lexically related to each other. The more the bonds formed between non-adjacent sentences, the more relevant the sentences are and the higher the degree of closeness of connection between sentences. Some studies, such as that of Skuja (1984) and Mountain (1987), suggest that the scope that different cohesive devices cover in native speakers' writing tend to be broader than that in non-native speakers' writing. This implies that bonding between non-adjacent sentences is a measure of mutual connection between sentences in a text. In his discussion on this matter, Hoey (2000) uses the term *a non-linear way of writing* to describe the way in which the non-adjacent sentences are relate to each by lexical repetition in writing. He claims that “one of the characteristics of

mature native speaker writing is that the writer's sentences are related to each other in non-linear ways" (p. 242). The ratio of the bonds formed between non-adjacent sentences to adjacent sentences in a text is an important index, which can reveal some of the features of the closeness of connection between sentences.

In Chapter 3, Figure 3.3 demonstrates the web-like interconnection of bonded sentences in a text. For the purpose of justification and convenience of explanation, it is presented again as Figure 4.7 to show the interconnection formed between non-adjacent sentences in text.

Figure 4.7 *The illustration of the interconnection of bonded sentences in the NCEE Text 4*



In this figure, there are bonds that are formed between adjacent sentences, such as 1-2, 4-5, 7-8-9, and 11-12. However, a greater number of bonds are formed between non-adjacent sentences, such as 1-4, 1-5, 1-8, 1-9, 1-12, 2-4, 2-8, 2-9, 2-12, 4-9, 7-11, 7-12, 8-11, 8-12, and 9-12. Clearly, it is the bonds formed between non-adjacent sentences that play an important role in the organization of the text, and contribute to the mutual relevance of the sentences, and consequently make the text intelligible to readers. Table 4.5 below shows the actual ratio of the number of the bonds formed between non-adjacent sentences to that between adjacent sentences in 50 texts in the NCEE Corpus. For the convenience of their

presentation in the table, the former is abbreviated as BNA and the latter BA.

Table 4.5 *The ratio of bonds formed between non-adjacent sentences (BNA) to that between adjacent sentences (BA) in the NCEE Corpus*

	Text Title	The number of BNA	The number of BA	C1	C2
1	能思考的苇草	29	7	3:1	3.4:1
2	情感理性认知	41	6	41:6	
3	勿以亲疏定善恶	15	4	15:4	
4	情感的权力	18	5	18:5	
5	诚信抉择	13	5	13:5	
6	偶然	12	3	4:1	
7	珍惜拥有和懂得舍弃拥有的方式	12	3	4:1	
8	让理性主持宣判	11	1	11:1	
9	兼听则明	5	3	5:3	
10	知之思之行之	5	2	5:2	
11	缤纷成就和谐	20	5	4:1	
12	谈意气	15	6	15:6	
13	让纪念闪耀理性光芒	29	4	29:4	
14	好奇心 1	26	3	26:3	
15	没有付出哪有累累硕果	30	5	6:1	
16	诗意地生活	15	9	15:9	
17	坚持就是胜利	20	9	20:9	
18	铭记八荣八耻	34	2	17:1	
19	好奇心 2	33	6	33:6	
20	好奇心 3	9	7	9:7	
21	好奇心 4	21	8	21:8	
22	清欢	32	10	3.2:1	
23	不要轻易说“不”1	9	4	9:4	
24	不要轻易说“不”2	21	4	21:4	
25	莫把情云遮慧眼	33	3	11:1	
26	公则生明	20	5	4:1	
27	勿以好恶论断之	11	3	11:3	
28	感情不是保险绳	36	10	3.6:1	
29	天平和七弦琴	7	5	7:5	
30	切莫走近	29	9	29:9	
31	不要让眼光带上感情	23	6	23:6	
32	切莫感情用事	20	7	20:7	
33	呼唤理智	32	4	8:1	

34	亲戚与明理	12	7	12:7
35	调准感情集距 拥有清晰世界	16	3	16:3
36	做一回理性自己	8	5	8:5
37	勿以亲疏定是非	16	6	16:6
38	莫让浮云遮望眼	27	10	27:10
39	独一无二	30	9	10:3
40	生命的意义	20	11	20:11
41	小杯子大道理	6	4	3:2
42	杯中窥人生	18	11	18:11
43	沉默的大多数	13	4	13:4
44	生命的痕迹	22	6	22:6
45	遭遇挫折，笑对痛苦	10	2	5:1
46	减负中的“朝三暮四”	31	12	31:12
47	变味的善良	28	8	14:4
48	角色转换之间	32	9	32:9
49	站对人生的舞台	10	9	10:9
50	怎一个“情”字了得	30	8	15:4

4.5.4 The ratio of sentences that form bonds with other sentences to those form no bonds with others

This section reports the ratio of the number of sentences that form bonds with other sentences to those form no bonds with others in each text and the average ratio of bond forming sentences to non-bond forming sentences in 50 texts in the NCEE Corpus. These two statistics are coded as D1 and D2 in Table 4.1.

In Section 4.3.2, some of the properties of the bonds formed between sentences are briefly discussed to justify the significance of these statistics. There is, however, one more aspect of the bond that we need to consider, which is how the bonded sentences contribute to the development of the theme of a text. In his discussion on the function of bonded and unbonded sentences serve in a text, Hoey (2000) uses two terms: *marginal sentences* and *central sentences*. For him, a sentence that forms no bonds with other sentences may be deemed marginal, because it “neither build[s] lexically upon what has gone before nor provides the lexis for subsequent statement” (p. 105). It, therefore, does not “directly contribute to that main theme” (p. 105). The role that it plays in the organization

of the text is to provide “some necessary ancillary information without which the main theme might be open to misunderstanding” (p. 105). Some marginal sentences mainly serve as signaling sentences, which function like sentence-length conjuncts to relate earlier sentences to later sentences, while others may be used to exemplify a point that is not further pursued or to reject a possible topic of discussion. By comparison, sentences that manifest bonds may be regarded as central sentences, because they make connections across the text. In other words, through three or more lexical repetitions, sentences are interconnected between what currently being said and what have been previously said and what is intended to be said later. Consequently, central sentences serve directly to the development of the main theme. Depending on their position in a text, some sentences have a clear role of setting up topics, while others have topic closing-down roles (p. 118). There are also some sentences demonstrating the function of connecting its preceding sentences with its following sentences.

On the basis of the lexical analysis of the texts in the NCEE Corpus, it seems that bonded and unbonded sentences in Chinese text play a similar role in the organization of the text as that Hoey’s claims in English. To exemplify the discussion above, the NCEE Text 4 情感的权力 (from the analysis of which Figure 4.7 is created) is presented here to illustrate how the bonded sentences serve the function of development of main theme and how the unbonded sentences provide the ancillary information for the reader.

Example 45 (taken from the NCEE Text 4):

情感的权力[The power of emotion]

(1) 在我所听闻过的悲剧中，情感犯下的错是占大多数的，因为情感的权力有时太大，大得等不及真相的告白，就将一切画上了句号。[Among all the tragedies that I have heard, the majority of the mistakes were caused by emotion. Because sometimes the power of the emotion is so excessive that it brings about tragedy before it allows the truth to manifest itself.]

(2) 翻开历史，尤其是中国的历史，君王的昏庸，王朝的颓败，忠臣蒙"莫须有之罪"，奸候飞黄腾达的段子一折接一折，这大概就是情感惹下的不小的祸吧。 [In history, especially in Chinese history, because of the incompetent and fatuous emperors and the declining dynasties, numerous faithful court officials were persecuted with fabricated charges, while those treacherous ones received rapid promotion. This is probably the consequences of the abuse of the power of emotion.]

(3) 侯臣们大抵也是科举之人，一试一试考上来的，也大抵有着灵活的头脑和如簧的巧舌，三言两句套套近乎，便成了亲信，成了近宠,他们的话，岂止是"一言千金"，更是"一言系千命"呀。 [The court officials were probably selected through the imperial examination and, therefore must had been very intelligent and persuasive. By always being cotton up to their masters, they became the favourite ones or the henchmen of the emperors. What they said had great impact and had cost many people's lives.]

(4) 君主们的情感被这些叵测之心所操纵，君主们的情感也随即获得了无人企及的权力可悲的权力，惨痛的教训。 [Having been manipulated by these treacherous hearts, the emperors therefore had indulged themselves in abusing their power. What a regrettable abuse of power, and what a harsh lesson to learn.]

(5) 莎翁的悲剧有一出，叫《麦克白》，性质也是一样的，即使妻子身出名门，即使自己相信妻子是忠心不二，然而我们也不得不承认，情感是脆弱的，信任在怀疑和虚荣面前是瑟缩发抖直不起腰的。 [Similarly, the same happened in *Macbeth*, a tragedy written by Shakespeare. Even though Macbeth's wife was from respected upper-class family, and he believed that she was committed to their marriage, as we have to admit that human emotion is vulnerable and delicate, his belief did not bear the test of doubts and vanity.]

(6) 于是亲信的部下有了可乘之机，让麦克白亲手杀妻抱憾而终。 [Consequently, his doubts and vanity played into the hands of his favourite followers, which caused him to kill his wife with his own hands and regretted all his life.]

(7) 情感的权力施展起来往往就是这么残忍而不可理喻，它封闭一切正义的外力作用范围，总是一意孤行，让非"者昌"是"者亡。[When one abuses his power of emotion, he becomes so cruel and unreasonable that he excludes himself from all the righteous might and has his own way, letting those who comply with him thrive and those who resist him perish.]

(8) 谬误总爱偷穿真理华贵高雅的衣裳，卖弄纯洁卖弄博学，所以情感会禁不起诱惑，近谬误而疏真理，情感会操纵当权者的法杖，颠倒黑白，所以说，"感情用事"很多情况下是有害无益的，在情感与正义面前，我们需要让情感"回避"，让情感站远，让真理敢于靠前。[Falsehood always likes disguising itself as truth, pretending having no fault and being knowledgeable. One therefore succumbs easily to the falsehood emotionally, which causes him to uphold the wrong and to depart from the right. It also manipulates those who are in power to overthrow right and wrong. Because of this, being swayed by emotion does more harm than good in most circumstances. Facing emotion and justice, we need to keep ourselves from our emotion and hold on to the truth.]

(9) 当然，历史上，现代社会中也有很多大义灭亲的事例，不会单凭着感情的亲疏胡乱判案模糊执法的。[Nevertheless, in both history and our modern society, there are many instances where people uphold justice and righteousness even at the sacrifice of blood relations, and do not make wild judgment or execute partial enforcement because of their being close with somebody else.]

(10) 古语云:太子犯法与庶民同罪，太子亦然，这是明君之所为，才不愧"光明正大，明镜高悬"八个铮铮大字。[The old saying says, the king must himself be subject to the law. Such behaviour reflects what a wise emperor should do, and therefore is worthy of the spirit expressed by the saying- an honest official hangs his clear mirror aloft- if he judges fairly and above board.]

(11) 而在现今社会，小到班级学校，教师们不因为个人与某些学生的亲疏而错判正误;大到一个国家不因与某些国家的邦交厚薄而任意进行军事政治支援纵容非正义战争的肆意扩大，伤及无辜;这些，都是正确处理情

感与真理的极佳明证，不论在什么时间，什么地点，真理永远是真理，我们不能掩盖或歪曲它，所谓“真金不怕火炼”么！ [In today's society, a simple example of this can be seen from classrooms in schools, where teachers do not act wrongly just because they are in favour of some of their students. A complex example can be taken from the relations between nations, where some countries do not provide military or political assistance randomly to overindulge those nations that have close relation with them, so that the unjust war could not be escalated and the innocent people could not be hurt. These behaviors, therefore, are excellent examples of handling emotion and truth correctly. Truth will never change with the change of time and location, and we cannot conceal and contort it, just as the old saying goes, pure gold fears not the fire.]

(12) 当情感的亲疏开始干涉正义时，当情感的亲疏开始混淆是非时，当惊觉情感的权力超出界限时，我们应该学会反省，保持冷静，“退一步海阔天空”，——“退一步，真相会昭然而现”，站远一点看事情，会更加公正一些。 [When emotion is intervening the justice, when emotion is starting to obscure the distinctions between right and wrong, when emotion is found to cross boundaries, we should learn how to reflect and keep calm, as the old saying goes: one step back for boundless sky. If we step back, the truth will become obvious, if we distance ourselves from the things, we may be able to handle things in a more just and fair way.]

The main theme of the text is the impact the emotion has on the use of power. The text begins with the introduction of the statement that the majority of tragedies are caused by the abuse of emotion. Following this, an example is presented that throughout the history many emperors had manipulated their power with emotion and had consequently caused great hurt to their dynasties. The text then goes on to support the argument with another example, which is taken from Shakespeare's tragedy *Macbeth*. By providing the two examples, one from Chinese history, and the other Western literature, the writer makes the theme explicit that emotion has great impact on one's power. In the next half of the text, the writer presents another statement: when facing emotion and justice, one should hold fast to the

justice and keep himself from the abuse of emotion. Examples of correctly handling the relationship between power and emotion are then given to support this argument.

The text has 12 sentences in total. Out of 12 there are two sentences - 3 and 6- that form no bonds with other sentences, hence they are marginal sentences, while the rest of the sentences are bonded with each other and therefore are regarded as central sentences. According to Figure 4.1, Sentence 3 is clearly a marginal sentence, the function of which is to provide ancillary information. In this case it is to give some background information about those treacherous court officials. If we omit this sentence, the text can still be regarded as coherent. Consider the following sentences:

Example 46 (taken from the NCEE Text 4):

(2) 翻开历史，尤其是中国的历史，君王的昏庸，王朝的颓败，忠臣蒙"莫须有之罪"，奸候飞黄腾达的段子一折接一折，这大概就是情感惹下的不小的祸吧。 [In history, especially in Chinese history, because of the incompetent and fatuous emperors and the declining dynasties, numerous faithful court officials were persecuted with fabricated charges, while those treacherous ones received rapid promotion. This is probably the consequences of the abuse of the power of emotion.]

(4) 君主们的情感被这些叵测之心所操纵，君主们的情感也随即获得了无人企及的权力,可悲的权力，惨痛的教训。 [Having been manipulated by these treacherous hearts, the emperors therefore had indulged themselves in abusing their power emotionally. What a regrettable abuse of power, and what a harsh lesson to learn.]

Omitting Sentence 3 does not affect the coherence of the passage. Sentence 2 introduces the example that in history many faithful court officials were backstabbed by those treacherous ones, and Sentence 4 provides an explanation that this is because the emperors were instigated by those treacherous court

officials to abuse their power emotionally. The only word that needs to be changed to make the passage more coherent is the proximal pronoun 这些 *these*, which is used to refer to the treacherous ones just mentioned in sentence 3. When Sentence 3 is omitted, *these* should be changed to distal pronoun 那些 *those* to refer back to the treacherous ones in history. In comparison, Sentences 1, 2, 4, 5, 7, 8, 9, 10 are central sentences. If we break down Figure 4.7 in to individual sentence connections, we will get the following Figure 4.8.

Figure 4.8 *Connections that Sentence 1 forms with other sentences*



As can be seen from the figure, Sentence 1 connects with Sentences 2, 4, 5, 8, 9, 11, and 12. Clearly, it contributes greatly to the setting up the main theme, and therefore can be regarded as a topic opening sentence. Likewise, Sentence 2 has six connections, with one link back with Sentence 1, and five with subsequent Sentences 4, 8, 9, 11 and 12, as shown in Figure 4.9.

Obviously Sentence 2 forms a connecting link between the preceding and following sentences, and still can be seen as a topic opening sentence. Indeed, in the sample text, Sentence 1 introduces the general statement while Sentence 2 exemplifies the statement; the argument in these two sentences is consistent. Sentences 4 and 5 provide further explanation and support for the argument. They therefore mainly link back with 1 and 2, and have fewer connections with

subsequent sentences. Sentence 4 provides an explanation for Sentence 2; hence it links back with Sentences 1 and 2. While Sentence 5 is an example different from that in sentence 2, it mainly links back with sentence 1. This relation can be seen in Figure 4.10.

Figure 4.9 *Connections that Sentence 2 forms with other sentences*

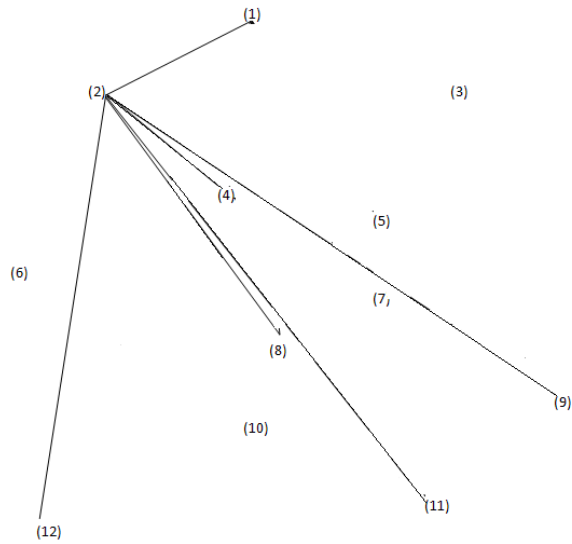
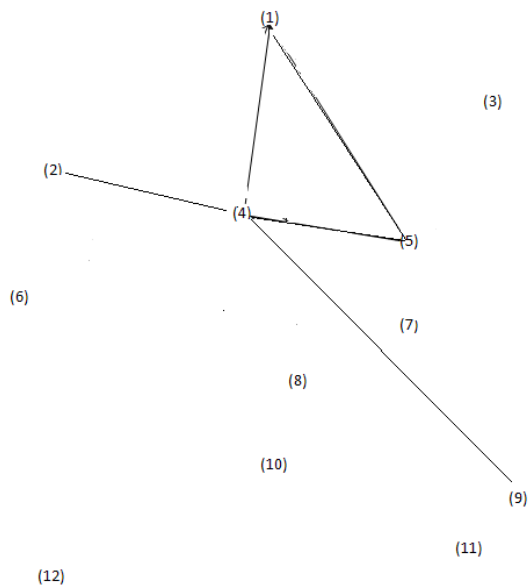
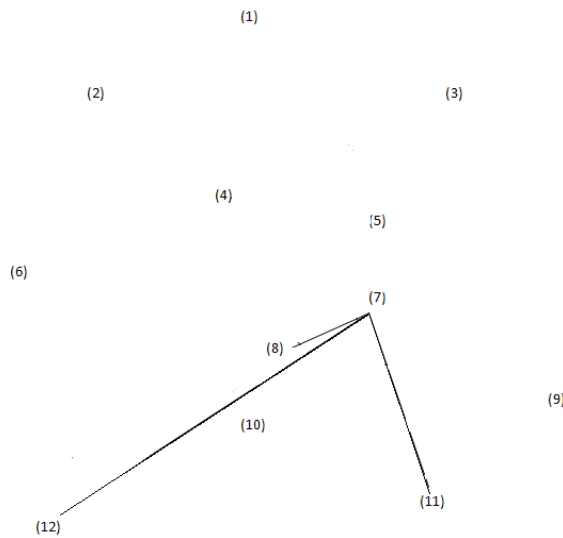


Figure 4.10 *Connections that Sentence 4 and 5 form with other sentences*



In the text, there is a transitional point in the argument, which is introduced by Sentence 7. From Sentences 1 to 6, the major argument is that the majority of tragedies are caused by the abuse of emotion. In Sentence 7, a new point is presented, which is that people should keep themselves from being affected by their emotion, and should uphold the truth. Sentence 7 is, therefore, a new topic opening sentence and mainly links with subsequent Sentences 8, 11 and 12 as shown in Figure 4.11.

Figure 4.11 *Connections that Sentence 7 forms with other sentences*



Similar to Sentence 2, Sentence 8 is also a connecting link between what precedes and what follows. It continues from Sentence 7 and opens up the topic further for discussion. It also links back with Sentences 1 and 2 to support the argument. The connections that it forms can be seen in Figure 4.12.

Sentence 9 also has strong connections with other sentences. First it links back to the general statement in Sentence 1, and then it uses the phrase *in history* to refer back to Sentences 2 and 4. It also has connections with its immediately preceding Sentence 8 and with its subsequent Sentences 12, which are within the scope of the second argument discussion. The connection it forms can be seen from the Figure 4.13.

Figure 4.12 *Connections that Sentence 8 forms with other sentences*

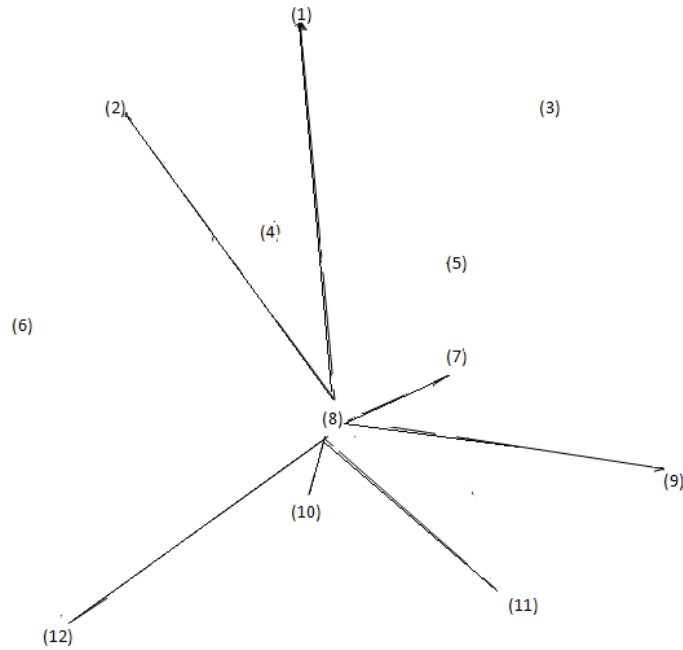
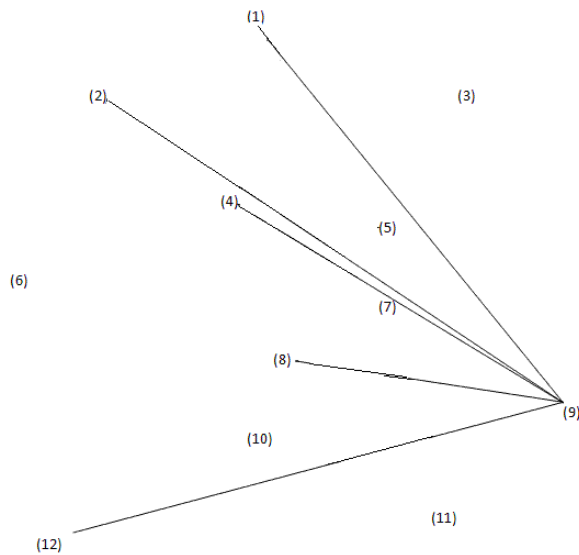


Figure 4.13 *Connections that Sentence 9 forms with other sentences*



Compared with Sentences 1, 2, 4, 5, 7, 8, and 9, Sentence 10 is a relatively weaker central sentence, as it only forms one connection with Sentence 8. Its function is to provide further support for the statement in Sentence 8.

Sentences 11 and 12 function as topic closing sentences, as it can clearly be seen that their connections are right across the text. Sentence 11 links with the topic opening sentences such as 1, 2, 7, and 8, and links with the closing Sentence 12. While Sentence 12 serves as a conclusion sentence, which connects with all the most important functioning sentences in the text: 1, 2, 7, 8, 9, and 11. Figures 4.14 and 4.15 show the connections of the two sentences form with other sentences respectively.

Figure 4.14 *Connections that Sentence 11 forms with other sentences*

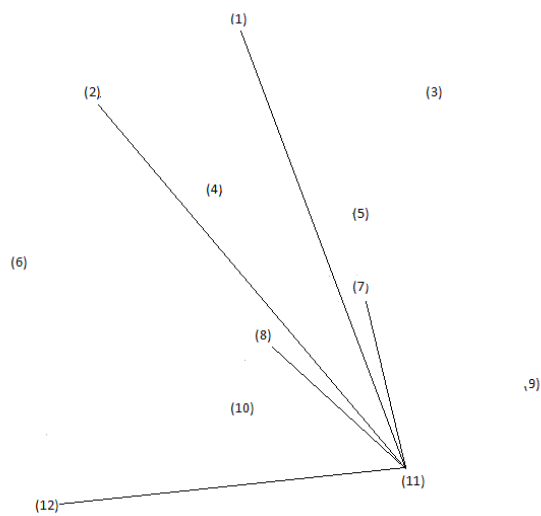
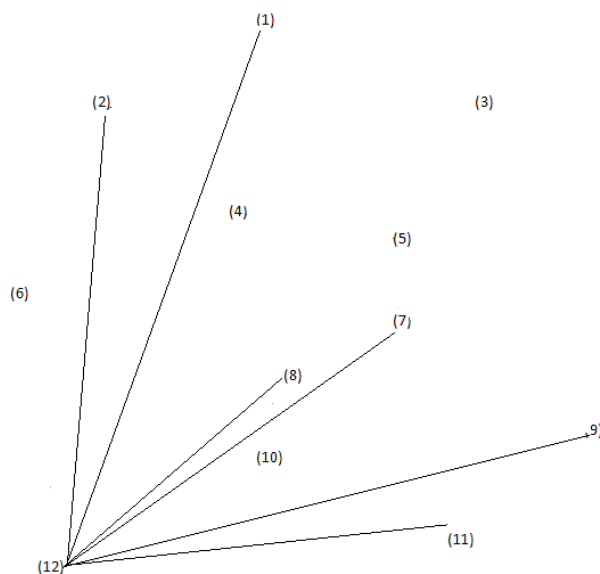


Figure 4.15 *Connections that Sentence 12 forms with other sentences*



To summarize, the above analysis explores the difference between sentences that form no bonds and sentences that forms bonds with other sentences and the function of these two types of sentences. An unbonded sentence is regarded as marginal and its function is to provide supplementary information to make the reading task easier but does not contribute directly to the development of the main theme. On the other hand, a central sentence is the sentence that forms bonds with others and it either serves as topic opening or closing sentence, or as connecting link between the preceding sentences and the following sentences. By introducing, connecting and concluding sentences, a central sentence contributes greatly to the development of the main theme of a text. It is, therefore, necessary to look at the number of sentences that form bonds with others and the number of sentences that form no bonds with others. By analysis of the ratio between the two, it is hoped that certain patterns in the use of bonds and the role that they play in the organization of text may be found. Table 4.6 below shows the actual ratio of the number of the sentences that form bonds with other sentences to those form no bonds with others in the 50 texts of the NCEE Corpus. For the convenience of their presentation in the table, the former is abbreviated as SFB and the latter SFNB.

Table 4.6 *The ratio of the number of the sentences that form bonds with other sentences (SFB) to those form no bonds (SFNB) in the NCEE Corpus*

	Text Title	The number of SFB	The number of SFNB	D1	D2
1	能思考的苇草	17	1	18: 1	1.6:1
2	情感理性认知	15	3	5: 1	
3	勿以亲疏定善恶	9	3	3: 1	
4	情感的权力	10	2	5: 1	
5	诚信抉择	15	9	15:9	
6	偶然? 必然!	13	5	13:5	
7	珍惜拥有和懂得 舍弃拥有的方式	12	14	6:7	
8	让理性主持宣判	13	17	13:17	
9	兼听则明[]	8	4	2: 1	
10	知之思之行之	7	16	7:16	
11	缤纷成就和谐	10	6	5:3	
12	谈意气	16	13	16:13	
13	让纪念闪耀理性光芒	15	9	3:1	

14	好奇心 1	18	11	18:11
15	没有付出哪有累累硕果	9	3	3:1
16	诗意地生活	14	9	14:9
17	坚持就是胜利	13	7	13:7
18	铭记八荣八耻	15	9	3:1
19	好奇心 2	14	10	7:5
20	好奇心 3	19	12	19:12
21	好奇心 4	12	5	12:5
22	清欢	15	11	15:11
23	不要轻易说“不”1	8	5	8:5
24	不要轻易说“不”2	12	7	12:7
25	莫把情云遮慧眼	14	5	14:5
26	公则生明	11	3	11:3
27	勿以好恶论断之	13	3	13:3
28	感情不是保险绳	10	7	10:7
29	天平和七弦琴	10	9	10:9
30	切莫走近	11	6	11:6
31	不要让眼光带上感情	14	11	14:11
32	切莫感情用事	11	8	11:8
33	呼唤理智	9	10	9:10
34	亲戚与明理	20	10	2:1
35	调准感情集距 拥有清晰世界	12	3	4:1
36	做一回理性自己	15	8	15:8
37	勿以亲疏定是非	10	5	2:1
38	莫让浮云遮望眼	16	9	16:9
39	独一无二	12	11	12:11
40	生命的意义	9	5	9:5
41	小杯子大道理	16	7	16:7
42	杯中窥人生	11	9	11:9
43	沉默的大多数	14	10	7:5
44	生命的痕迹	11	13	11:13
45	遭遇挫折，笑对痛苦	18	6	3:1
46	减负中的“朝三暮四”	18	10	9:5
47	变味的善良	19	5	19:5
48	角色转换之间	11	10	11:10
49	站对人生的舞台	14	7	2:1
50	怎一个“情”字了得	19	16	19:16

4.5.5 The average number of words that form repetition with other words in one sentence

This section reports the statistics of the number of words or phrases that form repetitions in each single sentence, regardless the different categories they are under. For the purpose of this study, it is important to compare the average number of items used to form connections in one sentence in the two corpora. By making such comparisons, it is hoped to find whether or not there is a significant gap in the use of the lexical cohesion between the two groups of language user. By recording the items that form lexical cohesion with other items in each single sentence and the total number of items that have connections with other items in the rest of the text, we can see the degree of complexity of the use of lexical cohesion in a text. The higher the total number of the items that form cohesion with other items in a single sentence, the more complex the use of the lexical cohesion it is. An example of matrix similar to Hoey's (2000) is presented in Table 4.7. Table 4.8 reports the statistical figure of the average number of words that form cohesion with other words in one sentence in the 50 NCEE texts.

Table 4.7 *The items that form connections (IFC) with other items in the NCEE Text 4 情感的权力*

7	1	悲剧	情感	错	犯下的错	权力	太大	真相											
7	2	段子	情感		惹下的祸				历史	君王	忠臣奸侯								
3	3									侯臣	亲信近宠								
5	4	可悲	情感			权力	无人企及			君主们									
4	5	悲剧	情感									麦克白	妻子						
3	6										亲信	麦克白	妻						
4	7		情感			权力								施展	正义				
9	8		真理情感	谬误						当权者				操纵	正义	近疏	颠倒黑白	感情用事	
7	9	事例	感情						现代	历史						亲疏		大义灭亲	社会
2	10									明君					光明正大				
9	11		情感真理	错	正确				现今						非正义	亲疏	歪曲		社会
7	12		情感			权力		真相							正义/公正	亲疏	混淆是非		
67																			
67/12-5.58																			

Table 4.8 *The average number of the items that form connections (IFC) with other items in each sentence in the NCEE Corpus*

	Text Title	The number of IFC	The total number of sentences	E1	E2
1	能思考的苇草	69	18	3.83	4.04
2	情感理性认知	73	18	4.06	
3	勿以亲疏定善恶	75	12	6.25	
4	情感的权力	68	12	5.67	
5	诚信抉择	78	24	3.25	
6	偶然? 必然!	79	17	5.67	
7	珍惜拥有和懂得 舍弃拥有的方式	82	26	3.25	
8	让理性主持宣判	57	30	1.90	
9	兼听则明	53	12	4.42	
10	知之思之行之	77	23	3.35	
11	缤纷成就和谐	79	16	4.94	
12	谈意气	118	29	4.07	
13	让纪念闪耀理性光芒	91	24	3.79	
14	好奇心 1	117	29	4.03	
15	没有付出哪有累累硕果	59	12	4.92	
16	诗意地生活	91	23	3.96	
17	坚持就是胜利	101	20	5.05	
18	铭记八荣八耻	89	24	3.71	
19	好奇心 2	95	24	3.96	
20	好奇心 3	122	31	3.94	
21	好奇心 4	81	17	4.76	
22	清欢	95	26	3.65	
23	不要轻易说“不”1	63	13	4.85	
24	不要轻易说“不”2	65	19	3.42	
25	莫把情云遮慧眼	85	19	4.47	
26	公则生明	57	14	4.07	
27	勿以好恶论断之	79	16	4.94	
28	感情不是保险绳	58	17	3.41	
29	天平和七弦琴	67	19	3.53	
30	切莫走近	87	19	4.58	
31	不要让眼光带上感情	110	25	4.40	
32	切莫感情用事	93	19	4.89	
33	呼唤理智	69	19	3.63	
34	亲戚与明理	114	30	3.80	
35	调准感情集距 拥有清晰世界	72	15	4.80	
36	做一回理性自己	98	23	4.26	
37	勿以亲疏定是非	51	15	3.40	

38	莫让浮云遮望眼	108	25	4.32
39	独一无二	99	23	4.30
40	生命的意义	49	14	3.50
41	小杯子大道理	93	23	4.04
42	杯中窥人生	98	20	4.90
43	沉默的大多数	80	24	3.33
44	生命的痕迹	96	24	4.00
45	遭遇挫折，笑对痛苦	99	24	4.13
46	减负中的“朝三暮四”	118	28	4.21
47	变味的善良	90	24	3.75
48	角色转换之间	87	21	4.14
49	站对人生的舞台	86	21	4.10
50	怎一个“情”字了得	145	35	4.14

4.6 Summary

This chapter reports the statistical results of various types of repetitions used in the NCEE Corpus, including a preliminary analysis the frequency of the use of non-lexical devices and the use of the lexical devices has been reported. Since these results are to be compared with the results of the analysis of the HSK Corpus to find out whether there is a gap between the two language groups, the next chapter focuses on the reporting of the result of the use of non-lexical and lexical devices in the HSK Corpus.

CHAPTER 5: STUDY 2 FINDINGS

5.1 Introduction

Chapter 4 identifies the various types of non-lexical and lexical cohesive ties in texts in the NCEE Corpus, and examines some possible lexical patterns in Chinese texts, such as the total number of types of lexical repetition, the number of each type, and the frequency of each type used in the sample texts. In this chapter, the two types of analyses carried out in Chapter 4 are undertaken in the same way to examine the use of different non-lexical and lexical cohesive ties in the HSK Corpus. The first type of analysis is a sole-rater manual analysis of the CFL Texts 1-10, followed by a computational analysis of all 50 texts in the HSK Corpus. As discussed in Chapter 3, when presenting the research methodology, the reference corpus (the NCEE Corpus) is used as a basis for comparison between the writing of native speaker Chinese and the writing of CFL learners. The criteria for the identification of the various types of non-lexical and lexical cohesion are the same as those defined in Chapter 4, Section 4.3. This chapter, therefore, does not include the definition of each type of lexical cohesive device. Instead, it only identifies the different types of cohesion in the HSK Corpus and provides examples for each type of lexical cohesion that occurs. Firstly, it examines the use of non-lexical cohesive devices in the HSK Corpus, and reports the results of analysis of the frequency and distribution of these devices used in the corpus. The second step of the analysis is to identify the use of lexical cohesive devices in the HSK Corpus. Using the concordancing programme NoteTab, this section examines lexical patterns in CFL learner texts by conducting a statistical analysis, which looks at the total number of types of lexical cohesive device, the number of each type, and the frequency of each type used in the learner texts respectively.

5.2 Non-lexical cohesion

In Chapter 4 Section 4.2, various types of non-lexical cohesion were defined and

discussed, including: thematic chunk, parallel structure, the use of idioms and proverbs, the use of quotations of classics, the use of summary words, and the use of metaphor. As discussed in Section 4.2, these different types of non-lexical cohesion are found in the NCEE Corpus, and they appear to play an important organizational role in Chinese argument texts that were examined. The analysis of these categories, however, does not replace the analysis of the use of the categories of non-lexical cohesion proposed by Halliday and Hasan (1976), namely reference, conjunction, ellipsis, and substitution, in the field of TCFL, as these non-lexical categories have become widely accepted by teachers and researchers in the field of TCFL. Many researchers have focused mainly on these devices in their studies of non-lexical cohesion in learner texts and consequently have contributed significantly to the literature. However, it is hoped that this study will broaden the scope of research in this area and further inform the teaching of CFL writing on the basis of the findings of the analysis.

Although the main focus of this study is to explore the patterns of lexical cohesion used in the argumentative texts written by native Chinese students and CFL learners and to identify the possible gap in the use of lexical cohesion between the two language groups, the complementary analysis of the use of non-lexical cohesion in texts written by the two groups contributes to a more complete understanding of the use of cohesion in argumentative text. As justified in the research methodology outlined in Chapter 3 Section 3.5.1, despite the fact that the number of these non-lexical cohesive devices used in native Chinese texts is not as large as that of lexical repetition, they still appear to play an essential role in the organization of the text. In addition, because of the role of traditional rhetoric in Chinese writing, control of non-lexical device has always been seen as one of the advanced skills of writing. For these reasons, it was considered useful to undertake a preliminary analysis of 10 CFL texts to examine how these non-lexical cohesive devices are used in CFL learners' writing. The result of this analysis is reported in the remainder of this section, in which the use of each device is compared with the findings of the pilot analysis of texts by ten native speakers of Chinese in Chapter 4.

In the following sub-sections (5.2.1 to 5.2.6), examples of each type of non-lexical repetition used in the HSK Corpus are provided and discussed. Similar to the approach in Chapter 4 Section 4.2, each Chinese example is followed by a corresponding English translation in square brackets to help the reader to understand the meaning of the Chinese text, and the translation approach adopted is more literal to highlight the structural and lexical features of the Chinese text. As these examples are taken from the HSK Corpus, some obvious errors, such as wrongly written Chinese characters, no punctuation between sentences, missing words, are corrected by the researcher to avoid confusion for readers.

5.2.1 Thematic chunk

There are no examples of thematic chunk found in the 10 chosen CFL learner texts. As summarized in Section 4.2, a thematic chunk has the following characteristics:

- 1) may be a phrase, or a sentence;
- 2) most serve as the title of the text;
- 3) repeated three or more times;
- 4) may be used independently or as part of the sentence;
- 5) may have variations but the meaning always remains the same; and,
- 6) contributes greatly to the organization of the text

In the ten CFL learner texts, although there are many chunks that have some features in common with that of thematic chunks in terms of form, they cannot be categorized as thematic chunk because they do not play an organizational role in texts. Since all the ten CFL learner texts have the same title 自己创业与进公司 [to start your own business or to find a job in a company], it seems that the a great number of learners used the whole title as a chunk or used it as two chunks after

breaking it into two parts 自己创业 [to start your own business] and 进公司 [to find a job in a company] in their writings. Consider the following example:

Example 47 (taken from the CFL Text 10)

(1) 再者，创业与进公司是两个很大的分歧，专业的领域也相去甚远，王强在不知道自己的个性、特长属于哪一方的情况下冒然创业必然自取死路，乖乖的进入公司，吸取更多的社会实经验才是上上之策。
[Moreover, to start your own business and to find a job in a company are two very different paths, and the two areas show little resemblance to each other. Without knowing his own personality and talent, Wang Qiang will inevitably go to a dead-end if he starts his own business rashly. The best strategy for him is to find a job in a company to learn more social experiences.]

(2) 所以我大胆地建议，你如果有勇气，可以负责一切的话，你自己创业也是不错的选择。我们还年轻，应该怀着大大的胸怀。 [Therefore I suggest boldly to you, if you are brave and able to be responsible for everything, starting your own business is a good choice. We are still young, so we should have a big heart.]

(3) 进一个公司过稳定的生活，虽然无忧无虑，但是没意思吧？ [If you find a job in a company and live a stable life, you may not have to worry about anything. However, it is rather boring, isn't it?]

In the examples, all the underlined phrases can be regarded as chunks. In Sentence (1), the writer uses 创业与进公司 [to start your own business and to find a job in a company] as a chunk, which is similar to the title 自己创业与进公司 [to start your own business or to find a job in a company]. In (2) and (3), the writer uses 自己创业 [to start your own business] and 进一个公司 [to find a job in a company] as chunks. These unanalysed chunks are more like lexical chunks, which are often used a sentence construction unit to compose a sentence. However, they do not contribute greatly to the organization of a text. As discussed in Section 4.2, a

thematic chunk often appears in the introduction part of the essay helping to open up the discussion, or in the beginning or the end of each paragraph in the main body paragraphs functioning as the topic sentence or concluding sentence. They can also be found in the conclusion part of the text serving as the concluding word or final touch. According to this criterion, there are no thematic chunks found in the ten learner texts.

5.2.2 The parallel structure

Similar to the results of the analysis of the use of thematic chunk, there are no examples of the parallel structure found in the ten learner texts. The main feature of a parallel structure includes having the same grammatical structure, sharing key words and structure marker, repeat two or more times, and may be consecutive or between distant paragraphs. In the ten learner texts, there are hardly any two sentences sharing two of the features of a parallel structure.

5.2.3 The use of idioms and proverbs

As defined in Section 4.2, an idiom in this study is a fixed expression whose meaning in the context does not match the literal meaning of its constituent words or phrases. There are a few examples of the use of idioms and proverbs found in the CFL learner texts. Consider the following example:

Example 48 (taken from the CFL Text 6)

(7) 你在大公司里不仅能准备未来还能得到薪水。 [When one works in a company, he can get paid while preparing himself for starting his own business in the future.]

(8) 难道这不是一箭双雕吗? [Isn't this a good opportunity for you to kill two birds with one stone?]

In this example, 一箭双雕 [kill two birds with one stone] is an idiom, which literally means in Chinese *to shoot two birds with one arrow*, and the meaning in this context is *to do two things at one time*. It forms a non-lexical repetition of the same idea with the phrase 不仅能准备未来还能得到薪水 [get paid while preparing himself for starting his own business in the future] in its preceding sentence.

Example 49 (taken from the CFL Text 8)

(18)自己创业可能会面临很多困难，或有可能你第一次搞的事业会失败，但是天上不可能掉馅饼，你也得学会挫折是使你发展的过程。[You may have to face many challenges if you start your own business, or your first business may end in failure. However, as a proverb goes, there's no such thing as a free lunch, you therefore have to learn that it is the defeat that makes you improve and grow up.]

(19)不要害怕失败，年轻人应该尝到人生的酸甜苦辣。[Therefore, don't be afraid of failure, young people should taste all the sweets and bitters of life.]

In this example, 天上不可能掉馅饼 [there is no such thing as a free lunch] is a proverb, which in Chinese means *it is impossible to have pies dropping down from the sky*. It forms a repetition with the phrase 年轻人应该尝到人生的酸甜苦辣 [young people should taste all the sweets and bitters of life] in its succeeding sentence. The results of the analysis of the use of idioms and proverbs in the ten CFL texts are shown in Table 5.1.

Table 5.1 *The results of the analysis of the use of idioms and proverbs in the ten CFL texts*

	Text Title	Total number of characters	Total number of idioms and proverbs	The frequency
1	自己创业 与进公司	475	0	0
2	自己创业 与进公司	486	0	0
3	自己创业 与进公司	507	0	0
4	自己创业 与进公司	514	0	0
5	自己创业 与进公司	515	0	0
6	自己创业 与进公司	517	1	$1 \times 100 / 517 = 0.19$
7	自己创业 与进公司	522	0	0
8	自己创业 与进公司	531	2	$2 \times 100 / 531 = 0.38$
9	自己创业 与进公司	534	0	0
10	自己创业 与进公司	593	0	0

5.2.4 The use of quotations from classics

As the name implies, the use of quotation from classics means a writer sometimes may quote some well-known sentences from classic literatures to illustrate his or her point. In the CFL learner texts, there are quite a few examples of the use of quotations from classics. Consider the following example:

Example 50 (taken from the CFL Text 3):

- (1) 古人云：“一子错，满盘皆输。”[As our ancestors said, a single careless move could cause the whole game to be lost.]

(2) 在人生这条又漫长又坎坷的道路上，抉择是断不可少的。 [On the long and rough road of life, one inevitably has to make crucial decisions.]

(3) 然而有时，看似一个微小的决定却带来了极大的影响。 [However, sometimes a seemingly small decision could have a significant impact.]

In this example, “一子错，满盘皆输。” [a single careless move could cause the whole game to be lost] is a quotation, which forms a repetition with the phrase 看似一个微小的决定却带来了极大的影响 [a seemingly small decision could have a significant impact] in Sentence 3.

Example 51 (taken from the CFL Text 2):

(8) 当然每件事物都具有双面性，创办公司需要很多时间何况还要进公司会拖延时间的。 [Surely every coin has two sides, starting your own business takes time, not mention that working in a company could delay your plan.]

(9) 可是我觉得当你创办公司时不要发生“万事俱备，只欠东风”的情况。 [However, when you start your own business, you should be cautious about the worst scenario, which is that you have everything but that one last piece of the jigsaw.]

(10) 因为创办一家公司不像开个小铺店，你自己也要考虑到自身有没有能引导大家的能力。 [This is because that starting a business is not like opening up a dairy shop, you have to consider whether you have strong leadership skills.]

In this example, 万事俱备，只欠东风 [you have everything but that one last piece of the jigsaw] is a quotation and it forms a repetition with the phrase 要考虑到自身有没有能引导大家的能力 [to consider whether you have strong leadership skills] in Sentence 10. Table 5.2 following summarizes the results of

the analysis of the use of quotation from classics.

Table 5.2 *The results of the analysis of the use of quotation from classics in ten CFL learner texts*

	Text Title	Total number of characters	Total number of quotations from classics	The frequency
1	自己创业 与进公司	475	0	0
2	自己创业 与进公司	486	2	$2 \times 100 / 486 = 0.41$
3	自己创业 与进公司	507	2	$2 \times 100 / 507 = 0.39$
4	自己创业 与进公司	514	2	$2 \times 100 / 514 = 0.39$
5	自己创业 与进公司	515	0	0
6	自己创业 与进公司	517	2	$2 \times 100 / 517 = 0.38$
7	自己创业 与进公司	522	0	0
8	自己创业 与进公司	531	2	$2 \times 100 / 531 = 0.38$
9	自己创业 与进公司	534	2	$2 \times 100 / 534 = 0.37$
10	自己创业 与进公司	593	1	$1 \times 100 / 593 = 0.19$

5.2.5 The use of summary words

As discussed in Chapter 4, Section 4.2, the main feature of summary words is that they are often used to refer to a certain part of writing, which could a sentence, several sentences, and sometime a paragraph or even couple of paragraphs. Summary words are usually words that have general meanings, such as *issue*, *problem*, *assessment*, *example*, *question*, *choice*, *things*. Consider the following example:

Example 52 (taken from the CFL Text 9):

(2) 到底是创业好还是进公司好? [To start your own business or to find a job in a company, which is better?]

(7) 在考虑这一问题时，请你先排除掉对钱财的顾虑。 [When you consider this question, please first ignore your worrying about money.]

(8) 虽然创业赚的钱都入自己的口袋，却有破产的危险性；虽然进公司只能得到自己固定的薪资，但你对公司的贡献可以让你得到相对应的高工资。 [Although you could make lots of money if you start your own business, you run the risk of becoming bankrupt; although you could only be paid at a fixed rate if you work in a company, you may get paid more gradually according to the contribution you make to the company.]

(9) 你最先顾及的应当是你对这两件事的渴望程度，你的性情爱好和你的能力水平，像王强这种情况，他对两个选择的渴望程度都比较高所以应该重点考虑：1. 哪一项更附合我的性格，2. 我最终目标是在哪儿。 [The first thing you need to consider is how desperate you are for the two things, then you have to think about your personality and you ability and proficiency. Like Wang Qiang, he has high expectations of the both choices, so he should consider in particular the following two things: firstly, which one suits my personality better? Second, what is my ultimate goal?]

In this example, there are three summary words which all refer to Sentence 2. The first summary word is 问题 [question], which refers to the question in Sentence 2: 到底是创业好还是进公司得好? [To start your own business or to find a job in a company, which is better?]. The second word is 这两件事 [the two things] in Sentence 9, referring back to 创业 [to start your own business] and 进公司 [to find a job in a company]. The third word is 两个选择 [the two choices], which also refer to 创业 [to start your own business] and 进公司 [to find a job in a company]. Table 5.3 summarizes the results of the analysis of the use of summary words.

Table 5.3 *The results of the analysis of the use of summary words in the ten CFL learner texts*

	Text Title	Total number of characters	Total number of summary words	The frequency
1	自己创业与进公司	475	0	0
2	自己创业与进公司	486	0	0
3	自己创业与进公司	507	0	0
4	自己创业与进公司	514	0	0
5	自己创业与进公司	515	0	0
6	自己创业与进公司	517		0
7	自己创业与进公司	522	0	0
8	自己创业与进公司	531	0	0
9	自己创业与进公司	534	6	$6 \times 100 / 534 = 1.12$
10	自己创业与进公司	593	0	0

5.2.6 The use of metaphor

There are some examples of the use of metaphor in the CFL learner texts. As defined in Section 4.2, it means a figurative use of words, phrases, or sentences to form a repetition with an aim to illustrate a point more clearly in the context. Consider the following examples:

Example 53 (taken from the CFL Text 3):

(13) 人生的道路往往是十字路口比较多的。 [On our road of life, there are quite lot intersections.]

(14) 抉择是十分重要的，要学会谨慎，正确地为自己选择适合的道路！
[Making decisions is crucial. One has to learn how to be cautious, and how to choose a suitable path for himself.]

In this example, 十字路口 [intersection] is a metaphor, which forms a repetition with 选择适合的道路 [to chose a suitable path].

Example 54 (taken from the CFL Text 5):

(2) 一般来说，自己创业的人是独立心强，有领导才能”的人，进自公司上班的则是脚踏实地，热爱学习的人，但我认为，既然是“开始就不必好高骛远，强求自己开启一番大事业，进公司扎实地立基础才是明智之举。
[Generally speaking, people who start their own business are usually independent and have strong leadership skills. While people who choose to work in a company are steady workers and keen learners. But I think since it is a start point for Wang Qiang, he therefore should not aim too high and force himself to do something big from the beginning of his career. Instead, he should find a job in a company to lay a good foundation for himself, which I think is a wise strategy.]

(11)如果你没有有鹰的翅膀，请用蜗牛那坚持不懈的精神，扎实的攀登方式，取得你梦想中的成功。 [If you don't have wings like that of eagles', then you just have to follow the spirit of a snail, climbing the mountain steadfastly until you achieve you ideal goal.]

In this example, 鹰的翅膀 [the wings of eagle] is a metaphor of 独立心强，有领导才能”的人 [people who are independent and has strong leadership skills], while 蜗牛 [snail] is a metaphor of 脚踏实地，热爱学习的人 [people who are steady workers and keen learners]

The results of the analysis of the use of metaphor are shown in Table 5.4.

Table 5.4 *The results of the analysis of the use of metaphor in the ten CFL learner texts*

	Text Title	Total number of characters	Total number of metaphor	The frequency
1	自己创业 与进公司	475	0	0
2	自己创业 与进公司	486	0	0
3	自己创业 与进公司	507	6	$6 \times 100 / 507 = 1.18$
4	自己创业 与进公司	514	2	$2 \times 100 / 514 = 0.39$
5	自己创业 与进公司	515	12	$12 \times 100 / 515 = 2.33$
6	自己创业 与进公司	517	0	0
7	自己创业 与进公司	522	0	0
8	自己创业 与进公司	531	0	0
9	自己创业 与进公司	534	0	0
10	自己创业 与进公司	593	8	$8 \times 100 / 593 = 1.35$

5.2.7 Summary of the analysis of the use of non-lexical repetition in the HSK Corpus

Compared with the results of the analysis of the use of non-lexical repetition in NCEE texts, the number of non-lexical cohesive devices used in the CFL texts is significantly lower. In particular, no instances of the use of thematic chunks and parallel structure are found in the ten analyzed CFL texts. Consequently, this remarkably low frequency of the use of non-lexical devices raises some issues for the TCFL researchers and teachers, such as:

- Why there is such a gap in the use of non-lexical cohesive devices between native and non-native writers of Chinese?
- What is the possible cause for this?

These questions are worthy of further investigation and discussion, and the findings may have benefit for the teaching of Chinese writing and the further investigation of the linguistic competences that relate to the writing of Chinese texts. This, however, is not the focus of this research and hence falls outside of the scope of this study.

5.3 Lexical cohesion

In this section, categories of the various types of lexical repetition in used in CFL texts are discussed and exemplified in detail. In order to identify and differentiate the different types of lexical cohesion, the heuristic flow charts, as discussed in Chapter 4 Section 4.3 are again used to (a) establish whether or not a lexical item is worth investigating; and, (b) to decide which type of repetition it is. However, to economize space in this report of the findings, these charts are not presented again in this chapter. In Section 5.4 following, the statistical data of the use of lexical repetition in texts written by the CFL learners is presented. However, preceding the presentation of the quantitative data, some examples of the use of the five types of lexical cohesive devices found in the HSK Corpus are presented, so that the reader may have an overall impression of the way in which these devices are used by second language speakers. These five types are: simple lexical repetition, complex repetition, simple lexical paraphrase, complex lexical paraphrase, and co-reference. The analysis showed that the other types (constituent repetition, superordinate and hyponym repetition) did not occur in the CFL texts. The statistical data showing the frequency of occurrence of these five devices are provided in Section 5.4 following.

5.3.1 Simple lexical repetition

As discussed in Section 4.3, simple lexical repetition can occur in two circumstances. Firstly, if a lexical item that has already occurred in a text is repeated, it can be said that they form a simple repetition. This is the most frequently used form of simple repetition in both corpora. The following is an example of this type of repetition.

Example 55 (taken from the CFL Text 10):

(1) 创业与就职，多少有点自知之明的人都应该能做出决定，创业要的不只是能力，还要有远见、胆识、眼光，因此很简单的就可以和到结论，王强应该进公司就职。 [In terms of whether to start one's own business or to find a job, anyone who somewhat knows himself should be able to make a decision, because to start a business does not only need capability, but also foresight, courage, and vision. It is, therefore, easy to draw a conclusion that Wang Qiang should find a job in a company.]

(2) 首先，他快毕业了，工作是决定自己人生的一大要事，在毕业的他没有做好思想、心理准备，证明他没有判断、决策的能力，创业而不能决断，自取灭亡之道。 [Firstly, although he is graduating soon, and knows that finding a job is one of the most important things in his life, he has not well prepared mentally. This proves that he lacks the ability to judge and to make decisions. If he wants to start a business yet cannot make right decisions, he is inviting his own destruction.]

(3) 再者，创业与进公司是两个很大的分歧，专业的领域也相去甚远，王强在不知道自己的个性、特长属于哪一方的情况下冒然创业必然自取死路，乖乖地进入公司，吸取更多的社会经验才是上上之策。 [Moreover, to start a business and to find a job in a company are two very different paths, and the two areas show little resemblance to each other. Without knowing his own personality and talent, Wang Qiang will inevitably go to a dead-end if he starts his own business rashly. The best strategy for him is to find a job in a company to learn more social experiences.]

In the above example, the word 创业 (to start a business) in Sentence (1) is repeated in Sentence (2) and Sentence (3); it therefore forms a simple lexical repetition with the word 创业 (to start a business) in Sentences (2) and (3) respectively.

Two lexical items can also be regarded as a simple repetition if the only alternation of the two is within a closed grammatical paradigm. In the case of Chinese, this means that if the only difference between two items is one has a suffix 们 (*a plural form of certain nouns*) while the other does not, then they can be seen as a simple repetition. As explained in Section 4.3.1, this type of lexical device is very rare in both corpora. Consider the following example:

Example 56 (taken from the CFL Text 14):

(10) 所以从整体上来说，刚大学毕业的学生根本没有维持自己公司的成本或防止破产的保险。 [Consequently, in brief, new graduates have no capital to operate their company and no insurance to prevent bankruptcy.]

(11) 因而，创业对大学生来说，是不切实际的，时为过早的举措。 [As a result, it is an unrealistic and premature move for college students to start their own business.]

(12) 其次，大学生们缺乏丰富的经验。 [Next, college students lack rich experience.]

In this example, 大学生 (college students) in Sentence (11) forms a simple lexical repetition with the word 大学生们 (college students) in Sentence (12), as the two shares 大学生 and the only difference between them is the later has a suffix 们。

5.3.2 Complex lexical repetition

Two items can be seen as an instance of complex repetition if they either share a lexical morpheme that is not formally identical or identical, but both have

different grammatical functions. A detailed explanation of this device is provided in Section 4.3.2. The following are some examples of the use of complex repetition in the CFL texts.

Example 57 (taken from the CFL Text 8):

(11) 如果你自己创业当老板的话，冒险性非常大。 [It is very risky if you want to start your own business and be your own boss.]

(12) 我不知道你是什么样的人，反正我很喜欢冒险，我受不了别人左右我的人生。 [I don't know what kind of personality you have, but I like taking risks, because I cannot stand the fact that my life is controlled by other people.]

In this example, the word 冒险性 (risky) in Sentence (11) shares the lexical morpheme 冒险 with the word 冒险 (risk) in Sentence (12). The only difference between the two is that the word 冒险性 (risky) has a suffix 性. The two therefore form a complex lexical repetition.

Example 58 (taken from the CFL Text 9):

(1) 就像高中时人们选择文科，理科一样，人初步社会时都会考虑王强的这一问题。 [Just like high school students, who have to choose between arts and science, every people who steps into the society has to consider the question that Wang Qiang asks.]

(2) 到底是创业好还是进公司好? [To start a business or to find a job in a company, which is better?]...

(5) 难道进公司是无法创业的人不得已的选择? [Is working in a company the only choice of those who cannot start their own business?]

In this example, the word 选择 (to choose) in Sentence (1) is a verb, while in Sentence (5) the word 选择 (choice) is a noun. These two words are identical in form but different in grammatical function. They therefore form a complex repetition.

5.3.3 Simple lexical paraphrase

When two lexical items are interchangeable in context without a discernible change in meaning and specificity, they can be regarded as an instance of simple paraphrase. Consider the following examples:

Example 59 (taken from the CFL Text 13):

(1) 首先，我的想法不是创业，也不是进公司，我的想法是一先进公司，再创业…[First of all, my thought is not to start a business, neither to find a job in a company. My thought is to find a job in a company first, then to start a business later. …]

(12) 所以，我的观点是，先进公司，准备好三方面的条件，待你一切准备好了，就可以一口气以稳当的，迅速的势力发展起来。[Therefore, my opinion is that he should find a job in a company first to prepare himself in these three areas. One he is ready, he can then develop rapidly and stably.]

In this example, the word 想法 (thought) has a similar meaning to the word 观点 (opinion) in Sentence 12. They are interchangeable in the context and there is no obvious meaning gain or loss. 想法 (thought) and 观点 (opinion) therefore can be seen as an example of simple paraphrase.

Example 60 (taken from the CFL Text 10):

(2) 首先，他快毕业了，工作是决定自己人生的一大要事，在毕业的他没有做好思想、心理准备，证明他没有判断、决策略的能力，创业而不能决断，自取灭亡之道。[Firstly, although he is graduating soon, and knows that finding a job is one of the most important things in his life, he has not well prepared mentally. This proves that he lacks the ability to judge and to make decisions. If he wants to start a business yet cannot make right decisions, he is inviting his own destruction.]

(3) 再者，创业与进公司是两个很大的分歧，专业的领域也相去甚远，王强在不知道自己的个性、特长属于哪一方的情况下贸然创业必然自取死

路，乖乖的进入公司，吸取更多的社会经验才是上上之策。 [Moreover, to start a business and to find a job in a company are two very different paths, and the two areas show little resemblance to each other. Without knowing his own personality and talent, Wang Qiang will inevitably go to a dead-end if he starts his own business rashly. The best strategy for him is to find a job in a company to learn more social experiences.]

In this example, 灭亡之道 (the way to death) in Sentence (2) and 死路 (the way to death) are different in form, but similar to each other in meaning, one can be replaced by the other but their meaning remains the same. These two words are, therefore, regarded as an instance of simple paraphrase.

5.3.4 Complex paraphrase

As defined in Section 4.3. 4, complex paraphrase occurs in two different situations. The first relates to antonymy, and most antonymous pairs in Chinese can be seen as potential instances of complex paraphrase. The second situation is that if one item is a complex repetition or simple paraphrase of another item and also a simple paraphrase of a third, then the second and the third can be regarded as a complex paraphrase link. The following are two examples of complex paraphrase:

Example 61 (taken from the CFL Text 4):

(5) 但，身无分文，欠缺对所创业的事情的了解，光靠信心跟热情是微不足道的。 [However, one cannot success only by relying on confidence and passion, if he is penniless and lacks understanding of the business world he is planning to step into.]

(6) 创业后要面对的是行内大公司的威胁、行间的挑战、越来越高的要求…所以，必须要有足够的，可以抵挡住这些压力的心境跟透彻理解的时候你才可以大胆地去创业。 [After starting your business, you have to face the intensive competition, challenge and higher demand in and between the

trades. Therefore, one can only start his own business bravely when he has a mental state that could handle huge pressure well and has a clear and enough understanding of the business.]

In this example, the meaning of the word 欠缺 (lack, not enough) is opposite to that of 足够 (enough). They form a complex paraphrase link.

Example 62 (taken from the CFL Text 11):

(6) 再者，要创业必须拥有一定的资金。[What's more, one has to have a certain amount of capital if he wants to start his own business.]

(7) 对于目前还是学生的王强来说，资金上的问题应该还是比较棘手的。[For Wang Qiang who is still a student right now, the capital issue is surely troublesome.]

(8) 所以说，我认为王强大可先进一家公司工作，有了一些积蓄之后再创业。[Therefore, I think that Wang Qiang should find a job in a company first, and start his own business once he has some savings.]

(9) 这样一来，创业初期的金钱上的困扰可以稍微缓解。[In so doing, the money issue at the initial stage of his business adventure could be addressed to a certain degree.]

In this context, the word 资金 (capital) in Sentence (7) forms a simple paraphrase with the word 金钱 (money) in Sentence (9), while the word 积蓄 (savings) in sentence also forms a simple paraphrase with the word 金钱 (money), in this case, 资金 (capital) and 积蓄 (savings) can be regarded as an instance of complex paraphrase.

5.3.5 Co-reference repetition

Co-reference occurs when two lexical items are unrelated but have identical referents in the context of a text. Co-reference can be demonstrated in the following examples.

Example 63 (taken from the CFL Text 2):

(2) 就王强的情况而言，应该进一家公司工作比较合适。[As far as Wang Qiang is concerned, I believe that finding a job in a company suits him better.]

(10) 对王强来说，或许还需要学习并培养自己承受失败的心理能力；还需要具备在困境中奋起的精神；而这一切对一个初出茅庐的青年学生来讲，太困难了。[For Wang Qiang, maybe he still needs to cultivate his ability to bear defeat or setback, and to foster the ability to rise up in adverse circumstances. However, this is very hard for a young student who is about to step into the society.]

In this example, the phrase 一个初出茅庐的青年学生 (a young student who is about to step into the society) in Sentence (10) does not relate to the name 王强 (Wang Qiang) in Sentence (2). However, the former clearly refers to the latter. Thus, the two form a co-reference repetition.

Example 64 (taken from the CFL Text 3):

(5) 而我的看法是，我认为对于王强而言，进公司比较适合他。[My opinion on this is that finding a job in a company suits Wang Qiang better.]

(6) 首先，对于一个刚毕业毫无任何工作经验的青年，倘若他要自己创业，领导别人，可能威望还不够，人际交往方面还有欠缺。[Firstly, for a newly-graduated young man who has no work experience at all, if he wants to start his own business and to supervise other people, he has not enjoyed any high prestige, nor does he has any interpersonal competence.]

In this example, again, 一个刚毕业毫无任何工作经验的青年 (a newly-

graduated young man who has no work experience at all) in Sentence (6) refers to 王强 (Wang Qiang) in Sentence (5), they are therefore an instance of co-reference.

5.3.6 Summary of the findings of the use of lexical cohesion in the HSK Corpus

This analysis of the CFL texts shows examples of the use of simple repetition, complex repetition, simple paraphrase, and co-reference, which constitutes the majority of the instances of lexical cohesive devices. The frequencies of these five types of lexical cohesive devices are reported in the next section. There are no occurrences of the other types of lexical repetition such as constituent repetition, superordinate hyponymic repetition in the texts. A detailed discussion about the absence of the use of these types of repetition can be found in Chapter 8.

5.4 The statistical corpus analysis

Following the same procedures employed in the analysis of the NCEE Corpus and outlined in Section 4.4, this section reports the statistical analysis of the use of lexical cohesive devices in texts written by the CFL learners. With the aim of exploring how lexical repetition helps to organize a text and to find whether there are gaps in the use of cohesive devices between the two corpora, the items that were examined statistically are the same as that in Table 4.1 as listed in Chapter 4.

5.4.1 The use of various lexical repetitions in texts of the HSK Corpus

As it is done in Section 4.4, an example of the data obtained from a sample text in the HSK Corpus is provided in Table 5.5, which clearly demonstrates the actual use of the various types of lexical cohesion in this particular text.

Table 5.5 *The number of the various types of lexical cohesive devices and their frequency per 100 characters*

The title of Text 507: 自己创业与进公司[Starting a business or finding a job in a company?]		
Total number of characters: 507		
Name of lexical cohesion used	Number of the items used	Frequency (per 100 characters)
Simple lexical repetition (type1 sr)	24	$24 \times 100 / 507 = 4.73$
Complex lexical repetition (type2 cr)	9	$9 \times 100 / 507 = 1.76$
Constituent repetition (type3 constituent re)	0	0
Simple paraphrase (type4 sp)	3	$3 \times 100 / 507 = 0.59$
Complex paraphrase (type5 cp)	3	$3 \times 100 / 507 = 0.59$
Superordinate and hyponym (type6 super-hypo)	0	0
Co-reference (type7 co-re)	2	$2 \times 100 / 507 = 0.39$

Although this table indicates the variety of cohesive devices that the writer used to organize the text and the writer's preferences of using different types of repetition, further statistical data need to be obtained to exploring the possible lexical patterns that are common in the HSK Corpus. To achieve this, the following formula is used to calculate the average value of the frequency of each type of repetition used in 50 sample texts.

$$AF_x = \frac{FT1x + FT2x + \dots + FT49x + FT50x}{50}$$

The parameters in the formula and the method of the use of this formula have been explained in Section 4.4. For reasons of economy of space, the same explanation is not repeated here. By following this formula, the average frequencies of the seven types of repetition were obtained and listed in Table 5.6 respectively:

Table 5.6: *The average frequencies of the seven types of repetition in the HSK Corpus*

Category	Average Frequency	Example
simple repetition	6.21	能力(ability)-能力(ability)
complex repetition	0.56	创业(starting an enterprise)-创业者(entrepreneur)
constituent repetition	0	
simple paraphrase	1.75	考虑(consider)-思考(think)
complex paraphrase	1.57	成功(success)-失败(failure)
superordinate-hyponymic	0	
co-reference	0.08	王强(Wang Qiang)-一个初出茅庐的青年(a young man who has just stepped into the society)

Table 5.6 clearly shows that the average frequency of simple repetition is the highest among the seven categories, which is almost four times as high as the numbers for simple paraphrase and complex paraphrase. The frequency of complex repetition is only 0.56, significantly lower than that of the above three devices. The lexical device with the lowest frequency among them is Co-reference, which is only 0.08.

In next chapter these figures will be compared with the set of figures (relating to the NCEE Corpus) listed in Table 4.3 in Section 4.5, and the significance of the difference between them will be discussed in detail in Chapter 7.

5.4.2 The number of bonds formed in the texts of the HSK Corpus

In this section, statistics of B1 and B2 are reported, which are the main indicators of the closeness of connection between bonded sentences and density of repetitions in a text. The justification for carrying out such an investigation is

provided in Section 4.3.2. The degree of closeness of connection in a text can be quantified and measured by the following formula:

$$\text{The bond density} = \frac{\text{the number of bonds} \times 100}{\text{the total number of characters in the text}}$$

The following Table 5.7 provides an overall picture of the distribution of the bond density in 50 sample texts in the HSK Corpus:

Table 5.7 *The individual bond density and the average bond density in texts in the HSK Corpus*

	Text Title	Total number of bonds	Total number of characters	B1	B2
1	自己创业与进公司	8	475	1.68	1.35
2	自己创业与进公司	13	486	2.67	
3	自己创业与进公司	2	507	0.39	
4	自己创业与进公司	3	514	0.58	
5	自己创业与进公司	16	515	3.11	
6	自己创业与进公司	9	517	1.74	
7	自己创业与进公司	5	522	0.96	
8	自己创业与进公司	0	531	0.00	
9	自己创业与进公司	8	534	1.50	
10	自己创业与进公司	4	593	0.67	
11	自己创业与进公司	15	456	3.29	
12	自己创业与进公司	2	442	0.45	
13	自己创业与进公司	5	424	1.18	
14	自己创业与进公司	3	580	0.52	

15	自己创业与进公司	9	576	1.56
16	自己创业与进公司	2	575	0.35
17	自己创业与进公司	3	571	0.53
18	自己创业与进公司	10	568	1.76
19	自己创业与进公司	4	566	0.71
20	自己创业与进公司	13	562	2.31
21	自己创业与进公司	2	552	0.36
22	自己创业与进公司	11	551	2.00
23	自己创业与进公司	9	550	1.64
24	自己创业与进公司	13	549	2.37
25	自己创业与进公司	1	548	0.18
26	自己创业与进公司	9	508	1.77
27	自己创业与进公司	4	505	0.79
28	自己创业与进公司	14	500	2.80
29	自己创业与进公司	8	499	1.60
30	自己创业与进公司	5	498	1.00
31	自己创业与进公司	4	495	0.81
32	自己创业与进公司	3	492	0.61
33	自己创业与进公司	12	479	2.51
34	自己创业与进公司	11	478	2.30
35	自己创业与进公司	4	467	0.86
36	自己创业与进公司	0	445	0.00
37	自己创业与进公司	9	493	1.83
38	自己创业与进公司	4	491	0.81
39	自己创业与进公司	5	490	1.02

40	自己创业与进公司	10	481	2.08	
41	自己创业与进公司	4	474	0.84	
42	自己创业与进公司	2	473	0.42	
43	自己创业与进公司	12	472	2.54	
44	自己创业与进公司	8	470	1.70	
45	自己创业与进公司	13	466	2.79	
46	自己创业与进公司	7	465	1.51	
47	自己创业与进公司	1	464	0.22	
48	自己创业与进公司	6	463	1.30	
49	自己创业与进公司	4	462	0.87	
50	自己创业与进公司	9	460	1.96	

5.4.3 The number of bonds formed between non-adjacent sentences to that between adjacent sentences in the HSK Corpus

As argued in Section 4.3.3, the ratio of the bonds formed between non-adjacent sentences to adjacent sentences in a text is a strong indicator of the degree of cohesiveness of a text. Table 5.8 following lists the number of bonds formed between non-adjacent sentences to that between adjacent sentences.

Table 5.8 *the ratio of bonds formed between non-adjacent sentences (BNA) to that between adjacent sentences (BA) in texts in the HSK Corpus*

	Text Title	Total number of BNA	Total number of BA	C1	C2
1	自己创业与进公司	3	5	3:5	1.4:1
2	自己创业与进公司	4	9	4:9	

3	自己创业与进公司	0	2	0:2
4	自己创业与进公司	2	1	2:1
5	自己创业与进公司	12	4	3:1
6	自己创业与进公司	7	2	7:2
7	自己创业与进公司	3	2	3:2
8	自己创业与进公司	0	0	0:0
9	自己创业与进公司	7	1	7:1
10	自己创业与进公司	2	2	1:1
11	自己创业与进公司	15	0	15:0
12	自己创业与进公司	1	1	1:1
13	自己创业与进公司	5	0	5:0
14	自己创业与进公司	2	1	2:1
15	自己创业与进公司	7	2	7:2
16	自己创业与进公司	1	1	1:1
17	自己创业与进公司	2	1	2:1
18	自己创业与进公司	4	6	4:6
19	自己创业与进公司	3	1	3:1
20	自己创业与进公司	11	2	11:2
21	自己创业与进公司	0	2	0:2
22	自己创业与进公司	8	3	8:3
23	自己创业与进公司	3	6	1:2
24	自己创业与进公司	10	3	10:3
25	自己创业与进公司	1	0	1:0
26	自己创业与进公司	4	5	4:5
27	自己创业与进公司	1	3	1:3

28	自己创业与进公司	5	9	5:9
29	自己创业与进公司	5	3	5:3
30	自己创业与进公司	3	2	3:2
31	自己创业与进公司	1	3	1:3
32	自己创业与进公司	2	1	2:1
33	自己创业与进公司	7	5	7:5
34	自己创业与进公司	9	2	9:2
35	自己创业与进公司	1	3	1:3
36	自己创业与进公司	0	0	0:0
37	自己创业与进公司	4	5	4:5
38	自己创业与进公司	3	1	3:1
39	自己创业与进公司	2	3	2:3
40	自己创业与进公司	9	1	9:1
41	自己创业与进公司	2	2	1:1
42	自己创业与进公司	2	0	2:0
43	自己创业与进公司	5	7	5:7
44	自己创业与进公司	7	1	7:1
45	自己创业与进公司	3	10	3:10
46	自己创业与进公司	2	5	2:5
47	自己创业与进公司	1	0	1:0
48	自己创业与进公司	4	2	2:1
49	自己创业与进公司	1	3	1:3
50	自己创业与进公司	2	7	2:7

5.4.4 The number of sentences that form bonds with other sentences to those form no bonds with others in the HSK Corpus

This section reports the ratio of the number of sentences that form bonds (with other sentences) to those that form no bonds with other sentences in each text, and the average ratio in 50 texts in the HSK Corpus. In so doing, it is hoped to understand how the bonded sentences contribute to the development of the theme of a text. (These groups of figures are coded as D1 and D2 in Table 4.1) Table 5.9 following presents the ratio of D1 to D2:

Table 5.9 *The ratio of the number of the sentences that form bonds with other sentences (SFB) to those form no bonds (SFNB) in texts in the HSK Corpus*

	Text Title	Total number of SFB	Total number of SFNB	D1	D2
1	自己创业与进公司	8	9	8:9	1:1.2
2	自己创业与进公司	8	5	8:5	
3	自己创业与进公司	4	9	4:9	
4	自己创业与进公司	6	10	6:10	
5	自己创业与进公司	9	2	9:2	
6	自己创业与进公司	11	2	11:2	
7	自己创业与进公司	7	8	7:8	
8	自己创业与进公司	2	20	1:10	
9	自己创业与进公司	8	7	8:7	
10	自己创业与进公司	5	3	5:3	
11	自己创业与进公司	12	5	12:5	
12	自己创业与进公司	4	8	1:2	
13	自己创业与进公司	6	6	1:1	
14	自己创业与进公司	8	10	4:5	

15	自己创业与进公司	5	6	5:6
16	自己创业与进公司	7	10	7:10
17	自己创业与进公司	10	9	10:9
18	自己创业与进公司	12	7	12:7
19	自己创业与进公司	14	7	2:1
20	自己创业与进公司	11	2	11:2
21	自己创业与进公司	14	8	7:4
22	自己创业与进公司	10	6	5:3
23	自己创业与进公司	10	9	10:9
24	自己创业与进公司	9	4	9:4
25	自己创业与进公司	8	2	4:1
26	自己创业与进公司	2	28	1:14
27	自己创业与进公司	5	6	5:6
28	自己创业与进公司	7	5	7:5
29	自己创业与进公司	11	14	11:14
30	自己创业与进公司	8	4	4:1
31	自己创业与进公司	4	0	4:0
32	自己创业与进公司	6	10	3:5
33	自己创业与进公司	4	1	4:1
34	自己创业与进公司	5	6	5:6
35	自己创业与进公司	6	13	6:13
36	自己创业与进公司	5	6	5:6
37	自己创业与进公司	4	18	2:9
38	自己创业与进公司	5	9	5:9
39	自己创业与进公司	6	8	3:4

40	自己创业与进公司	4	8	1:2
41	自己创业与进公司	3	23	3:23
42	自己创业与进公司	8	6	4:3
43	自己创业与进公司	7	8	7:8
44	自己创业与进公司	6	15	6:15
45	自己创业与进公司	6	9	6:9
46	自己创业与进公司	7	3	7:3
47	自己创业与进公司	9	7	9:7
48	自己创业与进公司	3	12	1:4
49	自己创业与进公司	4	7	4:7
50	自己创业与进公司	7	9	7:9

5.4.5 The statistics of the average number of words that form lexical cohesive links with other words in one sentence in the HSK Corpus

The final section of the corpus analysis reports the statistics of the average number of items that connect with other items in each single sentence in 50 CFL texts. The purpose of this quantification is to find out whether there is a significant gap in this use of the lexical cohesion between the two groups of language users. A detailed justification for doing this is provided in Section 4.3.4. Table 5.10 lists all the figures of the group coded as E1 and E2.

Table 5.10 *The average number of the items that form connections (IFC) with other items in one sentence in texts in the HSK Corpus*

	Text Title	The number of IFC	Total number of Sentences	E1	E2
1	自己创业与进公司	45	17	2.6	2.6
2	自己创业与进公司	43	13	3.3	

3	自己创业与进公司	27	13	2.1
4	自己创业与进公司	40	16	2.5
5	自己创业与进公司	48	11	4.4
6	自己创业与进公司	44	13	3.4
7	自己创业与进公司	46	15	3.1
8	自己创业与进公司	34	22	1.5
9	自己创业与进公司	50	15	3.3
10	自己创业与进公司	40	8	5
11	自己创业与进公司	60	17	3.5
12	自己创业与进公司	25	12	2.1
13	自己创业与进公司	32	12	2.7
14	自己创业与进公司	46	18	2.6
15	自己创业与进公司	38	11	3.5
16	自己创业与进公司	50	17	2.9
17	自己创业与进公司	49	19	2.6
18	自己创业与进公司	35	19	1.8
19	自己创业与进公司	54	21	2.6
20	自己创业与进公司	41	13	3.2
21	自己创业与进公司	57	22	2.6
22	自己创业与进公司	31	16	1.9
23	自己创业与进公司	36	19	1.9
24	自己创业与进公司	37	13	2.8
25	自己创业与进公司	21	10	2.1
26	自己创业与进公司	42	30	1.4
27	自己创业与进公司	36	11	3.3

28	自己创业与进公司	25	12	2.1	
29	自己创业与进公司	49	25	2.0	
30	自己创业与进公司	38	12	3.2	
31	自己创业与进公司	27	4	6.8	
32	自己创业与进公司	44	16	2.8	
33	自己创业与进公司	33	5	6.6	
34	自己创业与进公司	34	11	3.1	
35	自己创业与进公司	29	19	1.5	
36	自己创业与进公司	34	11	3.1	
37	自己创业与进公司	32	22	1.5	
38	自己创业与进公司	40	14	2.9	
39	自己创业与进公司	39	14	2.8	
40	自己创业与进公司	32	12	2.7	
41	自己创业与进公司	46	26	1.8	
42	自己创业与进公司	43	14	3.1	
43	自己创业与进公司	35	15	2.3	
44	自己创业与进公司	49	21	2.3	
45	自己创业与进公司	44	15	2.9	
46	自己创业与进公司	32	10	3.2	
47	自己创业与进公司	41	16	2.6	
48	自己创业与进公司	47	15	3.1	
49	自己创业与进公司	27	11	2.5	
50	自己创业与进公司	29	16	1.8	

5.5 Summary

This chapter reports the statistical results of various types of lexical cohesive device used in the HSK Corpus, and includes a preliminary analysis the frequency of the use of non-lexical devices and the use of the lexical devices. Because the focus of the study is to explore the possible patterns of the use of lexical repetition, the following groups of figures are reported in particular: the frequency of the use of different types of repetition, the bond density, the ratio of sentences that form bonds to those that form no bond, the ratio of the bonds formed between non-adjacent sentences to adjacent sentences in a text, the average number of items that form repetition with other items in one single sentence in a text. In Chapter 6, these results will be compared with the results of the findings from the same analysis of the NCEE Corpus to find out whether there is a gap between the two writer groups.

CHAPTER 6: COMPARISON OF THE DATA

6.1 Introduction

Chapters 4 and 5 identified the various types of lexical and non-lexical cohesive devices used in both the NCEE and the HSK Corpora. In comparing the findings from these two corpora, it is aimed to identify firstly the respective patterns of use of these devices in the two corpora, and secondly to discover whether or not there are significant differences in their use by the two writer groups. To provide empirical data to fulfill these aims, the results of statistical analyses of the use of each type of device were reported in each section. As stated in the methodology chapter, the importance of the study lies in the identification of any noticeable difference between the statistical results of the two sets of data. On the basis of these findings, it is hoped to identify some challenges that CFL learners might have in their study of writing in Chinese. This chapter focuses on a comparison of the two sets of data and in so doing, provides an overall summary of the use of the lexical and non-lexical devices in the two corpora. Using various tables and diagrams, the summary is designed to show the similarities and differences of the use of these devices as a basis for the discussion of these features in the next two chapters.

6.2 The comparison of the use of non-lexical devices in the two corpora

As indicated in Chapter 4, although the analysis of non-lexical devices is not the main focus of the study, their consideration here provides valuable and supplementary insights when examining the use of the different types of cohesive device that have a text-organizing role, and it contributes to a more holistic view of both the NCEE and the CFL learner texts. Therefore, because these non-lexical devices are not the central focus of this research, their analysis here is only preliminary, involving examination of their occurrences in 10 texts from each corpus. Tables 6.1 and 6.2 present the comparison of the use of non-lexical devices in the two corpora. This comparison is also exclusive of four types of non-lexical devices, namely reference, ellipsis, conjunction, and substitution, the

justification for their exclusion from this study has been provided in Chapter 4.

Tables 6.1 shows the numbers of the non-lexical devices used in each of the ten texts from the two corpora and their average values, and Table 6.2 lists the results of the comparison of the two set of values. From these two tables, the difference between the two sets of data can be summarized as below:

Table 6.1 *The number of the non-lexical devices used in each text and their average value*

Corpus	Text1	Text2	Text3	Text4	Text5	Text6	Text7	Text8	Text9	Text10	Total	Average
The NCEE Corpus												
thematic chunk	1.33	0	0.58	1.93	0.45	0.21	0	0.83	0	0.22	5.55	0.56
parallel structure	0.24	0.29	0	0.83	0	0	0.71	0.83	0.19	0.44	3.53	0.35
idiom and proverbs	0	0.58	0.58	0.26	0.9	0	0.71	0.28	0	0.22	3.29	0.33
quotation	0	0	0.58	0	0.9	0	0	0	0.38	0.87	2.73	0.27
summary words	0.24	0	0.29	0.26	0.22	0.42	0.47	0	0	0	1.9	0.19
metaphor	0.48	0.29	0.87	0.83	0.45	0.42	0.47	0	0.58	0	4.39	0.44
The HSK Corpus												
thematic chunk	0	0	0	0	0	0	0	0	0	0	0	0
parallel structure	0	0	0	0	0	0	0	0	0	0	0	0
idiom and proverbs	0	0	0	0	0	0.39	0	0.38	0	0.34	1.1	0.11
quotation	0	0.41	0.39	0.39	0.39	0.39	0	0	0	0.17	2.14	0.21
summary words	0	0	0	0	0	0	0	0	1.12	0	1.12	0.11
metaphor	0	0	1.6	0.39	2.3	0	0	0	0	1.35	5.64	0.56

Table 6.2 *The comparison of the use of non-lexical devices and their average values*

Corpus (10 texts)	Thematic chunk	Parallel structure	Idioms and proverbs	Classic quotation	Summary words	Metaphor
the NCEE Corpus (per 100 words)	0.55	0.35	0.33	0.27	0.19	0.44
the HSK Corpus (per 100 words)	0	0	0.11	0.21	0.11	0.56

Firstly, the most obvious contrast that can be seen when comparing Tables 6.1 and 6.2 is the difference in the frequencies of use of these devices. For example, the devices of the thematic chunk and the parallel structure are used frequently in the NCEE texts, but none is found in CFL texts. Table 6.1 shows that among the ten NCEE texts, seven used thematic chunks and seven parallel structures, while there is no CFL text that uses either of the two devices. The second marked difference relates to the use of idioms and proverbs in the two sets of data. The average number of the idioms and proverbs used in the ten NCEE texts is 0.33 per 100

words, which is three times the number of the CFL texts. The third difference of frequency, interestingly, is the numbers of metaphors. The average number of metaphors in the CFL texts is 0.56, which is slightly higher than that in the NCEE texts.

Apart from the frequencies of occurrence of these non-lexical devices, another important difference that emerges relates to their overall distribution in the two samples of texts. While these devices seem to be distributed more or less evenly in the NCEE texts, in the CFL texts, their use is much more sporadic. For example, although the average number of summary words in the NCEE and the CFL texts is similar, with 0.19 in the NCEE texts and 0.11 in the CFL texts, the distribution of this device is significantly different. Among the ten texts, six NCEE texts used this device whereas it is found in only one CFL text, which may indicate that this is an instance of overuse of the device. Similarly, although the average number of the metaphors in the CFL texts is higher, there are only four texts that use this type of device, with Text 5 having a very high average value at 2.3. In contrast, in the NCEE Corpus there are eight texts that used metaphor, and their average numbers are quite evenly spread.

From the above comparisons, it appears that CFL learners do not have the capacity to employ two devices, which are the thematic chunk and the parallel structure. On the other hand, it seems that they overuse the some of the other devices such as metaphor and summary words. The significance of these differences is discussed in detail in next chapter.

6.3 The comparison between the use of lexical devices in the two corpora

In Chapters 4 and Chapter 5, five groups of data are obtained to investigate the actual use of various types of lexical cohesive device in the two corpora. The categories and descriptions of the five groups of data can be seen in Chapter 4, Table 4.1. This section reports the results of the comparison of them between the two sets of data respectively.

6.3.1 Comparison of the A2 data in the two corpora

As indicated in the above table, the A 2 data demonstrate the actual frequency of each type of lexical device in the two corpora. By comparing the frequency, the pattern of the use of these devices may be identified and any apparent differences may be revealed. Table 6.3 reports the comparison of the NCEE and the HSK Corpora. As can be seen from Table 6.3, several differences emerge when comparing the average frequencies of the use of the lexical devices in the two corpora. Below is the brief description of the result of comparison.

Table 6.3 *Comparison of the average frequency (per 100 characters) of each type of lexical cohesive device used in 50 texts*

Categories \ Corpus	the NCEE Corpus	the HSK Corpus
simple repetition	6.22	6.21
complex repetition	0.47	0.56
constituent repetition	0.3	0
simple paraphrase	2.75	1.75
complex paraphrase	3.19	1.57
superordinate-hyponymic	0.9	0
co-reference	0.11	0.08

The most noticeable gap between the two sets of data is that there is no instance of constituent repetition and superordinate-hyponymic found in the CFL texts, while in the NCEE texts the average numbers are 0.3 and 0.9.

The next difference lies in the frequency of the use of paraphrase. In the NCEE texts, the frequency of complex paraphrase is 3.19, which is two times more than that in CFL texts (1.57). Similarly, compared with the average value of simple paraphrase in CFL texts (1.75), the frequency of these devices in the NCEE texts is much higher (2.75).

It is worth noting that there are some similarities between the two groups of data. The frequencies of the simple repetition are nearly the same, and that there are

only a slight difference between that of complex repetition and co-reference.

The results suggest that the TCFL may have little difficulty with using the simple repetition, complex repetition, and co-reference. They however appear to be incompetent in the use of constituent repetition, superordinate-hyponymic, and the two paraphrase devices. It is hoped to investigate these differences and similarities in Chapter 8 and to provide a detailed discussion accordingly.

6.3. 2 Comparison of the B2 data in the two corpora

As discussed in Chapter 4, bond density is an important indicator of the degree of closeness of connection between sentences. By comparing the bond density in the two corpora, it is aimed to discover whether or not there is a significant difference in the use of lexical cohesive devices as a textual organization device between the two writer groups. Table 6.4 presents the figures of the bond density in the two corpora, which clearly show that the average number of bonds formed per 100 Chinese characters in 50 NCEE texts is almost triple of that in the CFL texts. The gap may suggest that the CFL learners have difficulty in employing the lexical repetition to build up connections between sentences. This, however, needs to be investigated and interpreted in specific contexts, which is discussed in Chapter 8.

Table 6.4 *Comparison of the average bond density in 50 texts in the two corpora*

Bond density \ Corpus	the NCEE Corpus	the HSK Corpus
Average bond formed (per 100 Chinese characters)	3.13	1.35

6.3. 3 Comparison of the C2 data in the two corpora

This section focuses on the comparison of the average ratio of the bonds formed between non-adjacent sentences to that between adjacent sentences. The results of the comparison are reported in Table 6.5. As can be seen, the average ratios in both the NCEE and the HSK Corpora are bigger than 1.0, which means that the

number of bonds formed between non-adjacent sentences is bigger than that formed between adjacent sentences. However, the difference between the two corpora is that the average ratio in the NCEE texts is much higher than that in the CFL texts. Again this result is discussed in Chapter 8.

Table 6.5 *Comparison of the average ratio of the bonds formed between non-adjacent sentences to that between adjacent sentences*

Ratio \ Corpus	the NCEE Corpus	the HSK Corpus
C2 (per 100 Chinese characters)	3.4:1	2:1

6.3. 4 Comparison of the D2 data in the two corpora

This section reports the results of the average ratio of sentences that form bonds with other sentences to those form no bonds with others in the two corpora. Table 6.6 presents the comparison of the two sets of results. The table shows clearly that in the NCEE texts, the average ratio (of bond-forming to non bond-forming sentences) is nearly twice the ratio in the texts of the HSK Corpus. The difference between the two may suggest that the CFL learners have difficulty with using lexical cohesive devices to organize the text and to develop the topic. Again, more detailed discussion about the significance of this gap and its possible causes are provided in Chapter 8.

Table 6.6 *Comparison of the average ratio of sentences that form bonds with other sentences to those form no bonds with others in the two corpora*

Ratio \ Corpus	the NCEE Corpus	the HSK Corpus
D2 (per 100 Chinese characters)	1.6:1	0.96:1

6.3. 5 Comparison of the E2 data in the two corpora

In this last section, the focus is on the average number of words that form lexical connections with other words within sentence in the 50 sample texts in the two

corpora, see Table 6.7. The purpose of this comparison is to compare whether or not there is an overall difference in the frequency of the use of bond-making items between the two groups of writers. Since the aim of the study is to identify the lexical patterns in the two corpora, E2 is, therefore, an important micro-level index that reflects the characteristics of the use of bond-making items in the NCEE texts and how the writing of the CFL learners differs in this area. It is hoped that this comparison can further inform the teaching of CFL writing, especially the teaching of lexis. It can be seen that on average, more words are used to create lexical repetitions in NCEE texts than that in CFL texts. This result may indicate that the CFL learners need to realize that the focus of the lexis study may lay in the study of relations between bond-making items in particular context. This may also suggest that learners' lexical competence should take the central place in CFL language syllabus. This, however, needs to be interpreted in Chapter 8.

Table 6.7 *Comparison of the average number of words that form repetition with other words per sentence in 50 sample texts*

E2 \ Corpus	the NCEE Corpus	the HSK Corpus
Average number of words that form with other words in one sentence	4.04	2.6

6.4 Summary

This chapter compares the two sets of data obtained from the NCEE and the HSK corpora. The comparison provides a comprehensive overview of the use of cohesive devices in both the NCEE and CFL texts. Through this comparison, the patterns of the use of cohesive devices in both corpora are made more explicit, findings that can provide the basis for discussion of the teaching of writing in CFL and, in particular, the role of lexical cohesive devices in text organization.

CHAPTER 7: DISCUSSION OF THE USE OF NON-LEXICAL COHESIVE DEVICES

7.1 Introduction

Chapter 6 compares the statistical results of the analysis of the NCEE and HSK Corpora. On the basis of this comparison, the similarities and differences in the use of both the lexical and non-lexical devices identified in Chapters 4, 5 and 6 are summarized and reported by means of tables, diagrams, and detailed descriptions. As stated in Chapter 3, the main aim of the study is to investigate the patterns of use of the linguistic devices used by the two writer groups to achieve cohesion in texts, and to establish whether or not there are significant differences in the use of these devices between the texts of the two corpora. The findings from the study presented in Chapter 6, now provide the basis for a discussion that addresses the main research aim of the thesis along with some possible reasons why the two language groups have similar frequencies of using certain types of lexical devices and why there are considerable differences in the use of other types. The discussion also explores implications of the findings for both linguistic theory and TCFL pedagogy. In this chapter, Section 7.2 discusses the differences and similarities in the use of non-lexical cohesive devices in the two corpora and their implications for TFCL pedagogy. As indicated in Chapters 3 and 5, the analysis of non-lexical devices is not the main aim of the study, the discussion is, therefore, only preliminary and brief.

7.2 The discussion of the patterns of the use of non-lexical devices in the two samples of ten texts

Chapter 6, Section 6.2 reports the results of a comparison of the use of non-lexical

cohesive devices in two samples of ten texts from each of the two corpora. The most significant differences are in the use of the devices of the thematic chunk and parallel structure, with occurrences of seven instances of each of the two devices in the NCEE texts and none in the CFL texts. This marked difference seems to suggest that the CFL learners lack the capacity to use these two types of device to organize texts in their writing. This section discusses this gap, including its possible causes, and explores some pedagogical implications that arise from this finding. In Sections 7.2.3 and 7.2.4, the differences in use of idioms and proverbs and the use of metaphors are also discussed respectively.

7.2.1 Use of the thematic chunk in the two corpora

Chapter 4, Section 4.2 provides a detailed description of the function of the thematic chunk with examples. Clearly, the thematic chunk plays an important role in the organization of the texts in native speakers' Chinese writing. The analysis of the use of thematic chunk where it occurs in the seven sample texts indicates that the employment of this device requires students to master the following skills:

- a. text-structuring skills
- b. meaning-making skills
- c. high level of Language competence

The following example is presented to explain these skills

Example 64 (taken from the NCEE Text 35):

调准感情焦距 拥有清晰世界

[Setting one's emotional focus correctly to have a clear perception of the

world]

(1) 曾听过这样一个故事，一位艺术家去参观一个人体艺术展，哪知道到了展馆却见馆前竖着一块月球表面的地质图。(2) 他怀疑自己进错了展馆，一问方知那块地质图原来是人体放大一千倍的作品，名曰：《人体的真谛》
[I once heard such a story: an artist went to see a body art exhibition. However, on arrival he only saw a framed geological map of lunar surface standing in front of the museum. He wondered he might have come to a wrong exhibition hall, so he asked someone. To his surprise, the geological map was actually the body art which had been magnified 1000 times, and had been named *the essence of the human body*.]

(3) 也许，人对万事万物的认识都如此，距离太近或太远都难以把事物认识全面。(4) 再受感情亲疏远近的影响，对真相的认知则更难了。(5) "君不见多少贪官污吏任人唯亲？(6) 君不见多少笔客学霸结帮成派，互相吹捧？(7) 感情犹如---层雾，遮住了事物的真相;感情犹如一层烟，模糊了我们的眼睛。(8) 我们只有把眼睛擦亮，把感情的" 焦距"调准，才不至于把人体看成地图。 [Man's knowledge of everything in the world may be just like this. It is very difficult to know things comprehensively if one looks at it too near or too far away. It is even harder to know the truth if his perception is affected by his own emotions of likes and dislikes. Can't you see those corrupted officials who appointed people by favoritism? Can't you see those academic bullies who band together to exchange favors and praises for their selfish purposes? Emotion is just like a veil of mist, which covers up the truth; emotion is just like a veil of smoke, which blurs our eyes. Only when we remove the scales from our eyes and set our emotional focus correctly will things like seeing human body as a geological map be avoided.]

(9) 其实，古往今来，不少人都调准了自己感情的"焦距"，为自己展现出了清晰的世界。 [As a matter of fact, through the ages, there are a few people

who had set their emotional focus correctly and therefore had seen a crystal-clear world around them.]

(10) 六祖慧能当初上山学道，五祖弘忍只称其为"南蛮"，而当"菩提本无树，明镜也无台，佛性常清净，何处有尘埃"的绝世偶子展现在弘忍眼前时，弘忍立即调准自己感情的焦距，把衣钵传给了关系较疏远的慧能，而不是自己一向欣赏的人室大弟子神秀。 [In the beginning, the sixth ancestor of Buddhism, Huineng went to learn *dao* on a mountain. The fifth ancestor Hongren regarded him as a barbarian from the south. But when Hongren came to know Huineng's outstanding talent through his conception, which was manifested from the poem he wrote — The linden is not a tree, the shining mirror does not have a stamp either. The essence of the Buddhism is the purification of mind, there is no dust in the world at all — He immediately adjusted his emotional focus and placed the mantle on Huineng, who was in a comparatively distant relationship with him, instead of his top disciple Shenxiu, whom he had always valued.]

(11) 于是，"教外别传，不立文字，见性成佛"，世人的眼睛在六祖禅宗的拂拭下分外清醒。 [As a result, people are inspired by what the sixth ancestor of the Buddhism had gained from his perception, which could be summarized as: teaching disciples in accordance with their aptitude, not being restricted by the doctrines and not sticking to just one pattern, and gaining insights of the nature of the Buddha and getting enlightened completely.]

(12) 康乾盛世，作为两朝"重臣"的张廷玉，当其弟弟张廷露在科举选拔申作弊而被问斩之时，他立即调整了感情的焦距，眼中看到的是国家的利益而非个人的感情，挥泪亲身监斩嫡亲胞弟。 [At the time of "Kong and Qian Era", Tingyu Zhang, the prime minister of two dynasties, immediately adjusted his emotional focus when his blood brother, Tinglu Zhang, was sentenced to be beheaded due to cheating in the imperial examination. What Zhang Tingyu saw

was not his personal emotion, but the nation's interests. It is because of this that he could oversee the execution of his blood brother in person with tears.]

(13) 于是，雍正王朝整顿吏治的榜样高高矗立。[As result, He set a shining example for rectification of official management in the time of Yongzheng Dynasty.]

(14) 还有太多太多的例子，诸葛亮挥泪斩马谩，包青天怒铡包勉，不都是这样吗？(15) 世间事物的真相，都在感情的准确焦距上得以完美展现。(16) 感情上的亲疏远近左右着对事物认知的正误深浅，我们必须以公正、纯洁、无私的心，准确把握感情，调准感情焦距。[There are still a great number of examples like this, such as Zhuge Liang executed his favourite general Su Ma in tears, Zheng Bao executed his nephew Mian Bao angrily with a guillotine. All the truth concerning everything in the world is perfectly manifested through the well-set emotional focus. Whether we can know the truth and how deep we can know the truth depend upon how we deal with our emotion and our likes and dislikes rationally. With an impartial, pure, and selfless spirit, we must, therefore, control our emotion sensibly and set our emotional focus rightly]

(17) 请抹清自己的眼睛，把握自己的感情吧！(18) 只有如此，我们在对世界的摸索中才能拥有一个被准确认识的世界！[Let us remove the scales from our eyes, so that we can control our emotion rationally. Only in this way, can we have a correctly-perceived world as a result of our exploration.]

Text-structuring skill

From the identification of the use of the thematic chunks in the above text, it can be seen the writer has the ability to organize the text structurally from a holistic perspective. As discussed in Chapter 3, the NCEE Corpus texts are task-based

compositions. This writing is a composition responding to the given topic – *On how the emotion of likes and dislikes affects our perception of the world around us*. As can be seen in this text, the phrase 调准感情焦距[setting our emotional focus correctly] is repeatedly used seven times, and the role that it plays in creating the texture indicates that it is a thematic chunk. It seems obvious that the title 调整感情焦距，拥有清晰世界 [setting emotional focus correctly to have a clear perception of the world] is uniquely conceived and constructed by the writer to achieve his or her specific argumentative purpose. In the course of argument, the chunk was firstly used to introduce the argument, and then was employed to develop examples, in which the chunk was again used to highlight the writer's point of view. In the summary it appeared again to stress the writer's opinion. The thematic chunk 调整感情世界 [setting emotional focus correctly], therefore, functions like a thread, weaving all the ideas into a well-organized essay. In the text, the variations of the phrase and their translations have been underlined to enable the reader to identify.

The meaning-making skill

Meaning making here refers to the writer's ability to communicate effectively by making skillful use of a range of cohesive devices. In other words, it is the ability to make a piece of writing coherent so that which the reader is able to infer the writer's communicative intentions. Although the exact relationship between cohesion and coherence is a matter of controversy, it is undeniable that, in this text, the effective use of the thematic chunk makes it easier for the reader to process and to make sense of what they read. From the diagram it can be seen that the writer of the text wants to use the thematic chunk 调整感情世界 [setting emotional focus correctly] to express his or her central idea, which is therefore used as a title. To argue for this theme, the writer uses three examples, argumentative body paragraphs, and summary paragraph to achieve his purpose.

Firstly, Example 1 (Sentences 1-2) tells the idea that in relation to the physical world, a wrongly-set focus cannot help us to know the world correctly. This idea is then further discussed in the following paragraph (Sentences 3-8), in the end of which the central idea is introduced by the use of the thematic chunk. To prove and illustrate the central idea, the writer then presents two more examples explaining the history has proved that setting our emotional focus can help us make rational decisions. Subsequently, in the next paragraph (Sentences 9-14), the thematic chunk was used again in the topic sentence (Sentence 9), and the two examples (Sentences 10 and 12). From a communicative perspective, the use of these chunks helps to order the ideas logically and to piece together the writer's ideas into a logically coherent argument. In the next paragraph (15-18), the writer presents another point, which is that this argument can be applied to the perception of everything in our world. The thematic chunk was again used to help the thoughts and ideas flow smoothly and to connect the ideas to one another.

High level of language competence

The use of the thematic chunk in this text also shows that the writer has high level of language competence. In the discussion of the thematic chunk in Chapter 4, it is mentioned that its form may vary depending on the different sentence patterns or structures within which it occurs. In other words, the thematic chunk is not only used rigidly in its base form; it can be changed so that it is both grammatically and semantically correct in new contexts. For example, in the title and in Sentences 10 and 15, the thematic chunk is in the form of 调整感情焦距 (set emotional focus correctly), while in Sentence 8, it was changed to a 把字句 (*ba construction*): 把感情的“焦距”调准; in Sentences 9 and 12, both variations have the particle 了 [le] 调准了自己感情的“焦距”, which indicates that the action was completed. In Sentence 15, the variation becomes a noun phrase by using a particle 的 [de] 感情的准确焦距. On the face of it, the differences among the

variations may not seem significant, however, the pattern and particles used are some of the most difficult for second language speakers to master, and they help to express very subtle meanings.

To summarize, the analysis of the sample text suggests that the effective use of the thematic chunk requires three skills, namely abilities in achieving the holistic organization of a text, sense-making and knowledge of the systems of the language. The device of the thematic chunk is often employed when the whole text structure is planned and organized, and therefore, it helps to create a well-organized text if it is used appropriately. As one of the non-lexical cohesive devices, it not only has to serve the purpose of making text cohesive, but also that of developing the main theme of the essay in ways that help the reader to process the information more efficient by logically by organizing the thoughts and ideas and presenting the material coherently. This, consequently, requires the writer to have a high level of skill in sense-making through the text. Lastly, to be able to use the device properly, the writer has to vary the form of the device so that both the form and meaning of the structure that it embeds is correct and appropriate in each of its occurrences.

Possible reasons for the knowledge gap between the two writer groups

To understand the gap in the use of thematic chunk in the two corpora between the two writer groups, it is necessary to review the study and teaching of positive and negative rhetoric in the TCFL field. In her article *Teaching Rhetoric in TCFL*, Yu (2004) points out that in the last few years, the theory that Chinese rhetoric includes both positive rhetoric and negative rhetoric has been commonly accepted. She further explains that positive rhetoric refers to the use of figures of speech, such as *pun, metaphor, irony, synaesthesia, allusion and euphemism*, while the negative rhetoric involves the use of phonetic rhetoric, lexical rhetoric, syntactic rhetoric, and textual rhetoric. According to Chen (1979), rhetoric, in a broad sense,

is any effort to convey messages effectively involving the use of language, such as arranging words or phrases. In the light of this, it can be said that positive rhetoric is an active effort that is made to enrich meaning, and to express it more vividly, economically and forcefully by using various figures of speech. The prerequisite for the appropriate use of positive rhetoric is ability to use language correctly. In contrast, negative rhetoric, including phonetic, lexical, syntactic, and textual rhetoric, mainly focuses on how to choose and to use (non-figurative) words, phrases, sentence structures and patterns correctly and appropriately in a given context. Textual rhetoric refers to textual elements, such as the emphasis on the main theme, the logical order of ideas, the proper arrangement of material, the effective use of transitions and cohesive devices, and the correlation between the beginning and the ending.

On the basis of this theory, the thematic chunk can be regarded as one element of textual rhetoric, able to be categorized under negative rhetoric. From the analysis of the above sample text, it can be seen that thematic chunk contributes greatly to the emphasis on the central theme, to the organization of the whole text structure, and to the logical arrangement of the ideas. However, despite the important textual functions of this device, it does not seem to have received adequate attention in the teaching of writing in CFL classrooms.

Yu notes that, in the teaching of writing in TCFL, the teaching of rhetoric, especially negative rhetoric, has been neglected by many teachers, and is a missing element in the teaching of writing. The rationale for adopting this approach, she claims, can be summarized by the two following justifications:

1. Since Chinese is not the CF learners' first language, many scholars claim that teachers should not set unattainable standards for these learners. Their writing should be accepted as long as there are no evident mistakes.

Because of this, there is no need to push them to achieve the perfection in the use of rhetoric devices.

2. As the majority of the CFL learners are of a mature age, they will have been trained in how to use rhetorical devices in their first language. Therefore, as long as their first language writing skills are acceptable, they should be able to grasp the use of Chinese rhetoric spontaneously as they make progress in their learning of writing.

These claims, as Yu pointed out, have proven problematic in the reality of TCFL writing instruction. Many students, especially those who have reached upper intermediate and advanced levels, may be able to produce sentences that have few major errors, and are completely understandable. These sentences, however, often feel awkward and inappropriate in the context. In addition, this type of writing tends to be dull, lacking of variety and polish. In terms of textual organization, many of the text structures are often chaotic, with no distinction between the central theme and minor points. There are also some other common errors, such as paragraphs that are inappropriately divided, beginnings and endings that do not conform to Chinese usage and argumentative patterns that are transferred inappropriately from their native culture.

From the above analysis and the literature review, it seems that the reasons for the existence of the gap in the use of thematic chunk between the two corpora could be for two reasons: the use of this device has not been taught to the CFL learners; and, CFL learners do not have the language competence to use the device. The following two paragraphs provide some discussion of each of these reasons for the non-use of these devices by CFL writers.

Yu's (2004) study, together with other research studies, such as those by Zhou and Peng (2005) and Zhou and Hong (2010) suggest that the primary reason for this

knowledge gap is that insufficient attention has been given to the teaching and practice of this device in the CFL writing classroom. Although Chinese rhetoric has long been one of the major research areas in the field of Chinese linguistics, along with Chinese literature and the contrastive study of Chinese and English, rhetorics has received little attention in the field of TCFL until in recent years. The recent literature also suggests that even now, the focus is still more on the figures of speech, rather than on the negative rhetoric such as phonetic, lexical, syntactic, and textural rhetoric. For instance, in the last two decades, although several studies have researched this area (Chen, 2003; Lu, 1998; Tan, 2004; Yang, 1999, 2001; Yu, 2004; Zhou & Hong, 2010; Zhou & Peng, 2005), only the study by Zhou and Hong (2010) used a large-scale corpus of CFL texts, similar to the one used in this study. However, their study focuses only on the use of common figures of speech and not on negative rhetoric. As a particular type of textual rhetoric, there appears to have been no research on the thematic chunk. This, as a result, suggests that, in the current teaching of CFL writing, no attention has been given to this textual organization device. Without exposure to this device through it being consciously taught, it is not difficult to understand why CFL learners make little or no use of it in their writing.

The second reason is that CFL learners' writing ability may also contribute to the non-existence of the use of the thematic chunk in their writing. Firstly, although the CFL texts that are chosen for this study received high scores, there are still many evident developmental errors in the sample texts, errors that can be found at every level in the texts, namely at the levels of vocabulary, sentence, paragraph and text. These errors indicate that the competence levels of these CFL writers are still insufficiently developed to compare with the native-speaker Chinese writers of the full mark texts that were also chosen for this study. As discussed previously, to be able to use the thematic chunk, the writer has to master three skills: the

advanced text-structuring skills, meaning making skill and high level of language competence. Compared with the native-speaker Chinese writers, the CFL writers can hardly to be said to have these advance skills and, therefore, lack the full capacity to use the device. Apart from their language competence, the CFL learners may also not have access to the traditional Chinese rhetoric, the rich classical literature, the Chinese cultural background, the schematic relation between the culture and the particular given writing topic. In the preliminary analysis of the use of the device, many instances of the thematic chunk are skillfully conceived and structured, either by changing well-known classic quotations such as 怎一个“情”字了得 [how a word emotion could express completely], or by inventing vivid metaphors such as 感情不是保险绳 [emotion is not a safety rope], or by constructing unique expressions such as 勿以好恶论断之 [Never make a judgment about others based on your likes and dislikes]. This capacity of writing skill, is even hard for many Chinese native speakers to master, not mention the CFL learners who are still at a developmental stage.

7.2.2 Use of the parallel structure in the two corpora

This section discusses the second non-lexical cohesive device, the parallel structure, that occurs in the sample of native speakers' texts from the NCEE Corpus, but was not found in the non-native speaker texts from the HSK Corpus. Unlike the rhetorical device of the thematic chunk, which falls under the category of negative rhetoric, the parallel structure is commonly regarded as one type of common figure of speech, which is classified as positive rhetoric. For example, Zhou and Hong (2010) conduct a corpus-based study that investigated how CFL learners at intermediate and advanced level use common figures of speech in their writings. In this study, the various aspects of the use of 16 common types of figures of speech are investigated and analyzed. Among these 16 types of figures

of speech, the parallel structure is one of the types of rhetorical device that is used the most. From the examples presented in their study, it can be observed that the parallel structure, however, seems to refer to those used within sentences. While in this study, parallel structure is more concerned about those used between sentences and paragraphs. Consider the following example

Example 64 (taken from Zhou and Hong, 2010, p. 542)

大家好像很喜欢那个地方:有的跑步, 有的打羽毛球, 也有的开气车上山兜风...那里的活动很丰富。(中级下)

[Everybody seems like that place: some are running, some are playing badminton, and some are going up to the mountain for a drive..., it seems that there are various activities happening out there. (Upper intermediate)

In this example, the three parallel structures are 有的跑步, 有的打羽毛球, 也有的开气车上山兜风 [some are running, some are playing badminton, and some are going up to the mountain for a drive], they are within the sentence. In contrast, in the examples presented in Chapter 4, Section 4.2.2, the parallel structures are all between sentences and paragraphs.

As discussed in this section, the parallel structures in this research are those used to form connections between sentences and paragraphs, not within sentences. Although in the current TCFL literature there are a few studies which have investigate the use of parallel structure, such as Zhou and Hong's (2010) study on the parallel structure used within sentences, their research interest is quite different from that in this research. Zhou and Hong focus more on the rhetorical effect that the parallel structure is used to achieve, such as enhancing the imposing effect of the writing, helping to express the writer's feelings completely, and making the writing more rhythmic. On the other hand, this study focuses more on

its cohesive function in the structure of a text; in other words, how it contributes to making connections between sentences and paragraphs, and as a result, how it helps to organize the writing as a whole. In this sense, the study is more concerned about this device in terms of its textual rhetorical function despite it being traditionally categorized as an element of positive rhetoric.

According to the preliminary analysis of the sample of ten texts from the NCEE Corpus, parallel structures seem to be commonly used across paragraphs, rather than within a paragraph. As a result, the organizational role that it has is similar to that of the thematic chunk. To be able to use this non-lexical cohesive device, CFL learners need to have the three skills as mentioned in Section 7.2.1, namely: text-structuring skills, meaning-making skills, and a high level of language competence. Firstly, since parallel structures used in the argumentative writing often serve the function of presenting main ideas or arguments, and by using the parallel structures, the reasoning of the argumentative essay is often made more powerful, thorough, and compelling, writers have to master the skill of organizing the text structure by constructing the parallel structures before they start to write. This type of pre-writing development requires students to have a high level of text-structuring skills. Secondly, well-constructed parallel structures can only help to strengthen the writers' arguments if they help to make more sense to the reader. In other words, the parallel structure and the meaning that it conveys must be an organic combination. The parallel structure as a form must make the logic of the arguments being presented valid and meaningful. Otherwise, the structure itself becomes a cluster of elaborate patterns that do not serve the purpose of developing the argument. As pointed out by Lu (2011), many essays written in the Chinese writing exam in NCEE are full of fancy words, phrases, and patterns that often convey no meaning of their own. Therefore, a writer has to master the meaning-making skill if he wants to employ the parallel structure in his or her

writing successfully. Lastly, parallel structures require that the elements of a sentence should be alike in meaning and function be alike in construction. To be able to construct a parallel structure, a writer has to attain high level of competence in a number of aspects, such as grammar rules, sentence structure, meaning and textual elements, such as signal words and lexical repetitions.

Based on the above analysis, it seems that the reason for the non-existence of the use of parallel structure in the HSK Corpus is similar to that of thematic chunk. To begin with, the CFL learners may not being taught and exposed to the use of parallel structure in their writing classroom simply because so far there appears to be little research that investigated this area. Secondly, students may be aware of the use of this device in some reading material, or may have made some efforts to learn how to use it. However, their language competence may stop them from trying to use it, especially in an examination environment.

7.2.3 Use of idioms and proverbs in the two corpora

The result of comparison in Chapter 6 also reveals a noticeable gap in the use of idioms and proverbs between the two sets of data. Table 6.2 indicates that the average number of the idioms and proverbs used in the ten NCEE texts is 0.33 per 100 words, which is three times the number used in the CFL texts. In this section, the discussion is dedicated to the exploration of the possible reasons for this gap.

Native-speaker Chinese students have extensive opportunities to acquire idioms and proverbs in Chinese school education

Idioms and proverbs, as two types of fixed expression, have a special expressive function, and therefore, have always been given great attention in both family environment and school education system. As Nida (2001) states, “[i]dioms

usually carry more impact than non-idiomatic expressions because of their close identification with a particular language and culture.” (p. 28). Indeed, idioms and proverbs, especially the former, contain the historical and cultural information of a specific nation, and are the reflection of the nation’s environment, life, history, thinking mode and characteristics. Because of this, the ability to use idioms and proverbs has been highly valued both in daily life environment and in the education system. As a result, many Chinese children as young as three to four years old are often provided the opportunities to be exposed to idiom stories, which are usually in the form of published books, TV programmes, online animations, and DVD multimedia. By the time they go to school, many of them have already accumulated a considerable number of idioms. Likewise, in school education, they are constantly exposed to idioms in textbooks, in conversation, and are taught how to use idioms and proverbs in compositions. This immersion style of learning enables them to have a deep understanding of idioms and proverbs.

As one of the essential parts of the NCEE, the current Chinese university entrance examination always assesses students’ ability to understand and use idioms and proverbs in reading and writing. In the reading test, students are often given materials with some wrongly used idioms and proverbs, and are asked to find the incorrectly-used ones and substitute them with appropriate ones. In the composition test, although there is no clear requirement that states students have to use idioms and proverbs in their writing, the appropriate use of some idioms and proverbs is generally considered likely to impress the composition marker, and to result in a higher score. As discussed in Chapter 3, Section 3.5.1, in the expert comments following the full-mark compositions, the phrase 恰当地使用成语 [be able to use idiomatic expressions effectively] is frequently found to prove that writers of compositions that achieve full marks possess advanced writing

skills.

Thus, emphasis on learning, understanding and the correct use of idioms and proverbs in both daily life situations and education provides native-speaker Chinese students with access to extensive opportunities to acquire the skill in the use of idioms and proverbs. Consequently, when writing compositions in the Chinese NCEE examination, Chinese students, especially those who possess advanced writing skills, are able to employ idioms and proverbs appropriately in their writing, and these fixed expressions play an essential role in creating non-lexical types of connections between sentences.

CFL learners may have inadequate exposure to idioms and proverbs in their learning environment

In contrast with native-speaker Chinese students who are constantly exposed to idioms and proverbs in their surrounding environments including family life and school education system, CFL learners may be disadvantaged by not receiving sufficient exposure to these two types of idiomatic expressions. In this section some of the possible causes of this inadequacy are discussed.

The primary cause for CFL students not using as many idioms and proverbs as native-speaker Chinese students in their compositions may be, in general, due to the lack of consensus in the teaching of rhetoric in the TCFL field. As discussed in Section 7.2.1, Yu (2004) lists two principal reasons why there is the misconceived notion that it is not necessary to teach rhetoric in the TCFL field. She claims that currently the TCFL field has not given enough attention to the teaching of rhetoric, and this is even more neglected in the teaching of CFL writing. This in turn has led to a situation where little effort has been made to

carry out studies on the teaching and acquisition of rhetoric in TCFL. Without a common view of teaching rhetoric and some in-depth studies in this area, it is impossible for teachers to equip themselves with both the necessary theoretical and practical knowledge about the teaching and learning of rhetoric. Similar to the argument presented in the previous section, the lack of realization of the importance of teaching rhetoric has resulted in CFL learners having an inadequate exposure to those commonly used rhetorical devices.

The immediate consequence of not having a consensus about teaching rhetorical devices to CFL learners arises from a lack of theoretical study in the teaching and acquisition of idioms and proverbs. According to Yang (2011), in the last 30 years of TCFL literature, there is a very limited body of research that has investigated idioms and proverbs from a TCFL perspective. Furthermore, studies that have focused on the actual teaching and acquisition of idioms and proverbs in TCFL classrooms are even scarcer; only a few, such as those by Yang (1996), Zhang (1999), Wang (2001) and Hong (2003) can be found in the current literature. As a consequence, little theoretical guidance can be provided to the developers of teaching and learning materials. Pan (2006) points out that although a large number of idioms and proverbs can be found in textbooks at intermediate and advanced levels, the criteria for their selection, the level at which particular ones should be introduced and in which sequence they should be taught are not clearly defined. Under such circumstances, the current situation that relates to the teaching and acquisition of idioms in TCFL classrooms can hardly be described as satisfactory.

Another impact of this scarcity of theoretical study is that there is no single idiom and proverb dictionary compiled specifically for CFL learners. According to Yang (2011), presently around five hundred idiom dictionaries have been

published for Chinese native speakers to use, among these the most popular ones are no less than 20. However, no idiom or proverb dictionary has been specifically developed for CFL learners. Yang further points out that those currently available idiom dictionaries are not suitable for CFL learners for several reasons. Firstly, the definition and annotation of idioms are often very general and, therefore, do not help learners to understand their meanings precisely. Secondly, the examples given are often taken from classical literature, and some of the words and phrases employed often prevent the learner from understanding the idioms themselves. Moreover, there is no explanation given about how the idioms should be used in context, and what the learner should pay attention to in terms of collocation and context. Yang concludes that many types of errors in the use of idioms found in learners writing can be attributed to this lack of reference sources.

In terms of teaching practice, the impact that the lack of theoretical guidance has is considerably evident. The CFL learners may be further disadvantaged by the current orientation of the teaching of idioms in the TCFL classrooms, which as Pan (2006) implied as ambiguous and confusing. As he mentions, the current teaching of idioms does not identify the levels at which they should be introduced, such as according to CFL learner competencies and the idioms' frequencies as well as their levels of difficulty. He suggests that the first differentiation be among the CFL learners. In terms of the former, he argues that, generally speaking, the appropriate level for a learner to learn idioms and proverbs is from upper intermediate to advanced level. Being exposed to idioms too early may actually create fear and anxiety among students; Students should also be differentiated between those who are from the background of Chinese character cultural circle such as Japanese, Koreans, and Singaporean, and those from Western cultures, as the former often have already had some knowledge and foundation of Chinese idioms, while the latter may have no idea about this type of rhetorical device in

Chinese. The second suggestion relates to use, frequency and level of difficulty. He notes that some idioms that are so rarely used that even Chinese native speakers cannot recognize them in some textbooks. This indicates that currently no clear criteria of use frequency have been defined based on the statistics of the current contemporary Chinese literature. The difficulty level of idioms, which should have been considered in the teaching practice, also needs to be carefully studied and categorized, according to Pan. Because of their special structure, meaning, and function in the context, idioms should be classified in detail and given an appropriate sequence in which they should be taught.

The last possible cause may lie in the fact that CFL learners do not have to be tested on idioms and proverbs in the same way as the native-speaker Chinese students. As discussed previously, native-speaker Chinese students have to spend certain amount of time on studying the meaning and the use of idioms proverbs, which, to a great extent, serves a form of consolidation of what they learned. In contrast, there is no a specific type of test questions on idioms and proverbs in the HSK. Therefore, there may be less specific effort made to prepare for this type of question, which in turn deprives them of more chances of consolidation of idioms and proverbs learned from teaching materials or textbooks.

In conclusion, the gap in the use of idioms and proverbs between the two sets of data may be caused by the difference between the amounts of exposure to these devices that the two language groups have received. From the analysis above, it seems that the native-speaker Chinese students are provided with extensive opportunities to acquire these two types of devices, for the reason of both they are the native speakers, intensified by the pressure to pass the NCEE. In contrast, the CFL learners are disadvantaged by not being provided with well informed teaching methodologies, a lack of specifically developed reference books, and not

well organized teaching materials. All of these issues may be due to the lack of a realization of the importance of the teaching of rhetoric in TCFL, and the inadequacy of current research on the teaching and acquisition of idioms and proverbs.

7.2.4 Use of metaphors in the two corpora

It is interesting to note that the numbers of metaphors used in the two samples of ten texts from the NCEE Corpus and the HSK Corpus do not conform with the findings that relate to the other five non-lexical devices. Among all the six types of non-lexical devices compared between the two sets of data, use of the other five devices is similar in that their use in the NCEE Corpus is considerably higher than in the HSK Corpus. The use of metaphors by the two writer groups, however, is the opposite. The average number for metaphor use in the CFL texts is 0.56, slightly higher than that in Chinese texts, which is 0.44. This section provides some brief discussion of this finding.

In their study on the use of figures of speech in the texts written by intermediate and advanced level of CFL learners, Zhou and Hong (2010) find that the frequency of their use of metaphor is lower than that of native speakers of Chinese, which seems contradictory to the findings in this study. However, the Chinese native speakers in their study are primary students in Grades 5 and 6, and the CFL learners are mainly adults. In contrast, in this study the Chinese native speakers are around 18 years old, and the CFL learners are the same age or above. Zhou and Hong explain that the gap in the use of metaphor in their study between the native Chinese speakers and the CFL learners may be due to the different thinking modes of the metaphor users. They go on to interpret that child psychology has indicated that children in Grades 5 and 6 are at the peak of their creative thinking

mode, showing tendencies to be more spontaneously associative and vibrantly imaginative. As a result, children at this age often use significant quantity of metaphors, likening people to objects and vice versa, or comparing one thing to another. However, they gradually grow out of this type of thinking and develop the ability to think about more abstract concepts. Correspondingly, in their later use of language forms as young adults, it can be demonstrated that they use less metaphorical language, and tend to use more abstract and logical thinking skills. In the light of this argument, the contradiction of the findings between this study and Zhou and Hong's (2010) may be explained by the age difference between the native Chinese speakers in this study and theirs. Since the native speaker Chinese in this study are primarily of adult age, the frequency of their use of metaphoric language, according to Zhou and Hong's discussion, may be logically inferred less than that those in Zhou and Hong's study.

This assumption, however, may need a further discussion with regard to the reason why the figure for the frequency of use of metaphor by the CFL learners is slightly higher than that of native Chinese speakers in this study, not the opposite. This may lie in the speculation that CFL learners might employ more approaches to producing metaphoric expressions when writing in the target language than native Chinese speakers writing in their own language. To begin with, the fact that metaphor as a type of figure of speech can be found in the majority of world languages means that it may be less difficult for second language users to acquire. Zhou and Hong (2010) point out that it is commonly accepted that the use of metaphoric language is a universal linguistic phenomenon, and this can be supported by the fact that nearly every language in the world has some types of figures of speech, which often includes the commonly used ones such as metaphor, personification, hyperbole rhetorical question, and parallelism (Zhou & Hong, 2010). These expressive approaches, according to them, are not specific to a

particular language; rather, they are cross-linguistic features.

Another reason why the CFL learners may use slightly more metaphors than native Chinese speakers in their writing is that they could adopt more approaches to producing metaphors, especially when they are still at the developmental stage. When native speaker Chinese students write compositions in their first language, they are more likely to use some commonly accepted metaphors in Chinese culture, or create some of their own that accustomed to the typical Chinese thinking mode. In contrast, when CFL learners write essays in their target language Chinese, they not only use the metaphors they have learned in classroom, or the surrounding environment, but also use other means and resources to help create metaphoric expressions. Zhou and Hong (2010) point out that CFL learners use four principal ways to produce common figures of speech in their writing-imitation, creation, L1 transfer, and the use of communicative strategy. CFL learners often imitate what they have learned in text books or in their daily life to produce figure of speech. According to Zhou and Hong, this is one of the fundamental ways of producing figures of speech and, in general, the expressions that they produce are acceptable, provided that the CFL learners understand the meaning and usage of these figures of speech. The second way is to create their own figure of speech if there is a need in the process of their writing. The acceptability and correctness of the expressions created in this way is dependent upon whether these expressions conform to idiomatic Chinese. L1 transfer could also contribute to the production of figure of speech. Since various languages may have some common cognitive styles, it is possible for learners to transfer some figures of speech that are from their first language to the target language. However, different languages may have different cultural preferences and national characteristics, and the direct borrowing of figures of speech from the L1 may cause some cultural misunderstandings. The last approach that CFL learners may

adopt, according to Zhou and Hong, is the use of communicative strategy. When they lack the target language resources to express themselves, they may be forced to avoid the use of some expressions, and this may unexpectedly produce some sort of figure of speech. As CFL learners are still at their developmental stage of learning, they could actively adopt approaches, such as these four ways, to produce metaphoric expressions in their compositions.

Another factor that may be relevant to this gap is the difference in the distribution of these figures of speech in the data of the two corpora. Overall and detailed discussion about this distinction can be found in next section, but for the convenience of discussion of the use of metaphor, the following points are listed here. Table 6.1 indicates that out of ten native speaker Chinese texts, eight contain metaphors, with two having frequency at about 0.85 and the rest at around 0.4. In contrast, out of the ten CFL learner texts, only four texts have metaphors, with frequencies' at 2.3, 1.6, 1.35, and 0.39 respectively. Clearly, metaphors in these native Chinese speaker texts, which received full marks, seem to just being one of the many devices that used in the texts. For example, the two texts, which use metaphors with frequency at 0.87 and 0.83 respectively, also contain other four other devices, such as thematic chunks, parallel structures, idioms and proverbs, and summary words. In addition, those texts with frequency at about 0.4 also have four other devices, or at least three extra devices used; while in the CFL learner texts that use metaphor, only one extra device can be found in each text except one with two extra devices. On the basis of this difference, it can be reasonably inferred that although both Chinese native speakers and CFL learners have the competence to use metaphors in their writing, the former use it in a more comprehensive way, which means they are competent to integrate multiple non-lexical devices into their compositions. These devices, together with lexical devices, contribute to the creation of the cohesion in the texts in an integrated way.

The latter, however, only uses one or two devices to achieve such cohesive connections, which means that they are probably still at a developmental stage, with the ability only to use two, occasionally three, devices to create cohesion in their texts.

In conclusion, the slightly higher frequency of metaphor used in the HSK Corpus may be explained by the following three factors: firstly, since the use of metaphor in written language is a universal phenomenon, the difficulty of acquiring this figure of speech may theoretically decrease to a certain extent, and this in turn may help to increase the use of metaphor in CFL texts. The other factor is that, as second language speakers, CFL learners may consciously try to adopt more approaches to practising this figure of speech in their Chinese writing than native Chinese speakers do when writing in their mother tongue. These approaches, according to Zhou and Hong (2010), include imitation, creation, L1 transfer, and the use of communicative strategy. The last factor may lie in the inference that native Chinese speakers use metaphor in a more integrated way, while the CFL learners use it at a developmental level. One point worth mentioning here is that the extent to which these approaches contribute to the frequency of the use of metaphor needs further investigation, however, this falls outside of the scope of this study.

7.2.5 The overall distribution of non-lexical devices in the two samples of ten texts

The results of comparison in Chapter 6 also show that there is a significant difference in the overall distribution of non-lexical devices in the two corpora. A close look at Table 6.1 reveals that the five non-lexical devices investigated in this study, namely thematic chunk, parallel structure, idiom and proverbs, summary

word, and metaphor, seem to be evenly distributed in the NCEE texts, while in the CFL texts they are much more sporadic. This section mainly centers on this clear distinction and the possible causes of this. In Chapter 6, Table 6.2 presents the average values of the use of the non-lexical devices and reveals some significant differences between the two corpora. These average values, however, cannot reflect the distribution of these devices in the two corpora. Table 6.1, on the other hand, demonstrates some particular tendencies that are of great significance to discuss.

The first noticeable difference between the two groups of texts is that range of the devices used in native Chinese speaker text is greater than that of non-native Chinese speakers. For example, in average native Chinese speakers use four non-lexical devices in each text, while the non-native speakers use only 1.4 devices. Out of ten native speaker texts, three texts have three devices, four texts four devices, and three texts five devices. While in non-native speaker texts, two texts have no devices, three texts have one device, four texts have two devices, and one has three devices. This indicates that there is only one non-native speaker text that is at the same level as that of native speakers whose writings have less these devices among the ten chosen texts. The other clear distinction is that the frequency of the use of particular device differs greatly in each text. For instance, although the average frequencies of the use of summary words in the two group of ten texts are similar, with native speaker texts at 0.19 and non-native speaker texts 0.11. However, out of the ten texts in each group, six are found to use this device in native Chinese speaker texts, with frequencies at 0.24, 0.29, 0.26, 0.22, 0.42, and 0.47 respectively; while in the non-native Chinese speaker texts, only one text uses this device, with the frequency at 1.12, remarkably higher than that in the other group. Similarly, the use of metaphor in the two groups presents the same trend, which has been described in detail in the previous section where the use of metaphor is discussed.

As discussed in the previous section, the difference in the range of non-lexical devices used and the frequency of each device use in each text in the two corpora reflects the two groups of writers' different levels of language competence. The native Chinese speakers have the ability to employ a wide range of non-lexical devices to create cohesion in texts, as indicated in the figures interpreted in last paragraph. The CFL learners seem to be at a lower level of competence, as they can usually only employ one or two of the devices to do that. Furthermore, it seems that the former use these non-lexical devices at a more comprehensive level. By meaning comprehensive, it means they are able to integrate multiple non-lexical devices into their compositions, and in so doing, these devices, together with lexical devices, all contribute to the creation of the cohesion in the texts in a balanced and integral way. For example, in some of the texts, out of six non-lexical devices discussed in this study, usually four or five of them are used to produce the connections between sentences throughout the text. Compared with this, CFL learners seem to lack the ability to employ a wide range of devices in a text. This may either because they have not mastered the usage of the devices or they do not have the competence to integrate all these devices in their writing in a test environment.

The difference in the frequencies of the use of non-lexical cohesive devices in the two corpora also suggests that those native Chinese speakers whose composition received full marks seem to be at a similar level of competency, as the majority of them use these non-lexical devices with a similar frequency. For example, among the ten NCEE texts, six texts employed at least four different non-lexical cohesive devices, which clearly demonstrate that the native speaker Chinese students are at a similar level of competency in the use of these devices. This, to a certain extent, reflects that well written compositions in Chinese may require constructing and organizing in a similar way in which some of these commonly used non-lexical

devices are employed. In contrast, the use of these devices only exists in very few of the CFL learner texts. For instance, out of ten texts, on average only 2.5 texts use each of the six non-lexical devices, which is significantly lower than the native speaker group. Also the use of these devices is often concentrated in very few texts, such as metaphors are used in four texts, with three at very high frequencies. This may indicate that, although the CFL texts received high scores, they are not at a similar level. Some of them contain relatively high numbers of a particular device, while others have a very low number of that or even zero instances of the same device. This gap may also suggest that the CFL writers of these texts are still at their developmental stage and have not gained both the competence to use and the awareness of using these non-lexical devices in their compositions.

7.2.6 Implications for the study of cohesion and for TCFL pedagogy

This section explores the implications of the findings that relate to the non-lexical cohesive devices for the wider study of cohesion and also for TCFL pedagogy. The former is more theoretical as it is concerned with the question that whether the non-lexical devices discussed in this study is language specific or one of the generic features of human language; while the latter is more practical as it makes some suggestions to the teaching of cohesion in the TCFL field.

Implications for the study of cohesion

On the basis of the findings that relate to these non-lexical devices, two questions relevant to this study arise: the first is whether or not these devices are unique to Chinese or common to other languages; and, secondly, whether or not they are genre specific or can be found across different types of texts. As discussed in

literature review and in the methodology chapter, Halliday's (1976) taxonomy of cohesion includes two types: grammatical cohesion and lexical cohesion. Under the former category, there are four sub-categories: reference, conjunction, ellipsis, and substitution. In the current TCFL literature, this framework has been adopted by many scholars, researchers, and teachers to conduct either research or help to teach writing in classrooms. Clearly, however, the findings of this study suggest that there are other types of device that can also be used to achieve textual cohesion. So the question arises here is that whether or not these devices are cross-cultural variations and cross-genre variations.

Since one of the theoretical frameworks this study adopts is contrastive rhetoric, it is informative to consider the extent to which the findings of investigations of non-lexical devices in the current CR literature relate to this study. If it is the case, that the non-lexical devices discussed in this section can be found to be employed in other languages as cohesive devices, then the commonly adopted taxonomy of cohesion (of Halliday) may be enriched and expanded. For example, Huang (2006) examined non-lexical devices that are employed to create cohesion in civil engineering English texts. His findings suggest that in these science and technology texts, parallel structures, together with verb tenses, reference, substitution, and ellipsis play important role in enhancing the cohesion throughout texts. Poškiene (2009) has also compared the role of parallel structures as a cohesive device between technical texts written by Lithuanian and English scientists. Some other studies, such as those of Ponterotto (2000) and Wei and He (2009) examine the role that metaphor plays in English and Chinese texts and conversation respectively. These studies seem to suggest that in some languages, such as English, Chinese, and Lithuanian, some of these non-lexical devices are important devices that help create cohesion in various types of text. In spite of this, few studies have focused on the frequency of the use of these devices in

argumentative essays in the way undertaken in the present study. However, further investigation may be needed to explore whether or not these devices contribute integrally to the organization of argumentative texts in other languages, the results of which may offer more depth in understanding the properties and functions of cohesion in texts, and therefore may help to expand the theory of cohesion.

In addition to this, the findings suggest the need for rethinking of some of the research methodologies employed in the study of cohesion in the field of TCFL. In the current literature, many studies that focus on the use of cohesive devices in Chinese texts use the taxonomy of Halliday (1976). As a result, the use of this taxonomy predefines the types of cohesive devices they intend to investigate, and following this approach, they proceed to exploring the features of the use of these devices. Although this approach to analyzing cohesion makes an important contribution to the field (given that the purpose of this type of research is to either quantify or qualify the characteristics of the use of these devices), the approach largely neglects the actual use in texts of various other types of cohesive device, especially if some of the devices fall outside of the researcher's framework or have not been identified by any previous research. In contrast, the research design in this study takes an opposite approach. At the initial stage, the investigation begins by analyzing the cohesive ties formed throughout the texts both quantitatively and qualitatively without predefining any categories of cohesive devices that are under examination. The statistics obtained from the quantitative analysis provide a basis for deciding whether or not certain patterns of the use of some devices could be classified as new categories, while the qualitative analysis helps to answer some critical questions such as how and to what extent some new category of device contributes to the creation of cohesion. As demonstrated throughout this thesis, this approach has helped to identify some of the non-lexical devices such as thematic chunk, and lexical devices such constituent repetition,

devices that have not been investigated in detail in previous studies of cohesion.

The approach adopted in this study also suggests that an interdisciplinary study may be necessary to help the researchers to account for the organization of texts in a more holistic way. As acknowledged in previous discussion, some non-lexical devices such as the use of metaphor, idioms and proverbs, classic quotation, and parallel structure have been studied principally from the traditional rhetorical perspective. Consequently, this approach has led to a lack of research on their cohesive functions in the organization of texts. The present corpus-based, quantitative study, however, helps to reveal that these traditional rhetorical devices actually play a vital role in enhancing the cohesion in text. The approach employed here is an example of interdisciplinary perspective (a perspective that draws upon text analysis and traditional rhetoric), a perspective which expands the comprehensiveness of the scope of the study.

Implications for the TCFL pedagogy

As discussed in last section, the findings that relate to the use of non-lexical devices in texts suggest that there is a lack of teaching these non-lexical devices in CFL classroom, stemming from a lack of theoretical study of these devices, especially devices like thematic chunk and parallel structure, in the TCFL field. It would, therefore, seem to be important for TCFL researchers to realize that further and more in-depth studies are needed for the purpose of understanding what constitutes a good Chinese argumentative essay, particularly in this important area of cohesion and coherence. Such an understanding is crucial for the both teachers and the developers of teaching materials. This is because materials and methodologies depend to a large extent upon the degree of understanding of how and what should be taught to learners, (such has how to achieve cohesion and

coherence when writing texts) and to achieve this pedagogic purpose, what materials should be designed and developed. Future studies may also need to use large-scale corpora and well-designed methodologies to uncover the critical features and elements of good essays written by Chinese native speakers. Some of the corpus parameters or variations are also need to be carefully taken into account, such as the type of essay, the writing environment, the age group, the proficiency level, the purpose of the writing, to name a few.

Once the features of competent essays are fully understood, the next step is to generate lists of elements of language knowledge by teaching experts or researchers in this field, lists which can be used to design teaching strategies and methodologies. In the L2 classroom, it is often the case that an element of language knowledge may be easy to describe or define at an academic level, but difficult to teach to learners. This is often because mastering a complex language element, such as a device used to achieve cohesion in texts, may involve factors, such as the meaning and tone that it may convey and the context in which it may be used. The need to understand all of these elements of accompanying knowledge make the use of this type of item difficult for learners to master. When introducing non-lexical, cohesive devices, such as the thematic chunk and parallel structure, it is also important that the learner already has a high level of language competence. Since it is difficult to decide at what level these devices should be taught and what outcomes should be expected, more studies should also be carried out on CFL learner texts to understand better their patterns of the use (and non-use) of these non-lexical, cohesive devices. Further studies could also be carried out to analyse which type of these devices comes u close to native writer's level, and those which are still at a more developmental level. Detailed analysis of factors that contribute to the achievement gap may also of great importance to the researchers, learning material developers, and teachers.

If comprehensive and detailed research studies are conducted, such as those proposed above, the research results could then be summarized as a list of selection principles and criteria that course and learning material developers could use as a reference. If it is found that some of the non-lexical devices, such as the thematic chunk and parallel structure, have not been taught to students, this could indicate that there is potential for improvement in the area of the selection of appropriate reading and writing materials, targeting materials that can be specifically used for learners to acquire these writing skills. In terms of providing exercises, these principles and criteria are also important for the reason that they need to be designed in particular for learners to familiarize, to explore, and to master, at the levels of both individual device and a more comprehensive level, such as in texts in combination with a range of other cohesive devices. For example, a writing course developer first needs to know that lexical chunk is an effective non-lexical device used in the textual organization of Chinese native speakers' writing. Then, when designing writing lessons and exercises for training learners to use this device, he or she can ensure that this device is covered in the learning materials. Texts that demonstrate effective and natural use of the device should be provided to the learner. Materials should also contain a clear explanation of its function, usage, and tailored exercises may also be useful to learners. The expected learning outcome should also encourage the use of the device in learners' outputs.

Teachers are also need to be informed by such research findings so that they can teach these devices effectively. It is logical to assume that a teacher may lack the competence to teach learners how to use these devices without clearly understanding of these devices themselves. The teaching of writing has always been a weak link in the TCFL field, as discussed in Nan (1994), He (2007) and

Peng (2010). The lack of understanding of these devices may be one of the reasons why the teachers find it difficult to plan their lessons or find hard to assess students writings at a more effective and specific level. The findings of this study may suggest that teachers of writing firstly need to be informed by more detailed and comprehensive research in this area. Secondly they also should be able to transfer the knowledge of these findings into practical teaching strategies and methodologies in order to train learners to use these skills at more advanced levels.

The findings of the use of non-lexical devices may also provide some insight into the development of the HSK automatic marking system. The HSK automatic marking system, as discussed in Li (2006), has been one of the major projects in the development of the HSK test system. In the process of essay assessment, it is a common practice for essay markers to have assessment criteria. However, for the automatic marking system, the criteria may be less comprehensive and effective if they are not informed by research findings in this field. The expert comments on the full score receiving essays written by native Chinese students, as provided in the methodology section in Chapter 3, indicate that essay markers often subconsciously use their more traditional knowledge of what are regarded as a good essay, this could be well explained by some of the wordings such as “引经据典 [to cite from the classics or ancient works to prove a point], 旁征博引 [to use well-documented and extensive evidence], 恰当使用成语 [to use idiomatic expressions effectively]. For human markers, the word *extensive*, *effective* do not pose a problem as there is actually a schematic understanding of the level at which use of these devices can be regarded as appropriate, and on the basis of which the assessors may judge the use of these devices well. However, for the automatic marking system, the assessment could be more accurate, effective, and valid if these criteria are qualified and quantified according to well-conducted research in

the TCFL field. The literature review preceding this study implies that few studies have been carried out in the use of these non-lexical devices in the CFL learner texts; it is, therefore, hard to imagine that the degree of the use of these devices has been taken into account in the setting up the assessment criteria employed in the automatic marking system.

7.3 Summary

This chapter discusses the differences in use of non-lexical devices between the samples of ten NCEE and CFL texts and some possible causes for such differences. Implications for the study of cohesion and for the TCFL pedagogy are also provided. In Chapter 8, the differences in use of lexical devices in both corpora and some possible causes for the gap are presented; a brief discussion of the limitations of the study is also included.

CHAPTER 8: DISCUSSION OF THE USE OF LEXICAL COHESIVE DEVICES

8.1 Introduction

Chapter 7 discusses the differences and similarities in the use of non-lexical cohesive devices in the NCEE and the HSK Corpora. Their implications for the study of cohesion and for TFCL pedagogy are also provided. In this final chapter, the discussion mainly focuses on addressing the main research questions asked in Chapter 3, namely whether or not there are significant differences in the use of lexical cohesive devices in the NCEE and the CFL texts, and if so, what the possible contributing factors are. In this chapter, Section 8.2 discusses the possible causes for the differences in the use of lexical cohesive devices in the two corpora. Section 8.3 discusses the differences in the two groups of writers' ability to create high density bonds in their texts and possible factors contributing to this gap. Section 8.4 is devoted to the discussion of the patterns of the average ratio of the bonds formed between non-adjacent sentences to that between adjacent sentences in the two corpora, Section 8.5 is of the patterns of the average ratio of sentences that form bonds with other sentences to those form no bonds with others in the two corpora, and Section 8.6 is of the patterns of the average number of words that form repetition with other words per sentence in the two corpora. In Section 8.7 the discussion is centered on the implications that the findings of this study have for the study of cohesion and the pedagogy in the TCFL field. A brief discussion of the limitations of the study is also included in Section 8.8.

8.2 The patterns of the use of each type of lexical devices

In Chapter 6, Section 6.3.1 summarizes the frequency of the use of each type of lexical device in the texts of two corpora (see Tables 6.3, 6.4, 6.5, and 6.6). The following section focuses on a discussion of the patterns of the use of these

devices, namely constituent repetition, superordinate-hyponymic repetition, simple and complex paraphrase, and the implications of these findings for the study of lexical cohesion and TCFL pedagogy.

8.2.1 The patterns of the use of constituent and superordinate-hyponymic repetition

Table 6.4 clearly shows that, the most noticeable gap between the two sets of data is the use of the lexical cohesive devices of constituent repetition and the superordinate-hyponymic relation. In the HSK Corpus, no instances of constituent repetition and superordinate-hyponymic relations are found, while in the NCEE texts the average numbers are 0.3 and 0.9 respectively.

The gap in the use of constituent repetition

As discussed in Chapter 4, Section 4.3.3, constituent repetition plays an important role in the organization of the text through breaking down words or idiomatic expressions into constituents that can be used as elements of other linguistic forms in that particular context. Constituent repetition is a special type of device that demands advanced language skills, requiring writers to have not only high level of lexical competence, but also the other two skills discussed in Section 7.2.1, namely skills relating to text-structuring and meaning-making.

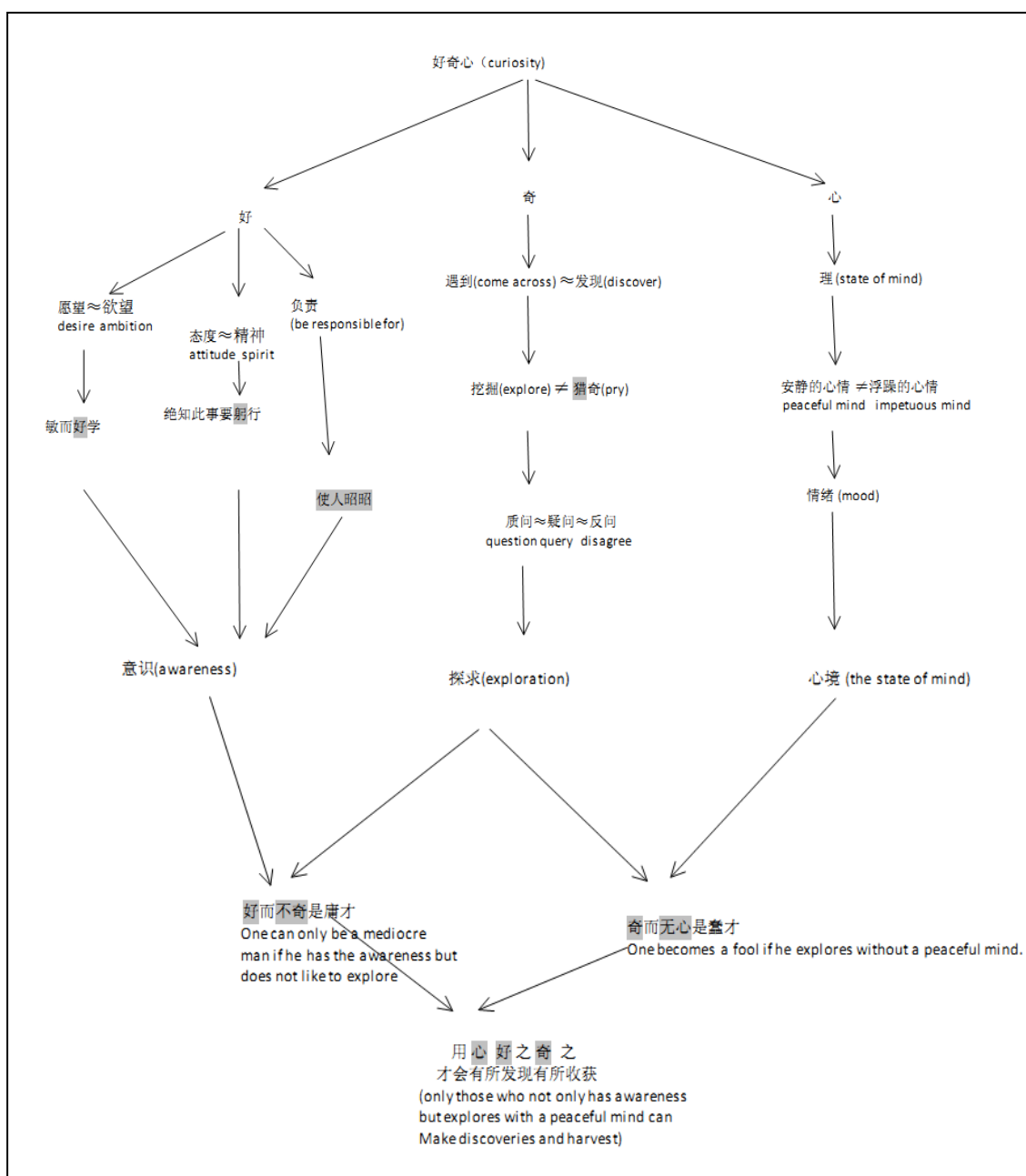
Although lexical competence is mainly used as a term to indicate how big a L2 learner's lexicon is and how well he or she uses the vocabulary (Meara, 1980, 1984, 1996), it is applied in this context to both native Chinese speakers and the CFL learners. As already stated, to be able to use constituent repetition, even a native speaker of Chinese needs to have a high level of lexical competence. To illustrate the point, the example used in Chapter 4 Section 4.3.3 is quoted here to help explain the prerequisite skills a writer should have. In the example, the word 好奇心 [curiosity] is used as the title of the essay, which presents the argument that curiosity enables both individuals and humankind to grow and flourish. To present his argument, the writer deconstructs the word 好奇心 [curiosity] into

three constituents: 好 [desire], 奇 [novelty and to explore the novelty], 心 [the state of mind]. Clearly the combination of the three equates approximately to what the word 好奇心 [curiosity] means: a state of mind in which people desire to explore the novelty or new things. Although each of these constituents, has a narrower meaning than that of the word 好奇心 [curiosity], they do have strong connections with the word. These three constituents are then used as three major points to support the main argument of the essay. Each constituent is defined within the context with explanations and quotations taken from Chinese classic literature. By using this type of repetition, the writer successfully creates unity within the text, while at the same time arguing effectively using the explanation of the three different but crucial aspects. In this example, the writer's lexical competence lies not only in his or her ability to break down words into constituents, but also in his or her ability to develop these constituents into complex points of an argument connected to classical ideas, texts, and quotations. The development of these constituents shows that the writer has mastered the skill of making connections between different semantic networks. Figure 8.1 is an illustration of how the writer uses the semantic works to develop the three sub points.

Figure 8.1 demonstrates a writer's high level of lexical competence. The word 好奇心 [curiosity] is firstly broken down into three constituents, upon which they are further extended to some much broader points that are developed with the complex argument that is communicated. The first part of the illustration is mainly an expansion: synonyms and antonyms are used to interpret the connotations of the three constituents. These are followed by quotations from classical literature to further elaborate the association between the main theme 好奇心 [curiosity] and its crucial elements 好[desire], 奇[novelty and to explore novelty] , and 心 [the state of mind]. In the later part of the illustration, which is more of a summarization, the semantic networks of the three constituents are gradually narrowed down to three key words, namely 意识 [awareness], 探索[exploration], and 心境 [the state of mind]. Within this context, the meanings of three key words are approximately the same as the three initial constituents. Here again, the writer

creates a textual unity by building up connections between these key words and the three constituents. This brief schematic analysis shows the writer's competence in using the semantic networks to achieve unity and a structured organization within the text. Consequently, it can be concluded that this capacity to break the word down into three constituents and to develop the main argument upon them requires that the writer have a high level of lexical competence.

Figure 8.1 *The semantic network of the constituents of the word 好奇心 [curiosity]*



Apart from a high level of lexical competence, the writer also needs to have high levels of skill in both structuring and in meaning-making in order to use this type of repetition effectively. These skills work in a way similar to the thematic chunk that is discussed in Section 7.2.1. In brief, if a writer has a good command of the text-structuring skill, s/he may be able to use constituent repetition to create a well-organized text, as constituents can help to present sub-points that support the main theme. To be able to make sense to readers, the skill of structuring or organizing the text has to be supplemented by the meaning-making skill, which requires the writer to use the constituent repetition device to organize the ideas logically, and present the material coherently, so that the reader can process the information with ease.

According to the comparison data (see Table 6.3) and the above analysis, it is reasonable to speculate that the reason why there is no instance of constituent repetition in the HSK Corpus may relate to the following factors: insufficient exposure to this type of lexical cohesive device, it not being taught in classroom, and hence the lack of capacity to use it. Firstly, the smaller number of instances of the constituent repetition in the NCEE Corpus indicates that this type of device is not among those frequently used by the non-native speaker writers. Compared with the average frequencies of other types of lexical repetition in the NCEE Corpus, such as simple repetition, simple paraphrase, and complex paraphrase, which are 6.22, 2.75, and 3.19 respectively, the average frequency of constituent repetition is only at 0.3. This low frequency may indicate that the likelihood of this type of repetition appearing in teaching materials such as, text books, supplement readings, or handouts in writing lessons is low. CFL learners, therefore, may have less opportunity to observe, imitate, or consciously learn to employ this type of device in their essay writing. For the same reason, teachers of essay writing may not be consciously aware of the existence or the effects of the use of constituent repetition in texts. Alternatively, even if they have an awareness of this device, they may not think that it is necessary or it is possible to teach learners how to use it in their essay writing. After all, learners need to have a very high level of language competence to master it successfully. This leads to the last

factor - CFL learners may lack the capacity to use it. These lacks of a high level of lexical competence, text-structuring skills, and meaning-making skills are very likely to pose barriers for learners to imitate or to consciously use this device in their writing.

The gap in the use of superordinate-hyponymic repetition

Similar to the pattern of the use of constituent repetition, the comparison of the use of superordinate-hyponymic repetition between the two sets of data also presents a significant difference. No instances of this type of repetition are found in the CFL texts while the average frequency of that in the NCEE texts is 0.9. Again, the reason why there is such a gap may mainly lie in CFL learners' level of lexical competence.

As defined in Chapter 4 Section 4.3.6, superordinate-hyponymic repetition occurs in situations where an earlier item is followed by a more general word in the context and the meanings of both items are approximately identical. In essence, the notion of superordinate and hyponym is about meaning inclusiveness. For example, in English, the meaning of *table* is included in that of *furniture*, and the meaning of *orchid* is included in that of *flower*. In other words, these words can be organized and associated to each other according to their semantic relations. When discussing the question of how lexical competence can be determined, Meara (1996) proposes that lexical competence can be measured by both the size of a learner's store of lexical items and the organization of those items. Nation (1990) and Nation and Waring (1997) points out that the organization of words means that an L2 learner understands the spoken and written forms, grammatical patterns and collocations, as well as their semantic, pragmatic, sociolinguistic and register connotations. Clearly, knowing how to organize words and to build up connections between words by sorting them according to their semantic relations is one component of lexical competence. In the NCEE texts, examples of the use of superordinate-hyponymic repetition indicate that these native speaker writers are highly capable of connecting words together within the context to achieve the textual cohesion.

In contrast, there are no instances of superordinate-hyponymic repetition found in the CFL texts. It is, therefore, important to consider the possible reasons for such a gap. On the face of it, understanding the semantic relation between different semantically associated words and its organizational role may not be that difficult for CFL learners, as the CFL writers of the sample texts are adults, which means they are mature enough to understand the logical relation between these words. However, using these words consciously to create lexical cohesive ties between sentences may pose a problem for the CFL learners. Gao (1997) points out that many CFL learners who are native speakers of English may be confused by some mis-matches between some Chinese words, which may be a superordinate in Chinese, and their English translations, which are translated as hyponym instead in English, or vice versa. For example, 山沟, in one text book is translated as *mountain areas*. One meaning of 山沟 is *remote mountain areas*, but it has the connotation of *undeveloped, poor or inaccessible places*. Mountain areas (山区) however does not necessarily have the meaning of *remote* and those similar connotations. Clearly, 山沟 is a co-hyponym of 山区 [mountain area], but they differ somewhat in terms of their referents. This type of mismatch between Chinese words and their English translations, as Gao (1997) argues, often creates barriers for CFL learners to use words correctly and, therefore, hinders them in reaching advanced levels of language competence. In addition to this issue, when discussing how errors arise in CFL learners' texts, Zhu (2009) points out that CFL learners may not know the exact contexts where superordinates and their related hyponyms are used. These factors, therefore, may contribute to the lack of the superordinate and hyponymic repetition device in the texts of the HSK Corpus.

Another reason for their absence in the texts of CFL learners may be similar to that for constituent repetition. Like the average frequency of the constituent repetition, the frequency of superordinate –hyponymic repetition is 0.9, which is higher than that of constituent repetition, but still is much lower than those of other more commonly used lexical cohesive devices, such as simple repetition, simple paraphrase, and complex paraphrase, which have frequencies of 6.22, 2.75,

and 3.19 respectively in the NCEE Corpus. The low frequency of the use of this cohesive device implies that superordinate-hyponymic repetition is not used as much as the three high frequently used ones in native speaker texts. As a result, CFL learners may have few chances to observe, imitate, or learn to use it. For the same reason, CFL teachers may either give little attention to this device, or even have little awareness of its function in organizing a text. This in turn would reduce the opportunities for CFL learners to actively learn how to employ the device in their writing.

8.2.2 The patterns of use of simple-paraphrase and complex paraphrase

From Table 6.4 it also can be seen that there are obvious differences between the texts of the two corpora in their use of the devices of complex paraphrase and simple paraphrase. In the 50 sample texts of the NCEE Corpus, the average frequency of use of complex paraphrase is 3.19 per 100 Chinese characters, while in the HSK Corpus it is 1.57 per 100 Chinese characters; the use of the device in the former corpus is more than two times greater than its use in the latter. Similarly, the average frequency of the use of the device of simple paraphrase in the NCEE Corpus is also much higher than that in the HSK Corpus although the gap between them is not as significant as that between the average frequencies of the use of complex paraphrase (see Table 6.3). The following discussion focuses on the analysis of the gaps between the uses of these two types of paraphrases.

The gap in the use of the two types of paraphrases

As discussed in Chapter 4, Sections 4.3.4 and 4.3.5, simple paraphrase is, in essence, mainly about the synonymous relation between two lexical items although this type of relation is strictly defined for the purposes of analysis in this study. A simple paraphrase only occurs when one item can be substituted by another in context without loss or gain in specificity and with no discernible change in meaning. Likewise, complex paraphrase refers to synonymous and antonymous relations identified on the basis of the established definition. To explore the reasons for this gap in the use of these two types of devices in the two

corpora, the following discussion firstly focuses on the research in the area of synonyms and antonyms in Chinese itself, then moves further to the research in the teaching of synonyms and antonyms in the TCFL field, followed by the discussion of the research in the use of synonyms and antonyms in the CFL texts.

Defining synonyms and the establishing criteria for defining of them have remained controversial issues in Chinese language studies for many years. This controversy has had both negative and positive effects on the study of the teaching of synonyms and antonyms in the TCFL field. According to the several writers (Qiao, 2010; Zhang, 2007; Liu, 2011), the controversy has centered on whether synonyms are *absolute synonyms* and *near synonyms*. Some scholars, such as Lǚ (1952), Hu (1995), and Zhang (1982), claim that synonyms have to be classified as absolute and near synonyms, and focus of the research on synonymy needs to be on near synonyms, as they constitute the majority of synonyms. However, this view has been challenged by other researchers in this field such as Hou (1963), Liu (1980, 1984), and Ge (2003), who both argue that near synonyms are not actually synonyms, and that they should be excluded from the research on synonymy. With regard to the criteria for the discrimination of synonyms, there are also controversies that have affected the teaching of synonyms in TCFL. For example, Shi (1961), Liu (1980) and Qian (1995) emphasize the role that meaning plays in the discrimination of synonyms. They believe that words that refer to same concepts, things or matters can be regarded as synonyms. Others such as Wang and Hou (1963), however, argue that both the form and the meaning have to be considered when assessing the relation between a pair of lexical items. In relation to the issue of whether synonymous lexical items should belong to the same group of part of speech, scholars such as Zhou (1956) and Sun (1982) claim that apart from having to share the same meaning, synonyms must belong to the same word class. On the other hand, Zhang (1956), Lǚ (1979), Xie (1982), and Mei (1987) contend that two lexical items can be defined as synonyms as long as their meanings are the same, and that their word classes do not necessarily have to be alike. Despite these long-standing controversies, the research field has generally reached a consensus about how to discriminate and analyze synonyms. In brief, it is commonly accepted that synonymy can be established in relation to

three aspects of lexical items: firstly, synonyms can be distinguished from the aspect of rationality and meaning. By asking whether their references are the same, what their emphasis is, and the degree or extent of their tones, one may find many differences between a pair of synonyms. The second aspect pertains to whether the synonymous words carry same implications or associations. Thirdly, one can also look at whether the two have a similar usage, which can be examined from the perspectives such as part of speech, collocation.

Similarly, different views also exist in the research of antonyms in Chinese. According to Li (2009), the differences lie in relation to the following aspects: whether or not the part of speech and the grammatical function of a pair of antonyms have to be the same; whether or not their meanings fall within the same semantic scope, for example, 长 [long] and 短 [short] both express the concept of length, and 开 [open] and 关 [close] the action; and whether they are within the same linguistic system, for example, 天 [heaven] and 地 [earth] can form an antonymous relation in Chinese as a result of Chinese philosophy, but it may not form same kind of relation in other languages. Thus, their use in the same pragmatic context identifies some pairs of words as antonyms for reasons based on convention and idiom.

The research on both synonyms and antonyms has had a considerable influence on the teaching and learning of Chinese in the Chinese education system. Synonyms and antonyms are taught to native speaker Chinese students from middle school through to tertiary level. In the Chinese language examinations for the NCEE, it is mandatory to assess whether or not students understand and have the capacity to use synonyms and antonyms. To prepare for this examination, students often spend a large amount of time memorizing these words, understanding them in both ancient Chinese and contemporary Chinese texts, and practising how to use them in the contexts of both everyday texts and examination compositions. Consequently, native speaker Chinese students have acquired the skills of understanding and using synonyms and antonyms naturally in their own cultural environment. As Chinese is their first language, they have gradually cultivated their sense of the language and improved skills in its use. This natural language environment also provides them with opportunities to understand the subtle

differences between synonyms and antonyms in various social, pragmatic contexts. This enables them to have a developed awareness of using synonyms and antonyms in their writing.

The controversy relating to the study of synonyms in Chinese has affected the research and teaching of synonyms and antonyms in TCFL both positively and negatively. From a positive perspective, although the scholars in this field have not reached a consensus on this controversial issue, the related discussion and debate help researchers and teachers in the TCFL field to realize that this is a complicated area, and that more studies are needed to explore further the nature of synonyms and antonyms. The theoretical and practical framework that the scholars adopt, especially in the area of discrimination of synonyms, provides a firm foundation for the teachers of TCFL as summarized by Qiao (2010). Thus, research in the discrimination of synonyms in Chinese itself, especially in how to distinguish the meaning and usage of synonymous lexical items, is the area that the teaching of synonyms in TCFL draws upon most intensively, and its theoretical framework for the discrimination of the meaning and usage has been kept to a large extent in the teaching of synonyms in the TCFL field.

The negative effects of these controversies are also very evident. Because the discussions and their applications are mainly oriented toward native speakers of Chinese, many of the definitions of synonyms and antonyms in teaching materials, textbooks, and methods of discrimination adopted for TCFL classrooms and their related teaching approaches are not suitable for teaching Chinese to second language speakers, which has caused confusion and difficulties for both teachers and students. In the early stage of the development of TCFL, teachers and researchers naturally inherited the tradition of the study of the synonyms and antonyms. They adopted the theoretical framework, the terminologies, and the methodologies of the previous research. However, with the further development of TCFL, many teachers and scholars found that the traditional studies in this area did not satisfy the needs of teachers and the actual requirement of students. For example, Guo (1988) points out that since native speaker Chinese students have a strong sense of the language, they, are hardly confused by those words that are

closely related in meaning, but yet are not synonymous. However, these are the words that pose a greater degree of difficulty for second language learners. Li (2010) points out that most entries of synonyms and antonyms that are seen as confusable in dictionaries are chosen by compilers and researchers, however, the most confusable words in the CFL classroom are those identified by CFL learners themselves, and these two groups of lexical items are very different. This means that the existing literature cannot provide practical help to either teachers or CFL learners. He further notes that the methodology traditionally used for the discrimination of synonyms does not tend to be helpful in CFL classroom teaching. Some teachers, who have been influenced by this approach, as he criticized, try to analyze the differences between words in a comprehensive and complicated way, which is like *carrying out academic research in CFL classroom*, the result of which only confuses the learners and makes it more difficult for them to understand and master this area of language knowledge.

On the basis of the discussion, Zhao and Li (2001), Zhou (2004) and Zhang (2007), propose that the scope of research on synonyms and antonyms in the TCFL field should be expanded, and the perspective of the research should be shifted from the needs of native speakers and oriented toward second language learners. They argue that for native speakers, many lexical items are clearly not synonyms, and will not cause any problems in understanding and using in writing. For CFL learners, however, they are confusable and difficult to distinguish. To address this problem, the definition of synonym should, therefore, be broadened. In this way, many lexical items that have been excluded from dictionaries of synonyms and the research scope of researchers could be included and studied.

They further propose that, since the term *absolute synonym* and *near synonym* cannot cover all the lexical items that pose difficulties for CFL learners. Therefore, these terms should be replaced by the new term *confusable words*. This approach may be beneficial by enabling researchers and teachers to choose those lexical items that cause real problems for CFL learners, not those taken from the synonym or antonym dictionaries by scholars who are unaware of the actual difficulties faced by CFL learners. Li (2010) also proposes new principles for the

teaching of synonyms and antonyms in CFL classroom. Firstly, teachers should only teach learners the findings of the analysis, not the process of the analysis. Secondly, the discriminatory process should be simple yet practical, and should be easy to carry out. Thirdly, CFL teachers should be able to predict the possible confusable words that student might be confused with, and the explanation should only concentrate on the relevant aspect that is central to the CFL students' discrimination problems.

In brief, both the research and the teaching of synonyms and antonyms in the TCFL field have gone through a dramatic change from the conventional research in Chinese to the TCFL discipline-oriented study. This transition not only helps the TCFL teachers and researchers to realize the special needs of the field, but also facilitates more efficient learning if the findings of the relevant research can be applied to the teaching appropriately. From the opposite perspective, however, the transition also means that the research on and the teaching of the synonyms and antonyms in TCFL is still far from comprehensive and mature, which in turn to a great extent affects the CFL learners' capacity to acquire the skills of using them in their writing.

Apart from reasons already discussed, some examination related-strategies may also contribute to the existence of the gap in the use of the two types of paraphrases. According to Luo (1999), students often have an examination-related mentality in which they would rather to use lexical items, grammatical structures that they are more familiar with or more certain about so that they make fewer mistakes or gain higher marks. Some of the strategies that can reflect this mentality are simplification or avoidance strategies that the CFL learners use, especially under test conditions. Since simple and complex paraphrase tend to involve the use of synonyms, antonyms, or lexical items that contain similar meanings within the same context, in a test environment, when a CFL learner wants to use an alternative vocabulary item, or to use synonyms or antonyms to create connections between sentences either consciously or subconsciously, he or she has to search synonyms or antonyms from a list of choices. However, if the student lacks the necessary vocabulary knowledge, he or she may have to

paraphrase a lexical item in his or her own words. Part of the challenge is that each synonym or antonym often carries its unique connotation, and has its own preferred collocations, which potentially makes the correct use of synonyms, antonyms, or paraphrases even more challenging. According to Wang (2009) and Zhang (2008), students often use simplification or avoidance strategies in such circumstances. For example, they use simpler lexical items or those they are more familiar with, or those have broader sense. For example, Wang points out that some Vietnamese students often use 懂 [understand] to replace words like 知道 [know] or 明白 [understand clearly], or use 想 [think] to replace 认为 [consider] and 希望 [hope]. Since 懂 [understand] and 想 [think] are usually taught at lower levels, their meanings are broader, and they are among the most frequently used words, students often use them rather than their synonyms, which tend to be more complex in meaning in different contexts. This suggests that, as a strategy for avoiding the potential misuse of more linguistically complex items, students substitute simpler and more familiar structures or lexical items. The other strategy students often use is avoidance. For example, by analysing his corpus, Fen finds that among the four groups of synonyms below, a considerable number of students would choose to use 懂 [understand], 感觉 [feel], 以为 [presume, think], 发生 [take place, happen] and avoid using other synonyms.

(1) 懂 [understand], 明白 [understand clearly] 理解 [comprehend], 了解 [find out, investigate]

(2) 感觉 [feel], 觉得 [think]

(3) 认为 [consider], 以为 [presume, think]

(4) 产生 [bring forth, exert], 发生 [take place, happen]

Such examples may indicate that the simplification and avoidance may contribute to the lower frequency of the use these two devices in the HSK Corpus.

8.3 The patterns of bond density in the two corpora

In Chapter 3, Section 3.5.2 and Chapter 4, Section 4.3.2 the notion of *bond* and the significance of the comparison of bond density were discussed. In Chapter 6,

Table 6.5 shows that the average number of bond formed per 100 Chinese characters in the 50 NCEE texts is almost triple of that in the 50 CFL texts. This section is devoted to a discussion of this gap between the two corpora and its possible causes.

As discussed in 3.5.2, *bond* is the term that Hoey (2000) uses to measure and examine the interconnectedness between sentences in a text. According to him, lexical items form links between sentences and sentences with three or more links are considered to form bonds. In the light of Hoey's (2000) claim that "density of repetition betokens closeness of connection" (p. 126), the term *bond density* was introduced in Section 4.3.2 to measure the extent to which sentences, especially sentences that are physically distant from each other in a text are interconnected in linguistic and semantic terms. Since "bonded pairs are almost invariably related and frequently coherent" (p. 183), a smaller number of bonds may indicate that sentences in a text are less closely related, and that the text has less coherence than a text with a greater number of bonds among its sentences, indicating that the sentences are more closely interrelated. Although there are variations in bond density in both the NCEE and the HSK Corpora, the average bond density in the two corpora differs markedly. In the NCEE Corpus, the bond density varies from 0.76 to 5.54 per 100 Chinese characters, the average bond density is 3.13 per 100 Chinese characters, while in the HSK Corpus, the highest bond density value is 3.29 and the lowest is 0, and the average value is 1.35 per 100 Chinese characters.

As quoted in last paragraph, Hoey (2000) claims that "density of repetition linkage between sentences is a measure of their closeness of connection" (p. 125), to understand the gap between the two average densities, however, it is important to discuss further the nature of the *closeness of connection*. Firstly, at the lexical level, closeness of connection indicates how closely the sentences in a text are lexically interconnected with each other. For example, a bonded pair (of sentences) that has the co-occurrence of three lexical cohesive devices (the number is frequently three, as discussed in Chapter 3 Section 3.5.2) is more lexically related to each other than a non-bonded pair of sentences, because the former has three pairs of lexical items that are repeated in a certain way, while the non-bonded

sentences may have two pairs or no pair at all. It is likely that a text that has high bond density is more lexically interconnected than a text that has low bond density because more sentences in the former can either build lexically upon what has gone before or provide the lexis for subsequent statements. Secondly, at the semantic level, closeness of connection indicates something of the level of the intelligibility of a text, in other words, how close the sentences in a text semantically interconnect to each other so that a reader can interpret those connected sentences as holding together in some way. Hoey (2000) suggests that “the co-occurrence of the requisite number of repetitions is sufficient to compel a reading of the pairs as intelligible, in other words, bonded sentences make sense together in their context without the intervening text” (p. 126). This implies that bonded sentences are not only lexically interconnected to each other, but also are semantically related to each other. To illustrate this point, Hoey (2000) used a parallelism to demonstrate the both lexical and semantic relatedness between bonded sentences. As an example from the data, the following two pairs of bonded sentences taken from Text 4 of the NCEE Corpus, one is formed between adjacent sentences, the other non-adjacent sentences.

Example 65: A bond formed between adjacent sentences

(1)	≈	(2)
在我所听闻过的 [among what I read and heard]	≈	翻开历史，尤其是中国的历史， [from the history, especially Chinese history, I read]
悲剧中， [among the tragedies]	≈	君王的昏庸、王朝的颓败、忠臣蒙"莫须有之罪" 奸候飞黄腾达的段子 [the plays about the fatuous emperors, the corrupted regime, and how the loyal court officials were under injustice, and how the treacherous ones got rapid promotions]
占大多数的， [in the majority]	≈	一折接一折， [a play after another]

情感	=	这大概就是情感
[emotion]		[this is probably the emotion]
犯下的错是_____	≈	惹下的不小的祸吧
[the mistake made_____]		[caused the big troubles]

The columnar layout demonstrates the parallels in the two sentences, and the dotted line and arrow indicates that the phrase 占大多数的 [in the majority] is originally in the sentence 情感犯下的错是占大多数的 [the majority of mistakes are made because of emotions], but for the purpose of displaying the parallels it is placed in the line that the arrow points to so that it can form a parallel with the phrase in the same line 一折接一折 [one play after another]. The words or phrases in bold indicate that they are lexical repetitions, and the sign = indicates that two are equivalent or approximately equivalent, and the ≈ sign indicates that the two are similar in meaning and, therefore, are closely related.

Sentence (1) and Sentence (2) form a bond. Firstly, three different types of lexical repetitions can be identified. In this context, 悲剧 [a play dealing with tragic events and having an unhappy ending] in Sentence (1) forms a complex paraphrase with 段子 [a play] in Sentence (2); 情感 [emotion] in (1) forms a simple repetition with 情感 [emotion] in (2); and 犯下的错 [made the mistakes] in (1) forms a complex paraphrase with 惹下的祸 [caused the troubles] in (2). In addition to this, because of these repetitions, the two bonded sentences are also semantically related. To explain this, it is necessary to compare the meanings of the phrases from line to line so that the covert semantic parallels can be revealed. In the first line, although the phrase 在我所听闻过的 [among what I read and heard] has a broader meaning than that of 翻开历史，尤其是中国的历史 [from the history, especially Chinese history, I read], the two can be said to be semantically related to each other to a certain extent. In the next line, although the phrase 君王的昏庸、王朝的颓败、忠臣蒙“莫须有”之罪、奸臣飞黄腾达的段子 [the plays about the fatuous emperors, the corrupted regimes, and how the loyal court officials were under injustice, and how the treacherous ones got rapid

promotions] does not contain a lexical item that means *tragic, tragedy, or unhappy ending*, however, the whole phrase does imply that the plays are tragic, because the loyal court officials are supposed to be promoted and awarded, yet they suffered a lot. Although the treacherous should be punished, they get rapid promotions. Hence it still has a semantic relation between the two phrases, and the reader can associate them as an intelligible pair. In the third line, the first appearance of the two phrases does not indicate any lexical or semantic similarity, however, the phrase 一折接一折 [a play after another] contains a syntactic structure 一...接一..., which, in this context, is used to exaggerate that there are a large number of plays. The phrase 一折接一折 [a play after another], therefore, to a certain degree has a close semantic relation to the phrase 占大多数的 [in the majority]. 情感 [emotion] and 情感 [emotion] in the phrase 这大概就是情感 [this is probably the emotion] in the next line are equivalent, so they are closely related semantically. The last line is also very clear, because the two phrases 犯下的错 [made the mistakes] and 惹下的不小的祸 [caused the big troubles] are approximately equivalent in this context, and they can be said to have a close semantic interconnectedness.

Example 66: a bond formed between non-adjacent sentences

(2)

翻开**历史**，尤其是中国的**历史**，
[from the history, especially Chinese]

君王的昏庸、王朝的颓败、忠臣
蒙“莫须有”的罪名、奸臣飞黄
腾达的
[the fatuous emperors, the corrupted
regimes, and how the loyal court
officials were under injustice, and how

(9)

当然，**历史**上，现代社会中
[however, in history, and in
modern society]

the treacherous ones got rapid
promotions]

段子一折接一折
[plays, one play after another]

≈

也有很多的大义灭亲的**事例**,
[there are also many examples of
how people sacrifice the ties of
blood to uphold righteousness]

这大概就是
[this is probably]

≈

不会
[(these examples) are not]

情感
[emotion]
closeness]

≈

单凭着**感情**的亲疏
[only based on their emotional
closeness]

惹下的不小的祸吧
[caused some big troubles]

≈

胡乱判案模糊执法
[carelessly deals with cases and
irresponsibly enforce the law]

In this example, there are three lexical repetitions, which are formed between the lexical items in bold in Sentences (2) and (9). These repetitions indicate that the two sentences are lexically closed to each other. At the semantic level, again there are similarities between each line. In particular, in line 3 段子一折接一折 [the plays, a play after another] and 也有很多的大义灭亲的事例 [there are also many examples of how people sacrifice the ties of blood to uphold righteousness] are closely related semantically, because 一...接一... [one ... after another] means 很多[many] , which is repeated exactly in the opposite phrase. In the next line 这大概就是 [this is probably is] forms a contrast with the phrase 不会 [(these examples) are not]. This contrast reflects that the former is presenting a statement, and the latter is offering a denial of the statement; they are, therefore, also a closely related semantic relation. The last line does not present to us a lexical closeness, however, the phrase 胡乱判案模糊执法 [carelessly deals with cases and irresponsibly enforce the law] implies that this kind of act would cause

serious trouble; the similarity is between the inference of their meanings.

The above two examples clearly show how bonded sentences are lexically and semantically interconnected to each other. The next examples taken from Text 4 of the NCEE Corpus are presented to show that non-bonded sentences are less or not connected to each other lexically or semantically.

Example 67: two sentences that are not bonded in a text

(2)

翻开历史，尤其是中国的历史，
一试一试考
君王的昏庸、王朝的颓败、**忠臣**
蒙“莫须有”的罪名、**奸臣**飞黄
腾达的段子一折接一折，这大概
是情感惹下的不小的祸吧
[From the history, especially Chinese
history I read, there are many plays
about the fatuous emperors, the
corrupted regimes, and how the loyal
court officials were under injustice,
and how the treacherous ones got
rapid promotions]

(3)

侯臣们大抵也都是科举之人，
上来的，也大抵有着灵活的头脑和如簧的
巧舌，三言两句套套近乎，便
成了亲信，
成了近宠。
[These court officials were
selected through
imperial examination system after
numbers
of selections, they therefore must
had been
very clever and persuasive. By
winning the
emperors' favour, they quickly
became the
emperors' trusted followers or
favourite ones.]

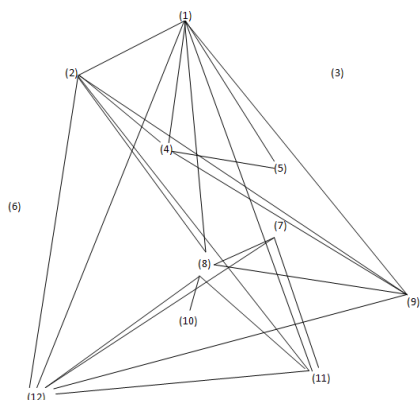
In this example, there are two lexical repetitions, **忠臣** [loyal court officials] and **侯臣们** [court officials], and **奸臣**[treacherous court officials] and **侯臣们** [court

officials]. However, it can be seen clearly that there is no close semantic relation between the two sentences.

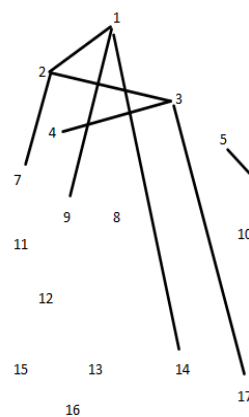
From the above three examples, it can be seen that the closeness of connection is at both lexical and semantic levels, and bonded sentences are more closely related through lexical repetitions, and are more intelligible semantically. This demonstration of the degree of closeness of connection between bonded and non-bonded sentences shows why bond density has considerable significance in contributing to the text organization. This is further illustrated through a comparison of the net of bonds of two texts taken from the NCEE and the HSK Corpora, with each has a bond density at about the average level of its own group, see Figure 8.2.

Figure 8.2 An example of the contrast between the average bond densities in the two corpora

The net of bonds of NCEE Text 4



the net of bonds of CFL Text 1



The two nets of bonds clearly show a marked difference in terms of the interconnectedness between sentences. In the NCEE text, the bonds are denser and there are far more sentences interconnected with each other. The higher bond

density in NCEE Text 4 reflects that the interconnectedness in the text is complex yet occurs in an organized way. By *complex*, we mean that there are bonds that are created in a linear way-between adjacent sentences, such as (1) - (2), (4) - (5), (7) - (8) - (9), (11) - (12), there are also bonds that formed in a non-linear way-between non-adjacent sentences, such as (1) - (4), (1) - (5), (1) - (9), (2) - (11), (2) - (12), (8) - (11), (8) - (12). The former contributes greatly to the development of a subtopic within a paragraph, while the latter often contributes more to the development of the central topic of a text. The two different kinds of bond, therefore, can be said to play a vital role in the organization of the text. In other words, the lexical relationships between adjacent and non-adjacent sentences help the writer to relate all the relevant ideas together as a complete proposition. By *in an organized way*, we mean that the bonds are not just created serving no purpose; instead they are structured in a way in which each bond has a certain role to play in the development of the main theme. For example, the bond between Sentences (1) and (2) is used to set up the topic, whereas the bonds between Sentences (8) - (9) and (8) - (10) support a claim within a paragraph, and that between Sentences (11) - (12) is employed to close down the discussion. It is obvious that it is necessary for the bond density to be at this level so that the writer can create a cohesive text and so that the reader can interpret the bonded sentences as belonging together and achieve a coherent reading and interpretation of the text.

In contrast, since CFL Text 1 has a low bond density, it seems that it is less well organized and the sentences are less interconnected to each other. Among the eight bonds found in the text, four are formed between adjacent sentences, namely Sentences (1) - (2), (2) - (3), (3) - (4), (5) - (6), while the other four bonds are formed between non-adjacent sentences, namely Sentences (1) - (9), (1) - (14), (2) - (7), (3) - (17). At the sentence level, out of the seventeen sentences, there are ten sentences that are lexically and semantically connected to each other, and the remaining sentences do not form bonds with other sentences in the text. As a result, they do not contribute directly to the development of the main theme. At the text level, it is clear that Sentences (1), (2) and (3) play important roles in organizing the text, as they each forms three bonds with other sentences. Sentence (1) serves as a topic-opening sentence, as it is the first sentence in the text, while

Sentences (2) and (3) may serve as a transition or sub-topic opening sentences. A close look at net of bond of the NCEE Text 4 shows that there are two important sentences at the end of the text - Sentences (11) and (12), because they make a large number of connections with previous sentences, some connections are adjacent, and some are at a great distance. The role of these connections is to conclude the discussion by summarizing or referring back to the main topic at the beginning. In contrast, the CFL Text 1 has only two sentences, (14) and (17), that refer back to previous sentences, which do not seem to contribute much to the summary of the discussion.

Possible causes for the low bond density in the HSK Corpus

The above discussion suggests that bond density of a text may directly affect the lexical and semantic interconnectedness between sentences, and consequently, may exert a considerable influence on the patterns of organization of a text. Consequently, the low overall level of bond density in the texts of the HSK Corpus appears to reflect the low capacity of CFL learners to create bonds between sentences in texts, which naturally lead to a number of less well organized essays in the corpus. Possible causes for this low capacity may relate to three major factors: the CFL learner's low level of lexical competence, text organizing skills and a lack of practice in organizing lexical items and semantic relations through pre-writing preparation activities.

The first possible cause of the low bond density in their texts may be related to the CFL learners' low level of lexical competence. As stated in this Chapter, Section 8.2, lexical competence is concerned with two factors: the size of a learner's vocabulary and his or her ability to organize that vocabulary. Hoey (2000) states that the most important connection between various language components, such as phonology, syntax, lexis, and text, is the connection between lexis and text, as "lexical selection affects or creates cohesive links that, as we have seen, help organize the text" (p. 217). Taking this viewpoint, *lexical selection* in this study concerns two aspects of learner writer competence: firstly, the extent to which the lexical items chosen by a learner are lexically and semantically related to each

other; and, secondly, the level at which a learner can use the various repetition devices effectively. It would appear that the ability to choose both lexically and semantically related lexical items is greatly dependent upon the size of a learner's vocabulary, and their effectiveness in the use of a variety of lexical cohesive devices is decided by their capacity to organize their vocabulary. For the first aspect, the ability to select both lexically and semantically related lexical items appropriate to a context, a detailed discussion will be presented in the following Section 8.6 when the last group data (see Chapter 6, Table 6.3) is interpreted. With regard to the second aspect, the data in Chapter 6, Table 6.4 have indicated that CFL learners have a lower level capacity to use some of the lexical devices. Among the seven lexical cohesion devices analyzed, namely simple repetition, complex repetition, constituent repetition, simple paraphrase, complex paraphrase, superordinate-hyponymic competition, and co-reference, CFL learners can only use simple and complex repetition and co-reference at a level that is similar to that of native Chinese speakers. On the other hand, they seem to have considerable difficulty in employing the other four devices in their writing, which indicates that CFL learners tend to lack the competence to use a range of lexical repetition devices. Therefore, their capacity to organize the vocabulary is low. This argument can also be supported by the comment made by Stotsky (1983) and quoted by Hoey (2000), which says that an important index of growth in English writing is the increase in the use of morphologically complex words rather than simple repetition of words or phrases (p. 244). Similarly, in terms of writing in Chinese, an important indicator of competency would be the level of the use of the lexical repetition devices frequently used by competent native Chinese writers. Detailed discussed about this gap has been presented in the previous section of this Chapter, where the CFL learner's ability to use constituents, superordinate and hyponymic repetitions, and synonyms and antonyms and other types of paraphrases are discussed and exemplified.

In addition, the ability to organize vocabulary in this study is also concerned with the CFL learner's semantic organization capacity in their Chinese writing. When discussing the theoretical implications of lexical analysis, Hoey (2000) claims that "a textual study is in a real sense a lexical study" (p. 221). In the same chapter, he

states that “given this partial inverse relationship between lexis and the semantic organization of text, it becomes imperative to investigate how lexical cohesion and text organization affect each other” (p. 220). Hoey suggests that lexical and semantic organization are closely related to each other and should be discussed together. To this end, I would like to refer to Figure 8.1 again to explain how lexical and semantic organization are intrinsically connected to and affect each other. In Figure 8. 1, the writer discusses the topic 好奇心 [curiosity]. At the semantic level, the Chinese native speaker writer first breaks down the key word 好奇心 [curiosity] into three constituents: 好 [desire], 奇[novelty and to explore novelty], and 心 [the state of mind]. These three key components serve as sub points and are further developed in detail separately. The discussion is then gradually narrowed down to three different key words 意识[awareness], 探索 [exploration], and 心境[the state of mind], which are synonymous with the three constituents 好[desire], 奇[novelty and to explore novelty], and 心 [the state of mind]. Finally in the conclusion, the writer joins the three sub-points together and makes a conclusion which refers back to the main topic in the beginning. In terms of lexical organization, the vocabulary items in Figure 8.1 suggest that they are appropriately chosen and, therefore, the lexical relations among them accurately mirror the semantic structure of the essay, in other words, semantic organization and lexical selections appears to mirror and to support each other.

The low bond density in the HSK Corpus, which is the linguistic form of the texts, therefore, firstly seems to reflect the CFL learners’ lower level of lexical competence, which in turn indicates their lower level of semantic organizational skill in their target language writing. Although CFL learners in this study are mostly mature adults, and therefore, it can be safely said that their reasoning skills are at the same level as that of native speakers Chinese, they may have difficulties with organizing their ideas and making appropriate choices of lexical items when writing in Chinese. This claim can be supported by Wu’s (2011) study, in which he states that when writing in Chinese, many CFL learners often feel frustrated. They are often unsure about what to write, and find it hard to recall the lexical items they want to use. As a result, they have trouble expressing themselves in

Chinese and their writing does not flow smoothly and coherently. This observation suggests that many CFL learners lack competence in organizing their ideas and lexical items when writing in Chinese. The contrast shown in Figure 8.2 is evidence for CFL learners' lower level of text-organizational skill. The high bond density in the native speaker Chinese text shows it to be well-organized both semantically and lexically, while the low bond density in the CFL text indicates that the CFL learner has difficulty in organizing the text at both the semantic and lexical levels. Evidence for this difficulty is that a large number of sentences do not form any bonds with other sentences, which means they are not interconnected with each other either semantically nor lexically.

Apart from the two above factors, the lack of practice in organizing lexical items and semantic relations may also contribute to the low level of bond density in the CFL text. Wu (2011) points out that writing competence is a comprehensive reflection of a writer's aptitude, such as mental and psychological abilities, and a group of learned skills, such as the language skills, analytical skills, and cognitive skills. Although CFL learners may have the same level of intelligence as native Chinese speakers, and they may have high level of writing competence in their first language, they still need to practise these writing-related skills in order to think and write effectively in Chinese. Providing this type of practice requires teachers to have a well-designed syllabus and effective activities, and students to understand the importance and effectiveness of these activities. On the basis of this theoretical framework, Wu (2011) conducted an experiment involving his CFL learners that involved introducing pre-writing preparation activities to better prepare them for writing texts. The result shows that this type of intervention helps learners to improve their writing skills, including the development of argument, the organization of ideas, the use of vocabulary, and other components such as grammar, punctuation, text style, and word limits. He summarizes the pre-writing preparation activity as four main stages:

- (1) Constructing scaffolding: CFL learners write down as many words or phrases as they can that are related to the given topic.
- (2) Entering the context: learners circle each key word and write down as

many words or phrases that are associated with them.

- (3) Continuing the organization: learners categorize the ideas expressed by the words and phrases they have generated and upon which make an outline of the topic.
- (4) Exploring the writing independently: learners write their own essays according to the outlines that they have generated (pp. 107-108).

Clearly the first two stages are mainly concerned with the lexical choices, and the third stage is a process of lexical and semantic organization, the last stage is mainly the process of jointing the ideas together and turning them into a well-organized essay. By doing this kind of exercise after a period of time, he found that the average score of learners in the experimental group is much higher than that of the learners in the control class. He also contends that the higher the learners level are, the more effective these types of exercise are, as more advanced students often have a larger vocabulary that they can draw upon. This means that they store more lexical items in contexts, and they have quicker access to them, which in turn may help the learners to organize the lexis and semantic relations so that they can make connections between what they are currently saying and what they have previously said and what they are going to say.

Although this approach of using staged pre-writing activities is effective, as Wu says, it is still in its infancy and needs to be improved in many ways. However, there are still many writing classes where students are only given the topic and are told to write the essay on their own without any guided pre-writing activities similar to Wu's approach. This implies that there are not many CFL learners have been taught to do this type of exercise, and therefore, this practice may take long time to be accepted by the majority of the CFL writing teachers. Without such practice, many CFL learners may still be struggling with organizing their text effectively in Chinese.

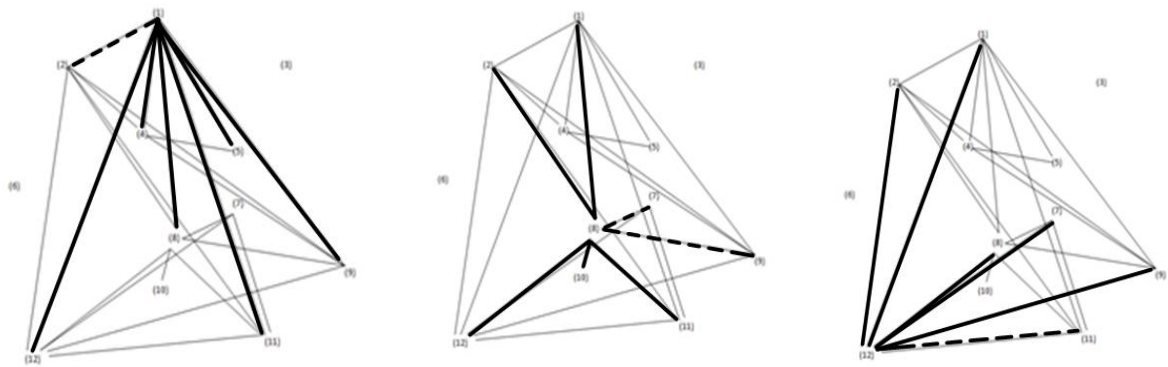
8.4 The patterns of the average ratio of the bonds formed between non-adjacent sentences to that between adjacent sentences in the two corpora

In Chapter 6, Table 6.5 presents data from the two corpora that show the difference between the average ratios of the bonds formed between non-adjacent sentences to those between adjacent sentences. In the NCEE Corpus, the average ratio of the 50 texts is 3.4:1, while in the HSK Corpus it is 2:1, which means that on average there are more bonds formed between non-adjacent sentences in each NCEE text than in each CFL text. This section discusses the significance of this difference and some of the possible factors that may contribute to the lower ratio of the bonds formed between non-adjacent sentences to that between adjacent sentences in the HSK Corpus.

When discussing the implications of lexical analysis for reading and writing, Hoey (2000) emphasizes the importance of writing in non-linear ways. By *non-linear*, he means using repetition devices to make connections between sentences across the text, in other words, between non-adjacent sentences. In contrast, *linear* means the connection between adjacent sentences. Although both non-linear and linear ways of connection in a text are important to create cohesion, Hoey's (2000) findings, as well as those of other scholars', such as Berber (1994) and Jin (2008), have shown that in argumentative texts, most bonded sentences are not adjacent. The statistical data in this study is also in accordance with their findings. Berber (1994) points out that "distant bonding is a very common, indeed natural feature of texts" (p. 1). In order to explore the possible causes for the smaller number of bonds formed between non-adjacent sentences by CFL writers, it may be helpful to understand why the number of bonds formed between non-adjacent sentences is generally greater than the number of bonds formed between adjacent sentences. In Section 8.3 of this chapter, Figure 8.2 illustrates the relative interconnectedness between sentences within the NCEE and the CFL texts. In Figure 8.3 following the network of bonds of the NCEE text is presented again with some highlighted features. There are three small diagrams in the figure, the left one is to demonstrate the bonds formed between Sentence (1) and other sentences, one is formed between adjacent Sentences (1) and (2), which is highlighted as a bold

dotted line, while the rest highlighted solid lines represent the bonds formed between non-adjacent sentences (1) - (4), (1) - (5), (1) - (8), (1) - (9), (1) - (11), (1) - (12). Likely, the middle one shows the bonds formed between Sentence (8) and other adjacent and non-adjacent sentences, and the right one shows Sentence (12)'s connections across the text.

Figure 8.3 An example of the number of bonds formed between adjacent and non-adjacent sentences in NCEE Text 4



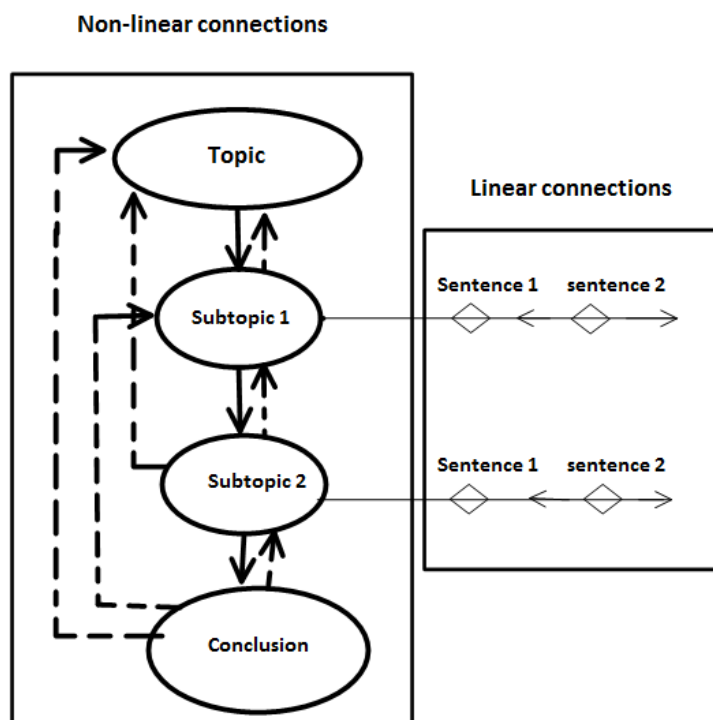
In each diagram, it can be seen that the number of the highlighted solid line is much greater than that of dotted lines, which signifies that bonds formed between non-adjacent sentences are more numerous than those between adjacent sentences. Previous interpretation of this network in Section 8.3 indicates that Sentence (1) serves as a topic opening sentence, Sentence (8) serves as a sub topic sentence, and Sentence (12) functions as a topic closing sentence. It is evident that these sentences play a far more crucial role in connecting the ideas and developing arguments across the text, consequently, they are typically identifiable in terms of the distance they are across over in a text and the number of sentences that refer to them. Because of this, it can be said that the number of bonds formed between non-adjacent sentences reflects the complexity and interconnectedness of a text. In contrast, the type of bond formed between adjacent sentences is usually transitional — to indicate the information shift from the previous argument to the following one, or supplemental — to help clarify or support a certain topic or subtopics in a linear way, therefore cannot connect to other non-adjacent

sentences. In contrast, the network of bonds in the CFL learner's essay in Figure 8.2 indicates that, although the number of bonds formed between non-adjacent sentences is similar to that between adjacent sentences, the organizational role that they play is weaker and less effective than the role the same type of bond plays in the NCEE text. Firstly, the bonds formed between non-adjacent sentences in CFL texts have a weaker organizational function. Figure 8.2 seems to suggest that the writer has difficulty with developing and concluding the discussion more fluently as there are fewer sub-topic opening sentences and topic-closing sentences, which either connect to some later text or refer back to earlier text through bonds formed between non-adjacent sentences. In addition, the function of the bonds formed between adjacent sentences in CFL Text 1 does not seem to be as effective as that in NCEE Text 4 either.

On the basis of the above analysis, it can be seen that the main reason for the lower ratio of the bonds formed between non-adjacent sentences to that between adjacent sentences in the HSK Corpus may be related to the low level competence in the use of topic sentence(s) in their writing. In his analysis of distant lexical bonds, Berber (1994) points out that "bonding keeps track of topics" (p. 1), and remote bonding in a real sense reflects how a text progresses, and how an argument is developed. In other words, remote bonding represents "the same topic being carried out in one part of the text is reflected a long way afterward" (Berber 1994, p. 3). Similarly, Hoey (2000) maintains that, "if topic-opening sentences are typically identifiable in terms of the number of sentences that later refer back to them, then it would seem advisable to writers to keep in mind their topic sentence(s) as they write, rather than allowing the difficulties of the composition process to swamp them" (p. 243). These views are helpful in terms of identifying the reason why CFL learners have difficulty using bonds formed between non-adjacent and adjacent sentences to organize their texts. When writing an argumentative text, it is common to introduce a topic at the beginning, and then further discuss the topic from different aspects or perspectives in detail. These aspects or perspective are actually sub-topics or subsidiary arguments. In order to well-organize a text both logically and lexically, one has to have a clear train of thoughts to follow. According to the researcher's observation, in the NCEE

Corpus, the argument developments in the texts can be summarized as the following typological Figure 8.4.

Figure 8.4 A typological pattern of the argument development in the NCEE Corpus



Since this is an only a typological diagram, the number of subtopics and sentences are only used to typify how an argument is progressed; it does not mean all the texts can be summarized as two sub-topics and two sentences in each subtopic development. It can be seen that in the non-linear connections rectangle, all the components are intrinsically interconnected to each other through the development of topics, while in the linear connections rectangle all the components are linearly link to each other so that they help to develop the subtopic within its boundary.

In the light of this discussion, it would appear that the fundamental reasons for CFL learners having difficulty with using the topic sentences can be attributed to the three factors discussed in previous section: low lexical competence, low organizational skills and lack of practice through pre-writing preparation activities. Pre-writing preparation activities would appear to be an effective way of

integrating the learner's competence in the use of vocabulary, the organization of ideas, and the outlining of their train of thought. Since this approach was discussed in the previous section, no repetition of the discussion is presented here.

8.5 The patterns of the average ratio of sentences that form bonds with other sentences to those form no bonds with others in the two corpora

Table 6.6 in Chapter 6 also shows a difference between the average ratios of sentences that form bonds with other sentences to those form no bonds with others in the two corpora. In the NCEE Corpus, the average ratio of the 50 texts is 1.6:1, while in the HSK Corpus the figure is 0.96:1, which means that on average, the NCEE texts have a majority of sentences that form bonds with other sentences, while in the CFL texts, less than fifty per cent of the total sentences form bonds with other sentences. Possible factors that may contribute to this gap are similar to those discussed in the previous section, especially the lack of pre-writing preparation activities.

The gap between the two average ratios in the two corpora may indicate that the CFL learner texts are less developed and well-organized due to a lack of central sentences. When discussing the properties of the net of bonds of text, Hoey (2000) notes that "sentences that manifest no bonds may be deemed *marginal*, in that what they have to say neither builds lexically upon what has been gone before nor provides the lexis for subsequent statements" (p. 105). He then further explains that marginal sentences do not help to develop the main theme of a text. However, the role they play in the organization is still very important, as they serve the purpose of providing necessary ancillary information to make a reader's task easier or to avoid misunderstanding. In the NCEE Text 4, there are two marginal sentences: Sentences 3 and 6, and in CFL Text 1, there are eight marginal sentences: Sentences 8, 10, 11, 12, 13, 15, 16 and 17. Their respective proportions to the total number of sentences are 17% (2/12) and 47% (8/17).

In contrast, Hoey proposes that "sentences that have an unusually high level of

bonding from the remainder...might be regarded as central” (p. 113), and the minimum number of bonds criterial for centrality varies from text to text. To help to identify central sentences, Hoey (2000) uses a tabular form to record the number of bonds that each sentence forms with previous and subsequent sentences. The following Tables 8.1 and 8.2 record the coordinates of each sentence of the NCEE Text 4 and CFL Text 1 according to the information of nets of bonds shown in Figure 8.2.

Table 8.1 Sentence coordinates of the NCEE Text 4

1 (0, 7)	7 (0, 3)
2 (1, 5)	8 (2, 5)
3 (0, 0)	9 (4, 1)
4 (2, 2)	10 (1, 0)
5 (2, 0)	11 (4, 1)
6 (0, 0)	12 (6, 0)

Table 8.2 Sentence coordinates for the CFL Text 1

1 (0, 3)	9 (1, 0)
2 (1, 2)	10 (0, 0)
3 (1, 2)	11 (0, 0)
4 (1, 0)	12 (0, 0)
5 (0, 1)	13 (0, 0)
6 (1, 0)	14 (1, 0)
7 (1, 0)	15 (0, 0)
8 (0, 0)	16 (0, 0)
	17 (0, 0)

In Table 8.1, Numbers 1-12 means the sentences in the NCEE Text 4. The two figures in brackets are coordinates. The first one indicates the number of bonds formed between that sentence and its previous sentences, and the second bonds formed with subsequent sentences. For example, 9 (4, 1) means for Sentence 9 in

the NCEE Text 4, it forms four bonds with its previous sentences, and one with its subsequent sentences.

When defining the term *bond*, Hoey (2000) proposes that if two sentences share three lexical links, they can be said to form a bond. As discussed in Chapter 3 Section 3.6.2, although the cut-off point between sentences with and without bonds is relative, three links between two sentences tends to be a minimum requirement. Therefore, in some texts the cut-off point may be four or higher if there are a high proportion of sentences linked by four or more lexical cohesive links. This principle also applies to the identifying the threshold for central sentences in a text. For example, if the average number of bonds formed between sentences is two, then it may be sufficient to say that sentences that form three bonds with other sentences are central sentences. Hoey (2000) proposes that central sentences have topic-opening, topic-closing functions in text, which is the case in the NCEE Text 4 and the CFL Text 1. As can be seen in Tables 8.1 and 8.2, the majority of the numbers indicated in the coordinates are 0, 1, and 2, hence it is reasonable to treat sentences that form three or above three bonds with other sentences as central sentences.

According to this criterion, in Table 8.1, it can be seen that there are seven central sentences in the NCEE Text 4: Sentences 1 (0,7), 2 (1,5), 7 (0, 3), 8 (2, 5), 9 (4, 1), 11 (4, 1), and 12 (6,0). The proportion of central sentences in this text is 58% (7/12), while in the CFL Text 1, only Sentences 1, 2 and 3 can be regarded as central sentences, and the percentage of central sentence is 18% (3/17). The following Table 8.3 summarizes the percentages of marginal and central sentences in both the NCEE Text 4 and the CFL Text 1:

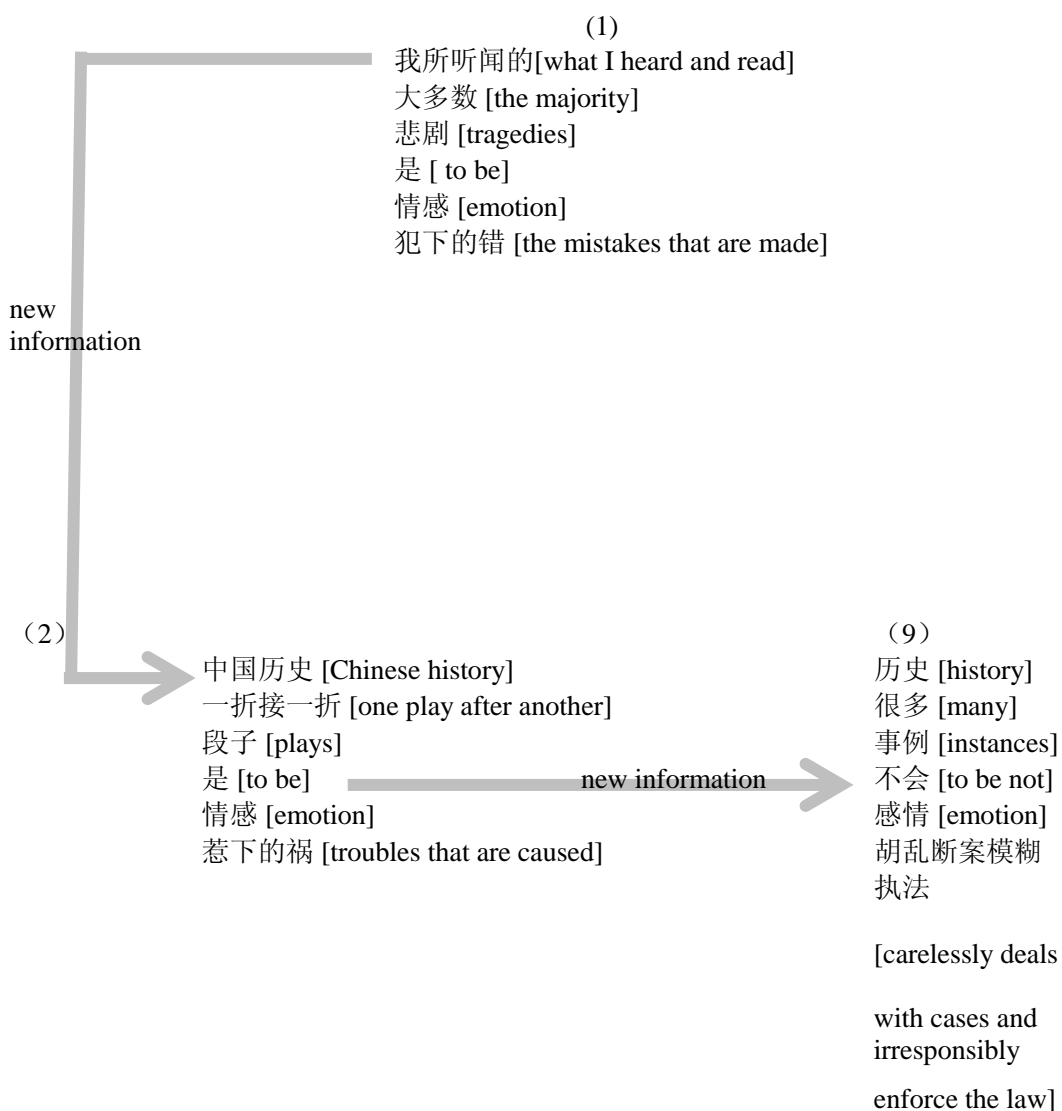
Table 8.3 The percentages of marginal and central sentences in both the NCEE Text 4 and the CFL Text 1

	NCEE Text 4	CFL Text 1
Marginal sentences	17%	47%
Central sentences	58%	18%

The above table shows that in the CFL Text 1, the percentage of central sentences is far lower than that in the NCEE Text 4. As central sentences play a crucial role in developing the main theme of a text, a significantly low proportion of them may result in a less developed and well-organized essay. According to the discussion presented in Section 8.3.2, low bond density means that sentences in text are less interconnected to each other both lexically nor semantically. Since CFL Text 1 has a much higher proportion of marginal sentences, which is indicated by its low bond density, this high percentage is, therefore, a reflection of low level of development and organization of the text. Although this table reflects only the difference between two individual texts taken from the two corpora, the average ratio of sentences that form bonds with other sentences to those form no bonds with others in the NCEE Corpus is nearly twice as many as that of the HSK Corpus, which implies that the contrast shown in Table 8.3 is fairly representative of the difference between the texts in the two corpora.

The most reasonable interpretation of this gap between the two ratios may lie in the lack of pre-writing preparation activities provided for CFL learners. In his discussion of the implications of his lexical repetition analysis for writing, Hoey (2000) points out that “writers who fail to connect what they are saying in any particular sentence to what they have said earlier are likely to be open to the charge of drifting from topic to topic” (p. 242). Indeed, as discussed in the CFL Text 1, there are only three central sentences that effectively contribute to the development of the main theme, while the majority of the sentences seem to drift away from the main ideas or topics, consequently, they contribute little to the development of the main argument. For such learners, Hoey suggests that they “will need telling that repetition usually provides the grounding for new information” (p. 244). In other words, such learners need to be aware that an adequate level of cohesion (of course, in various forms by using a variety of lexical cohesive devices) is essential to provide a necessary context or framework for explaining what they present as new in subsequent sentences. To illustrate this point, the example presented in Section 8.3 is presented here in a slightly different way to show how repetition provides grounding for new information in terms of topic development.

Example 68:



In the above example, it can be seen that from Sentence (1) to (2), there are enough repetitions to provide the context, such as 悲剧 [tragedies] - 段子 [plays], 情感 [emotion] - 情感 [emotion], 犯下的错 [the mistakes that are made] - 惹下的祸 [the troubles that are caused]. From Sentence (1) to (2), it can be seen that some new information- 中国历史 [Chinese history] is introduced. In Sentence (1) it talks about 我所听闻的 [what I heard and read], which is rather general in scope, while in Sentence (2) the meaning that 中国历史 [Chinese history] contains becomes much narrower. This makes more sense to the reader as the

subsequent sentences are talking about some examples that happened in Chinese history. Similarly, from Sentence (2) to (9), again new information 不会 [to be not] is introduced on the basis of the repetitions such as 历史 [history] - 历史[history], 段子[plays] - 事例[instances], 情感[emotion] - 感情[emotion]. In Sentence (2) the statement is an affirmative one as indicated by 是 [to be], which is negated in Sentence (9) by 不会 [not to be]. The denial of the statement in Sentence (9), therefore, helps to shift the argument from the negative perspective to the positive side. This example illustrates that, while the new information helps to develop the main theme (in this case it first shifts the discussion from a general scope to a specific one, and then goes further to shift the discussion from an affirmative statement to a negative one), the function that the use of lexical cohesive devices plays is crucial. Without the lexical framework that they provide, the new information may not make clear sense to the reader.

Hoey's (2000) suggestion, thus, sheds in-depth light on the interpretation of the gap between the two corpora in terms of the existence of bonds in texts. As discussed in Section 8.4, even if CFL learners have already had a high level of writing competency in their first language, pre-writing preparation practice for L2 writing is still needed, this will enable CFL learners to learn how to use repetitions to organize their ideas and how to develop their argument in a clear and coherent way. Wu (2011) presented an example of outline composed by a Japanese student in a pre-writing preparation exercise.

Example 69:

1. 今年春节我才知道 H1N1 很厉害;
[It is until this spring festival that I realized how serious the H1N1flu is.]
2. 暑假时, 有好几个同学感染了 H1N1, 以后我也在饭店被隔离了;
[In the summer holiday, a few classmates were infected by H1N1, and later I was quarantined in a restaurant.]
3. 我有急事, 要求批准我出去;
[I requested to be released because I had something urgent at that time]
4. 秋天我终于打疫苗, 才放心了。

8.6 The patterns of the average number of words that form repetition with other words per sentence in the two corpora

This last section is devoted to the discussion of the difference between the patterns of the average words that form lexical cohesive links with others per sentence in the two corpora, as shown in Table 6.8. The average number in the NCEE Corpus is 4.04, which means on average there are 4.04 words per sentence that form links with other words in the same sentence, while the figure in the HSK Corpus is 2.6, much lower than that in the NCEE Corpus.

Firstly, the difference between the two patterns may suggest that CFL learners may need to improve the way in which they acquire their vocabulary. This lexical knowledge, to a great extent, decides how competent they are in using bond-making items in their writing. This ability to use bond-making items can be examined in relation to two aspects: the number of different types of lexical device they are competent to use and the number of lexical items they can use to create links per sentence in text. The first aspect has been discussed at length in Section 8.3, where the statistical data suggests that CFL learners seem to have difficulty with using some certain types of lexical cohesive device in their writing, such as constituent repetition, simple and complex paraphrase, superordinate-hyponymic repetition. The second aspect is the focus of the discussion in this section.

Tables 8.4 and 8.5 show a significant difference between the two mean values. In the NCEE Text 4, the average number of repetitions that each sentence forms is 5.58, while in the CFL Text 1 this number is 2.8. To discuss this gap, two attributes of writing have to be taken into consideration: the average sentence length and the correlation between sentence length and the number of repetitions. In the NCEE Text 4, the average number of characters in each sentence is 64, varying from 25 to 133; while in the CFL Text 1, the average number is 25, with the largest 49 and smallest 11. This contrast shows us that sentences in the CFL text are much shorter. A preliminary comparison of the sentence structures and the use of various syntactic devices, such as dependent and independent clauses,

embedded clauses also reveal that the sentence complexity in CFL text is much

Table 8.4 NCEE Text 4 words that form repetition with others

7	1	悲剧	情感	错	犯下的错	权力	太大	真相												
7	2	段子	情感		惹下的祸				历史	君王	忠臣奸侯									
3	3										侯臣	亲信近宠								
5	4	可悲	情感			权力	无人企及			君主们										
4	5	悲剧	情感									麦克白	妻子							
3	6											亲信	麦克白	妻						
4	7		情感			权力									施展	正义				
9	8		真理情感	谬误						当权者					操纵	正义	近疏	颠倒黑白	感情用事	
7	9	事例	感情					现代	历史								亲疏		大义灭亲	社会
2	10									明君							光明正大			
9	11		情感真理	错	正确			现今									非正义	亲疏	歪曲	社会
7	12		情感			权力		真相									正义/公正	亲疏	混淆是非	
67																				
67/12=5.58																				

Table 8.5 CFL Text 1 words that form repetition with others

4	1	人	不一样	人生	道路														
6	2	人	不同	人		创业	进	公司											
5	3					创业开		公司	希望	经验	开								
2	4							厂子			开	爸爸							
6	5							公司		经验			机器	手机	运营	倒闭			
2	6									经验					管理				
2	7						进	公司											
3	8	人											机器						四年
1	9																		四年
2	10							工厂			开			手机					
3	11							公司					爸爸						好起来
2	12									经验			爸爸						
1	13									经验									
3	14	人		人生									爸爸						
0	15																		
2	16				路					经验									
3	17							公司		经验	开								
46																			
46/17=2.8																			

lower than that in the NCEE Text 4. Because of this, one possible argument would be that the gap is attributable to the difference in the average length of sentences: since the CFL learners are at an earlier developmental stage in their writing, their lower ability to write long and complex sentences hinders them from creating adequate repetitions between sentences. The results of correlation analysis of the sentence length and the number of repetitions in the NCEE Text 4 and the CFL Text 1 are supportive of this argument (see the correlation diagram 8.1). For the NCEE Text 4, the correlation coefficient $R=0.70$, which means that the relationship between sentence length and the number of repetitions formed is considerably positive and strong, while in the CFL Text 1, the index is $R=0.42$, which is positive but weak. One possible reason for this may lie in the number of central sentences that each of the two texts has. As in the CFL Text 1, out of seventeen sentences, there are only three central sentences, while in the NCEE Text 4, among twelve sentences, seven are central sentences.

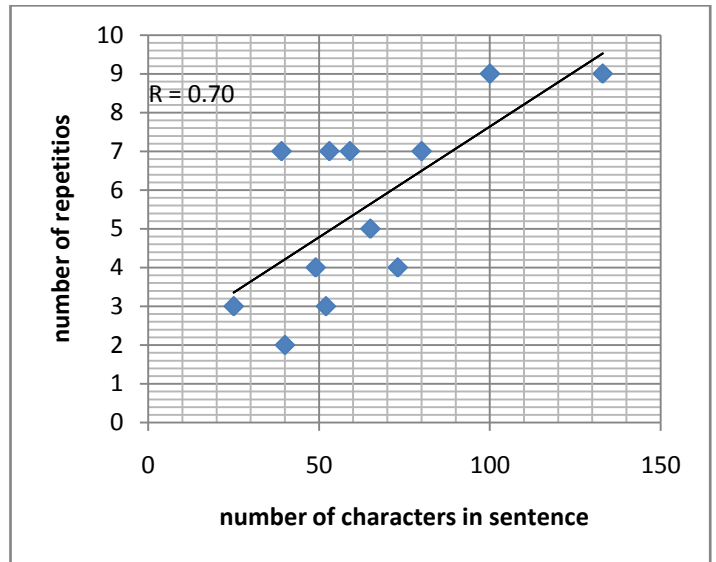
As discussed in the previous section, central sentences form more repetitions as they contribute greatly to the development of main theme. Since the majority of the sentences are marginal in the CFL Text 1 (in that they do not form repetitions between sentences), this results in a correlation coefficient at a weaker level. On the basis of this statistical analysis, it is, therefore, possible to say that simple and relatively short sentences may partially contribute to the small average number of words that form repetitions in each sentence in the CFL text.

Although the above analysis points out that the length and complexity of sentence may, to some extent, have an impact on the quantity of repetitions created between sentences, it does not necessarily mean that CFL learners who are still at developmental stage are not able to create adequate repetitions between sentences. In previous sections, it has been proposed that semantic networks formed between bond-making items are closely related to the textual level organizations. The relation between them can be illustrated in Diagram 8.2.

Diagram 8.1 *The correlation between the sentence length and number of repetitions*

NCEE Text 4

Sentence no.	Number of characters	Number of repetitions
1	53	7
2	59	7
3	52	3
4	65	5
5	73	4
6	25	3
7	49	4
8	100	9
9	39	7
10	40	2
11	133	9
12	80	7



CFL Text 1

Sentence no.	Number of characters	Number of repetitions
1	18	4
2	32	6
3	19	5
4	19	2
5	49	6
6	16	2
7	11	2
8	61	3
9	13	1
10	15	2
11	29	3
12	16	2
13	32	1
14	17	3
15	14	0
16	44	2
17	20	3

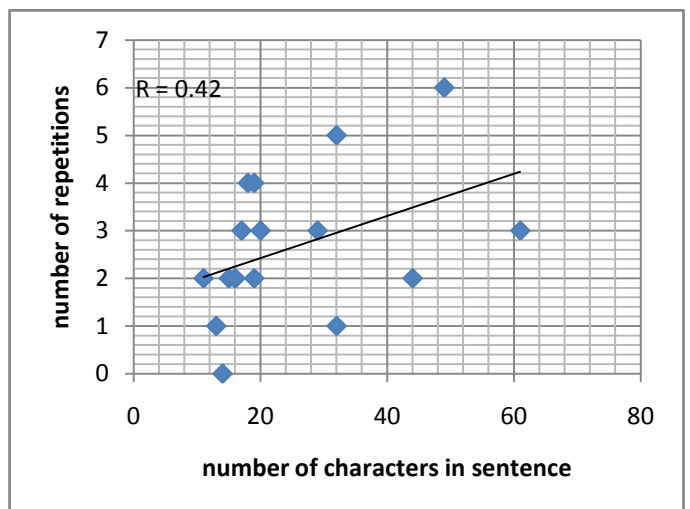
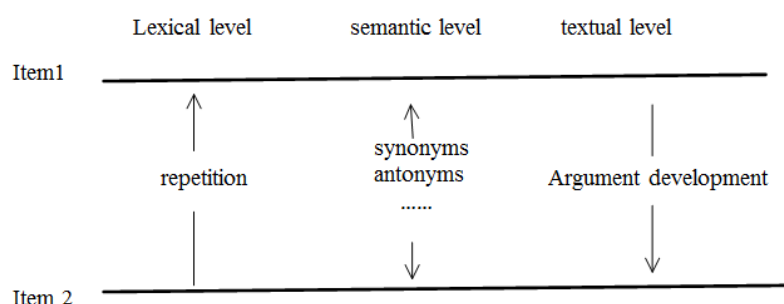


Diagram 8.2 *The relation between lexical repetition, semantic network, and textual organization*



On the basis of this understanding, it may be helpful to teach CFL learners to employ consciously bond-making items in the planning process, even at the outlining stage when they are engaged in pre-writing preparation activities. By doing this, students may realize that they are able to create cohesive connections between sentences even they are not able to write long and complex sentences as the native Chinese writers do. Taking the topic used in the HSK Corpus as an example, it could be outlined as the following:

Example 70:

大学毕业后进公司好还是自己创业好？

[Working in a company or starting up your own business after graduation?]

1. 大学毕业生应该先进公司学习管理经验和培养社交能力，然后自己创业

[University graduates should gain management experience in a company before starting up their own business.]

2. 刚从学校出来的年轻人没有经营公司的经验，直接创办企业可能很快就会失败。

[Young people who have just come out of university do not have experience in operating a company; as a result, they may fail quickly if they found their own business straight after graduation.]

3. 刚完成学业的大学生也不太懂得与人打交道，马上开公司遇到挫折。

[Graduates who have just finished their education may meet setbacks if they immediately start their own business without gaining socializing experience first.]

4. 在公司工作，可以学会如何打理业务和如何与人交往，这时再开创自己的事业，成功的可能性更大。 [Obtaining a job in a company may enable graduates to gain management and socializing experiences. It is more likely to lead them to success if they start their own career by then.]
5. 所以大学生毕业以后应该先提高自己负责业务的能力和交际能力，再发展自己的事业
[University graduates, therefore, should gain management experience, and practice their social skills before developing their own career.]

The lexical cohesive links in the above outline can be summarized as in Table 8.6.

Table 8.6 *Lexical cohesive links in the above outline*

Idea no.	Lexical item					Lexical level	Semantic level	Textual level
1	大学毕业生 [university graduates]	管 理 [management]	社 交 [socializing]	然 后 [afterwards]	创 业 [start a business]	repetitions	Synonyms antonyms	Developing arguments
2	刚从学校出来的年轻人 [Young people just out of university]	经 营 [operating]		直 接 [straight- after]	创 办 企 业 [start a enterprise]			
3	刚完成学业的大学生 [students who have just finished study]		与 人 打 交 道 [interact with people]	马 上 [immediately]	开 公 司 [Start a company]			
4		打 理 业 务 [dealing with business]	跟 人 交 往 [Dealing with people]	这 时 [until this time]	开 创 自 己 的 事 业 [start one's own career]			
5	大学生毕业以后 [after the university students graduate]	负 责 业 务 [be responsible for business]	交 际 [communicate]	再 [after this]	发 展 自 己 的 事 业 [start one's own career]			

It can be seen that although the sentences in the above outline are not long and complex, at least four repetitions are created between every two ideas, which indicates that the use of bond-making items is the key to improve the organization of text. Wang (2010) points out that the current approach to teaching CFL writing is still the same as the traditional approach, which has long been used for teaching native speaker Chinese students in secondary education. This approach first involves students writing about a given topic, normally chosen by their teacher, and then their essays are marked and corrected by their teacher. The last step is that the teacher reads two or three good essays as models to the whole class. The

main disadvantage of this approach is that the learners play a very passive role. The approach does not help students to expand actively their productive vocabulary and develop their organizing skills through activating known semantic associations, planning and outlining. To counter this problem, Wang (2010) suggests that teachers encourage students to explore a certain topic thoroughly in the form of group discussion. In so doing, students first gather enough information that is closely related to the topic, and then organize the ideas into a cohesive chain of arguments. This process in essence is similar to what Wu (2011) proposed as the pre-writing preparation activities. The goal that Wang (2010) and Wu (2011) are trying to achieve is similar to what is proposed in Table 8.6. By consciously using bond-making items, the ideas are well-organized at three levels, the use of lexis is cohesive, the semantic relations are meaningfully associated, and the development of arguments is fluent and effective. The strength of this type of practice, as manifested in the above example, is that CFL learners are given the opportunity to learn actively and consciously the lexical items that are able to form repetitions within a context. It is, therefore, can be concluded that even though CFL learners are not competent enough to write long and complex sentences, they may still be able to create adequate repetitions between sentences to achieve effective textual cohesion, if they have opportunities to acquire vocabulary that are contribute greatly to creating various types of repetitions within contexts by practicing pre-writing activities such as topic exploration, key words association, idea organization, and essay outlining.

8.7 Implications for the study of cohesion and the TCFL pedagogy

It is hoped that the discussion of this concluding chapter has shed some further light on the study of cohesion in the field of TCFL. Firstly, the discussion has led to the suggestion that more in-depth studies on the use of lexical cohesion in both native speaker Chinese writing and CFL writing are needed. The chapter has also considered at length issues identified in relation to CFL student writing, issues such as insufficient exposure to certain types of lexical cohesion (as a result of such devices not being taught in classroom), low levels of lexical competence and a lack pre-writing preparation activities. It may be the case that these issues arise

for the reason that research on lexical cohesion in Chinese text is still at an early stage.

In the current TCFL literature, few corpus-based research studies have been carried out to explore the types of lexical cohesion and their actual use in native speaker Chinese writing. Similarly, the number of studies on the use of lexical cohesion and error analysis of the use in CFL learner texts is also very limited. The majority that examine cohesion, such as those of Xiao (2001), Yang (2004), Zhao (2005), focus on the study of the use of grammatical cohesion, namely reference, ellipsis, conjunction, and substitution in CFL writings. As a result, a lack of in-depth study on lexical cohesion has led to the lack of understanding of the role that it plays in the organization of text in CFL learners' writing. Sun (2010) suggests that this situation has had some impact on the development of teaching materials and the competence of teachers of CFL writing. For instance, many of the current teaching and learning materials for writing used by teachers and CFL learners in language institutes do not cover the aspects of textual organization and the practice of such skills. Similarly, there may be a misconception among teachers and researchers about the level at which the awareness of lexical cohesion should be introduced to CFL learners. Many claim that it is inappropriate to practice text organization at beginner level writing. However, some researchers such Wang (2009) and Tao (2010) believe that the concept of lexical cohesion can be introduced to beginners for the reason that even a very short writing task that contains a few sentences may manifest various types of repetition.

In addition, the lack of study in this area also has led to the consequence that many teachers of CFL writing lack awareness of the role the lexical cohesion plays in textual organization. However, without this knowledge and relevant training, it is not difficult to imagine that they lack the consciousness to teach CFL learners how to use the lexical repetition to achieve text cohesion. With regard to the assessment of textual organization, in particular for HSK essay writing, Nie (2010) provides a historical account of the progress of the establishment of the assessment criterion for HSK writing. She points out that, although there were improvements from the 1995 version to the revised 2004 and 2007 version, some

of the assessment criteria, such as “the structure is clear and complete”; “the argument flows naturally”; and, “express one’s ideas fluently and coherently” still need to be improved so that the assessment can be more objective and scientific. Since these items are closely related to textual organization, and the current research findings in this area are neither comprehensive nor thorough enough to provide guidance and reference for HSK administrators to establish the assessment criteria, more in-depth research in this area is, therefore, needed to reveal more textual features of the native Chinese writing, upon which, more objective and scientific criteria can be established to improve the level of the validity, reliability, and operability of the assessment of the HSK writing.

Implications for TCFL pedagogy

The lack of in-depth studies on the use of lexical cohesive devices in both native Chinese writings and CFL writings, as discussed in previous section, has a direct impact on the development of teaching and learning materials, the compilation of CFL learner dictionaries, and research of effective methodologies for teaching writing. The research findings in this study, therefore, may provide some further light on the improvement of these three areas.

Firstly, the discussion about the gap in the use of superordinate and hyponyms suggests that when developing text books, consideration should be given to the contexts in which the lexis are used, the specific connotations that they have, their accurate translations in learner’s first language, and how they should interact with other words that within the same semantic network. The example presented in Section 8.3. 1 suggests that mismatches in translation may confuse CFL learners and consequently create a barrier to them being able to create a well-organized semantic network among a group of semantically related lexical items. Other common examples that demonstrate this kind of mismatch can be as the Chinese word 参观 [to visit a place], 访问 [to visit a country, a person], 看望 [to visit a person]. In many text books the Chinese words 参观, 访问 and 看望 are all simply translated as *visit*. It is true that in English the word *visit* can be either followed by a location noun such as *museum*, or by an animate noun such as a

friend; however, each of these three Chinese words has their own specific collocations according to context. Such inaccurate translations are very confusing, and consequently hinder CFL learners from grasping various aspects of these words such as the context in which they should be used, the semantic meaning, the collocation, the interactive relation with its synonyms. In current TCFL literature, many studies have been carried out to address this issue, such as those by Lu (2010), Wu (2010), Zhang (2011). These studies have provided much insightful advice on how to improve the presentation of new vocabularies in textbooks. The suggestions made for this area may be beneficial for learners to improve their lexical competence by not only helping them to master the meaning and usage of the lexis taught in textbooks, but also by introducing them to the organization of those semantically-related words in terms of their spoken and written forms, grammatical patterns and collocations, as well as semantic, pragmatic, sociolinguistic, register, connotations. Such an organizational ability, as has been argued throughout this thesis, is the core of a writer's lexical competence and is, therefore, crucial for raising their level of textual organization in their writing.

As discussed in Section 8.3. 2, one of the benefits of the discussion about the various aspects of synonyms and antonyms relates to the compilation of CFL learner dictionaries. The traditional method of compilation of Chinese dictionaries of synonyms and antonyms dictionary is not able to meet the demand of CFL learners' actual needs. For example, many words and idioms that used in the explanation part of the lexical items in dictionaries compiled for native Chinese speakers are not appropriate, even for advanced CFL learners, as those expressions may be beyond the CFL learner's language level. In addition to this, traditional dictionaries usually provide little explanation about items and their usage features, such as context, connotations, pragmatic and register, information which is crucial for CFL learners. To help CFL learners to improve their lexical competence, the compilers of learner dictionaries need to take these aspects into account, together with other considerations, such as the scope of the vocabulary, the frequency of lexis, and the organization of the item. For the synonyms and antonyms dictionary, it is vital to consider that the actual needs of CFL learners, as

Li (2010) suggests. Instead of choosing the items from their own perspective, the compilers need to consider the most confusable words in the TCFL classroom, as well as some of the relevant aspects that are central to the CFL students' discrimination problems. Again the implementation of these suggestions would be greatly beneficial to the improvement of CFL learners' lexical competence for the reason that simple paraphrase and complex paraphrase are two of the mostly used lexical devices that writers usually employ to achieve textual cohesion in their writing. A well-designed CFL learner dictionary may, to a great extent, help learners to master various aspects of the lexis in their learning process. As a result, it may provide one way of facilitating their learning of the use of synonyms and antonyms in their writing to make connections between sentences.

Lastly, the findings also have implications for the methodology of teaching CFL writing. The discussions in Sections 8.3.4 and 8.3.5 have indicated that pre-writing preparation activities are a positive and productive way of helping CFL learners to improve their lexical competence as well as their organizational ability. The examples presented in these two sections have shown that the approach may be both effective and achievable in terms of helping CFL learners to explore the given topic, preparing them to activate their vocabulary (and their semantic relations), encouraging them to organize their ideas, as well as leading them to composing the essay in an organized and coherent way. Since this process involves a range of activities, such as group discussion, key words and ideas association, outlining and organization, it is an active and productive way of arousing students' interest and for bringing learners' initiatives into play. As a result, learners may become more aware of the semantic relations between the vocabulary they have learnt, the inter-connections between them, and the cohesive relations they may form in a context. By learning vocabulary in this type of productive and systematic way, it is hoped that not only students' vocabulary size will be increased, but most importantly, their organizational ability will improve at both the lexical and textual level. Introducing this type of methodology to the teaching of CFL writing, however, involves consideration of a number of factors, including the learners' different language levels, backgrounds, ages, motivations, and learning purposes. Since these factors are beyond the scope of this research, it

is suggested that TCFL researchers need to conduct more in-depth studies in these areas.

8.8 Limitations of the study

There are two limitations to the present study. The first relates to the relative length of the texts of the two corpora. Although every effort is made to have the two sets of data match at various aspects as much as possible, however, it is difficult to achieve this due to the specific purpose of this study. For example, there is an overall gap between the average length of the NCEE Corpus texts and the HSK Corpus texts. The average number of Chinese characters in the NCEE Corpus text is about 800, while in the HSK Corpus text is about 450. To avoid possible impacts this gap may have, the study adopts the strategy that compares the various types of comparing items in per 100 characters. It is unknown that whether or not this gap may have some influence on the statistical results. Although it is possible to achieve cohesion even in a short outline writer by the researcher, as shown in the Example 70 in this Chapter, Section 8.6, further studies may need to take this factor into account. The second limitation lies in the fact that the topics of the NCEE Corpus texts are different, while that of the HSK Corpus texts are the same. This difference may also affect slightly the findings, as the use of lexical cohesion may vary depending on the topics. The ideal situation would be that the NCEE Corpus texts and the HSK Corpus texts share the same topic so that any possible factors that may influence the results would be eliminated. The study has endeavored to collect texts that share the same topic as many as possible. Efforts have also been made to ensure that all the texts are typical argumentative essays so that the various properties of the texts in NCEE texts are similar.

8.9 Summary

This chapter discusses the differences in the patterns of the use of lexical cohesion between the NCEE and the HSK Corpora. It focuses on the discussion of the possible reasons for the differences, and identification of the factors that may

contribute to the gap. The implications for future studies in the TCFL field, as well as implications for TCFL pedagogy and development of teaching and learning materials are also discussed. The chapter also includes a short brief discussion of the limitations of the present research.

References

- Al-Jubouri, A. (1984). The role of repetition in Arabic argumentative discourse. In J. Swales & H. Mustafa (Eds). *English for specific purposes in the Arab world*. Birmingham, UK: Language Studies Unit, University of Aston in Birmingham. Annual Convention of TESOL. Long Beach, CA
- Aston, G. (2002). The learner as corpus designer. In B. Kettemann (Ed.), *Teaching and language corpora*. Amsterdam: Rodopi.
- Aston, G., & Burnard, L. (1998). *The BNC handbook: Exploring the British National Corpus with SARA*. Edinburgh, UK: Edinburgh University Press.
- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing*, 7(1), 1-16.
- Atkinson, D. (1997). A critical approach to critical thinking in TESOL. *TESOL Quarterly*, 31, 71–94.
- Atkinson, D. (2002). Writing and culture in the post-process era. *Journal of Second Language Writing*, 12, 49-63.
- Atkinson, D. (2004). Contrasting Rhetorics /Contrasting Cultures: Why Contrastive Rhetoric Needs a Better Conceptualization of Culture. *Journal of English for Academic Purposes* 3, 277-289.
- Barlow, M. (2002). Corpora, concordancing, and language teaching. Proceedings of the 2002 KAMALL International Conference. Daejon, Korea.
- Berber, S. (1994). *Bonding and the organization of business text* [PowerPoint slides]. Presented at the Organization in Discourse conference, University of Liverpool.

- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243-257.
- Biber, D., S. Conrad & R. Reppen (1994). Corpus-based approaches to issues in applied linguistics. *Applied Linguistics*, 15, 169-189.
- Biber, D., S. Conrad & R. Reppen (1996). Corpus-based investigations of language use. *Annual Review of Applied Linguistics*, 16, 115-136.
- Brown, G. & Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.
- Canagarajah, A. S. (1999). *Resisting linguistic imperialism in English teaching*. Oxford: Oxford University Press.
- Casanave, C. P. (2004). *Controversies in second language writing: Dilemmas and decisions in research and instruction*. Ann Arbor, MI: The University of Michigan Press.
- Chastain, C. (1975). 'Reference and context' in (Ed.) K. Gunderson *Language, Mind and Knowledge*. Minnesota Studies in the Philosophy of Science, Vol VII.
- Chiu, M. F. F. (1993). *Lexical Cohesion in Expository Writing: Will a Study of the Similarities Between an English and Chinese Paragraph be Helpful to ESL Students?* San Bernardino: California State University.
- Connor, U. (1996). *Contrastive Rhetoric: Cross Cultural Aspects of Second Language Writing*. Cambridge, UK: Cambridge University Press.
- Connor, U. (1997). Contrastive rhetoric: Implications for teachers of writing in multicultural classrooms. In C. Severino, J. C. Guerra & J. E. Butler (Eds.), *Writing in Multicultural Settings* (pp. 198-208). New York: MLA.

- Connor, U. (2004). Intercultural rhetoric research: beyond texts. *Journal of English for Academics*, 3, 291-304.
- Connor, U. (2002). New directions in contrastive rhetoric. *TESOL Quarterly*, 36(4), 493-510.
- Connor, U., Davis, K., & DeRycker, T. (1995). Correctness and clarity in applying for overseas jobs: A cross-cultural analysis of U.S. and Flemish applications. *Text*, 15, 457-475.
- Connor, U., & Moreno, A. (2005). Tertium comparationis: A vital component in contrastive rhetoric research. In *Festschrift in Honor of Robert B. Kaplan*. Clevedon, England: Multilingual Matters.
- Connor, U. (1984). A Study of cohesion and coherence in English as a second language students' writing. *Papers in Linguistics*, 17 (1-4), 301-316.
- Corbett, J. (2001). Contrastive rhetoric and resistance to writing. In C. G. Panetta (Eds.) *Contrastive rhetoric revisited and redefined*, 31-46.
- Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning*, 39, 81-141.
- de Beaugrande, R-A. & Dressler, W. (1981). *Introduction to Text Linguistics*. London: Longman.
- Dodd, A. (2000). *SARA98*. Oxford, UK: Oxford University Computing Services.
- Engwall, G. 1994. Not chance but choice: Criteria in corpus creation. *Computational approaches to the lexicon*, ed. by B.T. Atkins & A. Zampolli, 49-82. Oxford: Oxford University Press.

- Gavioli, L. & Aston, G. (2001). Enriching reality: Language corpora in language pedagogy. *ELT Journal* 55(3), 238-246.
- Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (Ed.), *Learning with corpora* (pp. 108-137). Houston, TX: Athelstan.
- Gavioli, Laura & Aston, Guy (2001): Enriching reality: language corpora in language pedagogy. *ELT Journal* 55(3), 238-246.
- Geluykens, R. (1994). *The Pragmatics of Discourse Anaphora in English: Evidence from Conversational Repair*. Berlin: de Gruyter.
- Grabe, W., & Kaplan, R. (1989). Writing in a second language: Contrastive rhetoric. In D. Johnson & D. Roen (Eds.), *Richness in writing: Empowering ESL students* (pp. 263-283). New York: Longman.
- Granger, S. (1998). *Learner English on Computer*. London & New York : Longman.
- Granger , S. , Huang, J . & Tyson, S. P. (2002). *Computer Learner Corpora , Second Language Acquisition and Foreign Language Teaching*. Amsterdam/ Philadelphia : John Benjamins Publishing Company.
- Gregg, N., Coleman, C., Stennett, R. B., & Davis, M. (2002). Discourse complexity of college writers with and without disabilities: A multidimensional analysis. *Journal of Learning Disabilities*, 34(1), 23-38.
- Gutwinski, W. (1976). *Cohesion in Literary Texts*. The Hague: Mouton.
- Halliday, M. A. K. (1985). *An introduction to Functional Grammar*. London: Edward Arnold.

- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Harabagiu, S. M. (1999). From lexical cohesion to textual coherence: - a data driven perspective. *Journal of Pattern Recognition and Artificial Intelligence*, 13(2), 247–265.
- Hasan, R. (1984). ‘Coherence and cohesive harmony’ in Flood, J. (Ed.) *Understanding Reading Comprehension*. pp. 181-219. Edlaware: International Reading Association.
- Hinds, J. (1983). Contrastive rhetoric: Japanese and English. *Text*, 3(2), 183-196.
- Hinds, J.(1980). Japanese expository prose. *Papers in Linguistics: International Journal of Human communication*, 13 (1), 117-158.
- Hinds, J. (1984). Retention of information using a a Japanese style of organisation. *Studies in linguistics*, 8, 45-69.
- Hinds, J. (1984). Topic continuity in Japanese. University Park, Pennsylvania: Pennsylvania University Press.
- Hinds, J. (1987). Reader versus writer responsibility: a new typology. In U. M. Connor & R. B. Kaplan (Eds.), *Writing across languages* (pp. 141-152). Reading, MA: Addison-Wesley. [Reprinted in Silva, T., & Matsuda, P. K. (Eds.). (2001). *Landmark essays on ESL writing*. Mahwah, NJ: Erlbaum.]
- Hinds, J. (1990). Inductive, deductive, quasi-inductive: Expository writing in Japanese, Korean, Chinese, and Thai. In U. M. Connor & A. Johns (Eds.), *Coherence in writing: Research and pedagogical perspectives* (pp. 87-109). Alexandria, VA: TESOL.
- Hinkel, E. (1997). Indirectness in L1 and L2 academic writing. *Journal of Pragmatics*, 27(3), 360-386.

- Hinkel, E. (2002). *Second language writers' text*. Mahwah, NJ: Erlbaum.
- Hoey, M. P. (2000). *Patterns of lexis in texts*. Shanghai, China: Shanghai Foreign Language Education Press.
- Jenset, G. B. (2008). *Basic statistics for corpus linguistics* [Handout for methods seminar in English linguistics]. Bergen, Norway: University of Bergen.
- Jin, K. (2008). Lexical cohesion patterns in NS and NNS dissertation abstracts in Applied Linguistics: A comparative study. *Linguistics Journal*, 3 (3), 132-144.
- Kafes, H. (2012) Lexical cohesion: An issue only in the foreign language? *English Language Teaching*. 5(3), 83-94.
- Kaplan, R. B. (1966). Cultural thought patterns in intercultural education. *Language Learning*, 16(1), 1-20. [Reprinted in Silva, T., & Matsuda, P. K. (Eds.). (2001). *Landmark essays on ESL writing*. Mahwah, NJ: Erlbaum.]
- Kaplan, R. B. (1967). Contrastive rhetoric and the teaching of composition. *TESOL Quarterly*, 1, 10-16.
- Kaplan, R. B. (1972). *The anatomy of rhetoric: Prolegomena to a functional theory of rhetoric*. Philadelphia: Center for Curriculum Development.
- Kaplan, R. B. (1987). Cultural thought patterns revisited. In U. Connor & R. B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text* (pp. 9-21). Reading, MA: Addison-Wesley.
- Kaplan, R. B. (1988). Contrastive rhetoric and second language learning: Notes towards a theory of contrastive rhetoric. In A. Purves (Ed.), *Writing across*

languages and cultures: Issues in contrastive rhetoric (pp. 275-304). Newbury Park, CA: Sage.

Kaplan, R. B. (2001). Foreword: What in the world is contrastive rhetoric? In C. G. Panetta (Ed.), *Contrastive rhetoric revisited and redefined* (pp. vii–xx). Mahwah, NJ: Lawrence Erlbaum Associates.

Kaplan, R.B., (1967). Contrastive rhetoric and the teaching of composition. *TESOL Quarterly 1*, 10-16.

Kirkpatrick, A. (1997). Traditional Chinese text structures and their influence on the writing in Chinese and English of contemporary Mainland Chinese students. *Journal of Second Language Writing*, 6(2), 223-244.

Kobayashi, H. & C. Rinnert. (1992). Effects of first language on second language writing: Translation versus direct composition. *Language learning*, 42, 183-215.

Kony, T. (2005). Increasing the vocabulary of Chinese writers through contrastive rhetoric. The Internet TESL Journal. Retrieved from <http://www.chineseowl.idv.tw/html/Jp-news4.htm>

Kubota, R. (1997). A reevaluation of the uniqueness of Japanese written discourse: Implications for contrastive rhetoric. *Written Communication*, 14(4), 460-480.

Kubota, R. (1998). An investigation of L1-L2 Transfer in Writing among Japanese University Students: Implications for Contrastive Rhetoric. *Journal of Second Language Writing*, 7(1), 69-100.

Kubota, R., & Lehner, A. (2004). Toward critical contrastive rhetoric. *Journal of*

Second Language Writing, 13, 7-27.

- Land, R., & Whitley, C. (1989). Evaluating second language essays in regular composition classes: Toward a pluralistic U.S. rhetoric. In D. Johnson & D. Roen (Eds.), *Richness in writing: Empowering ESL students* (pp. 284-294). New York: Longman.
- Lay, N. (1982). Composing process of adult ESL learners: a case study. *TESOL Quarterly*, 16, 406-407.
- Leech, G ., Weisser, M. (2003). Generic Speech Act Annotation for Task-Oriented Dialogue. In *Proceedings of the Corpus Linguistics 2003 Conference* (Eds.), D. Archer, P. Rayson, A. Wilson and A. McEnery. Lancaster: UCREL Technical Papers.
- Leki, I. (1991). Twenty-five years of contrastive rhetoric: Text analysis and writing. *TESOL Quarterly*, 25(1), 123-143.
- Leki, I. (1992). *Understanding ESL writers: A guide for teachers*. Portsmouth, NH: Boynton/Cook.
- Leki, I. (1997). Cross-talk: ESL issues and contrastive rhetoric. In C. Severino, J. C. Guerra, & J. E. Butler (Eds.), *Writing in multicultural settings* (pp. 234–244). New York, NY: The Modern Language Association of America.
- Liebman, J. (1988). Contrastive rhetoric: Students as ethnographers. *Journal of Basic Writing*, 7(2), 6-27.
- Liebman, J. (1992). Toward a new contrastive rhetoric: Differences between Arabic and Japanese rhetorical instruction. *Journal of Second Language Writing*, 1(2), 141-165.
- Lorenz, G. (1999): *Adjective intensification - Learners versus native speakers: A*

corpus study of argumentative writing. Amsterdam: Rodopi.

Lyons, J. (1979). 'Deixis and anaphora' in (Ed.) T. Myers *The Development of Conversation and Discourse*. Edinburgh University Press.

Mahlberg, M. (2006). Lexical cohesion: Corpus linguistic theory and its application in English language teaching. *International Journal of Corpus Linguistics*, 11 (3), 363-383.

Makuta, M.A. (1997). *A computational model of lexical cohesion analysis and its application to the evaluation of text coherence* (PhD dissertation). University of Waterloo, Ontario, Canada.

Matalene, C. (1985). Contrastive rhetoric: An American writing teacher in China. *College English* 47(8), 789-808.

Matsuda, P. K. (1997). Contrastive rhetoric in context: A dynamic model of L2 writing. *Journal of Second Language Writing*, 6(1), 45-60. [Reprinted in Silva, T., & Matsuda, P. K. (Eds.). (2001). *Landmark essays on ESL writing*. Mahwah, NJ: Erlbaum.]

Matsuda, P. K. (2001). On the origin of contrastive rhetoric. *International Journal of Applied Linguistics*, 11(2), 257-260.

McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge; New York: Cambridge University Press.

McCarthy, M., & Carter, R. (1995). Spoken grammar: What is it and how do we teach it? *ELT Journal*, 49(3), 207-218

Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching and Linguistics Abstracts*, 13, 221-246.

- Meara, P. (1984). The study of lexis in interlanguage. In A. Davies & C. Cramer & A. P. R. Howatt (Eds.), *Interlanguage* (pp. 225-235). Edinburgh: Edinburgh University Press.
- Meara, P. (1996). The classical research in L2 vocabulary acquisition. In G. Anderman & M. Rogers (Eds.), *Words, words, words: The translator and the language learner* (pp. 27-40). Clevedon: Multilingual Matters.
- Meara, P. (1997). Towards a new approach to modelling vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 109-121). Cambridge: Cambridge University Press.
- Meara, P. (1998). [Review of Second language vocabulary acquisition: A rationale for pedagogy]. *Applied Linguistics*, 19, 289-292.
- Meara, P. (Ed.). (1989). *Beyond words*. London: British Association for Applied Linguistics, in association with Centre for Information on Language Teaching and Research.
- Mel'cuk, IA. and Pertsov, NV. (1987). *Surface Syntax of English*. John Benjamins.
- Mohan, B., & Lo, W. A-Y. (1985). Academic writing and Chinese students: Transfer and developmental factors. *TESOL Quarterly*, 19(3), 515-534.
- Moreno, A. I. (2004). Retrospective labeling in premise–conclusion metatext: An English–Spanish contrastive study of research articles on business and economics. *Journal of English for Academic Purposes*, 3(4), 321-339.
- Mountain, J. (1987). *An investigation of some textual properties of discursive compositions written by native speakers and L2 (Italian) advanced learners* (M.A. Project, University of Birmingham).
- Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists.

In Schmitt, N. and M. McCarthy (Eds.): *Vocabulary: Description, Acquisition and Pedagogy*: Cambridge, Cambridge University Press. 6-19

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, Mass.: Heinle & Heinle Publishers.

Nida, Eugene A. (2001). *Language, Culture and Translating*. Shanghai: Shanghai Foreign Language Education Press.

Ostler, S. E. (2002). Contrastive rhetoric: An expanding paradigm. In J. Flowerdew (Ed.) *Academic Discourse*, 167-181.

Ostler, S.E. (1987). English in parallels; a comparison of English and Arabic prose. In U. Connor, R.B.Kaplan (Eds.) *Writing Across Languages: Analysis of L2 Text*. Reading, MA: Addison-Wesley 1987,169-185.

P éry-Woodley, M-P. (1990). Contrasting discourses: contrastive analysis and a discourse approach to writing. "State of the Art article". *Language Teaching* 23(3), 143-151.

Phillips, M. (1985). *Aspects of Text Structure: an investigation of the lexical organization of text*. Amsterdam: North-Holland.

Ponterotto, D. (2000). The cohesive role of cognitive metaphor in discourse and conversation. In: Barcelona, A Metaphor and Metonymy at the Crossroads. A Cognitive Perspective. Berlin: Mouton De Gruyter. 283-298.

Poskiene, A. (2009). The role of parallel structure as a cohesion device in a technical text. *Studies about Languages*, 14, 5-12.

- Prodromou, L. (1996). Correspondence. *ELT Journal*, 50(4), 371-373.
- Purves, A. C. (1988). *Writing across languages and cultures: Issues in contrastive rhetoric*. Newbury Park, CA: Sage.
- Reynolds, D. (1995). Lessons in lexical aspect: A teacher perspective. Contribution to a colloquium entitled "What teachers can tell researchers." The Twenty-Ninth Annual Convention of TESOL. Long Beach, CA
- Reynolds, D. (1995). Repetition in nonnative speaker writing: More than quantity. *Studies in Second Language Acquisition*, 17, 185-209.
- Reynolds, D. (1996). Repetition: From the lexically impoverished to the effective writer. Paper presented at The Thirtieth Annual Convention of TESOL. Chicago, IL.
- Scarcella, R. (1984). Cohesion in the writing development of native and non-native English speakers. Unpublished doctoral dissertation, University of Southern California, Los Angeles.
- Scollon, R. (1997). Contrastive rhetoric, contrastive poetics, or perhaps something else? *TESOL Quarterly*, 31(2), 352-358.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (1997). Corpus Evidence in Language Description. In Wichmann, A et al (Eds.), *Teaching and Language Corpora*. (pp. 27-39). Longman: London and New York.
- Skuja, R. (1984). *An analysis of the organization features of argumentative*

compositions written by teachers and pupils in Singapore (M.A. Project, University of Birmingham).

Stubbs, M. (1996). *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture*. Oxford and Cambridge, Mass.: Blackwell.

Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Blackwell.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. New York: Cambridge University Press.

Teich, E. & Fankhauser, P. (2003). WordNet for lexical cohesion analysis. *Proceedings of the 2nd Global WordNet Conference*, Masaryk University, Brno, Czech Republik, January 2004, pp. 326—331.

Tierney, R. J., & Mosenthal, J. H. (1983). Cohesion and textual coherence. *Research in the Teaching of English*. 17, 215-229.

Tyler, S. A. (1978). *The Said and Unsaid* New York: Academic Press.

van Dijk, T. A. (1972). *Some Aspects of Text Grammars: a study in theoretical linguistics and poetics*. The Hague: Mouton.

Vechtomova O., Karamuftuoglu M. & Robertson S.E. (2006). On document relevance and lexical cohesion between query terms. *Information Processing and Management*, 42(5), 1230-1247.

Volk, M (2002). Using the web as corpus for linguistic research. In: Pajusalu, R; Hennoste, T. *ihendusepüüdja. Catcher of the Meaning. A Festschrift for Professor Haldur Õim*. University of Tartu, 1-10.

- Wallis, S. A. & G. Nelson. 2001. "Knowledge discovery in grammatically analysed corpora". *Data Mining and Knowledge Discovery*, 5(4), 305-336.
- Wang, S.P. (2005). Corpus-based approaches and discourse analysis in relation to reduplication and repetition. *Journal of Pragmatics* 37 (5), 505–540.
- Webber, B. L. (1978). *A Formal Approach to Discourse Anaphora*. Report No.3761. Cambridge, Mass.:Bot, Beranek & Newman.
- Wei,X. & He, X. (2009). Study on metaphor and its function in text coherence and cohesion. *Journal of Changchun University of science and technology*, 22(3), 24-30.
- Widdowson, H.G . (1991). 'The description and prescription of language' in J. Alatis (Ed.): *Georgetown University Round Table on Languages and Linguistics 1991*. Washington, D.C.: Georgetown UP, 11-24.
- Winter, E.O. (1974). 'Replacement as a function of repetition: a study of some of its principal features in the clause relations of contemporary English'. Ph.D. Thesis, University of London.
- Winter,E.O. (1979). 'Replacement as a fundamental function of the sentence in context'. *Forum Linguisticum*, 4 (2), 95-133.
- Witte, S. P. (1992). Toward a constructivist semiotic of writing. *WrittenCommunication* 32(2), 189-204.
- Witte, S. P., Faigley, L. (1981). Coherence, cohesion and writing quality. *College Composition and Communication*, 3, 189-204.
- Zhang, C. H. (2010). A comparative corpus-based study of semantic prosody.

- 曹慧. 从留学生作文看篇章层面的词汇教学[J]. 语言文字应用, 2002, (2).
- 常馨. 基于 HSK 动态作文语料库研究留学生习得程度副词“很”的偏误[J]. 教师, 2011, (8).
- 陈萍. HSK 中国汉语水平考试(高等)中的修辞问题[J]. 云南师范大学学报(对外汉语教学与研究), 2003, (1).
- 陈晨. 英语国家学生学习汉语在篇章连贯方面的常见偏误[J]. 四川大学学报(哲学社会科学版), 2005, (3).
- 陈望道. 修辞学发凡[M]. 上海: 上海教育出版社, 1979.
- 陈小荷. 跟副词“也”有关的偏误分析[J]. 世界汉语教学, 1996, (2).
- 崔希亮. 日朝韩学生汉语介词结构的中介语分析[J]. 中国语言学报, 2003, (11).
- 高书贵. 关于对外汉语教材生词与英文对译词的语义错位问题[J]. 天津师范大学学报, 1997, (6).
- 葛本仪. 汉语的词和词汇探析[M]. 山东大学出版社, 2003.
- 桂孟秋. 面向留学生的“被”字句语用教学—基于 HSK 动态作文语料库的研究[J]. 江汉大学学报(人文科学版) [J]. 2011, (1).
- 郭纯洁, 刘芳. 外语写作中母语影响的动态研究[J]. 现代外语, 1997, (4).
- 郭哥. 英汉语篇口译中的衔接与连贯[J]. 洛阳师范学院学报, 2007, (6).
- 何起凤. 对外汉语写作教学中常见的问题及应对措施[J]. 广东农工商职业技术学院学报, 2007, (4).
- 洪波. 对外汉语成语教学探论[J]. 中山大学学报论丛, 2003, (2).
- 胡裕树. 现代汉语[M]. 上海: 上海教育出版社, 1995.
- 胡壮麟. 语篇的衔接与连贯[M]. 上海外语教育出版社, 1994.
- 黄玉花. 韩国留学生的篇章偏误分析[J]. 中央民族大学学报(哲学社会科学版), 2005, (5).

- 纪蓉琴. 中英思维模式差异对英汉学术论文写作的影响[J]. 华东交通大学学报, 2006. (6).
- 柯汉琳(2010). 高考题缺乏想象力, 优秀作品逐年减少, retrieved from <http://zt.rednet.cn/c/2010/06/08/1976398.htm>
- 李绍林. 对外汉语教学词义辨析的对象和原则[J]. 世界汉语教学, 2010, (03).
- 李文中. 基于英语学习者语料库的主题词研究[J]. 现代外语, 2003. (2).
- 李文中. 语料库、学习者语料库与外语教学[J]. 外语界, 1999, (1).
- 李亚男. 汉语作为第二语言测试的作文自动评分研究[D]. 北京: 北京语言大学, 2006.
- 李占平. 关于反义词研究的几个问题[J]. 西南大学学报 (社会科学版), 2009,(4).
- 刘宝. 对外汉语教学中的同义词研究综述[J]. 四川职业技术学院学报, 2011, (2).
- 刘俊玲. 留学生作文中的篇章偏误类型[J]. 语言文字应用, 2005, (9).
- 刘宽平, 周业芳. 英汉思维差异对中国学生 EFL 写作的影响[J]. 外语学刊, 2004. (5).
- 刘叔新. 同义词和近义词的划分[C]. 词汇学和词典学问题研究, 天津: 天津人民出版社, 1984.
- 刘叔新. 同义词和近义词的划分[C]. 语言研究论丛, 天津人民出版社, 1980.
- 刘燕燕. 基于 HSK 动态作文语料库的留生成语语义偏误分析[J]. 文教资料, 2011, (36).
- 刘哲、王佳平. 对负语用迁移影响英语写作的分析与对策[J]. 东北大学学报 (社会科学版), 2003, (3).
- 卢桂芝. 有关对外汉语教材中生词英文注释问题[J]. 暨南大学华文学院, 2010, (1).
- 陆敏. “滥情作文”何以泛滥”——谈福建高考作文命题与评卷[R]. 厦门: 2011.
- 陆庆和. 对外汉语教学中的修辞问题[J]. 语言教学与研究, 1998, (2).

- 吕叔湘, 朱德熙. 语法修辞讲话[M], 北京: 开明书店, 1952.
- 吕叔湘. 汉语语法分析问题[M]. 商务印书馆, 1979.
- 罗青松. 外国人汉语学习过程中的回避策略分析[J]. 北京大学学报(哲学社会科学版), 1999, (6).
- 梅立崇. 词性不同的词能否构成同义词[J]. 逻辑与语言学习, 1987, (1).
- 南勇. 留学生汉语写作教学刍议[J]. 汉语学习, 1994, (6).
- 聂丹. 汉语水平考试(HSK)写作评分标准发展概述[J]. 云南师范大学学报, 2009, (6).
- 潘春雷, 吴格奇. 英语写作中的语用迁移与语用失误[J]. 杭州师范学院学报, 2006. (06).
- 潘先军. 简论对外汉语教学中的成语问题[J]. 汉语文教学与研究, 2006, (1).
- 彭婷. 图示理论与对外汉语写作教学[J]. 科教导刊, 2010, (2).
- 彭小川. 关于对外汉语语篇教学的新思考[J]. 汉语学习, 2004. (2).
- 郗志良. 对外汉语教学中词义辨析的几个问题[J]. 世界汉语教学, 1988, (1).
- 钱乃荣. 汉语语言学[M]. 北京语言出版社, 1995.
- 乔俊. 二十一世纪以来对外汉语同义词教学研究述评[J]. 中外教育研究, 2010, (2).
- 石安石. 关于词义的概念[J]. 中国语文, 1961, (8).
- 孙汝建. 同义词说略[J]. 语言文学, 1982, (2).
- 孙西谨. 从复现矩阵看越南留学生复现手段的使用特点[J]. 现代语文(语言研究), 2007, (4).
- 孙晓华. 近年来留学生篇章衔接偏误研究综述[J]. 语言教学研究, 2008, (2).
- 孙晓华. 论近年来留学生篇章衔接偏误研究综述现代语文[J]. 语言研究, 2008, (2).
- 谭汝为. 修辞文化与对外汉语教学[J]. 绍兴文理学院学报, 2004, (2).

- 陶岩. 基于语料库的留学生写作衔接偏误分析及对写作教学的启示[D]. 东北师范大学, 2010.
- 田然. 留学生限定话题语篇中词汇衔接状况考察[J]. 云南师范大学学报(对外汉语教学与研究版), 2006, (1).
- 王分年. 中级阶段越南留学生使用同义动词的偏误分析[D]. 广西民族大学, 2009.
- 王理嘉, 候学超. 怎样确定同义词[C]. 语言学论丛, 第五辑, 1963, 商务印书馆.
- 王若江. 留生成语偏误诱因分析—词典篇[J]. 暨南大学华文学院学报, 2001, (3).
- 王姝蘅. 美国学生习得"了"的偏误—基于 HSK 动态作文语料库[J]. 重庆与世界, 2011, (5).
- 王文宇, 文秋芳. 母语思维与二语写作—大学生英语写作过程研究[J]. 解放军外国语学院学报, 2002, (4).
- 王晓辉. 对外汉语写作课面临的问题及对策[J]. 吉林广播电视大学学报, 2010, (12)
- 文秋芳, 郭纯洁. 母语思维与外语写作能力的关系—对高中生英语看图作文过程的研究[J]. 现代外语, 1998. (4).
- 吴裴. “是…的”句式的偏误分析及教学策略—基于自建小型汉语中介语语料库[J]. 青年作家: 中外文艺, 2010, (9).
- 吴双. 留学生汉语写前构思活动对其作文质量的影响[J]. 世界汉语教学, 2011, (1).
- 吴艳. 对外汉语教材的生词注释研究[J]. 沈阳农业大学学报 (社会科学版), 2010, (5).
- 吴颖, 阙紫江. 学习者语料库在外语教学研究中的应用—以错误分析的数据收集为例[J]. 外语电化教学, 2005, (6).
- 吴玉花. 中美英语新闻词汇衔接手段的统计分析法[J]. 洛阳师范学院学报, 2007. (1).
- 肖华. 对外汉语教材翻译问题初探[J]. 商情, 2010, (25).

- 肖静, 王惠莲. 以英语为母语的外国留学生"了"的偏误分析—基于 HSK 动态作文语料库[J]. 大众文艺, 2009, (1).
- 肖奚强. 外国学生照应偏误分析—偏误分析丛论之三[J]. 汉语学习, 2001, (1).
- 谢文庆. 同义词[M]. 湖北人民出版社, 1982.
- 杨彬. 基于 HSK 动态作文语料库的"巴不得—恨不得"研究[J]. 青年文学家, 2011, (4).
- 杨春. 英语国家学生初级汉语语篇照应偏误考察[J]. 汉语学习, 2004, (3).
- 杨德峰. 试论修辞教学在对外汉语教学中的地位[J]. 修辞学习, 2001, (6).
- 杨德峰. 也谈修辞在对外汉语教学中的作用[C]. 汉外语言对比与偏误分析论文集, 北京: 北京大学, 1999.
- 杨梅, 王岩. 语料库语言学研究评述[J]. 长春师范学院学报, 2005, (1).
- 杨晓黎. 由表及里, 形具神生—对外汉语成语教学探论[J]. 安徽大学学报, 1996, (1).
- 杨玉铃. 留生成语偏误及《留学生多功能成语词典》的编写[J]. 辞书研究, 2011, (1).
- 于宏梅. 对外汉语写作教学中的修辞教学[J]. 乐山师范学院学报, 2004, (6).
- 俞燕明. 中介语对比分析法及其对词汇教学的启示[J]. 国外外语教学, 2004, (2).
- 张宝林. "HSK 动态作文语料库" 简介[J]. 测试研究, 2003, (4).
- 张博. 第二语言学习者汉语中介语易混淆词及其研究方法[J]. 语言教学与研究, 2008, (6).
- 张博. 同义词、近义词、易混淆词: 从汉语到中介语的视角转移[J]. 世界汉语教学, 2007, (3).
- 张德禄. 论衔接[J]. 外国语, 2001, (2).
- 张晶. 英汉部分新闻语篇衔接手段对比分析[J]. 煤炭高等教育, 2005, (5).
- 张丽平, 顾飞荣. 国内语料库语言学研究评述[J]. 韶关学院学报(社会科学版), 2007, (5).

- 张榕. 对外汉语教材中的英文翻译偏误分析及解决方法[J]. 海外华文教育, 2011, (3).
- 张瑞芳, 杨伊生. 留学生汉语学习焦虑与 HSK 考试成绩的相关分析[J]. 内蒙古师范大学学报(自然科学汉文版), 2011, (2).
- 张世禄. 词义和词性的关系[J]. 语文学习, 1956, (7).
- 张永芳. 外国留学生使用汉语成语的偏误分析[J]. 语言文字应用, 1999, (3).
- 张永言. 词汇学简论[M]. 武汉: 华中工学院出版社, 1982.
- 赵成新. 留学生汉语语篇衔接偏误目的语因素考察[J]. 周口师范学院学报, 2005, (7).
- 赵成新. 外国留学生汉语语篇衔接方式偏误分析[J]. 台州学院学报, 2005, (2).
- 赵成新. 外国留学生汉语语篇衔接方式偏误分析[J]. 台州学院学报, 2005, (4).
- 赵新, 李英. 对外汉语教学中的同义词辨析[J]. 暨南大学华文学院学报, 2001, (2).
- 周健, 彭彩虹. 中高级汉语教学应突出修辞能力培养[J]. 汉语学习, 2005, (3).
- 周莉. 对外汉语教学中的近义词辨析[J]. 理论观察, 2004, (1).
- 周小兵, 洪炜. 中高级留学生汉语中介语修辞使用情况考察[J]. 世界汉语教学, 2010, (4).
- 周祖谟. 汉语词汇讲话(八) [J]. 语文学习, 1956, (2).
- 朱其智. 偏误生成学的范围和方法[J]. 学术研究, 2009, (8).
- 朱永生. 韩礼德的语篇连贯标准[J]. 外语教学与研究, 1997, (1).
- 朱永生. 衔接理论的发展和完善[J]. 外国语, 1995, (3).
- 朱永生. 英汉语篇衔接手段对比研究[M]. 上海外语教育出版社, 2001.

Appendix I: Sample of Chinese composition test in NCEE of 2007

【作文题】

根据下面的材料，自选角度，自拟题目，联系实际，写一篇不少于 800 字的议论文。

一只蜗牛，很想做成一番惊天动地的大事业。开始它想东游泰山，一直爬到山顶，可一计算，要实现这个计划，至少需要 3000 年时间，只好忍痛放弃这个打算。后来它又想南下爬到长江边上，看一看奔腾的江水，可一计算，至少也需要 3000 年时间。蜗牛知道自己的生命非常短暂，不禁十分悲哀，于是什么也不肯做，最终死在了野草丛中。

(Sample taken from: <http://www.duyuxie.com/dyx/gz20070401/20434821.htm>)

Appendix II: Sample of Chinese composition test in HSK.

HSK 汉语水平考试试卷(高等)样题-作文 (2007)

1.HSK(高等)作文试卷样题

考试要求

- 1 考试题目：《一封求职信》(请根据招聘启事的要求写)。
- 2 书写要求：全部用汉字书写(也可以用繁体字)，每个空格写一个汉字。汉字书写要清楚工整。标点符号要正确，每个标点占一个空格。
- 3 字数要求：400~600 字。
- 4 书写格式：书信体格式。
- 5 考试时间：30 分钟。

精美服装公司招聘启事

本公司由于业务发展的需要，特招聘以下人员：

服装设计师：2 名。

广告设计师：2 名。

推销员：2 名(具有两年以上推销经验)，精通一门外语。

服装模特：男，2 名，身高 1 80 米以上，相貌端正。

女，2 名，身高 1 70 米以上，相貌端正。

应聘人员请将本人简历、从事这一工作的经验、成绩、本人特长等情况写明，寄往我公司，并注明联系地址。

联系人：季翔

精美服装公司

(sample obtained from:

<http://www.foreignercn.com/bbs/viewthread.php?tid=4309>)

Appendix III: Letter requesting permission to access data texts

Dear Sir or Madam:

I wish to request a sample of Chinese interlanguage writing texts from the HSK Composition Corpus and the Chinese Interlanguage Corpus administered by the centre for Studies of Chinese as a Second Language.

I am a PhD student of the Department of General and Applied linguistics in Faculty of Arts and Social Sciences at the University of Waikato in New Zealand, under the supervision of Dr. Ian Bruce, Dr. James McLellan (Department of General and Applied Linguistics), and Dr. Xiaoning Wang (Department of Humanities). I am interested in investigating the patterns of lexical cohesion in Chinese non-narrative text and providing teaching strategies and advice for the teaching of Chinese writing and the development of writing textbooks and teaching materials.

The proposed research will adopt quantitative and computational methodologies to analyze and compare the use of repetition devices in compositions written by both native Chinese students and learners of Chinese as a foreign language. Consequently, the project will involve establishing a Chinese inter-language corpus for its own specific purpose of research. Specifically, the texts I am requesting will be having the following attributes:

The grade the text given by examiner:	equivalent to A-, A or A+
The task assigned to the writer:	giving opinions on specific topics
The length of text:	400 or more words
The total number of writings requested:	100

I should like to assure you that the research will adhere strictly to the University of Waikato Human Research Ethics Regulations (2008). The data provided will only be used for the academic purposes of this research thesis and any resulting journal or conference presentations, where the data will be appropriately referred to through in-text citation and in reference lists (bibliographies).

The data received will be coded and stored in the format of plain text in the hard drive of the Computer at my office in the Department of General and Applied Linguistics, to which only the chief supervisor and I have access. It will be kept safe there for five years for the purpose of academic examinations and reviews. After that, all data related to the research project will be destroyed.

This research has been approved by the Human Ethics Research Committee of the University of Waikato, and any questions regarding the ethical conduct of this project may be addressed to the Secretary of the Committee (fass-ethics@waikato.ac.nz). If you have further enquires about the project, please contact me directly by e-mail: sl205@students.waikato.ac.nz or contact my supervisor Dr. Ian Bruce by Email: ibruce@waikato.ac.nz, Dr. James McLellan mclellan@waikato.ac.nz, and Dr. Xiaoning Wang xnw@waikato.ac.nz.

In return for your offer of the data requested, I undertake to share my research findings when the project is completed.

Thank you in advance for your attention to this request and I look forward to your reply.

Yours sincerely,

Shouji Li

Appendix: IV: Chinese version of Letter requesting permission to access data texts

尊敬的 HSK 动态作文语料库管理专家：

您好！

本人是新西兰怀卡托大学社会科学学院普通语言学与应用语言学系的在读博士生。本人的研究课题是汉语非叙述性语篇中词汇衔接模式的研究及其在对外汉语写作教学中的应用。该课题拟采用中介语对比分析法，通过比较中国学生和母语为英语的汉语学习者在使用词汇衔接上的异同而找出母语为英语的汉语学习者在写作过程中所面临的困难与挑战以及可能的解决办法。为达到此目的，该研究需要建立两个微型语料库，一个为中国学生作文语料库，语料来源是近几年来中国高等考试语文学科考试中的满分作文；另一个为母语为英语的外国留学生汉语水平考试（HSK）作文语料库，语料来源拟定为向贵语料库申请获准使用的留学生考试作文。

基于上述研究计划与设计，本人特向贵语料库管理专家们提出申请，希望能够得到贵语料库的支持，批准本人使用贵语料库收取的部分语料。根据本研究的实际情况，本人将需要符合以下要求的语料：

- 作文分数： 相当与 A-, A, 或者 A+ 的水平
- 写作任务： 对所给的观点或看法进行评论（议论文）
- 作文长度： 不少于 400 字
- 作文篇数： 100 篇

本研究已经获得怀卡托大学人文伦理研究委员会的批准，并将严格遵守该委员会制定的各项管理条例与规定。您如果对这些条例及规定有什么疑问，可以与该委员会联系，其电子邮箱地址是：fass-ethics@waikato.ac.nz。从贵语料库取得的语料将只用于该课题的学术研究以及相关的学术研究报告及会议论文。该语料将在未来 5 年内由本人妥善保存，除本人及指导教师，其他人均无权使用与查阅。5 年后该部分语料将由本人销毁。

该申请书已经获得以下三位博士生导师的审阅与批准，如有任何问

题，请与他们联系。他们的姓名及联系地址是：Ian Bruce 博士：ibruce@waikato.ac.nz，James McLellan 博士：mclellan@waikato.ac.nz，王效宁博士：xnw@waikato.ac.nz。本人的电子邮箱地址是：s1205@students.waikato.ac.nz。

为感谢贵语料库对该研究的支持，本人在研究完成后将向您汇报这一项目所取得的成果。

恳请贵语料库专家考虑批准。

此致

敬礼

李守纪

2009年2月23日

Appendix V: Application for Ethical Approval

FASS HUMAN RESEARCH
ETHICS COMMITTEE
The University of Waikato
Private Bag 3105
Hamilton, New Zealand

Phone +64 7 856 2889
www.waikato.ac.nz



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Shouji Li ←
Dr Ian Bruce
Dr James McLellan
Dr Xiaoning Wang

2 March 2009

Dear Shouji Li

Application for Ethical Approval: **Different or Similar: Patterns of Lexical Cohesion in Chinese Non-Narrative Text...**

Thank you for submitting an Application for Ethical Approval to the Faculty of Arts and Social Sciences Human Research Ethics Committee. I received your Application on 26 February. It is a very clear and well-written Application.

I have looked at the methodologies of information collection that you are proposing to use for your PhD research project. None of them involve gaining information from live human participants or using personal identifying data that is not readily available from a public or semi-public source or archive. As you note, the web sites from which you will obtain native Chinese compositions fall within the public domain; and the source of the compositions by learners of Chinese as a foreign language is the HSK Composition Corpus at the Beijing Language and Culture University. In both cases, no individually identifiable participants are involved.

For these reasons, there is no requirement for your project to be submitted to the Human Research Ethics Committee. The University's Ethical Conduct in Human Research and Related Activities Regulations (2008) apply only to research involving live human participants.

Of course, research ethics in a broad sense do apply to your project. Your Application provides plenty of evidence that you are intending to follow very high standards of ethical practice by, for example, obtaining appropriate permissions for access to research material and acknowledging your sources.

With best wishes for a successful research project,

John Paterson
Chair
FASS Human Research Ethics Committee