A Study of the Use of Lexical Cohesion in Chinese Postgraduate Writing at a UK University

XUANHONG GUO

A Study of the Use of Lexical Cohesion in Chinese Postgraduate Writing at a UK University

Xuanhong Guo

Abstract

As a key feature in the creation of coherent texts (Tanskanen, 2006), lexical cohesion is of critical importance for students' academic performance. Chinese students, whose writing is the subject of this thesis, have been identified as lacking awareness of lexical cohesiveness in English academic writing (Zhang, 2000). In order to inform EAP pedagogy for them, this thesis used a corpus-based approach to conduct in-depth investigations of lexical cohesive devices used in Chinese postgraduates' writing at a UK university. Based on Halliday and Hasan's model (1976), an analytical framework for the analysis of lexical cohesion was developed in two corpora, incorporating a new sub-category of lexical cohesive device alongside modifications of existing categories. One corpus consisted of 52 module assignment samples (17,538 words) allocated into four marking-scale groups (failed, pass, merit and distinction), the other corpus comprising 45 dissertation excerpts (19,148 words) divided into five functional-section groups (introduction, literature review, methodology, findings/discussion, and conclusion). Applying this framework, manual analysis of the corpora identified homogeneities of lexical cohesion as context sensitivity, topic-based use of lexical cohesion, dominant use of repetition, and use of modifiers to indicate lexical cohesive relations, suggesting the value of context-based pedagogy and the need to teach complex lexical cohesive devices with exemplars. The results of the ANOVA test and the Kolmogorov-Smirnov test suggested a significant difference in the use of lexical cohesion between the marking-scale groups due to fewer repetition pairs identified in the merit group, and no statistically significant difference in overall the use of lexical cohesion among the functional-section groups although the function of each section influences the use of certain lexical cohesive devices. Several factors are proposed as influencing the use of lexical cohesion: topic variety, writers' choice and function of texts, indicating the complexity of both applying and teaching lexical cohesion in academic writing.

A Study of the Use of Lexical Cohesion in Chinese Postgraduate Writing at a UK University

By

Xuanhong Guo

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

School of Education
Durham University

July 2019

Table of contents

List of tables

List of figures

Declaration

I declare that this thesis, which I submit for the degree of Doctor of Philosophy at Durham University is my own work.

Xuanhong Guo
School of Education
Durham University

Acknowledgements

I would like to express my appreciation to Dr Philip Nathan for his comments, suggestions and support throughout the research process and write-up.

I also would like to express my thanks to Shihui Bi and Huiru Yuan, who assisted with the rater work described in this thesis, as well as acknowledging Professor Martha Young- Scholten who made suggestions during the research process.

Thanks are also due to my supportive parents and partner. Without them, this PhD life would be lonely and more challenging.

Chapter 1 Introduction

Cohesive devices help create the connectedness of texts (Halliday & Hasan, 1976, p. 2). They contribute to developing the meaningfulness of texts and impact upon communicative effectiveness (Tanskanen, 2006, p. 1). Based on the forms of expressing cohesive relations, cohesion is divided into grammatical cohesion and lexical cohesion (Halliday & Hasan, 1976, pp. 5-6).

Several models of cohesion analysis have been developed, showing how cohesion is manifested in different genres of discourse due to its context-sensitive nature (e.g. Halliday & Hasan, 1976; Hoey, 1991; Martin, 1992; Tanskanen, 2006). As with this mainstream practice of putting models of cohesion to use, this thesis explores the importance of the use of lexical cohesive devices in creating cohesive academic writing within a specific academic discipline, in order to provide EAP (English for Academic Purposes) pedagogical implications to Chinese students as EFL (English as a Foreign Language) learners.

Halliday and Hasan (1976) created the original model of cohesion in their seminal book, *Cohesion in English*, which provides the basis for subsequent developed models (Hoffmann, 2012; Martin, 1992; Tanskanen, 2006), including the model of lexical cohesion developed in this thesis. According to their original model, including the two types of cohesion, lexical and grammatical, grammatical cohesion comprises four categories, i.e. reference, ellipsis, substitution and conjunction. Lexical cohesion, the focus of this thesis, is defined as relations in which two or more lexical items connect to each other and to other cohesive devices, in order to "build the continuity of the text" (Flowerdew & Mahlberg, 2009, p. 1). Lexical cohesion in general consists of two recognised key categories: reiteration and collocation, while other sub-categories within these two categories are varied among models of lexical cohesion in previous studies. These various categorisations which underpin the research in this thesis, will be discussed in detail in chapters 2, 3 and 4 of this thesis.

Compared with the investigation of grammatical cohesion, it is more challenging to explore lexical cohesion, as grammar has a closed system with limited elements, such as pronouns and elliptical syntax, while vocabulary is an open system which comprises unlimited items (Halliday & Hasan, 1976, p. 274). This feature of vocabulary makes lexical cohesion in general a problematic object to study, and therefore, researchers have tended to focus more on grammatical cohesion rather than lexical cohesion. However, lexical cohesion is necessary for effective communication and in particular effective writing. As Hoey (1991) points out, "lexical cohesion is the only type of cohesion that regularly forms multiple relationships. If this is taken into account, lexical cohesion becomes the dominant mode of creating texture[1]" (p.10). Therefore, despite of its complexity, as a key tool for creating texture, lexical cohesion is seen here as an essential focus of investigation, this central role in effective communication and the creation of texture providing a prime motivation for the current research.

In terms of the application of cohesion in teaching and learning, several researchers have focused on the role of cohesion in EAP pedagogy. They have studied the use of cohesive devices across a variety of EAP contexts as well as in both native and non-native writing in foreign and second language settings (e.g. Ong, 2011; Sinicrope, 2007; Zhang, 2000). These researchers without exception believe that lexical cohesion is vital in textual cohesion, playing a crucial role in text interpretation whether the language user is a native or a non-native speaker (henceforth NNS) of English. As such an important component of communication, these authors believe that lexical cohesion should be studied in its own right. In addition, in particular with respect to the NNS students of EAP, lexical cohesion is valuable for their acquisition of foreign language (Bensemmane, 1985, p. 3) and their performance in an academic context.

However, although lexical cohesion has been the subject of study by such researchers in relation to EAP pedagogy, and given the fact that the use of lexical cohesion in EFL students' writing clearly merits careful investigation, surprisingly,

[1] Texture is a term to denote the property of 'being a text' (Halliday & Hasan, 1976, p. 2).

few studies solely focus on lexical cohesion in higher education students' actual on course writing when such studies have the potential to inform EAP pedagogy.

Moreover, most studies of cohesion investigate both grammatical and lexical cohesion (e.g. Todd, Khongput, & Darasawang, 2007; Zamel, 1983) or only concentrate on grammatical cohesion (e.g. Alarcon & Morales, 2011; Khider & Othman, 2019). When studying both grammatical and lexical cohesion, researchers are prone to simplifying examination of each subcategory of lexical cohesive devices, sometimes excluding some problematic categories or subcategories, such as collocation (e.g. McGee, 2009; Wang & Zhang, 2019) and general nouns[2] (e.g. Yang & Sun, 2012). Other studies focus only on particular types of grammatical cohesion, such as conjunctions (Lei, 2012; Mohammed, 2015) and reference (e.g. Lindgren & Vogels, 2018) .

The range of studies of lexical cohesion tend to be limited in that they do not analyse texts at a deep and precise level. This thesis study aims to both focus on lexical cohesion in real samples of on course student writing, and remedy some of the limitations found in many previous studies of lexical cohesion by conducting a detailed analysis of lexical cohesion in students' written assignments.

Specifically, this thesis study focuses on examining lexical cohesion in the academic writing of Chinese postgraduate students. A small number of studies of the use of lexical cohesive devices in Chinese postgraduates' academic writing have been undertaken. However, most of these studies have been conducted in Chinese academic contexts (e.g. Liu & Qi, 2010; Zhou, 2007), and one in the US context (Jin, 2001), which focus not only on lexical cohesion but also grammatical cohesion. On the other hand, there have not been any studies of lexical cohesion in the writing of Chinese postgraduate students in the UK higher education context. However, since there is consistent rise in numbers of Chinese students in UK universities, Chinese students have already been the largest subgroup of international students in the UK academic settings, in which the majority are

[2] General nouns are regarded as a subcategory of the reiteration category in Halliday and Hasan's (1976) model of lexical cohesion (p. 275).

Chinese postgraduates (British Council, 2017). Due to the lack of related studies of lexical cohesive devices used in Chinese postgraduates' writing in the UK context and the importance of Chinese postgraduates within the international student group, the features of lexical cohesive devices used in Chinese postgraduates' academic writing in the UK context is worthy of investigation, which motivates the current research to focus on this specific issue.

Several studies have been conducted that demonstrate a range of challenges faced by Chinese learners in using lexical cohesion (e.g. Green, Christopher, & Mei, 2000; Tickoo, 2002; Lake, 2004; Ong, 2011; Tarawhiti, 2016), such as students having limited vocabulary, repetition of the same words, errors in lexical choice, misuse of collocations and overuse of general nouns (Tarawhiti, 2016, pp. 21-22). The underlying reasons underpinning these challenges in terms of lexical cohesion have been discussed in several of these studies. For example, the almost ubiquitous method of learning lexical items in isolation within Chinese educational context renders Chinese students more prone to inappropriateness or misuse of lexical items, especially misuse of collocations in authentic contexts; and studying writing at the sentential level rather than at the textual level creates challenges for Chinese students in constructing a cohesive English text (Zhang, 2000, p. 61).

Such dilemmas for Chinese students regarding learning lexical cohesion might be seen as suggesting that a particular approach in EAP pedagogy may need to be designed for teaching Chinese students the use of lexical cohesion. The results of this thesis hopefully will lead to further suggestions as to how to fill the need.

Notably in the context of academic writing, there is no accepted understanding of the relation between writing quality and the use of lexical cohesive devices in students' written work (Zhang, 2004). Some researchers have compared writing quality of students' assignments and the frequency of grammatical cohesive devices (e.g. Alarcon & Morales, 2011). Some prefer to include both grammatical and lexical cohesion in one study and therefore may not have looked at lexical cohesion in sufficient depth (Zhang, 2000; McNamara, Crossley, & McCarthy,

2010). Other investigators suggest that it is not the quantity of cohesive devices used in written work, but the appropriate use of the devices that is relevant to writing quality (e.g. Ong, 2011, p. 45; Yang & Sun, 2012, p. 31). Due the lack of research focusing on the correlation between the use of lexical cohesion and the quality of Chinese students' academic written assignments, this thesis aims to fill this gap by examining the correlation between the use of lexical cohesion and the marking scales (i.e. failed, pass, merit and distinction) of Chinese postgraduates' module assignments.

A final key and under-researched area regarding lexical cohesion in academic writing is the relation between discourse as a process (mainly defined as the flow of text in this research) and the use of lexical cohesion. In the course of this thesis research, one study was identified addressing the comparison of the frequencies of several types of lexical cohesive device used in four sections of research articles (i.e. introduction, methodology, results and discussion/conclusion) (Wang & Zhang, 2019). According to Hyland (2018), discourse "is a process in which writers are simultaneously creating propositional content, interpersonal engagement and the flow of text as they write" (p. 31). In cohesive writing, ideas seem to flow from one sentence to the next throughout the whole text. New ideas build on previously mentioned ideas, which creates flow of text (William H. Hannon Library, 2019). In addition to the use of cohesive devices to help connect ideas, writers produce the components of cohesive pairs in a linear order when editing their written texts (Keevallik, 2009), thereby the cohesive devices contribute to the flow of text. This thesis aims to explore how Chinese students, as producers of MA dissertations, deploy lexical cohesive devices during this long linear text, from chapters of introduction, literature review and methodology to findings/discussion and conclusion chapters, to examine producers' thinking flow.

In order to investigate the characteristics of lexical cohesive devices used in Chinese students' English academic writing at the UK research site, and the areas of research identified already in this chapter, using the lens of text analysis, this thesis conducts an examination of two corpora of excerpts from Chinese students' module assignments and MA dissertations submitted to MA TESOL and MA Applied Linguistics for TESOL programmes at a UK university, with the goal

of adding to the body of knowledge concerning current international postgraduates' academic writing.

Three key process stages were involved in this research: firstly, a suitable model of lexical cohesion was developed based on previous studies for the corpus-based manual analysis of lexical cohesive devices used in the MA dissertation and MA module assignment samples. This was necessary, as empirical studies of lexical cohesion analysis have shown the contextual sensitivity of cohesion such that certain cohesive devices adopted in one discourse cannot be found in other discourses (Xi, 2010, p. 143). Secondly, initial quantitative analyses focused on comparing frequencies of different types of lexical cohesive device used in the four MA assignment marking-scale groups and in the five dissertation-section groups, exploring whether there is correlation between marks of Chinese students' module assignments and the frequency of lexical cohesive devices' usage, and whether there is significant variation during Chinese students' writing process in dissertation samples regarding the use of lexical cohesion. Thirdly, linked to the lexicogrammatical features of lexical cohesive devices included in the model of this thesis, it was also decided to conduct qualitative analyses of specific lexical cohesive relations identified in the corpora.

In summary, this thesis attempts quantitative analyses of frequencies of lexical cohesive devices identified in two corpora of MA dissertations and module assignments produced by Chinese postgraduates at the UK university research site, and explores the relationship between the use of lexical cohesion and the writing quality of module assignments as well as the use of lexical cohesive devices during the flow of MA dissertation texts. This is combined with in-depth linguistic characterisation of lexical cohesive devices involved in the corpora. The final component of the thesis is a discussion of EAP pedagogical applications with regard to the findings of the present study, focusing particularly on the teaching of certain types of lexical cohesive devices with their corresponding typical modifiers, combined with examples in specific contexts.

Following this introduction in chapter 1, the outline of the remaining chapters in this thesis is shown as follows: chapter 2 develops the classification of lexical

cohesion used for the lexical cohesive analysis in this thesis via reviewing different approaches adopted in previous studies, including corpus-based and corpus-driven analysis, manual analysis with or without co-raters, and computer-assisted large corpora investigation.

Chapter 3 and chapter 4 are elaborations of two categories of lexical cohesion – signalling nouns and collocation respectively. There are two reasons for these chapter-length elaborations: Firstly, signalling nouns and collocation are researched in several areas or used under different terms. It is necessary to clarify the definitions of these two categories in the present study to avoid misunderstanding. Secondly, compared with other categories of lexical cohesion, these two categories are more subtle to be detected in analysis, which need to be elucidated with ample examples to assist readers' general understanding of signalling nouns and collocation in this study.

Chapter 5 firstly describes the complexity of identifying Chinese students – their similarities and differences; and then moves to previous studies about lexical cohesion in the academic writing of Chinese students, mainly focusing on postgraduates' writing which is the object of the present study.

Chapter 6 is an explanation for all aspects of methodology adopted in this study, including two corpora, research procedures and manual analytical strategies.

Chapters 7 and 8 comprise the quantitative and qualitative analysis of the module assignment corpus and dissertation corpus respectively. Both of these chapters will give general descriptions of the quantitative analysis firstly in order to answer the research questions regarding the correlation between the use of lexical cohesion and the quality of written assignments, and the connection between the use of lexical cohesion and the writing flow of the dissertation. After that, these two chapters will demonstrate detailed qualitative analyses of excerpts selected from two corpora respectively in order to investigate features of different lexical cohesive devices. A further discussion of the results of the analyses with comparison with previous studies will also be presented.

Chapter 9 is the final section which will firstly discuss and summarise the results of lexical cohesion analyses presented in chapters 7 and 8, and then illuminate the limitation and contributions of this study as well as implications for EAP pedagogy based on the research results and discussions.

Chapter 2 Classificatory system of lexical cohesion

2.1 Introduction

This chapter mainly concentrates on summarising existing models of lexical cohesion and developing the classification of lexical cohesion to support the sample analysis in the present study. Specifically, section 2.2 reviews the models which are central to the analyses conducted in the present study. Having compared different models of lexical cohesive relations highlighting their similarities and differences, section 2.3 then focuses on a detailed elaboration of the key categories of lexical cohesion which are relevant to the corpus-based analysis in the present study, while another two major categories, signalling nouns and collocation, are discussed separately in the subsequent chapters. The discussed lexical cohesive categories are elucidated with examples from previous studies in order to illustrate the corresponding lexical cohesive relations as clearly as possible.

2.2 Classifications of lexical cohesive devices

Several models of lexical cohesion have been identified and these will be introduced in chronological order. The reason for selecting these models is that all of these models contribute to the development of the operational framework of lexical cohesion in the present study from different perspectives, which will be further explained with the review of each model respectively. The first model is Halliday and Hasan's classification of lexical cohesion which was introduced in their book *Cohesion in English*. This book sets a systematic theoretical foundation for related concepts and classifications of lexical cohesion, some of which become the foundation for this thesis. Based on this fundamental model, Hasan's (1984) work and Halliday's (1985) made several modifications respectively, after which two renovative perspectives from Hoey (1991; 1994) and Martin (1992) will be discussed. Finally, two more operational models from Tanskanen (2006) and Hoffman (2012) will be illustrated.

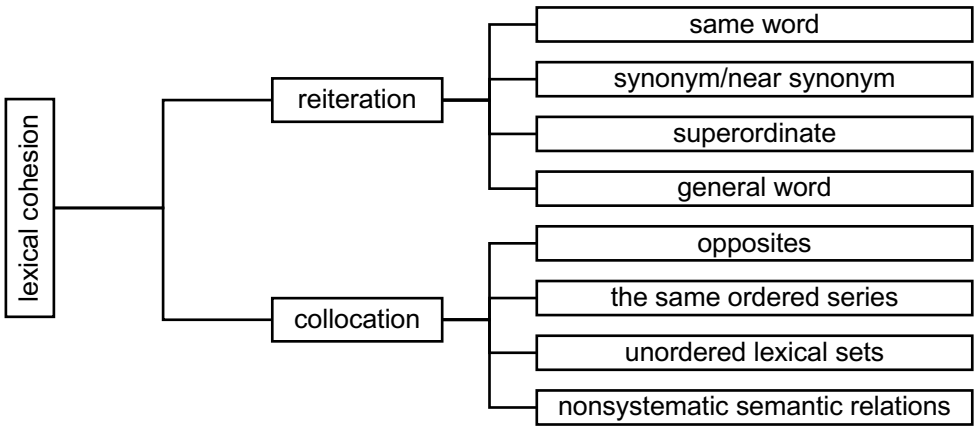
2.2.1 Halliday and Hasan: Fundamental model

As the present study regarding lexical cohesion is based on Halliday and Hasan's model, the categories of lexical cohesion mentioned in their model are also included in the model of lexical cohesion developed in this thesis. However, the inclusions of some lexical cohesive relations are slightly different in their model and the model in the present study. Therefore, it is necessary to firstly introduce the categories and related lexical cohesive relations involved in Halliday and Hasan's model before comparing their classification of lexical cohesion with that in the present study.

Although Halliday and Hasan (1967) mainly focus on grammatical cohesion, they also argue that without cohesive patterning of a lexical kind, a text will not be formed (p. 292), and, in fact, almost half of the cohesive ties analysed in their book are lexical (pp. 340-355). In this respect, lexical cohesion plays an important role in creating the texture of text, though the space allocated to lexical cohesion does not match its importance in their work. Despite their lack of focus on lexical cohesion, Halliday and Hasan developed a basic model for lexical cohesion, which provides a solid basis for further related research.

Specifically, their model of lexical cohesion (Halliday and Hasan, 1976, p. 338) is based on two main categories – reiteration and collocation. The detailed classification of this model is demonstrated in figure 1 as follows:

Figure 1 Classification of lexical cohesion in Halliday and Hasan's study (1976, p. 288)



Because the subsequent examples from *Cohesion in English* are all related to reiteration, these examples can illustrate related lexical cohesive relations in the reiteration category above more precisely:

2-1 There was 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** …

2-2 Accordingly … I took leave, and turned to the **ascent** of the peak. The **climb** is perfectly easy …

2-3 Then quickly rose Sir Bedivere, and ran,
And leaping down the ridges lightly, plung'd
Among the bulrush beds, and clutch'd the **sword**
And lightly wheel'd and threw it. The great **brand**
Made light'nings in the splendour of the moon …

2-4 Henry's bought himself a new **Jaguar**. He practically lives in the **car**.

2-5 Henry seems convinced there's money in dairy farming. I don't know what gave him that **idea**.

(Halliday & Hasan, 1976, pp. 275, 278)

Example 2-1 shows the repetition of the lexical item *mushroom*. In Example 2-2 *climb* relates back to *ascent*. As defined by the *Oxford English Dictionary*, *climb* is "an ascent" (OED[3] Online, 2019), which indicates that *clime* and *ascent* share similar meanings, and therefore, these two lexical items can be regarded as synonyms in the context of example 2-2. In example 2-3, *brand*[4] is used in this specific context as a synonym of *sword*, which illustrates that from the perspective of word meaning, *brand* and *sword* will not be regarded as synonyms. This results in classifying the relation between *brand* and *sword* into the near-synonym sub-category. Example 2-4 is an example in which the word *car* is used as a

[3] Oxford English Dictionary
[4] Generally a *brand* refers to a *sword* in poetic literary works (OED Online, 2019a).

superordinate of *Jaguar*. Superordinate refers to a more general class of an item. In this case, *car* is the name of a more general class which includes *Jaguar* as a member. Example 2-5 uses a general word *idea* to replace the whole previous clause, which is an economical method of avoiding redundant repetition. From examples 2-4 and 2-5, it is worthy of noting that the only difference between superordinate and general word is the level of generality in meaning (Halliday & Hasan, 1976, p. 278).

Overall, the examples discussed above have one feature in common, which is that the latter item refers back to the former item in a lexical cohesive pair, i.e. both items share the same referent or the relation between these two lexical items is co-referential. That is why these relations are called 'reiteration'. However, although sometimes the cohesive effect of lexical pairs formed by lexical items is strengthened by the co-referential relation between two items, this is not necessary for the items to form a cohesive pair. Pairs can also be cohesive without co-referentiality. Consider the example below:

2-6 There's a *boy* climbing that tree. Most *boys* love climbing trees.

(Halliday & Hasan,1976, p. 283)

In example 2-6, *boy* and *boys* do not share the same referent but still form a repetition relation as *boys* refers back to *boy* by repeating the form of *boy* itself. As Halliday and Hasan (1976) point out, "the cohesion exists as direct relation between the forms [of the lexical items] themselves" (p. 284).

Within the broad category of reiteration, there are several sub-categories which are frequently used in lexical cohesive analysis, as semantic relations between two lexical items in these sub-categories are relatively straightforward and thus are easily recognised. In contrast, another main category, collocation, is more elusive to be identified in text. It is somewhat unfortunate for Halliday and Hasan to call this category 'collocation' (Tanskanen, 2006, p. 33). Firstly, there is connection between 'cohesive' collocation and 'lexicographic' collocation in terms of definitions of this collocation category. Specifically, 'collocation' is widely used in the areas of lexicography and lexical semantics, in which the general definition

of collocation is “the relations of a word with other words it tends to occur with” (Tanskanen, 2006, p. 33). On the other hand, in *Cohesion in English*, collocation is defined as “cohesion that is achieved through the association of lexical items that regularly co-occur” (Halliday & Hasan, 1976, p, 284). It is noticeable that both of these definitions of collocation emphasise the ‘co-occurrence’ of lexical items as the prerequisite for these items being collocates. This similarity regarding the definitions of collocation may cause a problem so that misunderstanding is not avoidable if the context for the use of collocation is not clear.

Secondly, there is a main difference between ‘cohesive’ collocation and ‘lexicographic’ collocation in terms of research perspectives. The focus of the former is on the cohesive function of collocation, which means the distance of two cohesive lexical items can be further, such as above the sentential level, and the items tie the clauses in which they appear together by their cohesive force; while for the latter, the collocational items need to be next to each other or within the sentential level. Thirdly, both of these definitions are inspired by Firth who proposed the concept of ‘collocation’ in his article *Modes of meaning* in 1957. In his work, collocation is related to the expression *meaning by collocation*, which means “the meanings of words depend upon their co-occurrence in texts” (Tanskanen, 2006, p. 33). Therefore, it is reasonable for Halliday and Hasan to adopt the term ‘collocation’ from Firth because of the ‘co-occurrence’ concept mentioned in Firth’s definition of collocation, though the term ‘collocation’ itself may cause confusion as this term had already been widely used in other research areas. The three reasons mentioned above indicates the fact that the understanding of ‘cohesive’ collocation in Halliday and Hasan’s model is challenging, and therefore, it is crucial to figure out the specific context in which the term ‘collocation’ is used.

Before discussing the most intricate relations in ‘cohesive’ collocation, there are other types of relations included in this category in Halliday and Hasan’s model, namely ‘opposites’, ‘the same ordered series’ and ‘unordered lexical sets’. Compared with non-systematic semantic relations which are more complicated, these three types of collocational relations are systematically semantic and more

easily recognised. The first type, 'opposites', is further divided into three sub-types (see further explanation in section 2.3.6):

1. complementaries (e.g. boy – girl)
2. antonyms (e.g. like – hate)
3. converses (e.g. order – obey)

(Halliday and Hasan, 1976, p. 285, based on Lyon, 1969)

The second type of collocational relations in Halliday and Hasan's model is 'the same ordered series', such as days of a week (*Tuesday* – *Thursday*) or months of a year (*January* – *February*). The third type is 'unordered lexical set', which includes

1. part-whole relation (e.g. car – brake)
2. part -part relation (e.g. mouth – chin)
3. co-hyponyms[5] (e.g. chair – table)

(Halliday & Hasan, 1976, p. 285)

As regards opposites, it is rarely possible for the items in an opposite relation to have the same referent but their proximity in texts and their semantic connection also contribute to the cohesion in texts; in terms of the other two types, the possibility of denoting the same referent in relations with these types is much higher than that with opposites, and the semantic connections between the lexical items in these relations are related to superordinate-subordinate relations (see p. 38), which definitely helps create cohesion.

Halliday and Hasan (1976) claim that the principle of dividing lexical cohesion into two main categories (reiteration and collocation) is based on whether the relations between two lexical items are co-referential or are "of the form of reiteration accompanied by *the* or a demonstrative" (p. 287). That is to say, relations which cannot meet this principle will be included in the collocation

[5] The names of these sub-types follow the original expressions in Halliday and Hasan's work. Part – whole relation is also called 'meronymy'; likewise, part – part relation is called 'co-meronymy'; co-hyponyms refer to "both members of the same more general class" (Halliday & Hasan, 1976, p. 285).

category, which leads to the inclusion of the most problematic and intricate relations in collocation.

In terms of these intricate relations in 'cohesive' collocation, "whose meaning relation is not easy to classify in systematic semantic terms" (Halliday & Hasan, 1976, p. 285), these collocational relations have been referred to as "non-systematic semantic relations" (Morris & Hirst, 1991), which was also the focus of the present study regarding the collocation category. Compared with the reiteration category, lexical items in the non-systemantic semantic relations do not repeat each other or have co-referential relations, but only have subtle associations between them. Because of this subtlety, these collocational relations are comparatively subjective relations. As Tanskanen (2006) points out, "what is considered as a valid relation will inevitably [vary slightly] from one communicator to the next" (p. 34). The reason for including this challenging category is that collocation is related to the connectivity of text-knowledge and world-knowledge, which is essential for the functioning of cohesion and thus was considered essential for lexical cohesive analysis in the current research. As de Beaugrande (1980) points out, "often, no special consideration is given to the underlying connectivity of text knowledge and world-knowledge that makes these [cohesive] devices possible and useful, except in the discussion on lexical cohesion by Halliday and Hasan" (p. 132).

The example below serves to demonstrate the connectivity of text-knowledge and world-knowledge:

> 2-7 ... and that is the effect of changes in the curriculum the ways of teaching in schools – this is not anything to do necessarily with **comprehensive schools** or the abolition of **the grammar school** – it is notable that in this country it is the middle classes themselves who have revolted against the conception of **the eleven plus** – but those of us who thought that you should postpone the age at which irrevocable decisions were taken about a child's education...
>
> (Tanskanen, 2006, p. 34)

Example 2-7 demonstrates a speech at a British university, talking about issues in the British educational system. The lexical items *comprehensive schools* and *the grammar school* are related by the repetition of *school*. This illustrates the cohesive function of text knowledge. By contrast, *the eleven plus* is not easily interpreted into this educational context unless the background knowledge is known by readers that *the eleven plus* is a test to assess students' aptitude for continuation to *the grammar school*. This background knowledge comes from readers' world knowledge, which is the prerequisite for the readers to correctly interpret the meaning of *the eleven plus* used in the succeeding discourse in order to avoid the interruption of their interpretation of other information in the speech.

Although there are some difficulties in the identification of collocation, Halliday and Hasan's model was the first one to use collocation as a cover term to discuss these complicated lexical cohesive relations mentioned above, and provides the basic classification of lexical cohesion. A large number of subsequent studies are based on their model but develop their models for specific contexts (e.g. Hasan, 1984; Halliday, 1985; Hoey, 1991; Morris & Hirst, 1991; Martin, 1992; Tanskanen, 2006; Hoffmann, 2012). Likewise, Halliday and Hasan's basic categorisation (reiteration and collocation) was also followed in the present study.

2.2.2 Hasan: 'Instantial category' as a new category of lexical cohesion

Based on Halliday and Hasan's model, Hasan (1984) developed a new model of lexical cohesion for her study of children's stories, which includes two main categories: *general* and *instantial*. The 'general' category refers to the relations which can be explained by the general semantic system of English, such as repetition, synonymy, hyponymy, meronymy and antonymy. Although two sub-categories which used to belong to the original collocation category in Halliday and Hasan's model are included in the general category, i.e. antonymy and meronymy, the whole collocation category is not incorporated in her model. Hasan's rationale for exclusion of some of collocational relations is that these relations are too subjective. For example, the collocational pair *the grammar school – the eleven plus* mentioned above will be left outside her model. The reason for this is that *grammar school* and *the eleven plus* only can be interpreted as collocates in the specific context which involves the discussion of school

systems in example 2-7, and the interpretation of their collocation relation is based on readers' subjective world knowledge rather than on general semantic relations between these two lexical items.

Hasan's main contribution to the present study is her introduction of a new category, i.e. the "instantial lexical cohesion" (Hasan, 1984, p. 201), in order to compensate for the exclusion of some 'subjective' collocational relations in her model. she referred this category as "a significant resource for textual unity" (p. 201). The 'instantial' category refers to the relations which are "text-bound", and the validity of these relations is "an artefact of the text itself, and does not extend to the system" (Hasan, 1984, p. 201). That is to say, the instantial lexical relations are created by the specific contexts in texts, rather than being general semantic relations, which includes three sub-categories:

1. Equivalence (the *sailor* was their *daddy*)
2. Naming (the *dog* was called *Toto*)
3. Semblance (the *deck* was like a *pool*)

(Hasan, 1984, p. 202)

The three instantial types of relations denote part of the relations included in a sub-category (i.e. other relations with identity reference) of reiteration in the present study, which will be discussed in section 2.3.7.

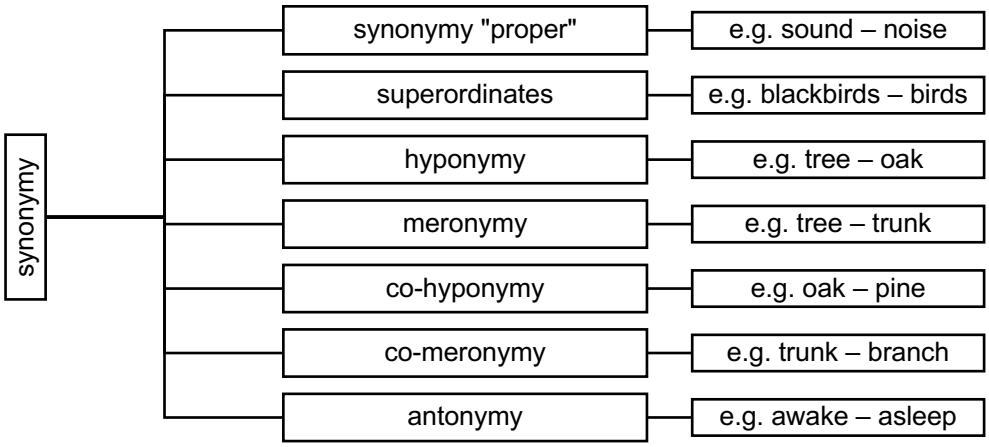
It is also worthy of mention that as Hasan focused on the analysis of narratives, i.e. children's stories, her examples are from this genre which include lexical cohesive relations at the intra-clausal level. However, as the present study focused on lexical cohesive relations at the inter-clausal level, the relations involved in Hasan's model were applied to a wider stretch of text for the investigation of lexical cohesion in this thesis.

2.2.3 Halliday: Revision of collocation and synonymy

A revised model of lexical cohesion based on Halliday and Hasan's original model was proposed by Halliday in his later work (1985; 1994). In Halliday's developed model, there are three main categories of lexical cohesion: repetition, synonymy

and collocation. Repetition (Halliday, 1994, p. 330) is largely the same as the one in Halliday and Hasan's model (1976). The synonymy category is expanded into a larger category encompassing several relations which Halliday regards as variations of synonymy, which are shown in figure 2 as follows:

Figure 2 Classification of the synonymy category in Halliday's study (1985; 1994)



The category of collocation is also included in Halliday's developed model but is reduced in scope compared to the original Halliday and Hasan's model (1976), and it includes only the relation between lexical items that depends on a certain association between them. The definition of collocation in this developed model simply refers to a tendency for items to co-occur, which helps readers anticipate what is to come next (Halliday, 1985, pp. 312-313). This modification regarding the boundary of collocation in Halliday's model narrows this collocation category into just non-systematic relations, which is exactly in line with the scope of the collocation category involved in the present study.

Furthermore, Halliday emphasised that collocational relations are tied to the type of texts in which they occur, i.e. collocation relations are context-oriented. For example, *hunting* can go with *shooting* and *fishing* in one text and with *souvenir* or *fortune* in another text (Halliday, 1985, p. 313). This example demonstrates the context-sensitive feature of lexical cohesion, which indicates that the research of lexical cohesion is more appropriate to focus on specific contexts in which the cohesive lexical items occur instead of being conducted in general. This context-

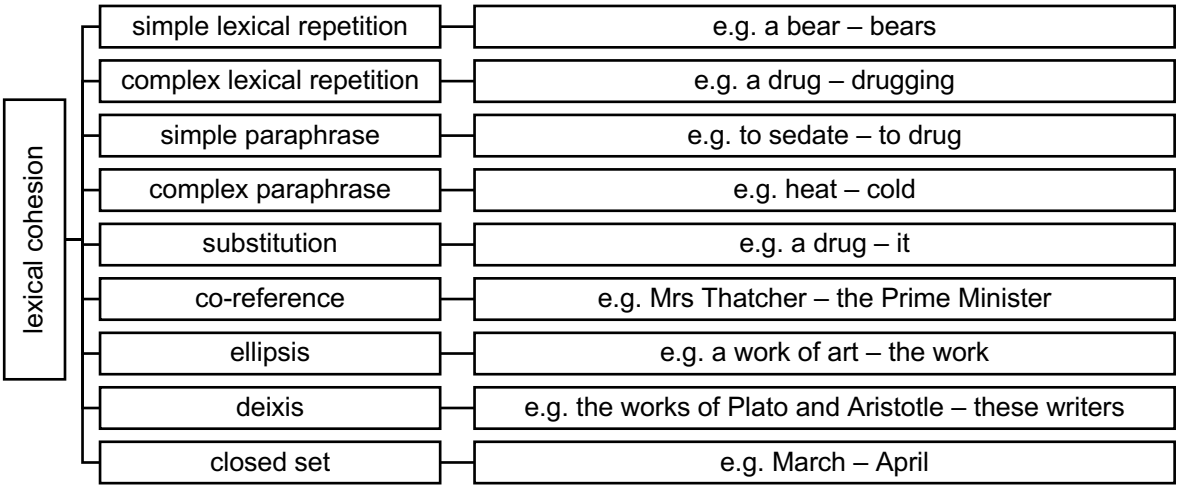
oriented view proposed by Halliday was also adopted in this thesis regarding the lexical cohesive analysis of the current corpora. Specifically, lexical cohesive relations, especially collocational relations, are identified based on the specific contexts in which related lexical items occur.

Another implication of Halliday's model for this study is that the context-sensitive nature of collocation in his developed model indicates that the corpus for lexical cohesion analysis used in the current study should comprise samples from the same discipline and as far as possible sharing similar contexts. As a consequence of Halliday's insights, the two corpora of texts used in this thesis were constructed from texts arising from similar tasks and the same broad disciplinary context.

2.2.4 Hoey: The prominence of lexical cohesion

Compared with previous models mentioned above, Hoey's classification of lexical cohesion is slightly different, as is shown through the classification and examples in figure 3 below:

Figure 3 Classification of lexical cohesion in Hoey's study (1991; 1994)



The similarity between Hoey's model and Hasan's model is that both of them do not include the collocation category. However, Hoey (1994) has mentioned some relations under the category of 'complex paraphrase' and 'closed set', which were found in the collocation category in Halliday and Hasan's model. Specifically,

‘complex paraphrase’ is similar to non-systematic collocational relations and ‘closed set’ is related to ordered set in Halliday and Hasan’s model respectively.

The new point in Hoey’s model is the inclusion of some grammatical cohesive relations (e.g. personal and demonstrative pronouns), as his focus is on repetition of previous items which contributes to the creation of texture in texts. However, these grammatical categories are treated less prominently than the actual lexical categories in his model, because “it is the lexical links that dominate the cohesive organisation” in non-narrative texts (Hoey, 1991, p. 74). This view is also shared by other researchers (e.g. Tanskanen, 2006) and is accepted in the present study which focuses on lexical cohesion rather than grammatical cohesion.

Another interesting point in Hoey’s model is his explanation of ‘complex paraphrase’ relations:

> ...imagine that we have three words in a text, **hot**, **cold** and **heat**: then if **hot** and **heat** form a complex repetition link, and **hot** and **cold** form an antonymous paraphrase link, then **cold** and **heat** will form a complex paraphrase.
>
> (Hoey, 1991, p. 65)

In this case, *hot* is the ‘mediator’ of the relation between *cold* and *heat*. What if the mediator is not in the text? Will the cohesive relation between *cold* and *heat* still be recognised? The solution to this question is to bring the missing mediator back to the text, as Hoey mentioned in another case which includes another two lexical items *instruction* and *teacher*:

> ...there is a missing item, **teaching**, that can substitute exactly for **instruction** in this context and which, of course, would be in a repetition link with **teacher** [...] This allows us to treat the relationship between **instruction** and **teacher** as one of complex paraphrase.
>
> (Hoey 1991, p. 67)

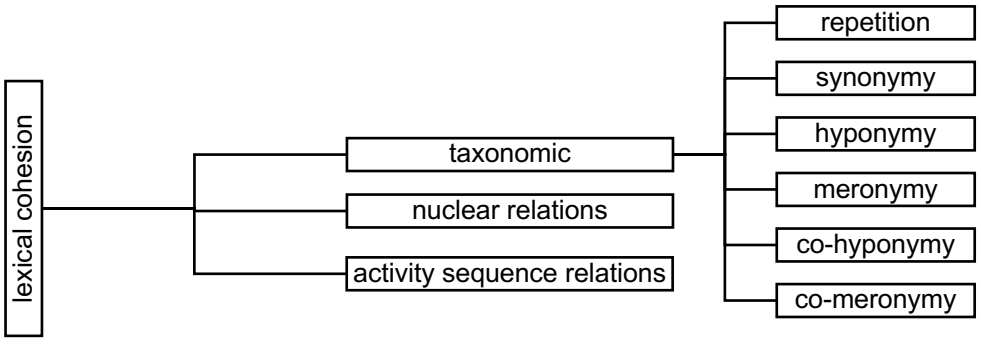
This substitution method is useful in recognising complex paraphrase relations (or collocational relations in Halliday’s model), though this category is described

as “a can of lexical worms” by Hoey (1991, p. 64). However, as trying to identify all lexical cohesive relations in a text is time-consuming, especially for collocational relations which are even harder to trace, the above approach at least provides a possible solution for the potential problems which are likely to be encountered in the present study.

2.2.5 Martin: Redefinition of Halliday and Hasan’s model

Martin’s (1992) framework of lexical cohesion redefines the categories in Halliday and Hasan’s model. There are three main categories in his framework: taxonomic, nuclear and activity sequence relations. The detailed framework is shown in figure 4 below:

Figure 4 Classification of lexical cohesion in Martin’s study (1992)



Taxonomic relations were already introduced in previous models (e.g. Hasan, 1984; Halliday, 1985) and have been discussed in the current chapter, so will not be discussed further here.

The creative and original ideas in Martin’s model centre round discussion about nuclear and activity sequence relations which redefine the collocation category in Halliday and Hasan’s model (1976) and Halliday’s model (1985). Nuclear relations are defined as relations reflecting the ways in which “actions, people, places, things and qualities configure as activities” (Martin, 1992, p. 309). For example, *serve* – *ace* in “Ben **serves** an **ace”** or in “Ben **serves**... that’s his fifth **ace** of the match” (Martin, 1992, p. 309). In the present study, the latter example

is accepted as lexical cohesion, as the relation between *serves* and *ace* happens beyond the clause. This concept of nuclear relations has inspired some subsequent studies (e.g. Tanskanen 2006; Hoffman 2012) and informs the present study. Based on this nuclear category, this thesis includes a sub-category of collocation called "activity-related collocation" (Tanskanen, 2006, p. 49) in the present framework of lexical cohesion (see section 4.4.3.2, chapter 4).

The second category of Martin's classification, activity sequence relations, refers to the way in which "the nuclear configurations are recurrently sequenced in a given field" (Martin, 1992, p. 321). For example, in the field of tennis, the players need to try various ways to get points to win a game. A typical sequence of activities regarding gaining points is as follows:

> player + serves
> opponent + returns
> player + volleys
> opponent + lobs
> player + smashes
> opponent + retrieves
> player + smashes
> opponent + misses
>
> (Martin, 1992, p. 321)

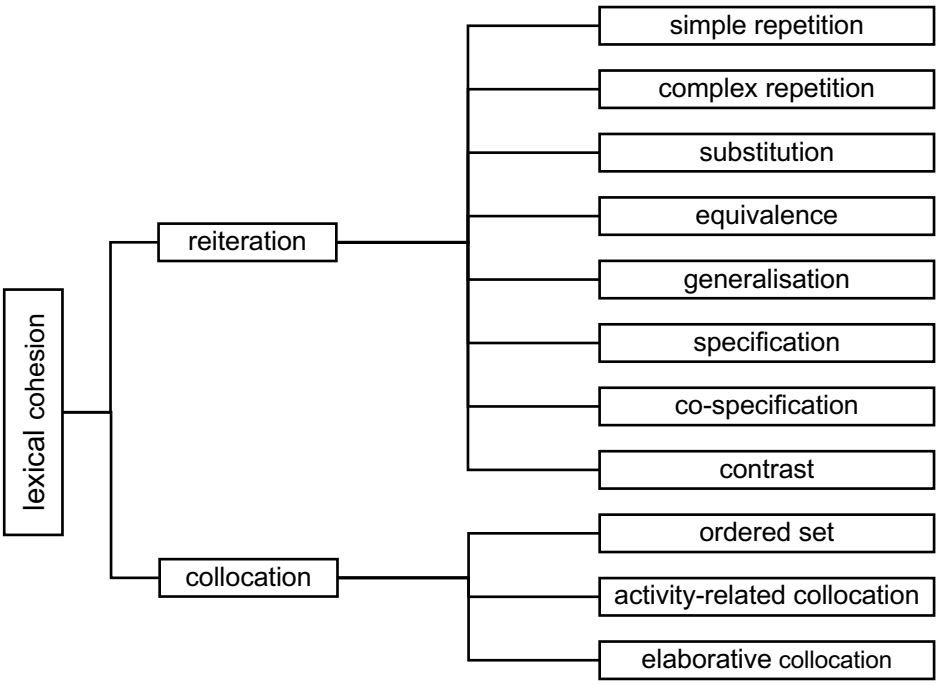
As can been seen from this example, "the activity sequences […] are themselves organised into composition [i.e. hyponymy] taxonomies" (Martin, 1992, p. 321), such as *player – opponent* as an opposite pair or co-hyponymic pair of 'game' as the superordinate. In this aspect, Martin's categories, taxonomic and activity sequence relations, sometimes overlap, which increases the challenges for the analysis of lexical cohesion. Additionally, as Tanskanen (2006, p. 46) has pointed out, the category of activity sequence relations makes it possible to identify almost every element in a text as cohesively connected, which also makes the analysis highly complicated. As a consequence of these two factors, this category of activity relations is not used in the present study.

2.2.6 Tanskanen: Activity-related and elaborative collocation

A discourse-specific perspective is adopted in Tanskanen's (2006) study of lexical cohesion, which leads to the use of different terminologies in her model compared with other previous models. The reason for adopting new terminologies in her model of lexical cohesion is that the discourse specificity of lexical relations is the focus of her study which adopts discourse-specific terminologies and regards discourse as the starting point of the analysis of lexical cohesion. The present study also entirely follows this discourse-specific approach, but for the sake of understanding, chooses to use traditional lexical semantic terminologies, rather than the discourse-specific terminologies in Tanskanen's model. The reason for this choice is that traditional terminologies, such as hyponymy, synonymy and meronymy, have been already widely accepted in related research areas, which helps readers understand the meaning of each lexical cohesive category without further explanation. In contrast, more elaborations on the new discourse-specific terminologies used in Tanskanen's model would be required to serve for readers' comprehension if such terminologies had been adopted in the present study.

Figure 5 below illustrates the classification of lexical cohesion in Tanskanen's study. A grammatical cohesive category, substitution, was included in her model as one sub-category of reiteration, as with Hoey's model. The reason for this inclusion is that in addition to lexical items, substitution "can [also] function in a very similar manner in repeating something" (Tanskanen, 2006, p. 49). However, for the present study, the definition of lexical cohesion requires that the examined items should be lexis in order to differentiate lexical cohesion from grammatical cohesion. Therefore, substitution is not included in the present model. As for the introduction of what Tanskanen refers to as discourse-specific terminology in her model, 'equivalence' replaces the corresponding lexical semantic term 'synonymy'; 'generalisation' substitutes 'superordination'; 'specification' is the new term for 'hyponymy'; and 'co-specification' replaces 'co-hyponymy'.

Figure 5 Classification of lexical cohesion in Tanskanen's study (2006, p. 49)



Tanskanen's biggest contribution to the present research is her explanation of two sub-categories of collocation, activity-related collocation and elaborative collocation, which are used to provide the system of classification adopted in the present study. Specifically, the sub-category of activity-related collocation is based on Martin's nuclear relations and the elaborative collocation is similar to Halliday's (1995) definition of collocation. Furthermore, Tanskanen's concept of 'trigger' and 'frame'[6] (developed from Jordan, 1984; Fillmore, 1985; Fillmore & Baker, 2001, p. 3) are used in the present study to help identify these two types of collocational relations in the thesis corpora (see chapter 4 for further elucidation).

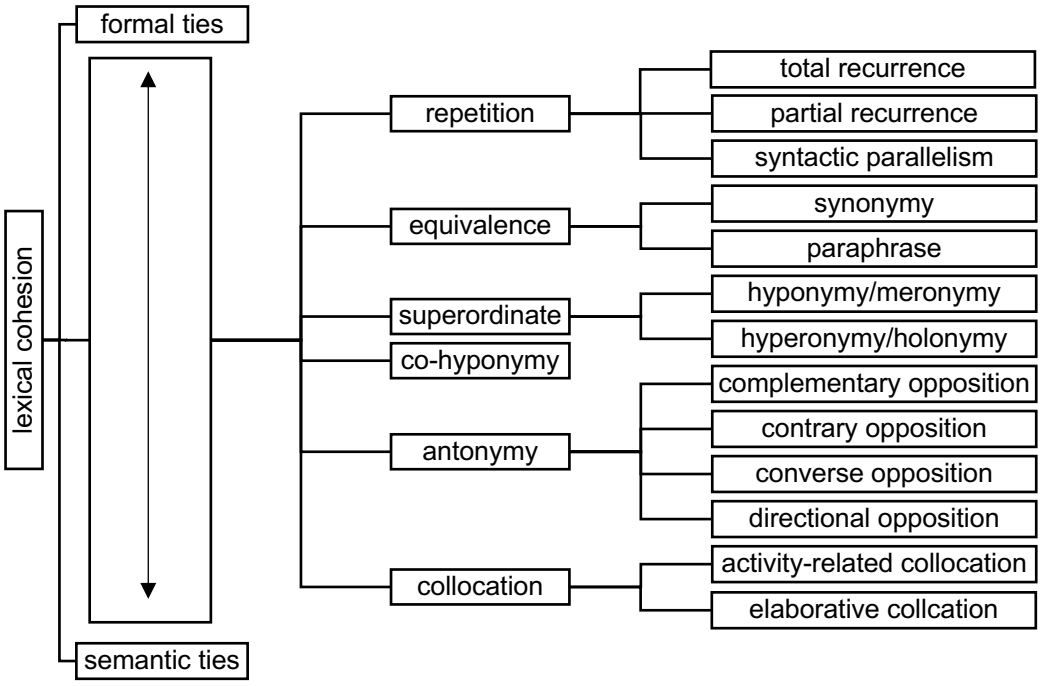
2.2.7 Hoffmann: Scale from formal to semantic relations

Figure 6 below demonstrates the classification of lexical cohesion in Hoffman's (2012) study which emphasises that lexical cohesion is not entirely a category of semantic relations as there are formal resemblances in relations included in the

[6] Frame is defined as "knowledge structures evoked by lexical items" (Tanskanen, 2006, p. 63); trigger is defined as a "repetition of the previous topic or a part of it, employed to make clear the association between an item and its re-entry" (Jordan, 1984, cited in Tanskanen, 2006, p. 38)

repetition category. This also indicates that lexical items themselves do not need identity of reference to create cohesion (Halliday & Hasan, 1976, p. 284).

Figure 6 Classification of lexical cohesion in Hoffman's study (2012, p. 94)



The valuable point of Hoffman's model for this thesis is its discarding of ordered sets in the collocation category while maintaining the sub-categories of activity-related collocation and elaborative collocation created by Tanskanen (2006). This adaptation informs the classification of collocation in the present study because collocation itself is already problematic and such simplification of this category facilitates the analysis of lexical cohesion. Another reason for this simplification adopted in the present study is that other sub-categories (i.e. ordered set, meronymy and antonymy) which have been included in the collocation category are similar to synonymy in nature regarding the systematic semantic relations involved in these sub-categories, and have already been allocated to the reiteration category by researchers (e.g. Halliday, 1985).

2.3 Categories of lexical cohesion in the present study

In this section, the classification of lexical cohesion in the present study will be discussed with examples provided of how this classification is applied in the

current research. To inform this discussion, a comparison between the models introduced above and the model developed in this thesis is shown in table 1 below. The comparison in this table shows that similarities outweigh differences between models. The differences seem reflective of the fact that each model is designed for specific contexts. In contrast, the similarities mainly lie in the fact that there are some elements included in all models, i.e. repetition, synonymy, superordinate and antonymy. The reason for their common occurrence in these models is that lexical cohesive relations involved in these categories are related to general semantic relations which are straightforward and easily identified in texts compared with other more complicated relations, such as collocational relations which are not connected with general semantic relations directly. Therefore, the collocation category or its corresponding category with different terms (e.g. 'complex paraphrase' in Hoey's model) is not included in all models, which is perhaps indicative of its complexity in the real analysis. However, because of its cohesive effect, collocation is of necessity still included in the present study in order to provide a complete picture of lexical cohesion.

Table 1 Overview of categories of lexical cohesion from previous studies and in the present study

Halliday & Hasan (1976)	Hasan (1984)	Halliday (1985)	Hoey (1991; 1994)	Martin (1992)	Tanskanen (2006)	Hoffmann (2012)	The present study
reiteration: same word (repetition)	general: repetition	repetition	repetition: simple /complex repetition	taxonomic: repetition	reiteration: simple/complex repetition	repetition: total recurrence /partial recurrence	reiteration: total repetition /partial repetition
			repetition: substitution		reiteration: substitution		
reiteration: synonymy/near synonymy	general: synonymy	synonymy: synonymy "proper"	repetition: simple paraphrase	taxonomic: synonymy	reiteration: equivalence	equivalence: synonymy	reiteration: synonymy
						equivalence: paraphrase	
reiteration: superordinate		synonymy: superordinate	repetition: superordinate	taxonomic: hyponymy	reiteration: generalisation	superordinate: hyperonymy	reiteration: hyperonymy
						superordinate: holonymy	
reiteration: general word							reiteration: signalling nouns
	general: hyponymy	synonymy: hyponymy	repetition: hyponymy		reiteration: specification	superordinate: hyponymy	reiteration: hyponymy

collocation: meronymy	general: meronymy	synonymy: meronymy		taxonomic: meronymy		superordinate: meronymy	reiteration: meronymy
collocation: co-hyponymy		synonymy: co-hyponymy/ co-meronymy		taxonomic: co-hyponymy /co-meronymy	reiteration: co-specification	co-hyponymy	reiteration: co-hyponymy/ co-meronymy
collocation: opposites	general: antonymy	synonymy: antonymy	repetition: complex repetition or paraphrase	taxonomic: contrast	reiteration: contrast	antonymy	reiteration: antonymy
	instantial: equivalence		co-reference				reiteration: other relations with identity of reference
	instantial: naming						
	instantial: semblance						
collocation: ordered/unordered set			repetition: closed set[7]		collocation: ordered set		
collocation: non-systematic semantic relations		collocation: associative relations	repetition: complex paraphrase	nuclear: extending and enhancing	collocation: activity-related collocation	collocation: activity-related collocation	collocation: activity-related collocation

[7] 'Closed set' is introduced by Hoey in 1994

					collocation: elaborative collocation	collocation: elaborative collocation	collocation elaborative collocation

2.3.1 Repetition

This category of lexical cohesion was used in the current study, which is the most straightforward relation in lexical cohesion in most previous research. The repetition category was separated into two types – simple and complex repetition in Tanskanen's study (2006). Simple repetition refers to the identical form or form with grammatical change of a lexical item. For example, singular – plural, or present tense – past tense. Complex repetition is defined as forms with derivational change or word class change. For example, in example 2-8 below, the repetition of *student* is an identical repetition as the grammatical form of *student* is the same; and the repetition of *grade* is an example of complex repetition, because the first lexical item *grade* is used as a verb while the second *grade* is a noun. However, for the sake of convenience and succinctness in the sample analysis, simple repetition in the present study is defined as the identical form of a lexical item, while other repetitions are classified as complex repetition relations. This categorisation of the repetition category is in line with Wu's (2010, p. 99) framework of repetition in her study of Chinese students' oral discourses.

2-8 Rosie, one option for dealing with _any_ conflict of interest with a **student** in your class is to ask a colleague who is familiar enough with the subject and your expectations to **grade** the **student**, or at least review with you the **grade** you give.
(Tanskanen, 2006, p. 50)

In example 2-9 below, the repetition relations between lexical items are a longer distance apart in the same text, with one complex repetition pair *cultural determinism – cultural determinist* and one simple and close repetition of *Freud*.

2-9 And **cultural determinism** is the idea that the way people think and act is largely determined by their culture, their upbringing, their socialization, their home environment, peer group pressure, this kind of thing, and is not to be looked for in natural causes, in their genes, for example, or in individual psychological experience, as was the prime focus of **Freud** in psychoanalysis. So the result is that, where people did take notice of **Freud**, and here Talker Parsons is the prime example, they interpreted Freud as if he too were a **cultural determinist**.
(Tanskanen, 2006, p. 51)

Another type of complex repetition is partial repetition. The following example illustrates this relation:

2-10 PVC insulation can be stripped with **V-Type blades** which can be adjusted to fit the conductor size. They can be used to strip wire sizes between approximately #12 to #32 AWG. The travel of **the blades** is controlled by micrometer nuts.
(Jordan, 1981, p. 54)

The second mention of *V-Type blades* is the short version *the blades.* Another example is *a work of art – the work* (Hoey, 1991), which illustrates a tendency of using the determiner *the* before the central word of the lexical item which has occurred already in the same text (e.g. *blades* in *V-Type blades*) to form a partial repetitive item[8].

What is worthy of pointing out here is that there is no necessity for the lexical items to share the same referent. Otherwise, verbs or adjectives would not be included in such lexical cohesive relations. On the other hand, it is necessary to avoid 'chance repetitions', i.e. homonymic repetitions whose meanings are completely different based on specific contexts (Hoey 1991, pp. 56–57), which is demonstrated in example 2-11. In this example, the speakers first mention the lexical item *bodies* in the context of giving papers in conferences. Then, they change to another topic about disabled people, in which *bodied* (physical structure of people) is used for another meaning from the first use of *bodies* (group of people). Therefore, the lexical item *bodies* and *bodied* cannot be regarded as simple repetitions as they have different meanings in this context.

2-11 C: I think I'd quite enjoy giving papers to **bodies** actually.
A little later, the speakers start discussing handicapped or disabled people, and the following exchange takes place:

[8] Although Hoey (1991) names this type of relation as 'ellipsis', to avoid confusions between the grammatical cohesive category 'ellipsis' in Halliday and Hasan's (1976) model and the 'ellipsis' category in Hoey's sense, this study will follow Jordan's (1982, p. 6) term to call this type of repetition relation 'partial repetition'.

A: ... we've got a new pedestrian crossing in Raynes Park and it's one of the kind where you press a button and it operates the lights and they've got a buzzer..

C: m ah yeah

A: .. which I think is a very good idea for people who can't see or can't tell the difference between green and red

C: oh God perfectly able-**bodied** but colour-blind person

(Tanskanen, 2006, p. 70)

2.3.2 Synonymy

The synonymy category was also included in the present study, as with all of the previous studies mentioned above. However, the names used for this category have involved some disagreements. Both Tanskanen (2006) and Hoffman (2012) have used 'equivalence' to replace 'synonymy' in their models for similar relations. 'Equivalence' is a term in the taxonomic terminology of McCarthy's (1988) study of conversations. McCarthy has proposed that "general semantic relations, such as synonymy or antonymy, may not be the best choices for describing lexical relations between items in use" (Tanskanen, 2006, p. 40), and therefore developed discourse-specific terms (e.g. equivalence) for the lexical cohesive relations involved in his study. In line with McCarthy, as mentioned above, the reason for this substitution regarding terminologies for lexical cohesion is to make the discourse-specific approach adopted in Tanskanen's and Hoffman's studies more explicit by using non-lexical semantic terminologies.

Basically, both of the terms, equivalence and synonymy, refer to the same category which denotes relations between two lexical items whose meanings are somehow synonymous. However, in a more discourse-specific approach to lexical cohesive relations, the equivalence of a lexical item may not be "semantically absolutely synonymous" with that item (Tanskanen, 2006, p. 55). According to Tanskanen (2006, p. 54), the synonymous relations between lexical items can only be justified and explained within the context of a specific text in which the lexical items are involved. Tanskanen (2006) correctly, in the view of this thesis, insisted on the conviction that "meaning is made in context" (p. 55). Hoffman (2012) further defined 'equivalence' as "[the relation which] involves two

lexical items [that] bear similar conceptual meanings in a given context of use" (p. 88). Furthermore, Halliday and Hasan (1976) discussed in detail the idea that the meaning of lexical items depends on the context:

> . . .each occurrence of a lexical item carries with it its own textual history, a particular collocational[9] environment that has been built up in the course of the creation of the text and that will provide the context within which the item will be incarnated on this particular occasion. This environment determines the 'instantial meaning', or text meaning, of the item, a meaning which is unique to each specific instance.
>
> (Halliday & Hasan, 1976, p. 289)

The examples below elucidate the difference between equivalence from a discourse-specific perspective and synonymy from a lexical semantic approach.

2-12 I spent a good hour talking to him about anti-Semitism and genocide, and the things that distinguished the Nazi **extermination** of the Jews from other forms of oppression in the world. I also told him that it was an issue that affected me deeply, that my extended family had lost many people to the Nazi **slaughter**.

2-13 . . .**Carbon dioxide** is the most soluble of the gases because as it dissolves it doesn't just go through a physical solution it goes through a chemical conversion [2 sentences omitted]. . .If you put more **C O two** into the system the concentrations of all these go up. . .

2-14 We are **pausing** on the road for no other reason than that we have been bounding ahead so rapidly and could all do with a **breather**.

(examples 2-12, 2-13 and 2-14 are from Tanskanen, 2006, p. 56)

2-15 Everyone **cheered**. The leader acknowledged the **applause**.

(Halliday & Matthiessen, 2014, p. 646)

[9] In the original Firthian sense (see chapter 4 for further explanation of Firth's definition of collocation)

Example 2-12 includes two synonyms *extermination* and *slaughter* which are synonymous in this particular context referring to what the Nazis did to the Jews. Example 2-13 involves a more straightforward pair of equivalence: *carbon dioxide* – *C O two*. Examples 2-14 and 2-15 illustrate another possibility of forming equivalent relations from the discourse-specific perspective respectively, i.e. two items from different word classes can also form equivalent relations: *pausing* (verb) – *breather* (noun), *cheered* (verb) – *applause* (noun).

It is clear from these examples that equivalence here refers to meaning in use, not abstract meaning without context. As McCarthy (1988) proposed, more discourse-specific terms should be introduced to describe lexical relations between items in use, in which the relations between items depend on the text in which they occur, instead of being an example of relations regarding abstract meanings (Tanskanen, 2006, p. 40). Tanskanen (2006) believed that choosing 'equivalence' as the term for the category in question is a small gesture but an important signal of the attitude "that the relations between the items should […] be clear even from the terms utilised [in her study]" (p. 57). However, in the present study, although the definition of this category is in line with Tanskanen's sense, to avoid misunderstanding of the meanings of lexical cohesive categories investigated here, the terms for these categories still follow the lexical semantic tradition, i.e. the categories are still named as synonymy, hyperonymy, hyponymy, meronymy and antonymy respectively.

2.3.3 Hyperonymy

The third category involved in the current classification of lexical cohesion is hyperonymy. A hyperonymic relation is defined as the "relation which holds between a more general, or superordinate, lexeme and a more specific, or subordinate, lexeme" (Hoffmann, 2012, p. 90). There are two types of hyperonymic relations in the present study: kind-whole and part-whole relations. The first type refers to the relations that the first item is a kind of the second item. The examples below illustrate the first type of relations:

2-16 Over the past decade or more, Western governments have taken action, individually and collectively, both to reduce dependence on **imported oil** and to

provide for an emergency should it arise. In particular, they have made considerable progress, some of it quite recent, in freeing internal markets for **energy products**.

2-17 Gordon: If **Labour** get in and they can't fulfil their promises...
Audrey: Well I can't well I mean there's an awful lot, I mean would, no no matter which **political party** it is, they all make promises, but they don't carry them all out.

2-18 It will be observed that, as is often the case, the most informal or 'slang' words are regionally restricted, being in this case unknown or unusual in **North American English**. It will also be observed that there are no strict cooccurrence restrictions here as there are in some **languages** – one can say 'long journey' and 'lengthy trip' just as well as 'lengthy journey' and 'long trip'.
(Tanskanen, 2006, p. 57)

In example 2-16, *energy products* refers to a more general category and *imported oil* is a type of *energy products*. Therefore, there is a hyperonymic relation between *energy products* and *imported oil*, or, *energy products* reiterates *imported oil* in a more general way. This is also the case for the cohesive pair *political party* and *Labour* in example 2-17, in which the former is the superordinate of the latter. In example 2-18, the first appearing item *North American English* is a kind of the later item *languages.* Therefore, *languages* is the hyperonym of *North American English*.

The following example illustrates the part-whole relations, i.e. the first item is a part of the second item:

2-19 Yeah, the **department** is still the way it was ,maybe worse, but the **company** I work in and the people I work for and with are the best.
(Hoffmann, 2012, p. 91)

In example 2-19, *department* is part of the whole *company*. That is to say, in this case, the holistic concept, *company*, refers to its constitutive part, *department*. This relation is called 'holonymy' by Hoffman (2012, p. 90). However, the relation

between the two items is still specific-general. Therefore, the present study still places this type of relation under the category of hyperonymy to simplify the classification of lexical cohesion.

In his later work on lexical cohesion, Halliday (Halliday & Matthiessen, 2014, pp. 645-646) still allocated superordinates to the category of synonymy, which is defined as synonyms of some higher level of generality. An example from his work is shown as follows:

2-20 Four-&-twenty **blackbirds**, baked in a pie.
When the pie was opened, the **birds** began to sing.

(Halliday & Matthiessen, 2014, p. 645)

Birds is at a higher level than *blackbirds* regarding generality. Therefore, *birds* is the superordinate of *blackbirds*.

Halliday also provided an example to show the change of generality of lexical items used in texts:

2-21 the **blackbirds** began to sing
the **birds** began to sing
the **creatures** began to sing
they began to sing

(Halliday & Matthiessen, 2014, p. 645)

This example shows a continuum of lexical items from specific to general, in which *blackbirds* is the most specific item while the reference item *they* is the most general of all. This is the continuum from subordination, superordination, general noun to reference regarding generality, which demonstrates the connection between lexical cohesion and reference items (e.g. *they*). This continuum also informs the classification of hyperonymy and general nouns (one sub-category of the signalling noun category, see chapter 3) in the present study that the hyperonymy category is one level below the category of general nouns regarding generality.

A further example is given below in order to show the interaction between lexical cohesive devices and reference items, which is used to track the participant's voice in texts.

2-22 As an added means of self-defense the **ankylosaur** had a club on its tail. **The creature** may have been able to swing the club with great force and aim a savage blow at an enemy.

(Halliday & Matthiessen, 2014, p. 645)

"The + a superordinate" is an example that demonstrates that the combination of lexical cohesive device and the reference item *the* can trace a participant through the text (e.g. *the creature*), and this structure also helps identify hyperonymic relations between lexical items in the current lexical cohesive analysis.

In particular, "*the* + general noun" is frequently used in the category of general nouns which refers to "a superordinate member of some taxonomic class of entity" (Halliday & Matthiessen, 2014, p. 647), such as *thing, stuff, creature, people, matter, move, place* and *idea* (Halliday & Hasan, 1976, pp. 274-275). These general nouns perform a strongly cohesive function when they refer to a more specific item or segments in texts. However, the general nouns themselves do not contain specific information. Their contextual meanings depend on their co-referential contents, and their co-referentiality relation is typically connected by the reference item *the* or other determiners, such as *this*. Examples 2-23 and 2-24 below demonstrate the use of general nouns.

2-23 Chen said he did not have the power to single-handedly determine the future of Taiwan and that there would have to be public consensus before Taiwan pressed ahead in trying to establish a confederation with the mainland. But he said it was an example of "new thinking that could bring a breakthrough." Chen added, "There's a lot of room for discussion of this **matter**."

2-24 Then somewhere in the middle of the desert – about six hundred miles later[,] I didn't see a connection with anything. He bangs on the side of the car and I let

him out. Now I know Indian people better, and I know that the **guy** probably didn't speak English, or if he did, he was ashamed of it.

(Halliday & Matthiessen, 2014, p. 647)

In example 2-23, the structure "determiner *this* + general noun *matter*" refers back to the previous underlined segment; in example 2-24, the noun phrase *the guy* replaces *He* and *Indian people*. Both *matter* and *guy* do not have specific meanings themselves but are given meanings in particular contexts.

2.3.4 Hyponymy

Another category included in the model of lexical cohesion in this thesis is hyponymy. In line with the hyperonymy category, this category also includes two types of relation: the second item is either a subclass or another class at the same level of classification of the first one (Halliday & Matthiessen, 2014, p. 646). These types of hyponymic relation are demonstrated as follows:

2-25 And do you know anything about medieval **literature**; have you ever heard of any other kinds of literature in the medieval period besides **Chaucer**?

2-26 Most limestone probably originates from **organisms** that remove calcium carbonate from sea water. The remains of these **animals** may accumulate to form the limestone directly, or they may be broken and redeposited.

2-27 Noah's wife and his sons' wives went to the fields to gather **fruit** and **grain** and **vegetables**.

(Halliday & Matthiessen, 2014, pp. 646-647)

In examples 2-25 and 2-26, the relations between the pairs *literature* – *Chaucer* and *organisms* – *animals* are superclass-subclass relations. Therefore, *Chaucer* and *animals* are hyponyms of *literature* and *organisms* respectively. In contrast, in example 2-27, *fruit, grain* and *vegetables* are co-hyponyms of a superclass of food.

As Tanskanen (2006) points out, it is not necessary for the general item (i.e. superclass) to appear in the text. In examples 2-28 and 2-29 below, the general

items *coin* and *world English* appear in the text respectively; while in example 2-30, the general item does not appear in the text: *RP speakers* and *Standard English speakers* are still identified as hyponyms without the mention of their general item *English speakers*.

2-28 The **farthing** has ceased to be a **coin** of the realm, the **halfpenny** is on its way...

2-29 C: no but the thing is if they use them you see and if you're describing **world English** one branch of it is **Indian English** because it's spoken by a very great many people A: yes
b: well I'm sure another branch is **South African English**

2-30 It is widely agreed, though, that while all **RP speakers** also speak Standard English, the reverse is not the case. Perhaps 9%–12% of the population of Britain […] speak Standard English with some form of regionial [regional] accent. It is true that in most cases **Standard English speakers** do not have 'broad' local accents (i.e. accents with large numbers of regional features which are phonologically and phonetically very distant from RP).

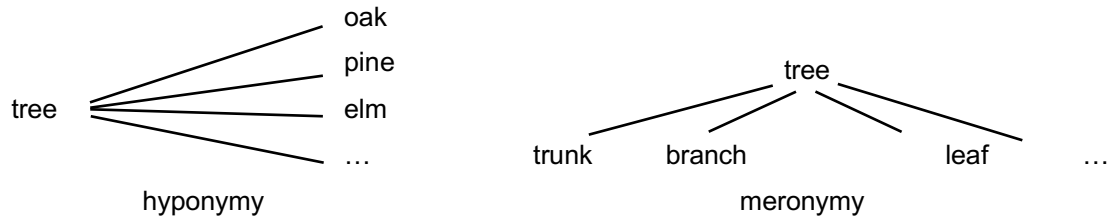
(Tanskanen, 2006, p. 59)

It is worth noting that in example 2-29, the appearance of *branch* is another signal of the hyponymic (type of an item) relation, acting between *Indian English* and *South African English*.

2.3.5 Meronymy

As with the other two categories which describe superordinate-subordinate relations mentioned above (i.e. hyperonymy and hyponymy), the meronymy category, representing another type of superordinate-subordinate relation, is also included in the present classification of lexical cohesion. According to Halliday and Matthessen (2014), hyponymy and meronomy are closely related as both of them have the function of extending a more general item. The comparison of these two lexical relations are demonstrated in figure 6:

Figure 6 Illustration of hyponymy and meronymy (Halliday & Matthienssen, 2014, p. 648)



In the hyponymic relations, *oak, pine* and *elm* are all 'kinds of' tree, while in the meronymic relations, *trunk, branch* and *leaf* are all 'parts of' tree. The occurrence of any pair of items within the category of tree are cohesive. Examples below instantiate the meronymic relations in detail:

2-31 On the left of the park lies the Exhibition **Centre** which covers a massive 25,000 square metres of column-free space under the one **roof**. Opened in January 1988, the Centre is designed to hold major international exhibitions. The glassed eastern **facade** is stepped back in five separate stages that can be partitioned off to form smaller **halls**. The fifth **hall** is linked by covered walkway to the Convention Centre.

2-32 Elfrida had a beautiful little glass scent-**bottle**. She had used up all the scent long ago; but she often used to take the little **stopper** out ...

2-33 She knelt down and looked along the passage into the loveliest **garden** you ever saw. How she longed to get out of that dark hall, and wander about among those beds of bright **flowers** and those cool **fountains**, ...

(Halliday & Matthiessen, 2014, p. 648)

In example 2-31, *roof, façade* and *halls* are meronyms of *centre*; in example 2-32, *stopper* is a meronym of *bottle*; and in example 2-33, *flowers* and *fountains* are meronyms of *garden*. It can be seen that the common feature of the three examples is to mention the general item first, and then extend to the parts. If the general item does not appear at all, and only the constitutive items are in the texts, the type of relation between the constitutive items is co-meronymic. This type of relation is still included in the meronymy category, as co-hyponymy is included in

hyponymy. For instance, *roof, façade* and *hall* are parts of *centre* in example 2-31 above. If *centre* was not in the text, then *roof, façade* and *hall* would form co-meronymic relations.

Halliday and Matthiessen (2014) further propose that there is no fixed boundary between meronymy and hyponymy, especially regarding abstract terms; and a set of items in question may be co-hyponyms of one superclass in one context, but co-meronyms in another. For example, "*chair, table* and *bed* are 'kinds' (hyponyms) of furniture, but 'parts' (meronyms) of furnishings; *forward, half-back* and *back* are 'kinds' of players but 'parts' of a team, and so on" (Halliday & Matthiessen, 2014, p. 648). This example also demonstrates the context-sensitive feature of lexical cohesion in terms of decisions made in the analysis of lexical cohesion in the thesis corpus, i.e. the decision as to which category specific lexical items belong to is dependent on the contexts in which the lexical items occur.

2.3.6 Antonymy

In common with the studies mentioned above, the antonymy category is also included in the current classification of lexical cohesion. What is worthy of mention is that this category has been designated with different category titles, however, 'antonymy' and 'contrast' are the most complementary. According to previous studies, this category has been included either into a sub-category of synonymy or collocation, or as a separate category of lexical cohesion. Since the present study chooses to simplify the classification of the collocation category, and the antonymic relations are relatively involved in the systematically semantic relation system, antonymy is treated as a sub-category of reiteration in this thesis. In addition to different titles used in this category, there are several types of antonymic relations discussed in previous studies, which are discussed in the following section in order to develop an operational definition and classification of the antonymy sub-category for the present study.

For example, in Halliday and Hasan's (1976) model, antonymy was defined as "pairs of opposites of various kinds" (p. 285), which was included in the collocation category and was divided into three types (see p. 14). Halliday and

Hasan's division of antonymy is based on Lyons' (1969) classification of opposite relations. In his later work, Lyons (1977) extended the scope of this classification from the structural semantics perspective by proposing that although "the standard technical term for oppositeness of meaning between lexemes is antonymy" (pp. 270-271), the term 'oppositeness' is also widely used by most authors.

Lyons (1977) also suggested that there are several kinds of lexical opposite. The first type are gradable opposites where "grading involves comparison" (Lyons, 1977, p. 271). When grading objects, it is appropriate to test "whether they have [the same certain] property to the same degree or not" (Lyons, 1977, p. 271). For example, when comparing the temperature of X and Y, whether it is correct to say that *X is as hot as Y* or *X is hotter than Y "depends on the gradability of 'hot'"* (Lyon, 1977, p. 271). By contrast, there are lexemes which cannot be graded. For example, 'female' is ungradable. It is not acceptable to say "*X is as female as Y* or *X is more female than Y"* (Lyons, 1977, p. 271). On the other hand, both of the two lexemes have their opposites: 'cold' and 'male' respectively.

There is an important logical difference between these two pairs, i.e. *cold – hot* and *female – male*, which is related to the distinction of contradictories and contraries, two terms for opposite relations proposed by logicians (Lyons, 1977). Ungradable pairs (e.g. *female – male*) denote contradictory relations, while gradable pairs (e.g. *cold – hot*) refer to contrary relations. It is important to note that this distinction is the consequence of gradability, not its cause. The logical conditions for these two types of opposite relation are demonstrated in the table below.

Table 2 Logical conditions of gradable/ungradable relations (based on Lyons, 1977, p. 272)

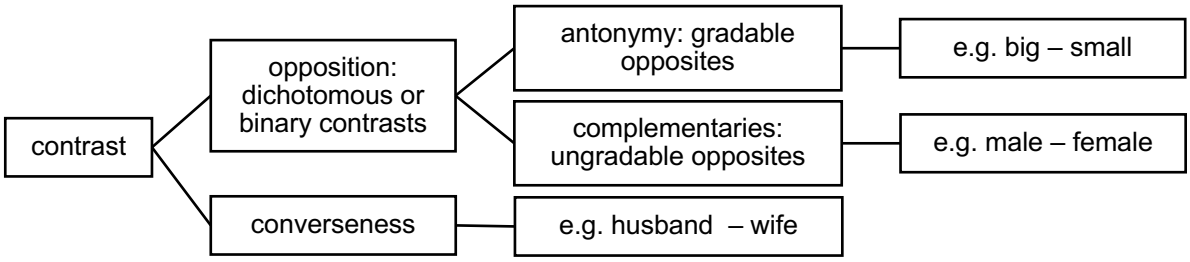
Ungradable/contradictory relation	*p* and *q* cannot both be true or both be false[10]
Gradable/contrary relation	*p* and *q* cannot both be true (though both may be false)

[10] *p* and *q* represent propositions.

As can be seen from table 2, propositions expressed by two lexical items in an ungradable/contradictory relation cannot be false at the same time, while in a logical condition of gradable/contrary relation, propositions denoted by two lexical items can be false simultaneously. This basic logical difference between these two kinds of opposite relations help identify and differentiate such relations in the lexical cohesive analysis in the present study.

In addition to the explanation of gradability of some types of opposite relation, Lyons (1977) proposed a classification which he believed to be workable and convenient. As with Halliday and Hasan, Lyon's classification of opposite relations have also inspired other studies related to lexical cohesion. The present study makes no exception. This classification is shown in figure 7 below.

Figure 7 Classification of opposite relations in Lyons' (1977) work



As figure 7 demonstrates, in Lyons' classification, 'contrast' is used as the cover term for different types of opposite relation. Dichotomous opposition refers to antonymy which includes gradable opposites, while binary contrasts represent ungradable complementaries. There is a third type called 'converseness' which expresses the converse relation between two lexical items. For example, as shown in this figure, *husband* and *wife* are two "reciprocal social roles" and the converse of the proposition expressed by the sentence "X is the husband of Y" is expressed by another sentence "X is the wife of Y" (Lyons, 1977, p. 280).

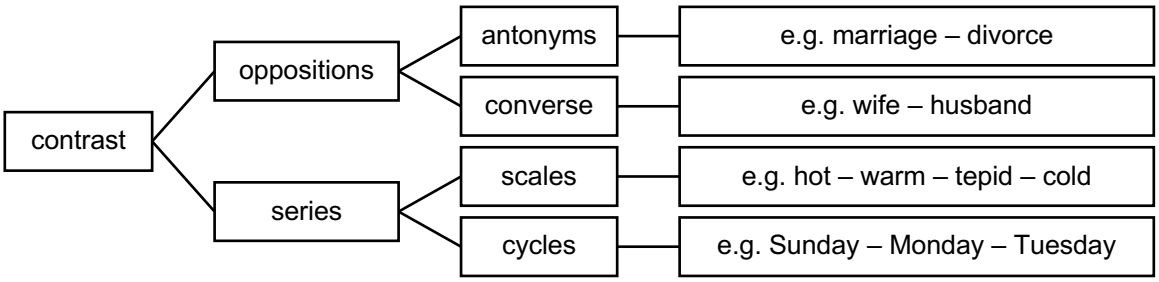
In terms of the classification of antonymy, compared with Halliday and Hasan's model, in Halliday and Matthiessen's (2014) framework of lexical cohesion, antonymy is included in the synonymy category, which is treated as a special case of synonymy, describing the cohesive relations between lexical items which are opposite in meaning, such as the pair *woke* – *asleep* in the following example.

> 2-34 He fell **asleep**. What **woke** him was a loud crash.
>
> (Halliday & Matthiessen, 2014, p. 646)

Asleep and *woke* are from different word classes but they still form an antonymic relation in example 2-34. Their relation provides an example of a sub-category of 'near-antonymy' in the antonymy category, which is also included in the present study, as with the sub-category, 'near-synonymy', in the synonymy category (see section 2.3.2).

Martin and Rose (2007) defined contrast relations as "elements [in such relations that] differ significantly in meaning" (p. 86) and in contrast with Lyons' structural semantics perspective, they divided contrast relations into two main kinds from the perspective of discourse semantics. Their classification of contrast relations are shown in figure 8.

Figure 8 Classification of contrast relations in Martin and Rose's (2007, pp. 86-87) study



As can be seen from this figure, in line with Lyons' framework of opposite relations, the cover term in Martin and Rose's model is also 'contrast'. The first type of contrast relation is called 'oppositions', which includes two sub-types – antonyms

and converses. Martin and Rose (2007) pointed out that 'antonymy' is the most familiar sub-type of contrast relation, and antonyms are always in pairs, such as *win – lose, married – single* and *quickly – slowly* (p. 86). As a sub-category of oppositions, antonymy is defined as the relation between two lexical items which have opposing meanings, like *marriage – divorce*. This definition is very close to Halliday and Matthiessen's (2014) definition of antonymy above.

Another sub-type of oppositions is converse which is almost the same as Lyons' definition of converse relations. However, Martin and Rose's classification suggests that converse relations in their study are not only involved with converse social roles, such as *victim – perpetrator and mother – son*; but also converse locations*, such as on top of – underneath* and *before – after* (Martin & Rose, 2007, p. 86), which inspires the present study regarding the identification of different converse relations.

The second type of contrast relation is series, which has two sub-types: scales and cycles. Scales are associated with the "outermost poles of meaning" (Martin & Rose, p. 80), such as

> 2-35 hot – warm – tepid – cold
> 2-36 pass – credit – distinction – high distinction
> 2-37 tutor – lecturer – senior lecturer – associate professor – professor
>
> (Martin & Rose, p. 80)

The first example of 'scales' is in line with Lyons' definition of gradable opposites. Examples 2-36 and 2-37 demonstrate co-hyponymic relations in the present study, as the lexical items in each example share the same superordinate, i.e. 'markings scales' and 'academic positions' respectively.

The second sub-type, cycles, includes relations in which lexical items are members of a set and put into orders. This sub-type is similar to 'ordered set', the sub-category of collocation in Tanskanen's (2006) model of lexical cohesion, which is included in the meronymy category in the present study. The examples of the sub-category of cycles from Martin and Rose's study are shown below:

2-38 Sunday - Monday - Tuesday - Wednesday - Thursday - Friday - Saturday

2-39 2000 - 2001 - 2002 - 2003...

(Martin & Rose, 2007, p. 80)

In common with other studies (e.g. Hoffman, 2012) and the present study, Tanskanen's (2006) model of lexical cohesion regards the contrast relations as a sub-category of the main category of reiteration, and defines it as "the relation between an item and another item which has an opposite meaning" (p. 59). Tanskanen summarised terms used in previous studies for the contrast relations: *antonymy, opposition, complex repetition* or *paraphrase*. Also, she pointed out that these relations do not have to be "strictly antonymous in the lexical semantic sense" (Tanskanen, 2006, p. 59), but the relations between two lexical items should contrast in a way within a specific context. This view is also shared by the present study in which the analysis of lexical cohesion identifies contrast relations in particular contexts in the samples of the thesis corpora. Two examples are used to explain what elements construct or enhance "the contrastiveness of the items" (Tanskanen, 2006, p. 59) within the context.

2-40 And the reason for this is that it belongs to a tradition, a fashion if you like, of writing which went dramatically **out of fashion** immediately after World War One. So, at the time when it was published most readers would have regarded it as completely **up to date** in its in its style and in its presentation.

2-41 Audrey: I mean where are they going, wherer are they going to get the extra money, form er to pay for **the old aged pensoners**' er eight pound rise?
Gordon: Mm.
Audrey: Or so they say, we'll get eight pound. Somebody's got to pay for it. So it'll be **the working people**...

(Tanskanen, 2006, pp. 59-60)

Example 2-40 is related to the enhancement function of the context, i.e. "the contrastiveness of *out of fashion* and *up to date* is enhanced by the use of, respectively, *dramatically* and *completely*" (Tanskanen, 2006, p. 59). Example 2-41 is a typical instance to address the constructive function of the context. The

opposite pair *old aged pensioners* and *working people* are related by contrast in this particular context. The *pensioners* get money which are from the *working people*. They have benefit conflicts in this context, which relate them in an opposing sense. Tanskanen explains that these two items can also be co-specifications in other contexts, for example representing various phases in people's lives. This evidences the fact that the contrast relation in example 2-41 is determined by the particular context which is given in this example, rather than being based in presupposed abstract semantic relations, and this shows the fluidity of cohesive relations. Similar to Tanskanen's study, this thesis is also interested in how lexical cohesive devices contribute to the process of making meaning in texts, rather than in abstract semantic relations between lexical items. As Tanskanen (2006) points out, "the fluidity is inevitable and necessary simply because each text brings with it a unique possibility of making meaning" (p. 60).

Hoffmann (2012) has summarised contrast relations in all previous studies mentioned above and proposed a classification of these contrast relations with four different categories. Therefore, the present study follows his comprehensive framework of contrast relations involved in the antonymy sub-category, which is shown in table 3 below with definitions of sub-types of antonymic relations.

Table 3 Classification of antonymic relations in Hoffman's study (2012, pp. 91-92)

category	definition	example
complementary antonymy	This category defines binary contrast relations between two lexical items which are not gradable.	dead – alive, inside – outside
contrary antonymy	This category refers to gradable contrastive relations between lexical items, and therefore, allows for comparative relations of the following kind: "A is smaller than B" or "B is larger than A".	small – big
converse antonymy	This category represents contrastive relations characterised by being "the result of a change of perspective" (Schubert, 2008, p. 49, cited in Hoffman, 2012, p. 91), which is usually expressed	buy – sell, come – go, lend – borrow, husband – wife, teacher – pupil

	through procedural verbs and nouns which express reciprocal social roles.	
directional antonymy	This category refers to lexical pairs which imply “a motion in one of two opposed directions with respect to a given place” (Lyons, 1977, p. 281).	up – down, north – south, forwards – backwards

2.3.7 Other relations with identity of reference

As mentioned in section 2.2.2, Hasan (1984) introduced the ‘instantial’ category. This category deals with relations which cannot be explained by the general semantic system, but are created by specific texts. This instantial category includes three types of relations, i.e. equivalence, naming and semblance. The three examples (see p. 17) which was used by Hasan to explain these types of instantial relations are within the clause boundary. However, this thesis focuses on lexical cohesive relations beyond the clause. Therefore, the relations expressed in the three examples provide good evidence for the categorisation of a new category named ‘other relations with identity of reference’ in the present study, while the examples per se are not suitable for the current lexical cohesion analysis.

The reasons for having ‘other relations with identity of reference’ (henceforth identity) as the category name are twofold. Firstly, relations in this category are co-referential, i.e. lexical items in a cohesive pair refer to the same subject. In Halliday and Matthiessen's study (2014), some lexical items which form cohesive relations in the repetition and synonymy categories have been described as having “identity of reference” (p. 645) which is the source of the name for this new category in the present study. Secondly, regarding lexical cohesive relations in texts, some relations in which two lexical items share the same referent and therefore create cohesive effects, have not been identified explicitly in previous studies. Some of the relations involved in this identity category have been illustrated in Hasan’s (1984) instantial category above, but her examples are within the clause. This thesis extends her analysis to the beyond-clause level and investigates further types of relations which can be included in this identity category.

In addition to Hasan's study, there are also other studies which have mentioned some examples that can be included in this identity category. For example, Hoey (1991) has one category called 'co-reference' (e.g. *Mrs Thatcher – the Prime Minister*); and Hoffman (2012) has mentioned one category called 'paraphrase' in synonymy. His examples are as follows:

> 2-42 Fiachna called, after spending most of the day (2pm–7:30pm) at the local hospital, helping out one of **the little Irish girls** – **a friend of his daughter**.
>
> 2-43 **The owners**, **Claude and Zaina**, have not updated its classic 19th century bistro style interior, …
>
> (Hoffmann, 2012, p. 89)

In example 2-42, *the little Irish girls* and *a friend of his daughter* refer to the same person; and in example 2-43, *the owners* and *Claude and Zaina* also represent the same group of people. Therefore, these lexical items form strong cohesive relations by their co-referentiality in specific contexts. This indicates that although the identity category does not belong to the general semantic system, the cohesive relations included in this category are still worthy of investigation within the classification of lexical cohesion.

Another reason for including 'identity' in the name of this category is that 'identity' itself has another semantic meaning which is "the quality or condition of being the same in substance, composition, nature, properties, or in particular qualities under consideration" (OED Online, 2019i). On the one hand, in some cohesive relations of the identity category, two lexical items express different qualities of the same referent. In Example 2-43, *the owners* and *Claude and Zaina* refer to the same referent group. On the other hand, the two items refer to different aspects of this group of referents: *the owners* reveals the social identity of the referents while *Claude and Zaina* provides the names of the referents. Thus, being included in the name of this category, 'identity' has been used in a double sense: firstly, 'identity' denotes that the lexical items in one pair share the same

referent; secondly, the lexical items can demonstrate different identities of the same referent.

2.3.8 Signalling nouns and collocation

Signalling nouns and collocation are the final two categories included in the current classification of lexical cohesion. As the category of signalling nouns includes several types of cohesive relations and is seen within the context of a broad complex signalling-noun-like phenomenon for investigation, in-depth discussion of this category will be presented in chapter 3. Likewise, the complex collocation category will also be described at length in chapter 4, based on Tanskanen's (2006) classification of collocation, i.e. activity-related collocation and elaborative collocation.

2.4 Summary and discussion

This chapter has examined models of lexical cohesion in previous studies with comparisons made between these models. Based on the discussion of these models, this chapter has developed a framework of lexical cohesion for the present study, which includes all the categories in Halliday and Hasan's (1976) original model with modifications, and explained definitions of these categories and types of lexical cohesive relations involved in each category, except the categories of signalling nouns and collocation which will be introduced separately in chapters 3 and 4 due to their complexity.

Since Halliday and Hasan's (1976) model is the foundation for the development of all the subsequent models and the current model of lexical cohesion in this thesis, all the categories in their model were introduced at first to set the broad context for the discussion of the following models. Other related studies were explored with the perspective of addressing modifications which were made in these studies based on Halliday and Hasan's original model, as well as their contributions to this thesis regarding the development of classification of lexical cohesion and the principles for the lexical cohesive analysis in the present study. In particular, Halliday's (1985) insights of the context-sensitive nature of collocation provides the guidance for the discipline-based selection of samples which formed the current corpora. Hoey's (1991) emphasis of the importance of

lexical cohesion in the creation of texture justifies the significance of this study which focuses on lexical cohesion. Additionally, his substitution method is of significant help with the identification of collocational relations in the present study. Martin's (1992) development of new categories of lexical cohesion, i.e. nuclear and activity sequence relations, inspires the operational classification of the collocation category in this thesis. Tanskanen's (2006) framework of collocation further sets the foundation of collocation investigated in the present study.

As for the examples included in this chapter, they were used to help the explanation of abstract definitions of lexical cohesive categories and related lexical cohesive relations included in the current study. Furthermore, these examples indicate the idea which has been mentioned in previous studies (e.g. Halliday, 1985; Tanskanen, 2006) that lexical cohesion is context-sensitive in general, and the identification of lexical cohesive relations are dependent on specific contexts. This discourse-specific perspective provides the key principle for the current corpus-based analysis of lexical cohesion.

Halliday and Hasan's (1976) explanation for the relation between the co-referentiality shared by two lexical items in a cohesive pair and the identification of lexical cohesive relations provides the guidance for the identification of lexical cohesive pairs in the present analysis, i.e. lexical cohesive relations in the reiteration category exist between lexical items in two conditions: co-referentiality and repetition of lexical forms. Furthermore, although their development of collocation as a separate category of lexical cohesion may bring challenges to the research of lexical cohesion, this collocation category is still worthy of discussion as it provides another perspective to investigate the lexical cohesive forces generated by the co-occurrences of two lexical items. Halliday's (1985) discussion about collocation informs the present study in the aspect of investigating collocation as a cover term for only non-systematic semantic relations which has been treated as a sub-category of collocation in Halliday and Hasan's (1976) original model. However, due to the fact that collocation has a long history regarding its usage in other research areas as well, the exploration of 'cohesion collocation' still remains complex and underinvestigated, which

provides further motivation for this thesis to take in-depth research of this category from a broader scope.

A new category called ‘other relations with identity of reference’ was created by the present study mainly based on the studies conducted by Hasan (1984), Hoey (1991) and Hoffman (2012). With regard to detailed investigation of lexical cohesive relations involved in this identity category, the fact that related relations addressed in these three studies indicate that lexical cohesive relations dealing with this category in this thesis have already been partially mentioned in previous studies, provides additional inspiration for characterising these relations as components in the current identity category as well as investigating further related relations in the thesis corpora. Other types of lexical cohesive relation were reclassified into different categories and sub-categories in order to develop an operational framework of lexical cohesion for the present study. This reclassification was based on the difference between systematic semantic relations and non-systematic semantic relations. Therefore, in the current model of lexical cohesion, seven sub-categories were reallocated to the broader reiteration category due to their representing different types of systematic semantic relations in nature, while two sub-categories were developed in the main collocation category from the perspective of non-systematic semantic relations (see pp. 27-28). This developed model serves as the fundamental framework of lexical cohesion for the corpus-based analysis in this thesis.

Chapter 3 Signalling nouns and lexical cohesion

3.1 Introduction

As addressed in chapter 2, as an important lexical cohesive device within a complex signalling-noun-like phenomenon (henceforth 'SN-like phenomenon'), discussing the category of signalling nouns requires a chapter of its own. Therefore, the purpose of this chapter is to identify and critically evaluate previous research on the SN-like phenomenon, and position the present work with regard to up-to-date research of this phenomenon in the scope of lexical cohesion in English.

The reason for giving the broad label SN-like phenomenon in the present study is that there are several closely-related terms which all refer to different features of this complex phenomenon and provide perspectives through which it can be explored in depth. These terms, either overlapping with SN-like nouns or forming a sub-group of SN-like nouns, reflect researchers' various interests and perspectives in their studies. The discussion of these terms and terminology used underpins key elements of the framework of analysis and terminology used in the present study.

The differences between the definitions and descriptions of SN-like nouns are regarded as disagreements on the same SN-like phenomenon, about which there is no universal definitional consensus (Flowerdew & Forest, 2015, p. 9). Therefore, SN-like phenomenon is an umbrella term which covers the range of related studies investigating this phenomenon from different perspectives. This chapter applies the results of these studies and subsequent analyses of these results to the present study, which focuses solely on lexical cohesion.

With regard to the term chosen to represent both the SN-like phenomenon and the subcategory of reiteration in the framework of lexical cohesion, the present study borrows the term 'signalling nouns' (henceforth 'SNs') from Flowerdew and Forest's (2015) study, as the 'signalling' function indicated by this term is seen as more closely related to the cohesive function of the group of nouns which are

investigated in this thesis than other features of these nouns indicated by other terms in previous studies. For clarity and convenience in the following discussion about the group of nouns in question, 'SNs' in Flowerdew and Forest's (2015) study will be referred to as 'discourse-based signalling nouns' (henceforth 'DBSNs') because DBSNs exist within the framework of a discourse-based view of language in their study, in which, specifically, the main approach is more oriented to the features of DBSNs at the discoursal level rather than at the sentential level in other studies (e.g. Huddleston & Pullum, 2002; Schmid, 2000).

This chapter outlines the claim made by Flowerdew and Forest (2015, p. 6) that SNs cannot be entirely defined from a merely lexicogrammatical viewpoint. This argument will be developed in the following sections, relying on previous literature, especially the works of Halliday & Hasan's (1976) research into general nouns (henceforth GNs), Schmid's (2000) shell nouns as well as Flowerdew and Forest's (2015) DBSNs.

3.2 The phenomenon of signalling nouns

As mentioned above, previous studies have explored the SN-like phenomenon from different perspectives using different terminologies. Overall, there have been four general areas of interest: SNs as lexical cohesive devices, the 'semantic generality' feature of SNs, the 'container' image of SNs, and the encapsulating function of SNs.

Halliday and Hasan's (1976) main interest in GNs relates to the contribution of these nouns to the cohesion of texts. GNs are regarded as one type of lexical cohesive device in Halliday and Hasan's work, and this forms the starting point for the present study in order to relate the SN-like phenomenon to lexical cohesion. Specifically, based on the definition of GNs mentioned in Chapter 2 (pp. 35-36), GNs denote a group of nouns which are general in meaning and whose meanings can only be defined in specific contexts by reference to other parts of the same text. By emphasising the co-referentiality between lexical items in the same text, as stated above, GNs are used to create the lexical cohesive force between these items and increase the lexical cohesiveness of the whole text.

Although Halliday and Hasan's main interest in GNs is in relation to their role as lexical cohesive devices, the 'semantic generality' (or, for Schmid (2000), 'unspecificity') feature of GNs is also mentioned. The typical examples of GNs are nouns like *people, creature, place* and *idea*, all of which are highly general in meaning (Halliday & Hasan, 1976, p. 274). This feature of semantic generality reveals the typical semantic feature of SNs in the present study, namely, as stated above, that the meaning of an SN is highly general in isolation and needs to be specified by other information when placed in specific contexts. This is related to the third feature of SNs, described in the next paragraph.

The third general area in the research of SN-like phenomenon is the 'container' image of SN-like nouns. The concept of 'container' image derives from the group of nouns named 'container nouns' in Vendler's (1968, pp. 72-82) study which exploits nouns denoting facts and events. Nouns, such as *fact, result, reason, cause* and *idea*, are understood as 'container nouns' because they can act as central parts of clauses which function as 'containers' or 'hosts' for other nominal clauses. Two of Vendler's (1968) examples are "that he died is a fact" (p. 73) and "it is an axiom that all men are equal" (p. 77). In the former example, the noun *fact* (together with the copular *is*) acts as a 'container' for the clause '*that he died*'; in the latter example, a case of extraposition involving an abstract noun *axiom* is regarded as another variant of a 'container noun' structure. The 'container' image of SNs illustrates the exclusive way in which SNs create lexical cohesiveness in texts compared with those of the other lexical cohesive devices which have been discussed in chapter 2.

As a container for specific information, SN-like nouns serve an encapsulating function in the organisation of texts, which is the fourth general area in related studies. The notion of 'encapsulation' originated in Sinclair (1993, p. 7), and was later used by Francis (1994) in her description of 'retrospective labels', inherently unspecific nominal elements which require the specification of their meanings in certain contexts. These 'retrospective labels' serve to encapsulate or package stretches of texts (Francis, 1994, p. 85), which demonstrates a certain economical way of creating lexical cohesiveness by encapsulating more complicated information in a text with a more condensed lexical item.

This encapsulating function of SN-like nouns corresponds to the description of 'shell nouns' in Schmid's (2000, p. 3) study, where shell nouns "create conceptual boundaries by casting larger chunks of information into nominal structures" (p. 12). In addition, shell nouns are claimed to also include the other three general features of SN-like nouns mentioned above.

Shell nouns are defined by Schmid (2000) as "an open-ended functionally defined class of abstract nouns that have, to varying degrees, the potential for being used as conceptual shells for complex, proposition-like pieces of information" (p. 4). The metaphor of 'shells' is similar to those metaphors of 'containers' or 'carriers' mentioned above, emphasising the container-like feature of shell nouns. Used as conceptual 'shells' for "elaborate chunks of information which are expressed in clauses or in longer stretches of the neighbouring text" (Schmid, 2000, p. 6), shell nouns are abstract in isolation and need to resort to their linguistic context to provide more specific information about their meaning, illustrating the semantic generality feature of shell nouns.

Although shell nouns include four key features of SN-like nouns, a structural perspective was adopted by Schmid (2000) to investigate the use of shell nouns in his large computerised corpus. This structural perspective was not adopted by the present study, and therefore shell nouns were not chosen as the cover term for this sub-category of lexical cohesion as discussed in this chapter. The reason for this consideration is that structural perspectives tend to use specific grammatical patterns as the defining criteria for the membership of shell nouns, providing the clearest route to the identification of members of the shell-noun category. However, the present study focuses on the role of SN-like nouns in contributing to textual cohesion, which means the SNs here are recognised during the actual analysis, and any fixed a priori theoretical commitment, such as their grammatical patterning, may obscure the identification of SNs as cohesive devices in texts. Therefore, the structural perspective cannot be adopted in this study, although grammatical patterns which involve SNs can be used as helpful tools for the identification of SNs.

As with shell nouns, DBSNs in Flowerdew and Forest's (2015) work also include the four features of SNs. However, compared with shell nouns, in Flowerdew and Forest's study of 2015, emphasis is placed on the cohesive function of DBSNs at the discourse level, and DBSNs are treated more as a semantic and discourse phenomenon, rather than confined strictly to a lexicogrammatical category. The term 'signalling' indicates both the anaphoric (referring backwards) and cataphoric (referring towards) function of SNs as cohesive devices. The reason for using the term 'SN' as the cover term for this category of lexical cohesion in this study, as noted, instead of the alternatives (e.g. shell nouns), lies in the fact that a discourse perspective on DBSNs is adopted in Flowerdew and Forest's study (2015, p. 6), emphasising that DBSNs are best understood from the perspective of discourse, which is in line with the discourse-specific perspective adopted in this thesis (see chapter 2, p. 50). Another term – 'lexical specification' – is also borrowed from their study (Flowerdew & Forest, 2015, p. 2) to denote the contents that the SNs refer to in the present study.

Among all the terms which have been mentioned within the above explanation of the SN-like phenomenon, this chapter will focus on three (GNs, shell nouns and DBSNs) in order to build the theoretical background of this phenomenon as it is involved in the current research and evaluate the grammatical patterns and semantic categories of SN-like nouns included in previous studies, aiding the identification of SNs in the lexical cohesion analysis in this thesis.

3.2.1 General nouns

As mentioned in chapter 2, the classification of lexical cohesion in the present study is based on Halliday and Hasan's work (1976, pp. 274-277), in which the GN category was originally proposed to discuss one type of lexical cohesive relation between lexical items. Therefore, at an early stage of the present study, GNs were regarded as a sub-category of the reiteration category of lexical cohesion. However, the signalling function and the semantic generality feature of GNs are also related to other nouns involved in the SN-like phenomenon, such as shell nouns and DBSNs. These similarities suggest that GNs represents a bridge between the SN-like phenomenon and lexical cohesion, a subject which needs to be discussed firstly.

Specifically, the co-referential function of the GNs as cohesive devices, as noted above, is the theoretical foundation for the SN-like phenomenon investigated in this study. Furthermore, as GNs emphasise the semantic generality feature of SN-like nouns, semantic categories of GNs can be created according to the classification of major nouns which are highly general in meaning (see table 4):

Table 4 Semantic category of general nouns (Halliday & Hasan, 1976, p. 274)

category	example
human	people
non-human animate	creature
inanimate concrete count	thing
inanimate concrete mass	stuff
inanimate abstract	business
action	move
place	place
fact	question

This semantic classification of the GNs is adopted in this study so as to identify lexical items which are used as GNs in order to create cohesive effects, as other terms of the SN-like phenomenon do not regard concrete nouns, such as *people* and *place*, as SN-like nouns, while the present study includes such nouns because of their cohesive function in texts.

As mentioned in chapter 2 (see p. 35), the GN category is one level above the hyperonymy category regarding the semantic generality. This semantic generality feature of GNs is further emphasised in Halliday and Matthiessen's study:

> There is a category of general noun that functions cohesively to refer to a more specific one that has preceded […] [the] membership is not exactly defined, and other items can be transferred into this category […] to function cohesively – provided they embody a move up in the scale of generality.

(Halliday & Matthiessen, 2014, p. 386)

This description clearly shows that from this semantic generality perspective, GNs are regarded as a functional category because this word class is open-ended and not strictly defined in nature. This means that any noun has the potential to become a GN if it plays a role in creating a cohesive relationship with more specific preceding information, and this information acts as an anaphoric referent. Furthermore, acting as GNs means that the nouns are highly general in 'inherent meaning' (Schmid, 2000, p. 189) or for Flowerdew and Forest (2015, p. 25), 'constant meaning', which increases the scale of generality of the anaphoric referent. These two aspects demonstrate the anaphoric generalisation function of GNs as lexical cohesive devices, which can be demonstrated by the following examples:

3-1 Chen said he did not have the power to single-handedly determine the future of Taiwan and that there would have to be public consensus before Taiwan pressed ahead in trying to establish a confederation with the mainland. But he said it was an example of "new thinking that could bring a breakthrough." Chen added, "There's a lot of room for discussion of **this matter**[11]."

3-2 Then somewhere in the middle of the desert – about six hundred miles later, "I didn't see a connection with anything" – he bangs on the side of the car and I let him out. Now I know Indian people better, and I know that **the guy** probably didn't speak English …

(Halliday & Matthiessen, 2014, p. 647)

In example 3-1, the GN is *matter*[12] and its anaphoric referent is the previous underlined stretch of text. In example 3-2, the GN phrase is *the guy*, which refers back to *he*, who is one of the *Indian people* in this context. It can be seen from these two examples that the referent of GNs can be of any length, from a single word to a longer chunk of segment in a text. The replacement of the longer chunk

[11] The phrases which include the SN-like nouns or the SN-like nouns themselves are bolded in the examples and their lexical specifications are underlined.
[12] The words taken from the examples are italicised in the text.

of segment is a particularly salient feature of GNs, indeed of all SN-like nouns, as noted above, which can encapsulate more complex information into much shorter single lexical items. Furthermore, the use of the GNs *matter* and *people* demonstrate the semantic generality features of these GNs, each of which requires more information to spell out its contextual meaning (or 'variable meanings' in Flowerdew and Forest's sense (2015)) in addition to its inherent meanings. The above two examples also reveal the typical grammatical items which GNs frequently go with, as described in chapter 2 (see pp. 36-37) : the reference determiner *the* (e.g. *the guy* in example 3-2) or demonstratives (e.g. *this matter* in example 3-1).

It is noticeable from Halliday and Matthiessen's statement and examples given above that their main interest in GNs is through their contribution to the anaphoric lexical cohesion of texts. The present study develops this view into both anaphoric and cataphoric lexical cohesion created by GNs and treats GNs as a sub-type of SNs.

3.2.2 Shell nouns

The word class 'shell nouns' was created by Schmid (2000) for the investigation of this class from the perspective of cognitive grammar and corpus linguistics. The prerequisite for becoming a shell noun is the yielding of a 'semantic information gap' filled by other parts (normally immediate co-text) of the same text (Schmid, 2018, p. 111) (see section 3.4.2.2 below for further information). This notion of 'semantic information gap' is adopted in this study as it reveals the nature of using shell nouns to create lexical cohesive force by connecting these nouns and related information which fills the semantic gap. As for the methodology of investigating shell nouns, a large computerised corpus and pre-fabricated grammatical patterns were used to determine whether a noun can function as a shell noun (Schmid, 2000, p. 38). The rationale for using pre-fabricated grammatical patterns as the determinant criterion for the inclusion of nouns as members of the class of shell nouns is explained by Schmid (2000). He points out that the co-interpretation of shell nouns and their shell contents (i.e. the referents of shell nouns) is vital for communicative success between speakers. The speakers trigger this co-interpretation by means of a limited number of

lexicogrammatical patterns which are used to connect shell nouns and their referents. Schmid (2000) proposes that there are four main types of lexicogrammatical patterns involved in the use of shell nouns, and these lexicogrammatical patterns, as well as their variants, are used to identify potential shell nouns in his corpus. This method – that the inclusion of a noun in the class of shell nouns depends on objective syntactic patterns – is partly used in the present study in which some lexicogrammatical patterns are established based on the existing literature. However, as the main perspective of this study is discourse-based, the final syntactic patterns are determined by the examples identified in the current corpora, not vice versa.

In terms of the connection between GNs and shell nouns, Schmid (2000) examined GNs in four types of grammatical patterns and pointed out that not all of Halliday and Hasan's examples of GNs were able to fit into these two types of grammatical patterns:

a. (premodifier) + General noun + *that*-clause/*wh*-clause/*to*-infinitive
b. (premodifier) + General noun + *be* + complementation (*that*-clause/*wh*-clause/*to*-infinitive)

(Schmid, 2000, p. 3)

Therefore, not all the GNs were included in the category of shell nouns in Schmid's study. As with Halliday and Hasan (1976), Schmid also suggested taking *shell nouns* as a functional instead of an inherent concept, as the term 'shell nouns' denotes abstract nouns based on their function of providing semantic information gaps in texts. The two examples below demonstrate this function:

3-3 **The reason** why the preceding sentences are written in such a lousy style is that I am trying to illustrate the way shell nouns work in actual use.

(Schmid, 2018, p. 114)

3-4 **The irony** of the anti-hunting councillors' action is that this motion will not have any significant effect on hunting in Hampshire.

(Schmid, 2000, p. 68)

In these two examples, both *reason* and *irony* are used as shell nouns because these nouns are vague in their inherent meaning, which provides semantic information gaps for the subsequent underlined clauses to fill in particular contexts.

3.2.2.1 Functional definition of Shell Nouns

As mentioned above, the quality of shell-nounhood is a functional quality, as shell nouns in Schmid's sense are not defined by inherent features but further functional features. That is to say, when a shell noun is used in a specific context to encapsulate chunks of information for certain purposes, the noun can be regarded as a shell noun, and it is more appropriate to regard shell nouns as one of many types of the usage of abstract nouns, "rather than as *shell lexemes* in their own right" (Schmid, 2000, p. 13). Several points were stressed by Schmid (2000, p. 14) based on this functional definition of shell nouns: first, nouns can only be viewed as shell nouns when seen in specific contexts; second, there is no complete list of shell nouns due to the fact that any noun has the potential to become a shell noun as long as the relevant contexts need it to function as a shell noun; third, the category of shell nouns is extremely varied in both semantic features and the extent of shell-nounhood that the use of a noun can reach. These three points indicate the context-related criterion for deciding the inclusion of abstract nouns in the category of shell nouns, as well as the open boundary and variety of shell nouns. This context-based criterion and open boundary of shell nouns are in line with the context-sensitive feature of lexical cohesion. To a certain degree, this similarity of being context-sensitive justifies the cohesive function of shell nouns in texts.

3.2.2.2 The shell metaphor for the shell noun

Although the methodology in Schmid's research was corpus-driven, and syntactic patterns were used as the key criterion to select shell nouns in his corpus, the shell metaphor he created for the definition of shell nouns is considered worthy of discussion in order to explain several functions shell nouns possess when used in context. Firstly, as mentioned above, the very reason for naming abstract

nouns as shell nouns is that these nouns can be illustrated by analogy with 'shells' (e.g. egg shells, nut shells), as shell nouns have the function of encapsulating other complex information in texts. The term 'shell nouns' is a short version of "use-as-shell nouns" (Schmid, 2000, p. 4). One important function of 'real' shells in reality is to be a container, a host or a shelter for contents which are vulnerable or easily scattered. This analogous encapsulating function of shells with a 'container-image' is of significance for creating cohesive discourse. Shell nouns act as 'conceptual' shells for complicated information, which makes it easier for readers to connect different parts of a text as well as carrying the concept along when they continue to read the text. The notion of 'conceptual shells' is elucidated in the following example:

> 3-5 **The Government's aim** is to make GP's more financially accountable, in charge of their own budgets, as well as to extend the choice of the patient. Under **this new scheme**, family doctors are required to produce annual reports for their patients ...
>
> (from a radio programme, cited in Schmid, 2000, p. 7)

There are two shell nouns in example 3-5. The first shell noun, *aim*, is involved in an intra-clausal cohesive relation with the succeeding underlined segments. Furthermore, *aim* is used in the noun phrase *the Government's aim* as the subject of the first clause. This shell noun *aim* semantically connects the whole subject and the underlined object part in the same clause to create the lexical cohesive force within a clause in two ways: it characterises this passage in the object position as an *aim* that the Government intends to achieve; and, secondly, it encapsulates this chunk of passage in this single shell-noun phrase[13]. The lexical cohesive force between the larger chunk and the smaller shell-noun phrase as a 'conceptual shell' helps the listener "keep the gist of the information active and […] reactivate it if this should be required as the discourse unfolds" (Schmid, 2000, p. 7).

[13] A term used in Schmid's (2000, p. 8) study to refer to the phrase in which the shell noun occurs.

The second shell noun *scheme* occurs in another shell-noun phrase *this new scheme*, and the whole shell-noun phrase acts as a lexical cohesive device across the clause boundary to refer back to the same underlined passage. That is to say, *the Government's aim* and *this new scheme* share the same referent. The difference is that the former acts as a cataphoric cohesive device to signal the following referent, while the latter functions as an anaphoric signal to re-activate the previous referent. Although this referent has already appeared before the introduction of this shell noun *scheme*, which may indicate *this new scheme* as given information, *scheme* in fact also provides new information to the listener regarding this referent, namely that *the government's aim* has already been developed into an organised and detailed plan. In addition, the use of the anaphoric demonstrative *this* before *scheme* also assists the listener to link the second shell noun, *scheme*, to the previous passage.

Overall, the two shell nouns *aim* and *scheme* act as two conceptual shells which create lexical cohesive effects within and across clauses by characterising the same referential content from cataphoric and anaphoric perspectives respectively. Their cataphoric and anaphoric functions are also taken into consideration by the present study when identifying potential shell nouns as a sub-type of SNs in the current corpora, although the use of *aim* in example 3-5 above is not included in the present analysis because it occurs with its shell content at the intra-clausal level.

3.2.2.3 Three defining functions of shell nouns

As mentioned in section 3.2.2.1 above, shell nouns are defined as a functional class. Certain functions of abstract nouns define these nouns when they are used as shell nouns. Among these functions, three main functions have been suggested to act as the context-related criteria to define the functional class of shell nouns because these three functions were identified as being involved in all uses of shell-content complexes[14] in Schmid's corpus (2000, p. 14). The three key criteria are:

[14] A term used by Schmid (2000, p. 8) to refer to shell nouns or shell-noun phrases as well as the content that they link up with.

The cognitive function of temporary concept formation

Shell nouns contribute to temporary concept-formation, i.e. shell nouns are used to encapsulate complex chunks of information which are conceptually unbounded into provisional nominal concepts "with apparently rigid and clear-cut conceptual boundaries" (Schmid, 2000, p. 14). This concept-forming function derives from the hypostasizing or reifying potential of nouns. More precisely, shell nouns single out different types of experiences and encapsulate these experiences as concepts to reify or hypostasize them temporarily as instances of classes of 'things' by suggesting that these experiences have "thing-like qualities" (Schmid, 1997, p. 3). The examples below can demonstrate this reifying process from 'experiences' denoted by the shell contents to 'concepts' represented by shell nouns in texts:

3-6 **The problem** is that the water companies are as loath since privatisation as they were before it to transfer the reservoirs of surplus water to where they are needed.

3-7 **The problem** was to safeguard the many civil radar sites round Britain from encroachment by property development.

(Schmid, 2000, p. 271)

As can be seen from examples 3-6 and 3-7, the shell noun *problem* is used in the two contexts to represent completely different experiences, which are underlined in the examples: the first use of *problem*, in example 3-6, refers to a water company's behaviour regarding surplus water transfer; and the second use of *problem*, in example 3-7, denotes the protection for civil radar sites around Britain from encroachment by property development. The same shell noun *problem* encapsulates these two types of experiences into a stable concept with 'thing-like qualities', which is "an unwelcome and therefore negatively evaluated state of affairs" (Schmid, 1997, p. 4). The reifying process between shell nouns and their referents also indicates that compared with other nouns, the hallmark of shell nouns is that the concepts created by them are variable, context-specific and therefore transient. The reason is that the contents of the concepts change with the situational and linguistic context in which they are used. What is

characterised as a *problem* changes from examples 3-6 to 3-7, depending on the propositional shell contents.

The two aspects mentioned in the reifying process suggest that the concepts created by shell nouns consist of a stable symbolic and a variable indexical part (Schmid, 2000, p. 18). One part of the meaning of *problem* in the two examples above remains stable, denoting that the perspective from the authors who produced the texts in examples 3-6 and 3-7 regarding the concept-formation of what *problem* refers to is, as mentioned above, 'an unwanted situation'. This semantic part of shell nouns, e.g. *problem*, is included in the second criterion of characterising function of shell nouns.

The semantic function of characterising complex contents

This function represents the writers' attitudes when they use shell nouns to perspectivise complex information which is normally in the form of clauses or even longer stretches in the linguistic co-text. In particular, the writers select particular shell nouns with modifiers to accomplish the semantic characterisation of the related contents. Below is an example to show this semantic characterising function of shell nouns:

3-8 **The problem** is that only a limited number of aspects can be dealt with, giving rise to **the challenge** to select the most important ones.

(Schmid, 2000, p. 112)

In example 3-8, the use of the shell noun *problem* in the main clause characterises the shell content '*that only a limited number of aspects can be dealt with'* as "an obstacle" (Schmid, 2000, p. 112); and in the following co-text, *challenge* is used as another shell noun to perspectivise the shell content '*to select the most important ones*' as "a difficult task" (Schmid, 2000, p. 112).

Schmid points out that the potential of shell nouns to serve the characterising function is not as transient as their potential to serve the encapsulating function. That is to say, the characterising potential is partly related to the inherent meaning of the shell nouns rather than being dependent entirely on their contextual

meaning, and their inherent meaning can vary from extremely generic to quite specific. For example, nouns like *thing* and *people* are extremely generic, while nouns like *disadvantage* and *reassurance* are quite specific (Schmid, 2000, p. 112). The inherent meanings of *problem* and *challenge* in example 3-8 are somewhere in the middle. This variation of inherent meanings is different from GNs, which are normally extremely general in their inherent meanings and entirely dependent on the contexts to provide contextual meanings to them. Therefore, as mentioned in section 3.1, for the sake of capturing a broader picture of the SN-like phenomenon within the scope of lexical cohesion, this study includes not only GNs, which were originally regarded as a lexical cohesive category, but also shell nouns, which depict the same phenomenon but focus on the above perspectives which are still related to lexical cohesion, especially the third main function of shell nouns, which is turned to now.

The textual function of co-referentiality

Shell nouns link nominal concepts with complex information, thereby helping the reader or listener construe different parts of a text into a whole unit, which is the nature of cohesiveness in a text. In the description of the shell-noun phrase *this new scheme* in example 3-5 above, Schmid (2000) indicates the cohesive function of shell nouns by saying that “in the second sentence the speaker uses the noun phrase as a signal for precisely such a re-activation” of the previous referential information (p. 7). It can be seen that shell nouns can function as signals of given or new information to re-activate or indicate such information respectively with a more general and attitudinal concept; at the same time, the information explains the shell nouns by imposing a temporary meaning on them. The whole interwoven interaction denotes a co-referential cohesive relationship between these shell nouns and the referential information. This function forms a core part of the current research in the present study in which the identity of reference in shell-content complexes is investigated from the perspectives of both grammatical patterns and semantic categories. The example below demonstrates the cohesive function of shell nouns, which have identity of reference with other parts in the same text:

3-9 **The advantage** is that there is a huge audience that can hear other things you may have to say.

(Schmid, 2000, p. 22)

In example 3-9, *advantage* is the shell noun with the pre-modifier *the*, and the whole shell-noun phrase *the advantage* refers to the following underlined clause. According to Schmid's framework, it is considered most convincing to use the shell nouns which are included in the pattern 'Noun-*be*-clause' to demonstrate this 'identity of reference' relation between the shell noun and its referent, because in this pattern the shell noun or the shell-noun phrase (e.g. *the advantage* in example 3-9) and its referent are linked by a form of the verb *be*. Such "equative expressions of the type '*A is B'* clearly suggest that *A* and *B* are identical" (Schmid, 2000, p. 27). It is also noticeable that in example 3-9, the identity of reference relation between *the advantage* and the underlined clause is reinforced by the use of the verb *is*, which clearly elucidates the lexical cohesive force between these two parts in the same sentence.

The combination of these three defining functions of shell nouns, as mentioned above, "turns shell nouns into such powerful communicative and cognitive tools" (Schmid, 2000, p. 19). This statement indicates that the textual function of co-referentiality, which is related to the cohesive function of shell nouns in the present study, is only one aspect of shell nouns that contributes to communicative and cognitive purposes. However, as these three functions are also involved in the use of SNs as lexical cohesive devices, shell nouns are regarded as a sub-type of SNs, and in this research the three special functions of shell nouns also help identify SNs in texts.

Overall, the shell metaphor for shell nouns brings to light one of the key features of SNs, which is how they act as 'shells' for other contents. The three key functions as defining criteria for shell nouns also indicate the nature of shell nouns as a functional class, and the fact that whether an abstract noun can be regarded as a shell noun depends on the context in which it is used to encapsulate complex information into cognitive concepts. This functional and context-related feature of

shell nouns is in line with the context-sensitive feature of lexical cohesion, which qualifies shell nouns as lexical cohesive devices.

However, Schmid's structural perspective and the method of using syntactic patterns to identify shell nouns in his corpus have been criticised by other researchers (e.g. Flowerdew & Forest, 2015; Jiang & Hyland, 2016). As Flowerdew and Forest (2015) point out, the decision to define shell nouns as a structural linguistic category leads to several "methodological and descriptive conventions that may obscure some aspects of the phenomenon under study" (p. 6). That is to say, using syntactic patterns to determine the inclusion of abstract nouns in the category of shell nouns is not an effective way of investigating a broader picture of the SN-like phenomenon because some parts of this phenomenon will be omitted if they do not fit into these patterns in Schmid's study. Therefore, although some of Schmid's findings are drawn on in the present study, his structural perspective and his use of grammatical patterns to select shell nouns in his corpus are not adopted.

3.2.3 Discourse-based signalling nouns

As previously discussed, Flowerdew and Forest's (2015) study of DBSNs mainly adopted a discourse-based approach, in contrast to the syntactic and conceptual one in Schmid's study. In the discourse-based approach, DBSNs were positioned in a discourse context to explore the boundaries of DBSNs and their roles in contributing to textual development. This approach is adopted in the present study.

DBSNs have been defined as abstract nouns with non-specific meaning in their own right and specific meaning in a particular linguistic context (Flowerdew & Forest, 2015). This definition emphasises that the meaning of DBSNs is twofold: constant (as mentioned in p. 58) and variable meaning. This twofold meaning of DBSNs corresponds to the two parts of the meaning of shell nouns – a stable symbolic part and a variable indexical part (Schmid, 2000, p. 18), both of which embody a combination of the stability and flexibility of the meaning of SN-like nouns. The difference between the two descriptions is that the dichotomic division of the meaning of shell nouns is used to explain their characterising function and

concept-forming function, while the twofold meaning of DBSNs contributes to determining whether a noun is used as a DBSN in texts. When one can find the variable meaning of an abstract noun according to the neighbouring co-text, the abstract noun can be regarded as a DBSN. This twofold meaning of DBSNs is also adopted as a deciding criterion in the present study to determine the inclusion of a certain noun into the class of SNs.

3.2.3.1 Basic features of discourse-based signalling nouns

In addition to the adoption of the deciding criterion for SNs from Flowerdew and Forest, their descriptions of features of DBSNs are also influential in the present study regarding the investigation of the features of SNs from the perspective of lexical cohesion.

The first relevant feature of DBSNs is the complementary relationship between a DBSN and its lexical specification. They affect each other in terms of meaning: the lexical specification provides the specific meaning for the DBSN, whereas the DBSN shows how the specific contents should be understood in relation to the neighbouring discourse (Flowerdew & Forest, 2015). These two elements are indispensable for using a DBSN in texts. Therefore, the study of DBSNs inevitably involves the lexical specification as well. This view was supported in Schmid's (2000) study in which the author states that "shell nouns and shell-noun phrases can only be studied appropriately if what they link up with is taken into account" (p. 8). As the present study includes DBSNs and shell nouns in the category of SNs as lexical cohesive devices, it is also essential to study both parts in a lexical cohesive pair, i.e. the SN and its referent.

The second relevant feature is the anaphoric and cataphoric cohesive functions of DBSNs, which are referred to as the encapsulating function and the prospecting function respectively in Flowerdew and Forest's study (2015, p. 5). Compared with GNs, which mainly focus on their anaphoric function in lexical cohesion, DBSNs focus on both anaphoric and cataphoric aspects in lexical cohesion. In discourse, DBSNs replace complex notions in stretches of discourse, as with the encapsulating function of shell nouns. By substituting for these notions, DBSNs contribute to the creation of the continuity of the text and maintenance of

the flow in the text by condensing substantial information into nouns in order to make complex meanings carry forward in the discourse, thus maintaining continuity. By acting as an anaphoric cohesive device, DBSNs condense complex meanings into a simple and general concept to reduce the burden on the reader and keep the text as succinct as possible. On the other hand, DBSNs also act as labels in advance to provide conceptual information about the nature of an utterance that will be discovered and explained in the forthcoming text.

In this co-referential aspect of endophoric relationships (anaphoric and cataphoric), DBSNs are important signals to help label and trace intricate notions in current discourse and avoid the repetition of the full complicated information in these notions. Thus DBSNs represent an economical way of making the discourse cohesive and neat, which is the main role of DBSNs in creating lexical cohesive relations in texts. This study elaborates this aspect of DBSNs when analysing examples in the results and discussion chapters (chapters 7, 8 and 9).

3.2.3.2 Research status: a discourse perspective

The discourse-based approach to DBSNs is of course strongly concerned with the discourse features of DBSNs, and in this manner this approach is significantly different when compared with the methods in other works where the emphasis has been on syntactic patterns and sentence-level realisation of the SN-like phenomenon (e.g. Hunston & Francis, 1999; Schmid 2000; Mahlberg, 2005). As mentioned above, the SN-like phenomenon in the present study is discussed from the perspective of lexical cohesion, which is context-sensitive. Therefore, discovering the features of SNs in a specific discourse is more practical and valuable, though it is recognised that it will be more challenging to adopt this perspective in contrast to using traditional structural categories of grammatical patterns. It has already been pointed out by Flowerdew and Forest (2015) that the discourse-based approach of DBSNs requires time-consuming manual analysis of every single sentence in a text to explore their appropriate use, as opposed to using computerised corpora and pre-fabricated grammatical patterns as the examining criteria to discover them (e.g. Schmid, 2000). Although the discourse-based approach is labour intensive, it emphasises the fact that a SN can not only be regarded as a ‘conceptual shell’ of lexical realisations found in a

stretch of text, but also needs to be regarded as a cohesive device from a semantic and discourse stance. Basically, Flowerdew and Forest's discourse-based approach is adopted by this study because it is seen as more accurate and insightful regarding the lexical cohesive analysis of SNs.

However, the discourse-based view in Flowerdew and Forest's study is broader than the structural approach in Schmid's study, which may bring the current research to the boundary of whether a lexical item can be counted as an SN. In order to avoid the vagueness of defining the boundary of SNs, this thesis focuses on SNs as cohesive devices and uses this cohesive function as the determining criterion for the identification of SNs, rather than following Flowerdew and Forest's (2015) approach to investigating the SN-like phenomenon from a broader discourse-based perspective.

3.3 Grammatical patterns of signalling nouns

There are no universal agreed-upon grammatical patterns common to all SNs and researchers have developed slightly different classifications of grammatical patterns based on their own corpora. For example, for the pattern 'SN + complement clause', in terms of adopting corpus-driven methodology, several researchers use this pattern to identify SN-like nouns in their corpus: Schmid's (2000, p. 41) aim was to identify frequently-used shell nouns in a large corpus, and therefore his strategy was to combine this pattern, which he named 'N-*cl*', with Vendler's (1967; 1968) 'container sentence', which he named 'N-be-*cl*'. This combination of two patterns was used as a discriminator for the identification of shell nouns in his corpus. Hunston and Francis (1999, pp. 185-188) narrowed the scope of grammatical patterns down to only rely on the pattern 'SN + *that*-clause' for their discussion of shell nouns. Although their study emphasises the significance of this pattern, such an approach restricts the exploration of grammatical features of SNs; which is not the focus of this study. This thesis only regards this pattern as a core member of grammatical features of SNs; it does not depend on this pattern for the identification of SNs.

Furthermore, the present study mainly borrows the classification of grammatical patterns of DBSNs from Flowerdew and Forest's (2015, pp. 13-24) work. It does

this for several reasons. Firstly, their work is considered insightful by the current researcher and is indeed the most up-to-date and detailed study of the SN-like phenomenon. Secondly, their study also compares previous related research and summarises a comparatively comprehensive category of grammatical patterns which either occurred in their corpus or existed in other works. Thirdly, the present study adopts their discourse-based approach, which means the grammatical patterns for DBSNs in their study and SNs in this study are not pre-fabricated, but are identified and summarised based on the corpora used in the studies. Overall, the grammatical patterns in Flowerdew and Forest's study are comprehensive enough as a starting point for the present study to investigate the grammatical features of SNs. Their grammatical patterns as well as some patterns in other studies are discussed in the following sections.

3.3.1 Noun phrase structure: Premodifiers + SN

In some of the previous literature, demonstratives have frequently acted as premodifiers of SN-like nouns (e.g. Francis, 1994, p. 85; Ivanic, 1991, p. 111; Flowerdew & Forest, 2015, p. 13). As Charles (2003) and Flowerdew (2006) point out, the structure '*this* + SN' is frequently used in academic writing, such as "*this case, this way, this model, this problem, and this point*" (Flowerdew & Forest, 2015, p. 13). Francis (1986, p. 28) further proposes that regarding the use of demonstratives as premodifiers of SNs, *this* and *these* are more frequently used than *that* and *those*.

Despite the popularity of demonstratives as premodifiers of SNs, it is their popular use in other situations that prevents them from becoming the grammatical discriminator of the SN status of nouns (Schmid, 2000, pp. 40-42). This study only treats demonstratives as one type of grammatical pre-modifier that SNs frequently occur with, instead of the discriminator of SN status of nouns.

In addition to the ubiquitous use of demonstratives as pre-modifiers of SNs, there are other types of pre-modifier, such as quantifiers and ordinals, adjectives, as well as indefinite and definite articles (Flowerdew & Forest, 2015, p. 14). These pre-modifiers also help confine the variable meanings of SNs in contexts. The

examples below demonstrate the use of different pre-modifiers with SNs to create cohesive force between lexical items in texts.

3-10 The defendant was abusive, unco-operative and possibly hostile to the officers, using obscene language calculated to provoke and antagonise the officers and tried to walk away. The justices were satisfied that **this conduct** amounted to an obstruction for the purposes of a charge under s 51(3) of the Police Act 1964.

3-11 whether people have these strong rights is **the critical point**.

3-12 **One commonly held view** after Condron and Condron was that the jury had to make a qualitative judgment about the defendant's silence.

(Flowerdew & Forest, 2015, pp. 49-50)

In example 3-10, the SN *conduct* is pre-modified by the demonstrative *this*; in example 3-11, the SN *point* is preceded by the definite article *the* and the adjective *critical*; and in example 3-12, there are several pre-modifiers for the SN *view*, which are the quantifier *one* and the adjective phrase, which consists of the adverb *commonly* and the adjective *held*. These three examples above demonstrate how different pre-modifiers are used to confine the variable meaning of SNs in specific contexts. Specifically, *this* is used to suggest that *conduct* is used as an anaphor in the text to refer back to previous underlined lexical specification in example 3-10. The use of the definite article *the* helps specify the referent of the SN *point*, while *critical* indicates the central role and importance of what *point* is linked with in example 3-11. As for the use of pre-modifiers in example 3-12, *one* is used to indicate that there might be other views apart from this *view* referred to in this context. The use of the phrase *Commonly held* reveals the popularity of this *view*. As can be seen from these examples, compared with the specification function of the referents of SNs, adding different pre-modifiers to SNs is another flexible way of confining the variable meaning of SNs in specific contexts.

3.3.2 Noun phrases: SN + complement clause

The pattern 'SN + complement clause' has been evaluated as "one of the most dependable phrase-level characteristics of [DB]SNs" (Flowerdew & Forest, 2015, p. 14). As mentioned above, this pattern has been taken as the discriminator of shell nouns in Hunston and Francis' study (1999) as well as in Schmid's 2000 study. The complement clauses in this pattern are the lexical specification of SNs, which include several types: *that*-clause, *wh*-clause, *to*-clause, and non-finite clauses with prepositions. The examples below show the use of these types of complementary clause with SNs:

N-that

3-13 my **belief** that there are a lot of other real ecologies all over the place…

(Flowerdew & Forest, 2015, p. 78)

N-wh

3-14 a **scenario** of ancient of events in which the earliest life form could have evolved more than 4 billion years BP (before the present)…

(Flowerdew & Forest, 2015, p. 80)

N-to

3-15 a **tendency** to saturate at higher doses…

(Flowerdew & Forest, 2015, p. 78)

N-prep.

3-16 this **idea** of a critique from within…

(Flowerdew & Forest, 2015, p. 78)

As the research scope of this thesis study is lexical cohesive relations beyond the clause boundary, this noun phrase structure of SNs, in which the SNs and their lexical specifications are included in the same clause, are omitted.

3.3.3 Clause structure: SN + *be* + nominalisation

Another structural discriminator applied to identify shell nouns is Vendler's (1967, 1968) 'container sentence' frame which only focuses on *that*-clause as the nominalisation. However, this study uses the term 'container sentence' to refer to a broader category of five types of clauses. In the structure 'SN + *be* + nominalisation', a SN is generally part of the subject of a clause; the verb *be* is the linking verb; and different types of clauses are regarded as nominalisations occurring in the object position. In this structure, as mentioned above (see pp. 16-17), the nominalisation element shares an equivalent status with the SN, as shown in the examples below:

SN + be + that-clause

3-17 The **difficulty** is that many members of the public regard the EU as remote until some threat to the British way of life is perceived.

(Flowerdew & Forest, 2015, p. 78)

SN + be + to-clause

3-18 The **method** that works very well is to take the vinyl derivative of another.

(Flowerdew & Forest, 2015, p. 80)

SN + be + deverbal noun

3-19 The transmission of the action potential is the **basis** of the nerve impulse that is transmitted along the axon.

SN + be + gerund

3-20 One of the **problems** of being a conservative is knowing what to do when a major change does take place.

(Flowerdew & Forest, 2015, p. 82)

3.3.4 Other grammatical patterns

Other grammatical patterns are based on the above two main patterns. The present study summarises six additional patterns from previous studies. Five of them are in accordance with Flowerdew and Forest's (2015) research, while one is from Schmid's (2000) classification.

3.3.4.1 Extraposition of lexical specifications

The explanation of this structure starts with two examples:

3-21 But **it** is a good **idea** to stop and think about it.

3-22 Agnelli is due to step down soon as head of Fiat and already **there** is **speculation** that he might move into politics.

(Schmid, 2000, p. 24)

In example 3-21, the underlined nominalisation is to offer the SN *idea* its contextual meaning, rather than modify it. That is to say, the underlined nominalisation is the real lexical specification of the SN *idea*, but does not appear in the subject position in the sentence. Similarly, in example 3-22, although the subject of the second clause is *there*, the underlined clause provides the lexical specification of the SN *speculation* instead of being a post-modifier of this SN.

The basic structure shown in the examples above is '*it/there* + *be* + SN + nominalisation', which is named the 'extraposition structure of SN' in the present study.

3.3.4.2 SNs and specifics in apposition

Apposition is used in a broad sense in Flowerdew and Forest's research (2015), which denotes that DBSNs juxtapose with their lexical specification "in a relation of identity" (p. 21). Their study identifies that only the lexical specifications can be

allocated to the apposition position in a sentence, while the present study expands their pattern to allow for both SNs and their lexical specifications to be put in the apposition position. This 'relation of identity' between the SNs and their lexical specifications in this structure is demonstrated in the following examples:

3-23 Pesticides move into the soil as a result of three **processes**: pesticide diffusion in the aqueous phase along a solute-concentration, diffusion in the gas phase in response to a gradient in gas-phases concentration, and convection (mass flow) of the pesticide.

3-24 The giraffe and the elephant have solved the same **problem**: how to drink without kneeling.

(Flowerdew & Forest, 2015, p. 21)

It can be seen that in examples 3-23 and 3-24, the underlined lexical specifications are in the apposition position of the SN *processes* and *problem* respectively, signalled by punctuation in the texts**.**

3.3.4.3 Comparative specifics

This is the borderline pattern regarding the use of SNs as lexical cohesive devices, as the lexical specifications exist but in a comparative construction. See the examples below:

3-25 Under severe **conditions**, such as occurred during 1950 in parts of the south when heavy rains followed heavy spray applications.

3-26 Other optimisation **criteria**, like maximisation of the smallest singular value, can be found in [cttn].

(Flowerdew & Forest, 2015, p. 21)

In example 3-25, the lexical specification of the SN *conditions* is by exemplification, signalled by the phrase *such as*. Although the underlined content cannot represent the whole content of *conditions*, it still functions as the lexical specification of *conditions* to fill some of the semantic gaps created by *conditions*.

Therefore, this structure can be regarded as a grammatical pattern of SNs in the present study. A similar case is shown in example 3-26. The SN *criteria* is partly specified by the underlined part in the same sentence via analogy, signalled by the preposition *like*. The two examples reveal the middle status of this 'comparative specifics' pattern on the scale of non-specific to fully specific lexical specification. This pattern was accepted in Flowerdew and Forest's study (2015), even though it is regarded as a borderline pattern. The current study follows their stance, as the key feature of identifying a SN is to find its corresponding lexical specification in the same text in order to create lexical cohesive effects.

Overall, the grammatical patterns mentioned above are found in most previous studies. This study follows the stance which Flowerdew and Forest (2015) take, to treat some of these patterns only as grammatical features of SNs when being used in contexts, and therefore as helpful tools for the identification of SNs.

3.4 Semantic features of signalling nouns

Apart from the discussion of grammatical patterns which involve the use of SNs, semantic features of SNs are also worthy of investigation, as SNs are "characterised by a number of semantic features" (Flowerdew & Forest, 2015, p. 25). The present study critically evaluates the semantic features of DBSNs and the semantic prerequisite for shell nouns, and connects them with the SNs (as defined by this thesis) and their lexical cohesive function.

3.4.1 Semantic features of discourse-based signalling nouns

Flowerdew and Forest (2015) proposed three main semantic features of DBSNs, which will be introduced one by one in the following sections.

3.4.1.1 Discourse-based signalling nouns as superordinates

When used as lexical cohesive devices in texts, SNs are generally considered as lexical superordinates which have a constant and a variable meaning (Flowerdew & Forest, 2015, p. 25). This status of DBSNs as lexical superordinates is closely related to GNs. However, Flowerdew and Forest (2015) point out that this superordinate status of DBSNs is not good enough to be a discriminator for labelling a noun as a DBSN. The first reason is that this superordinate status is

too ubiquitous and wide-ranging to determine whether a lexical item can neatly fit in the category of DBSNs due to the fact that any lexical item can take on superordinate status in certain contexts. Their second reason is that many superordinates used as GNs are concrete nouns (e.g. *people, animal, plant* and *tree*) while an important feature of DBSNs is their abstractness.

As mentioned above (see p. 71), the present study focuses on the lexical cohesive force that SNs can create in texts, Flowerdew and Forest's stance regarding the connection between DBSNs and superordinates as well as GNs is not entirely accepted in this study. The reason for this is that even concrete nouns have a constant and variable meaning in specific contexts when they act as lexical cohesive devices in texts, such as the SN *man* in example 3-27 below:

> 3-27 Didn't everyone make it clear they expected the minister to resign? – They did. But it seems to have made no impression on the **man**.
>
> (Halliday & Hasan, 1976, p. 275)

The concrete noun *man* is regarded as a GN in Halliday and Hasan's study (1976, p. 275), and the GN *man* in example 3-27 has a constant meaning – "an adult male human being" (OED Online, 2019k) – as well as a variable meaning in this context, i.e., *the minister*. As GNs are one sub-type of SNs in the present study, concrete nouns can also be regarded as SNs when they are used as lexical cohesive superordinates.

However, the present study also agrees with Flowerdew and Forest's claim that superordinates are not appropriate to be the discriminator for identifying DBSNs because there is another category called hyperonymy in this thesis which specially deals with superordinate cohesive relations. SNs can be regarded as superordinates in texts, but not all superordianates can be regarded as SNs, as SNs have other semantic features in addition to the feature of being superordinates, such as creating unspecific semantic gaps. Therefore, superordinates can be regarded as a semantic feature of SNs rather than being a discriminator to identify SNs in texts.

3.4.1.2 The constant and variable meaning of discourse-based signalling nouns

The second related semantic feature of DBSNs is that a lexical item possesses both a constant and a variable meaning when used as a DBSN, which has been mentioned in section 3.2.3. This feature has been regarded "as a key criterion for assigning SN status to an item" (Flowerdew & Forest, 2015, p. 26), which can be readily illustrated by the use of the noun *theory* in the two different examples below:

> 3-28 all the **theory** you need is to remember, that there are these things called inertia forces and that you can plug them in to the beam bending formula that you have been using for a long while
>
> 3-29 the **theory** of perfectly contestable market is saying that you have free entry and free exit you will have in equilibrium prices equal to marginal cost you'll have all firms producing at minimal unit costs they will the industry output will be produced at minimal cost and they will the market will clear all consumers will be satisfied at a price which is equal to marginal cost even if you only have two firms or three or four firms in the market.
>
> (Flowerdew & Forest, 2015, p. 26)

As can be seen from these examples above, the constant meaning of *theory* in these examples is that of 'explanatory framework'. The variable meaning of *theory* in each example is "the representation of the specific explanatory framework of relevance to each classroom discussion" (Flowerdew & Forest, 2015, p. 26). It is these two different combinations of the constant and variable meaning of the noun *theory* that determine the use of *theory* as a DBSN in two different contexts.

3.4.1.3 Extended reference of discourse-based signalling nouns

'Extended reference' was created by Halliday and Hasan (1976, p. 52) to refer to a reference which is more than just a person or object. This concept is opposed to 'text reference', in which the referent is a person or object. In the grammatical

patterns, ‘extended reference’ is normally expressed by a clause or clauses, while ‘text reference’ is displayed by a single nominal (Halliday & Hasan, 1976, p. 52).

Francis (1994) articulated this feature well when discussing her category of retrospective labels:

> A retrospective label serves to encapsulate or package a stretch of discourse. My major criterion for identifying an anaphorically cohesive nominal group as a retrospective label is that there is no single nominal group to which it refers: it is not a repetition or a ‘synonym’ of any preceding element. Instead, it is presented as equivalent to the clause or clauses it replaces, while naming them for the first time. (Francis, 1994, p. 85)

According to Francis’ description of the criterion for identifying a retrospective label, it is necessary to add the caveat to the lexical specification to which the retrospective label refers should not be a single nominal group, which requires the lexical specification to be clause or a number of clauses. This criterion was modified by Flowerdew and Forest when they discussed the boundary of the lexical specifications of DBSNs. In their study, single nominal groups can also be included in the group of lexical specifications if “such a nominal group exhibits clause-like semantics” (Flowerdew & Forest, 2015, p. 28), such as in the form of a deverbal noun, deadjectival noun or gerund as the head of the nominal group. Their reason is that “the semantics of these lexical [specifications] reflect a process of grammatical metaphor which has encoded a clausal structure in the form of a nominal structure. This clause-like-ness provides the justification for including [such nominal groups]” (Flowerdew & Forest, 2015, p. 28).

The present study has a slightly different stance from Flowerdew and Forest’s opinion regarding this semantic feature of DBSNs. ‘Extended reference’ is only a feature of some of the lexical specifications of SNs in this study. Lexical specifications can also be single nominal groups regardless of whether they have the ‘clause-like-ness’ feature. As the focus of this study is on the lexical cohesive aspect of SNs, as long as they are used as lexical cohesive devices in texts and

their lexical specifications can be found in the same text, it is not necessary to confine the form of the lexical specifications. As is the case in example 3-27 above, *the minister* and *the man* form a lexical cohesive pair in which *man* is regarded as a GN in a sub-type of SNs. *The minister*, as the lexical specification of the SN *man*, is a single nominal group but without the 'clause-like-ness' feature.

Overall, 'extended reference' is treated as a semantic feature of some of the lexical specifications of SNs in the present study, rather than a major criterion for identifying SNs in texts, which is different from previous studies (e.g. Francis, 1994; Flowerdew & Forest, 2015).

3.4.2 Semantic prerequisites for shell nouns

There are two semantic prerequisites for the classification of shell nouns and they need to be addressed before the discussion of this classification in section 3.5.

3.4.2.1 The abstractness of shell nouns

The first semantic prerequisite is the abstractness of shell nouns. There are two main factors which are considered to influence the feature of abstractness in shell nouns. First, the referent of shell nouns is untouchable and invisible, known as the 'extensional abstractness' of shell nouns. Second, the use of shell nouns in different ways contributes to the 'stylistic abstractness' of the whole text.

<u>The extensional abstractness of shell nouns</u>

Schmid (2000) suggests that the most common way of testing the abstractness of shell nouns is to examine their referent, which is referred to as an extensional research perspective, and in which the abstractness is called "extensional abstractness" (p. 63). Furthermore, Schmid (2000) also claims that the levels of abstractness of the referents of different shell nouns vary (p. 64). Quite a number of nouns denote notions and ideas which are already abstract, while in the case of some other nouns the level of abstractness is not quite clear. For example, when referring to places and locations, shell nouns like *area, region* and *site* represent some parts of the concrete physical world. It becomes more debatable when the shell nouns denote activities, such as *campaign, mistake, refusal* and *reaction*, as these activities happen in the physical world (Schmid, 2000, p. 64).

To flesh out the concept of 'extensional abstractness', another concept developed by Lyons (1977) is proposed, namely the tripartite distinction among first, second and third-order entities existing in the world (Schmid, 2000, p. 64). According to Lyons (1977), first-order entities are persons, animals, other organisms and physical objects which exist in space and possess "fairly constant perceptual properties" (p. 443). Because of these characteristics, the first-order entities are regarded as bad candidates for temporary concept-formation (Schmid, 2000, p. 64). Second-order entities are events, processes and situations. They occur in time rather than exist in the world. In contrast, third-order entities are 'abstract' entities such as concepts, propositions and ideas, which are outside place and time. The three types of entities are on a continuum from the least abstract end to the most. Second-order entities are in the middle because

> events, states, processes and activities are more abstract than persons and things in the sense that they have no stable existence in the dimension of time, but less abstract than ideas and propositions because they take place in the physical world. (Schmid, 2000 , p. 67)

Because these three types of entities are at different levels of abstractness, only second- and third-order entities are treated as the referents of shell nouns and categorised into smaller groups. *Event* was used as the cover term for the categories of second-order entities, and the classification of situation types created by Quirk, Greenbaum, Leech, & Svartvik (1985, p. 201) were adopted, which are [±DYNAMIC] and [±AGENTIVE]. Based on these two dimensions, the categories of activities, processes and states in second-order entities were further classified into different types, as table 5 demonstrates:

Table 5 Situation types of second-order entities (based on Schmid, 2000, p. 66)

	dynamic	agentive
activities	+	+
processes	+	-
states	-	-

Compared with the categorisation of second-order entities, it is harder to divide third-order entities into subcategories, as there is no superordinate term for this class. Schmid (2000) explained that at first he had the intention to choose 'proposition' as the superordinate term, but did not do so because 'proposition' had been used in different contexts, something which might lead to confusion if he had adopted it in his study. The term '*abstract relation'* was chosen instead as the cover term for the third-order entities. Under this cover term, abstract relations were divided into three subcategories: facts, ideas and utterances. Facts are conceived of as 'being the case' without any epistemic or truth-conditional claims; ideas are related with propositions and thoughts expressed by single nouns, like *idea, thought* or *assumption*; utterances are "linguistic expressions of ideas" (Schmid, 2000, p. 67).

Overall, the classification of second- and third-order entities as referents of shell nouns underpins the semantic framework of shell nouns which will be discussed in section 3.5.2.

Stylistic abstractness of shell nouns

Schmid (2000, p. 70) borrows a particular type of 'construal of a conceptualisation' (Langacker, 1987; 1991) in Cognitive Grammar to explain the notion of 'stylistic abstractness'. The speaker/writer chooses a linguistic item to express their construal of a conceptualisation. One aspect is the choice of word class. Another aspect is the combination of three elements: "the distribution of figure and ground[15] within the clause, the perspective taken on a situation, and the allocation of attention to various aspects of the situation" (Schmid, 2000, p. 70). Compared with cognitive linguists who focus on how conceptualisation is reflected in language, Schmid is more interested in the stylistic effects of linguistic construals of conceptualisations.

[15] Matrix clauses code foreground information while subordinate clauses code backgrounded information in order to "support, enrich, or comment on the events of the main narrative" (Diessel, 2004, p. 45). That is to say, subordinate clauses encode the ground for the figure event encoded in the matrix clause (Diessel, 2004).

As with Flowerdew and Forest's (2015) study, the concept of 'grammatical metaphor' proposed by Halliday (1994) was also adopted in Schmid's study (2000). The core idea of this concept is that events and facts can be expressed in two ways: with 'literal' expressions or with 'metaphorical' expressions. In terms of shell nouns, 'grammatical metaphor' was used to explain the difference between the two methods of expressing concrete events (i.e. experiences) – 'concrete' and 'abstract' methods. In Halliday's examples, this difference is clear:

3-30 Mary **saw** something wonderful. (literal statement/concrete method)

3-31 Mary came upon a wonderful **sight**. (grammatical metaphor/abstract method)

(Halliday, 1994, pp. 343-344)

Schmid provides three examples to differentiate the two methods further:

3-32 his Buddhist mind-training exercises led him to hold **the firm belief** that the mind has unrecognized powers.

3-33 A very minor flaw lay in **the fact** that Tom Wood, alone among his travelling companions, had met Markham face to face.

3-34 Among his travelling companions Tom Wood alone had met Markham face to face, and **this** was annoying.

(Schmid, 2000, p. 71)

In example 3-32, the reporting expanded predicate *hold the firm belief that* … is compared with its corresponding reporting verbal expression *believe firmly that*… by Schmid (2000, p. 71). He claims that both of the *that*-clauses in these two phrases are the projections of ideas, i.e. these *that*-clauses do not function as direct representations of experience but as representations of representations (Halliday, 1994, pp. 250-273, 263-264). There are two ways available to express other people's mental states: one is nominal and another is verbal. The 'stylistic abstractness' is involved in the writer's choices of the two ways of projecting ideas.

In terms of example 3-33, the paraphrasing task of the phrase '*the fact that Tom Wood* […] *had met Markham face to face*' is not straightforward. A possible paraphrase version is given by Schmid (2000), shown in example 3-34. The pronoun *this* is used as an 'extended reference' to replace '*the fact*' and split the original expression into two parts. It is hard to say that this paraphrase is precisely in line with the original version in example 3-33. One may find it possible to insert *fact* in example 3-34 but not *event*. This attempt suggests that the function of *fact* in example 3-33 is unveiled in the expression *this fact* in example 3-34 which shifts the conceptual status of the same experience from *event* to *fact*. Consequently, the shell noun *fact* is less easy to paraphrase compared with the shell noun *belief* in example 3-32. This suggests that grammatical metaphor plays a more important role in example 3-32 than in 3-33.

The result of this comparison is that some shell nouns (e.g. *fact*) create a 'conceptual transformation' (Schmid, 2000, p. 73) from event to abstract relation (e.g. fact) while not causing more stylistic abstractness than is required by the experience that the nouns denote. In contrast, some other shell nouns (e.g. *belief*) keep the same conceptual status but contribute more to the stylistic abstractness. In a word, nouns are more effective than verbs to express the abstractness of information by being the predicates of a clause and contributing to the verbalisation of second-order entities.

Overall, Schmid proposes that, firstly both shell nouns and shell contents can represent events or abstract relations. Shell contents alone cannot reveal the distinction. Second, when a shell noun and its corresponding shell content both denote abstract relations, deciding whether stylistic abstractness can play a role here or not depends on whether the same experience can also be expressed by verbal methods (e.g. *belief* – *believe*). Third, when a shell noun represents an abstract relation while the shell content represents an event, stylistic abstractness cannot be involved, as the shell noun functions to turn the event into an abstract relation, in which case the shell noun cannot be replaced by a non-abstract paraphrase (e.g. the use of *fact* in example 3-33). Finally, when the shell nouns and shell contents represent events, stylistic abstractness is at work while extensional abstractness does not change (Schmid, 2000, p. 73).

3.4.2.2 Unspecificity of shell nouns

In addition to the abstractness feature of shell nouns mentioned above, another semantic prerequisite for shell nouns is the unspecificity of shell nouns. Some shell nouns, representing second-order entities, are considered to be highly unspecific, such as *event, state, action, act* and *process.* They belong to the category of GNs in Halliday and Hasan's framework of lexical cohesion. In linguistics, 'unspecificity' is regarded not as an extensional but an intensional[16] term. The meaning of a noun is unspecific if this meaning is determined by only one or a very small number of semantic dimensions (Schmid, 2000, p. 74).

Schmid proposes that many shell nouns possess the semantic property of 'intensional unspecificity' (i.e. conceptual unspecificity). These nouns can be paraphrased by three groups of phrases: 'something happens' (*event, process*), 'someone does something' (*act, action*) and 'something is the case' (*state*) (Schmid, 2000, p. 74). For example, the noun *time* has one semantic dimension, 'temporal'; for the noun *place*, the sole semantic dimension is 'locative'; and the only semantic dimension of *fact* is 'factual'. These shell nouns possess the very nature of having mono-dimensional senses by acting as GNs.

However, not all shell nouns are unspecific in this sense and some highly unspecific nouns cannot be used as shell nouns (e.g. *creature, person,* or *object* in Schmid, 2000, p. 75) because, as argued by Schmid, nouns with concrete meanings cannot be regarded as shell nouns, a claim that occurs with Flowerdew and Forest's stance on the issue of excluding GNs as DBSNs (see p. 79). Schmid discusses solutions offered by previous researchers of shell-noun-like issues. For Ivanic (1991), the unspecificity of carrier nouns is from a syntactic perspective. She explained that the unspecificity of the carrier noun *purpose* is "like a syntactic specification with slots to be filled according to each new context in which 'purpose' plays a part" (Ivanic, 1991, p. 94). Francis (1994) offers another

[16] There is a distinction between 'intension' and 'extension' when describing the meaning of a linguistic expression. 'Intension' denotes the property or concept of the linguistic expression, while 'extension' represents objects in the world which this expression refers to (Allwood, Lars-Gunnar, & Östen, 1977, p. 126).

interpretation of unspecificity in her work on advance and retrospective labels: “the main characteristic of […] a label is that it requires lexical [specification] […] in its co-text: it is an inherently unspecific nominal element whose specific meaning in the discourse needs to be precisely spelled out” (Winter, 1982, 1992, cited in Francis, 1994, p. 83).

As Francis suggests, the above explanation refers to Winter’s work. In his earlier work, Winter mentioned a type of clause called *unspecific clauses*. These clauses need other co-text clauses to help complete their meanings. He (Winter, 1992) subsequently examined nouns which are semantically unspecific, noting that these nouns talk about “the nature of the clause or sentence as a message in the text itself” by serving as a “metalanguage for the clause” (p. 133). These nouns need to be “lexically unique” (Winter, 1992, p. 153) to be communicatively effective and successful (Schmid, 2000, p. 75).

Based on these suggestions, as mentioned in section 3.2.2 above, Schmid concludes that one semantic prerequisite for shell nouns is that they have one or more ‘semantic gaps’ which need to be filled by shell contents. He claims that a shell noun has two parts in the semantic structure: “a stable and a relatively well-determined” (Schmid, 2000, p. 76) part, and one which has structure-inherent ‘semantic gaps’ which need to be filled with other information. However, the nature of these gaps is still determined by the shell nouns, a point which defies explanations offered by other researchers of the intensional unspecificity of shell-noun-like lexical items. As mentioned above, Schmid suggests that the examining factor of shell nouns should be the patterns N-*cl* and N-*be*-*cl.* His reason for this is that the semantic characteristics of shell nouns are sometimes hard to examine (Schmid, 2000, p. 79). This view is not supported by the present study as the approach adopted is to identify the use of SNs in different contexts, while the grammatical and semantic features of SNs only act as tools for the identification of SNs rather than determining factors.

3.5 Semantic classification of signalling nouns

Based on the semantic features of SN-like nouns noted above, there are diverse semantic types of SNs which need to be classified within a coherent and unified

theoretical framework. Two semantic frameworks are elucidated below from previous studies: shell nouns in Schmid's study (2000) and DBSNs in Flowerdew and Forest's study (2015). After the introduction of their frameworks, an operational semantic framework is developed for helping the identification of SNs explored in the present study. For each of the categories introduced below, at least one example of its prototypical members in use as an SN-like noun will be given. This, hopefully, will link this rather abstract semantic discussion of SN-like nouns to their actual use in texts.

3.5.1 Semantic framework of shell nouns

It is claimed by Schmid (2000) that the semantic categorisation of shell nouns is necessary because it is helpful to obtain an overview of the data in his large corpus and for a more thorough differentiation and description of shell nouns (p. 87). This view is shared by the present study.

According to Schmid (2000), shell nouns are categorised into five groups. What is worthy of note here is that this categorisation of shell nouns allows the same noun to be allocated into several groups, which is termed "multiple category membership" in Schmid (2000, p. 87). Therefore, all the categories are named based on the 'uses' of their shell noun members instead of the 'meanings' of the shell nouns. This 'use-based' approach, or, in Schmid's (2000) term, "pragmatic perspective" (p. 87), is more appropriate for expressing highly unspecific shell nouns with several uses rather than with several meanings, e.g. *point* (see examples 3-37 and 3-38 below).

Schmid (2000) also proposes that although 'intensional unspecificity' is not a necessary feature of shell nouns, this semantic prerequisite for shell nouns still contributes to their classification as the majority of shell nouns do not command many semantic features (p. 88). For example, highly specific nouns such as *irony* or *gripe* are much less frequent than other unspecific shell nouns in Schmid's corpus. Since the links between shell nouns and shell contents are essential features of shell nouns which need to be identified, and most shell nouns are intensionally unspecific, it is important to clarify the order of their relational semantic features as they may share similar or even the same semantic features,

but are only different in terms of orders of those features. For example, *aim* and *attempt* share the same components of semantic features, namely [MENTAL][17], [FUTURE], [ACTION] and [ACCOMPLISHMENT]. As *aim* is typically used to "shell ideas or mental states that are directed towards future accomplishments" (Schmid, 2000, p. 87), the order of its semantic features should be [MENTAL], [FUTURE], [ACTION] and [ACCOMPLISHMENT]. Comparatively, *attempt* is used to denote people's activities which have specific aims to be achieved. Therefore, the order of the features of *attempt* needs to be changed to [ACTION], [MENTAL], [FUTURE] and [ACCOMPLISHMENT]. It can be seen from this example that it is necessary to determine the order of semantic features when two highly unspecific shell nouns share the same set of semantic features.

On the other hand, it is due to the characteristic of 'intensional unspecificity' that the number of semantic features of shell nouns is limited. The five classifying semantic features summarised by Schmid according to the data in his corpus are listed in the table below with a brief introduction of the types of experiences that the features describe:

Table 6 Six semantic features functioning as classifiers for shell nouns (Schmid, 2000, p. 89)

semantic feature	type of experience being described
factual	facts, states of affairs
mental	ideas, cognitive states and processes
linguistic	utterances, linguistic acts and products thereof
modal	possibilities, abilities, permission, obligations, etc.
eventive	activities, processes, states
circumstantial	situations, times, locations, manners of doing things and conditions for doing things

Although these features are considered to be neither primitive nor precise, they are adopted to constitute an economical classification of experiences, and used

[17] Following Schmid's convention, the semantic features of shell nouns are marked by insertion in brackets and capital letters for the purpose of readers' convenience.

as descriptive devices of shell nouns. An explanations for each feature, with examples, is given in the following sections.

3.5.1.1 Factual shell nouns

The first category, factual shell nouns, describes abstract relations between two entities which express the meaning of "to be the case" (Schmid, 2000, p. 89) , but, in fact, the truth of the expressions does not matter at all. As Schmid (2000) claims, "[any] experience, with the exception of first-order entities, can be construed as a fact by means of an appropriate shell-content construction" (p. 92). An example is provided below:

> 3-35 While this is a marvellous development, and will provide a much-needed boost to research into the biology of Amazonia, it will not contribute significantly to the fundamental problems. **The fact** is that neither of these institutions have significant social or political research arms...
>
> (Schmid, 2000, p. 97)

In example 3-35, *fact* is used in the pattern 'the fact is + *that*-clause' as an emphatic gesture to emphasise the claim that the content in the *that*-clause is true (Schmid, 2000, p. 97). Furthermore, *the fact is that* has been paraphrased as "*the thing in disconformity with what is believed/apparent is that"* (Tuggy, 1996, p. 722). This paraphrase points out that the contrastive meaning carried by the word *disconformity* is indicated in the use of this pattern "the fact is + *that*-clause". As shown in this example, *neither* is used in this *that*-clause to express that *the fact* is in disconformity with what people used to believe regarding the development in these institutions.

3.5.1.2 Linguistic shell nouns

Linguistic shell nouns are used to portray linguistic activities and their contents as well as products (Schmid, 2000, p. 131). The shell nouns themselves in this category contain the reporting speakers'/writers' characterisation of the utterances[18] that they report, as demonstrated in the example below:

[18] Utterance is used in a broader sense to refer to "linguistic expressions of ideas" (Schmid, 2000, p. 66). Halliday (1994) uses another term, 'locutions', to refer to similar notions (pp. 263-264.

3-36 To counter the apparent dynamism of the Democrat team, the best the White House has been able to conjure up is **the tired accusation** that they are liberals and lackeys of special interest groups.

(Schmid, 2000, p. 131)

In example 3-36, the author uses the shell-noun phrase *the tired accusation* to characterise the underlined utterance *they are liberals and lackeys of special interest groups*.

Linguistic shell nouns are divided into two sub-categories: propositional and illocutionary. The majority of linguistic shell nouns are used to portray utterances as either propositional or as illocutionary acts[19], especially the latter. Two examples are given to demonstrate the use of these two types of linguistic shell nouns:

3-37 Not only was it Jane who raised Stephen's morale when faced with **the depressing news** that his illness would eventually leave him paralysed and speechless, but it was she who physically helped him to write his blockbuster.

3-38 The Association will give **a warning** that poll tax bills in some Conservative districts will exceed government guidelines unless they receive a higher proportion of the money being made available to local authorities.

(Schmid, 2000, p. 136)

The shell contents of both examples are in the post-modifying *that*-clauses which denote the original utterances produced by other people. The difference between the use of the shell noun *news* and *warning* lies in their different perspectives of profiling the original utterances. The use of *news* in example 3-37 emphasises the propositional content in the original utterance, while the noun *warning* is used to allow the speaker who produced the content of example 3-38 to impute his/her

[19] Propositional acts refer to the acts of referring and predicating, and illocutional acts denote acts that have the intention of communicative aims (Schmid, 2000, p. 135).

communicative intentions to the original utterances produced by original 'experiencers'.

3.5.1.3 Mental shell nouns

The third category, mental shell nouns, refers to a human's cognitive states or processes, and is divided into two sub-categories: conceptual shell nouns and psychological-state shell nouns. This division is the counterpart of the division of linguistic shell nouns (i.e., propositional and illocutionary shell nouns). More precisely, the division between propositional and illocutionary uses of shell nouns in the linguistic domain can be transferred to the mental domain. Analogous to propositional linguistic shell nouns, conceptual shell nouns contain the propositional contents of ideas; and in analogy to illocutionary shell nouns, psychological-state shell nouns allow speakers/writers to characterise their attitudes towards the original ideas (Schmid, 2000, p. 185).

Schmid (2000) points out that conceptual shell nouns, together with factual and propositional linguistic shell nouns, form the "prototypical core of the whole class of shell nouns" (p. 188). Specifically, conceptual shell nouns are qualified as shell nouns in almost all aspects: firstly, they are 'original' nouns, not derived from or related to verbs or adjectives morphologically; secondly, they represent the exclusive way of shelling events and abstract relations as ideas; and thirdly, they turn physically observable events and abstract relations into mentally perceivable ideas, which is the precise sense of projection. Therefore, the mental shell nouns epitomise the metaphor of 'conceptual shells' which contain mental entities. This is demonstrated by the following example.

3-39 **The thought** that that man [Charles] will one day be King.

(Schmid, 2000, p. 188)

In example 3-39, the noun *thought* creates a conceptual shell and its shell content is provided readily in the following clause. The result of the connection between the shell noun *thought* and the underlined shell content is a temporary concept. The significance of conceptual shell nouns is that "the concepts they create are

not just concepts in the metalinguistic sense in which all shell nouns create concepts, but also from the perspective of the object-language"(Schmid, 2000, p. 188). Another feature of conceptual shell nouns is that they highlight the propositional content of ideas at the cost of backgrounding the fact that ideas tend to be conceived and experienced by 'experiencers'.

The use of *point* as a shell noun is a good starting-point because the use of *point* marks the transition zone between propositional linguistic shell nouns and conceptual shell nouns. Uses of the shell noun *point* in example 3-40 and 3-41 below can demonstrate its transitive use between these two types of shell nouns:

3-40 I repeat my *point* that the Agency would be acting illegally.

3-41 Except for the *point* that he may be acting unconstitutionally.

(Schmid, 2000, p.189)

Example 3-40 shows the linguistic use of *point*, while example 3-41 indicates the conceptual use of the same noun, which also demonstrates the 'multiple category membership' feature of shell nouns.

As for the other sub-category of mental shell nouns, psychological-state shell nouns allow speakers/writers to include a characterisation of the psychological state which they want to attribute to 'experiencers' in their shells for the ideas of the 'experiencers'. Three examples are given below to demonstrate the characterisation function of psychological-state shell nouns:

3-42 At any one of Lord McAlpine's five clubs [...] members can speak up for absent friends like Mr Botnar or Mr Asil Nadir in **the comfortable knowledge** that their views have been spoken out loud by a former treasurer of the Tory party on the Breakfast With Frost programme.

3-43 The current recession has finally punctured **the belief** that house prices can rise forever.

3-44 **The impression** that events are fast slipping from their control is confirmed by an order from the Ministry of Defence authorising the withdrawal of all military units from the streets of Moscow, (BBC)

(Schmid, 2000, p. 197)

The shell nouns *knowledge, belief* and *impression* in the three examples above differ from each other with respect to a dimension of imputed strength of commitment (Lyons, 1977, p. 794). In example 3-42, *knowledge* and its pre-modifier *comfortable* are used to attribute to the 'experiencers' the strongest commitment that the members' beliefs are true representations of past events. In example 3-43, the shell noun *belief* also portrays a fairly strong commitment to the idea that *house prices can rise forever*. In contrast, in example 3-44, the commitment imputed by *impression* is much weaker. In fact, the use of *impression* suggests the uncertainty of the truth of the shelled belief.

3.5.1.4 Modal shell nouns

As for modal shell nouns, the case is more complicated than other categories of shell nouns. Most elements in this group are related to possibilities either in the epistemic or concrete aspects. Several aspects are mentioned in Schmid's (2000) analysis: "ability, volition, permission, obligation, as well as possibility and necessity" (p. 89). As the shell contents of modal shell nouns represent either facts or events, it is not appropriate to address that modal shell nouns shell 'modalities'. The fact is that the speakers/writers use modal shell nouns to express their subjective judgments about the degree to which it is possible, probable or certain that the shelled facts are true or the shelled events will take place.

Obviously, the nouns in this category are related semantically to modal verbs which also express modalities. The point is that modal nouns are easier to handle than modal verbs, as the meanings of the former are less ambiguous and indeterminate than those of the latter (Palmer, 1990, pp. 20-22). For example, the verb *must* has a deontic reading and an epistemic reading in the examples below, respectively. If the utterances in these two examples are paraphrased with

modal shell nouns, the difference between the two readings emerges automatically:

3-45 Deontic: John **must** be home by ten; Mother won't let him stay out any later.

3-46 Epistemic: John **must** be home already; I see his coat.

3-47 Deontic: John has **the obligation** to be home by ten.

3-48 Epistemic: There is **a good chance** that John is at home already.

(Schmid, 2000, p. 232)

A general frame is suggested by Schmid for the category of modal shell nouns, consisting of a 'modality indicator' and a 'proposition'.

3-49 But it is **the possibility** that some of Iraq 's vast 10 billion Dollars nuclear weapons programme has escaped which provokes the most unease.

(Schmid, 2000, p. 236)

In example 3-49, the modality indicator is *the possibility*, and the proposition is the underlined fact.

On the other hand, the classifications of modal verbs are also helpful to divide modal shell nouns into different sub-categories. Based on the previous frameworks of modality regarding modal verbs, Schmid proposes that epistemic and deontic modality constitute the core of the modality domain expressed by modal verbs, which can also be borrowed as two categories of modal shell nouns. Furthermore, a new sub-category – 'dynamic modality' – was added to fit into Schmid's corpus data. Three examples below demonstrate the use of modal nouns in the three sub-categories.

3-50 The Granada move has also focused attention on **the possibility** that other companies will become targets for stake-building.

(Schmid, 2000, p. 237)

In general, epistemic modality focuses on “judgments about the possibility, probability or certainty that something is or is not the case” (Schmid, 2000, p. 235). Example 3-50 demonstrates the use of the shell noun *possibility*, which creates a neutral conceptual shell for the possible fact that *that other companies will become targets for stake-building*.

3-51 He was not embarrassed by the leak of this letter, but the Attorney-General told him **permission** to publish all his correspondence "might give him some difficulty”.

(Schmid, 2000, p. 246)

In Schmid’s (2000) definition of deontic shell nouns, the deontic modality they express should be connected to futurity and the necessity or possibility of acts of giving permission and laying an obligation (p. 249). In example 3-51, the noun *permission* is used to expressed the deontic modality of giving permission *to publish all his correspondence*.

3-52 It’s really **an opportunity** to concentrate on something different from work and home.

(Schmid, 2000, p. 251)

Example 3-52 demonstrates the use of dynamic modal shell nouns developed by Schmid himself. The dynamic modality expressed by the dynamic nouns is “at work in combinations of nouns and infinitive clauses [as used in example 3-52] that are used to indicate what can possibly, probably or necessarily happen under certain circumstances” (Schmid, 2000, p. 251). In example 3-52, the use of *opportunity* indicates that, in a certain situation, it is possible *to concentrate on something different from work and home*.

3.5.1.5 Eventive shell nouns

The term ‘eventive’ was borrowed from Generative Grammar to denote a situation in which the concept of events is distinguished from facts, which,

regarding the use of shell nouns, helps discriminate events from facts regarding the use of shell nouns. The term *event* itself is used "as a superordinate term for mental entities representing physically observable event which have a temporal duration" (Schmid, 2000, p. 261). The category, 'eventive shell nouns', refers to three types of events – activities, processes and states. An example is shown below:

3-53 **Sarah's next move** was to block what she thought were attempts to draft her daughter into munitions work.

(Schmid, 2000, p. 262)

In example 3-53, the shell-noun phrase *Sarah's next move* is used to characterise the underlined event.

3.5.1.6 Circumstantial shell nouns

The circumstantial shell nouns refer to situations, times, locations, manners of doing things and conditions for doing things. Members in this category can create conceptual shells that include circumstantial information about events (Schmid, 2000, p. 276). An example of the circumstantial use of shell nouns is demonstrated below:

3-54 We now have **a situation** where thrombolytic therapy is very widely accepted and very widely practised.

(from the BBC, Schmid, 2000, p. 279)

The propositional content of the *where*-clause in example 3-54 describes a state in which *thrombolytic therapy is very widely accepted and very widely practised*. The word *where* suggests a locative perspective on the passage, which indicates that the speaker is more interested in the circumstances of the events than in the events themselves.

What is worth noting here is that although some circumstantial nouns, such as *time* and *place,* can function as shell nouns in texts, "they mark a transition zone between shell noun uses and other uses of nouns" (Schmid, 2000, p. 14).

Schmid's reason for questioning the status of such temporal nouns (e.g. *time* and *stage*) and locative nouns (e.g. *place* and *area*) in the circumstantial nouns category is that, strictly speaking, they do not have extensional abstractness, especially for the locative nouns, which refer to places in the real world, and therefore are ontologically concrete. Secondly, it is doubtful that these nouns possess structure-inherent semantic gaps, which is another important semantic prerequisite for shell nouns.

However, these nouns can still be included in the class of circumstantial shell nouns because of three aspects. Take the shell noun *time* and *place*, for example. Firstly, the grammatical patterns '*time* + *when*-clause' and '*place* + *where*-clause' still evoke the link of experiential identity between what the shell nouns (*time* and *place*) and the clauses refer to. Both the shell nouns and the shell contents share the semantic dimensions 'temporal' and 'locative' respectively. Secondly, although *time* and *place* do not possess inherent semantic gaps, they are still intensionally unspecific in meaning, requiring additional information to be informative for communicative purposes. Thirdly, *place* and *time* have the potential to be used with the conceptual encapsulating function. The temporal dimension of *time* and three-dimensional space of *place* are regarded as "unstructured continua" (Schmid, 2000, p. 276). The locative and temporal expressions in the shell contents contribute to the division of these continua by selecting specific portions from them. In this aspect, *time* and *place* are more analogous to abstract nouns like *fact* and *aim*, which partition parts of "the amorphous abstract domain" (Schmid, 2000, p. 276), than to concrete GNs like *person*, *creature* and *object*, which do not help with conceptual division by referring to individuals rather than continua. Based on the three reasons sketched above, locative nouns (e.g. *time*) and temporal nouns (*place*) are still regarded as circumstantial shell nouns.

3.5.2 Semantic framework of discourse-based signalling nouns

The second framework of the SN-like phenomenon is provided by Flowerdew and Forest's (2015) work on DBSNs. Halliday's term 'grammatical metaphor' is adopted by Flowerdew and Forest to classify the origin of DBSNs into two parts. 'Grammatical metaphor' extends the definition of nominalisation into both nouns

from historical derivation, such as nouns derived from verbs, and nouns which are associated with other grammatical units, such as conjunctions, verbs and adverbs, in order to realise their meanings. Similarly, DBSNs are also divided into two types: one type of nouns which derive from verbs (e.g. *belief, argument*); and another type derived from nouns with other grammatical units. The second type expresses "abstract mental constructs ([e.g.] *idea*) and logical relationships ([e.g.] *fact, evidence, problem*)" (Flowerdew & Forest, 2015, p. 28). Overall, the concept of 'grammatical metaphor' provides a coherent theoretical position for the discussion of semantic types of DBSNs within a unified concept.

Another borrowed notion is from Halliday and Matthiessen's (2014) classification of four types of embedded and dependent clauses, which represents process types, and the construal of a process is a fundamental characteristic of the lexical specification of DBSNs. The terms in that classification are *act, locution, idea* and *fact*. Based on the classification, two modifications were made in Flowerdew and Forest's study (2015): firstly, *fact* is divided into *modal fact* and *non-modal fact*. Secondly, another sub-category, referred to as *circumstantial fact*, is added within the *fact* category of DBSNs. Altogether, the semantic classification of DBSNs was developed into six categories during the process of data analysis in their study, which shows significant similarities to Schmid's framework outlined above.

However, as the bases of the two frameworks are different (i.e. Schmid's work is based on Lyons' tripartite distinction of concreteness and abstractness, while Flowerdew and Forest's study is based on Halliday and Mattiessen's process type classification), some lexical items are allocated into different categories in the two frameworks. For example, *agreement, deal, contract, compromise* and *pact* are regarded as locution nouns in Flowerdew and Forest's study, while they are mental nouns in Schmid's work (Flowerdew & Forest, 2015, p. 29). Furthermore, as with Schmid's 'multiple category membership', a similar caveat is added to the semantic framework of DBSNs, which states that "there is no one-to-one relation between [DBSN] and semantic category" (Flowerdew & Forest, 2015, p. 33). Their examples for this 'multiple membership' of DBSNs are also from the multiple uses of the noun *point*:

3-55 Normally of course you make a **point** of having oxygen there.

3-56 In general, this is a general **point**, the ionic selectivity of such channels is much less than for voltage gated channels.

(Flowerdew & Forest, 2015, p. 33)

The noun *point* in example 3-55 is used as a circumstantial DBSN while in example 3-56, it is regarded as a locution DBSN instead. The difference between the opinions regarding this 'multi-membership' feature of SN-like nouns is that Schmid proposes that it is typical for shell nouns to have the feature of 'multiple category membership', while Flowerdew and Forest (2015) claim that most DBSNs fit into consistent semantic categories, though some DBSNs also appear in several semantic categories (p. 33). The six semantic categories of DBSNs are detailed in the following section.

3.5.2.1 Act nouns

The first category is act nouns. The DBSNs in this class represent events and actions in reality. Generally, these nouns derive from material processes. Some elements during this process, such as grammatical actor (e.g. human agents) and goals (e.g. outcomes), are denoted by these act nouns. An example is shown below:

3-57 When one photon of the appropriate energy is captured by 11-cis-retinal, its configuration is changed to 11-trans-retinal (Fig. 12.26). **This conversion** occurs in a few picoseconds.

(Flowerdew & Forest, 2015, p. 30)

The grammatical actor in example 3-57 is *11-cis-retinal* and the goal is to change the configuration of *11-cis-retinal* into *11-trans-retinal*. Both of these elements are denoted by the act noun *conversion* in the subsequent clause.

3.5.2.2 Idea nouns

Members in this category, idea nouns, usually denote the cognitive states and mental process of human beings. DBSNs in this category include normalisations

of mental process verbs, such as *assumption, prediction* and *belief*, as well as nouns which indicate complex mental constructs but which are not necessarily derived from particular verbs, such as *theory, idea* and *concept*. Two examples with uses of idea nouns are shown below:

3-58 It is **my belief** that there are a lot of other real ecologies all over the place but they happen not to be on this planet but on other planets other aqueous planets around in the galaxy.

3-59 and **the idea** behind it is to prevent the head lice from becoming resistant to one particular chemical.

(Flowerdew & Forest, 2015, p. 30)

In example 3-58, *belief* is the nominalisation of its corresponding mental process verb *believe*. Therefore, *belief* denotes the believing mental process of human beings. In example 3-59, although the DBSN *idea* does not have its cognate mental process verb, this DBSN still indicates complex mental constructs of human beings, represented by its following underlined lexical specification.

One feature of mental process clauses including 'idea' DBSNs is that this type of clause always includes two parts (Flowerdew & Forest, 2015, p. 30): a 'human' participant labelled the 'sensor'; and a grammatical element labelled the 'phenomenon'. This grammatical element represents what is "felt, thought, wanted or perceived" (Halliday & Matthiessen, 2004, p. 203), and typically acts as the lexical specification of the 'idea' DBSNs. However, the human participant in this case is not visible but implicit, as seen in the following example:

3-60 The **rationale** for encouraging the used market is that a high resale value will raise the prices buyers will pay for new goods.

(Flowerdew & Forest, 2015, p. 30)

In example 3-60, in the lexical specification part of the DBSN *rationale*, it is people who think of the reasons for the phenomenon of *a high resale value*. However, the agent of the action of thinking, i.e. the people, is not present in this sentence.

3.5.2.3 Locution nouns

DBSNs in this group project the content of verbal activities, such as *argument, question* and *discussion*, and signal the product of the verbal activities, such as *chapter, section* and *paper*. Examples are:

3-61 Should marketers maintain retail stores in low-income areas, ensure supplies for uses of limited amounts of the firm's product, or keep up locations in declining rural markets? **These questions** are difficult to resolve, because they involve individuals rather than broad segments of the general public.

3-62 **The following section** outlines the types of reaction observed . . .

(Flowerdew & Forest, 2015, p. 31)

In example 3-61, the locutionary nominal phrase *these questions* is used to represent the content of the previous verbal activity. In example 3-62, the nominal phrase *the following section* is used to signal the product of the verbal activity *the types of reactions observed*….

3.5.2.4 Fact nouns

The members of this category, fact nouns, report information about the world without framing the information with mental or verbal activities (Flowerdew & Forest, 2015, p. 31). Two examples are shown below to demonstrate the use of fact DBSNs in their corpus:

3-63 **A nice example** that Provotzki uses is that in Germany, in France and in Britain in the inter-war period, social democratic parties did hold political power, at er, at some er crucial points in time.

3-64 **The overall effect** of this step is to oxidise the FeIIT2MPyP and to make the system catalytic in FemT2MPyp.

(Flowerdew & Forest, 2015, p. 31)

In example 3-63, *example* is used as a fact DBSN to refer to some factual information about western European politics without addressing any mental or verbal activities. However, the adjective modifier *nice* is used to frame the information expressed in the following clause, which suggests that *example* itself alone does not frame the information but the phrase in which *example* is included does frame the information. Similarly, in example 3-64 the phrase *the overall effect* is used to refer to the whole result of *this step* in the following chemical reactions, instead of, for example, some part of the result caused by *this step*. The framing function of fact DBSNs is in line with Schmid's description of the use of neutral factual shell nouns in reality, which points out that these factual shell nouns do characterise related shell contents in various ways.

The categorisation for fact clauses in Halliday and Matthiessen's (2014) study is adopted here to divide the category of fact DBSNs. In the framework of fact clauses, there are four types: cases, chances, proofs, and needs. This classification is based on the types and degree of modality. The first type, cases, represents non-modalised facts, which is equated with the fact category in the semantic classification of DBSNs here. The other three types (chances, proof and needs), which exhibit some forms of modality, are included in the category of modal fact nouns below.

3.5.2.5 Modal fact nouns

Halliday's dichotomy of *modulation* and *modality* is used to name two sub-classes of modal fact DBSNs:– the former refers to "doing with permission and obligation", and the latter with "possibility and doubt" (Flowerdew & Forest, 2015, p. 32). The reason for separating fact nouns into fact nouns and modal fact nouns is that in academic discourse, modal DBSNs are typically used in hedging, which needs to be a separate category for further investigation. The examples below show the use of modal fact DBSNs in these two sub-classes:

3-65 The organization does not only have a **responsibility** to make profits, but it also has a **responsibility** for a fair distribution of profits and other outcomes of the organization among all persons.

3-66 It is therefore a **possibility** that Harris had to apply a higher dose of quarts and to compensate for increased adsorption.

(Flowerdew & Forest, 2015, p. 32)

Responsibility is used twice in example 3-65 to demonstrate the obligation or modulation of the organisation; while *possibility* is obviously used for addressing modalisation or possibility of the following underlined lexical specification in example 3-66.

3.5.2.6 Circumstantial facts nouns

For the same reason as the inclusion of modal fact DBSNs, the sixth category, circumstantial facts nouns, is also added in Flowerdew and Forest's framework of DBSNs, as members in this category also occur highly frequently in academic discourse, especially in the hard sciences, where they represent circumstantial facts which present information about where, when, how, and why something is presented (Flowerdew & Forest, 2015, p. 32). The example below illustrates the use of *way* as a circumstantial facts noun:

3-67 An alternative **way** of restraining this type of structure is to have some form of buttressing system where you have inclined support.

(Flowerdew & Forest, 2015, p. 32)

In example 3-67, *way* is used as a DBSN to refer to the underlined information about how *this type of structure* is restrained in this example.

3.5.3 Summary of the two semantic frameworks of the SN-like phenomenon

The two frameworks detailed above only attempted to differentiate nouns into groups which share 'family resemblance'. There are various semantic features distributed amongst the members of a group but not all are shared by all of them. Some members may also share certain features which are typical of other classes. This is in line with what Schmid called the 'multiple membership' of shell nouns. Therefore, the boundaries between various categories of SN-like nouns are fuzzy. What the semantic frameworks can assure us is that "for each class there is a

unifying concept and a central core of prototypical members, while the remaining members conform to the prototype more or less" (Francis, 1986, p. 9). This view is shared by the present study, which treats such semantic frameworks as guiding tools for the recognition of SNs in use. As can be seen from the two frameworks of the SN-like nouns given above, there are, to different degrees, certain overlaps (see table 7):

Table 7 Comparison of three semantic frameworks of the SN-like phenomenon

Schmid (2000)	Flowerdew and Forest (2015)	
factual	fact	
mental	Idea	
linguistic	locution	utterance[20]
		text
eventive	act	
modal	modal facts	
circumstantial	circumstantial facts	

As can be seen from Table 7, these two frameworks of SN-like nouns have a similar set of groupings. What is worth noting here is that the locution nouns category of DBSNs comprises two types – text nouns and utterance nouns – while the linguistic shell nouns category only includes utterance nouns. This thesis study adopts the locution nouns category as a type of SNs, the reason being that text nouns as a type of locution nouns, such as *chapter, section* and *paragraph*, can still be used as lexical cohesive devices in texts when they refer to other parts in the same text.

Based on the comparison demonstrated in table 7 as well as the categorisation of GNs (see table 4), the semantic framework of SNs in the present study is developed and illustrated in the following table:

[20] Utterance nouns and text nouns are two categories of anaphoric nouns in Francis' (1986) study. The former is similar to the definition of linguistic shell nouns while the latter is defined as nouns "simply refer to stretches of preceding segments in text" (Francis, 1994, p. 93).

Table 8 Semantic framework of SNs in the present study

name	definition
factual nouns	report information about the world, mainly states of affairs and facts.
mental nouns	represent cognitive states and mental process, as well as the products of the process.
linguistic nouns	represent illocutionary acts semantically, project the content of verbal activities and signal the product of the verbal activities.
eventive nouns	refer to activities, processes and states in the reality.
modal nouns	represent modality, modulation and dynamic modality.
circumstantial nouns	refer to situations, times, locations, manners of doing things and conditions for doing things.
first-order entity nouns	refer to experience involving persons, organisms and objects.

As for the conformity of the names for each category, the structure of each category name is 'an adjective + nouns', except for the last category which derives from the categorisation of GNs and the definition of first-order entities in Lyons' (1977) tripartite taxonomy of experiences. Furthermore, the linguistic nouns category of SNs in the present study equate to the locution nouns category of DBSNs, as mentioned above, which include both the text nouns category of A-nouns and the linguistic nouns category of shell nouns. As with the grammatical patterns sketched above, this semantic framework of SNs is also used as a tool to contribute to the identification of SNs rather than as a determining criterion.

3.6 Summary

Overall, the previous studies mentioned above have two main approaches to the SN-like phenomenon: one is to place more weight on the structural features of SNs and use grammatical patterns to test whether a lexical item is included in the class of SNs (e.g. Hunston & Francis, 1999; Schmid, 2000); the second approach is to place more weight on the semantic and discourse features of SNs and tend to use semantic features to identify the SN status of a lexical item (e.g. Francis, 1986; 1994; Ivanic, 1991; Flowerdew & Forest, 2015). As the nature of lexical cohesion is context-sensitive, the present study adopts the second discourse-based approach as the main method for exploring the SNs in the corpora. At the

same time, grammatical patterns and semantic frameworks act as aids to identify the SNs.

As for the other features of SN-like nouns mentioned above, the extended reference of DBSNs also contributes to differentiating SNs in use, though not all SNs possess this feature in the present study. Furthermore, the concept formation function of shell nouns is a key tool for distinguishing SNs from other lexical cohesive devices, because only SNs can both act as a container which encapsulates much more complex chunks of information into a single noun or a nominal phrase, and at the same time conceptualise the information with 'thing-like qualities'. This is a unique and economical way of creating lexical cohesiveness in texts. In addition, the semantic generality of GNs as superordinates of the major noun classes also helps make SNs a separate type of lexical cohesive device, though other superordinates in the hyperonymy category (see chapter 2) also share the unspecificity feature to various extents.

In terms of the membership of SNs, except for GNs, which do mention some of the concrete nouns, the majority of the terms for SN-like nouns mentioned above do not include concrete nouns or nouns denoting first-order entities. The category of SNs in the present study chooses to include concrete nouns because such nouns can also function as lexical cohesive devices in texts with unspecific semantic gaps which need to be filled by their lexical specifications in the same text. Overall, the category of SNs in lexical cohesion includes the SN-like nouns mentioned in this chapter, especially the GNs, the shell nouns and the DBSNs.

Chapter 4 Collocation

4.1 Introduction

As mentioned in chapters 1 and 2, collocation is an important lexical cohesive device, but has also been researched in other areas or used under different contexts. There is necessity for clarifying the definition of collocation in the scope of lexical cohesion to avoid misunderstanding. Furthermore, the detection of collocational relations in analysis involves more subtleties compared with that of other lexical cohesive categories. Therefore, this collocation category of lexical cohesion needs chapter-length elaborations to differentiate the use of collocation as a type of lexical cohesive device from that of collocation in other areas.

This chapter focuses on the nature of collocation. As Bahns (1993) points out, collocation is a cover term for various phenomena (p. 57). The crucial goal of this chapter is to critically evaluate previous definitions of collocation and the corresponding approaches to these definitions, and then develop an operational definition of collocation and a tailored framework of collocational relations for this study. In order to achieve this goal, firstly, different definitions of collocation will be discussed, in particular 'coherence collocation' and 'neighbourhood collocation'. Secondly, based on different definitions, three main approaches to collocation analysis will be analysed and their consequent classifications of collocation subject to critical analysis: the psychological approach, the Firthian approach and the phraseological approach. This will be followed by the developed classification of collocation in the present study according to the critical evaluation of classifications of collocation in previous studies. Finally, based on these discussions and critical analysis, the operational definition of collocation and approach to collocation analysis in this study will be introduced.

4.2 Definitions of collocation

The only universal aspect of collocation in most definitions is their reference to types of syntagmatic[21] relation of words (Nesselhauf, 2005, p. 11). With his work

[21] According to Saussure (1916, cited in Rapp, 2002), there are two fundamental types of associations between words that correspond to basic operations of humans' brains: syntagmatic and paradigmatic associations. Syntagmatic associations refer to the relations between two

dating back to the 1930s, Palmer has been regarded as the pioneer of studies of collocation by corpus linguists (Kennedy, 2014; Mitchell, 1975; Sinclair, Jones & Daley, 2004), since he discussed frequently occurring word combinations (Palmer, 1933; 1938). His simple definition of collocation – word combination – is exactly the same as that which current corpus linguists accept (Léon, 2007, p. 10). Specifically, Palmer (1933) described collocation as "a succession of two or more words that must be learnt as an integral whole and not pieced together from its component parts, [such as] *to strike while the iron's hot*" (title page). However, his definition only offered a general concept of the nature of collocation without the presence of any classification criteria. Despite his initial work, it remained unclear what particular kinds of word combination could be regarded as collocation. Therefore, since the publication of Palmer's ideas, several interpretations have been developed located within the framework of his general definition of collocation.

The notion of collocation was formally explained by Firth in the 1950s. Firstly, collocation was included in his term, 'meaning by collocation' (Firth, 1957), which is described as follows:

> the statement of meaning by collocation and various collocabilities does not involve the definition of word-meaning by means of further sentences in shifted terms. Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of *night* is its collocability with *dark*, and of *dark*, of course, collocation with *night*. (Firth, 1957, p. 196)

What can be implied from Firth's definition of 'meaning by collocation' is that the two lexical items do not juxtapose randomly but follow an order of mutual expectancy (Firth, 1968, p. 181), and the focus of collocation in Firth's framework is words such as *night* and *dark*. Secondly, the statement that 'meaning by collocation' lies at the syntagmatic level implies that the analysis of collocation

(continuation of footnote 21 on previous page) words if they co-occur in a certain language more frequently than expected from chance and if they have different grammatical roles in the sentences in which they co-occur. For example, there are syntagmatic associations between the two words in the following pairs: *coffee – drink, sun – hot,* or *teacher- school* (Rapp, 2002, p. 1).

should occur at the textual level, as the meaning of one lexical item depends on co-text lexical items, which reveals the cohesive nature of collocational relations. Furthermore, 'meaning by collocation' is not directly related to the conceptual approach to the meaning of words.

Firth's (1968) formal definition of collocation was presented in his later work, which described collocation of a word as "statements of the habitual or customary places of that word" (pp. 181–182). However, he did not propose any specific criteria to define the meaning of 'habitual or customary'. Researchers in different linguistic fields have been trying to interpret and develop Firth's definition of collocation for application in their own research.

Within the context of discourse analysis, Halliday and Hasan (1976) defined collocation as "a cover term for the cohesion that results from the co-occurrence of lexical items that are in some way or other typically associated with one another, because they tend to occur in similar environments" (p. 287). They provided some examples to instantiate their definition, such as *letter, stamp,* and *post office, or hair, comb, curl,* and *wave* (Halliday and Hasan, 1976, p. 287). This definition is somewhat vague in nature, particularly as no specific criteria are provided for collocation identification in use. However, they emphasise the fact that the characteristic of collocational relations is the tendency to co-occur in similar environments. The reason for this co-occurrences of items is that "they describe things or happenings that occur in similar situations" (Tanskanen, 2006, p. 60). This rationale indicates a conceptually based association between lexical items. This conceptual definition is different from Firth's 'meaning by collocation' which, as mentioned above, emphasises that meaning by collocation is not directly related to the conceptual approach to the meaning of words.

In terms of corpus linguistics, collocation has been defined as "words that actually co-occur with a word in text (e.g. the co-occurrence of *my, this*, and *a* with *letter*)" (Xiao, 2015, p. 107). Scott (2010) differentiated Halliday and Hasan's definition of collocation in discourse analysis and Xiao's definition of collocation in corpus linguistics by calling the former 'coherence collocation' and the latter

‘neighbourhood collocation’, but he also stated that the research for collocation has never differentiated these two definitions at a satisfactory level (p. 121).

Sinclair (2004) further differentiated the two definitions of collocation based on the following two angles: in the sense of lexicography, collocates are next to the node (the given word, see p. 113 below) within the span of four to six words typically; while in the sense of cohesion, collocates can appear within a larger stretches in the text, since cohesion occurs at the textual level. That is to say, the principle difference between lexicographical collocation (i.e. neighbourhood collocation) and cohesive collocation (i.e. coherence collocation) is that, in general, the distance between cohesive collocates is further than lexicographical collocates . For example, looking at Firth’s (1957) famous collocation *night* and *dark*, “if they occur next to each other we deal with a lexicographic collocation, but if they appear separately in a longer text we deal with cohesion collocation” (Cerban, 2010, p. 2). This differentiation of definitions is important in this study, as the approaches adopted in this thesis derive from one of the two definitions, i.e. cohesive collocation.

In terms of practical analysis of collocation, it is difficult to use statistical tools to measure coherence collocation, as its identification requires a conceptual interpretation, or as Scott (2010) commentated, “[detecting] coherence collocation is very tricky, as once we start looking beyond a horizon of about 4 or 5 words on either side, we get so many words that there is more noise than signal in the system” (p. 121).

By contrast, computational methods are used pervasively in research for neighbourhood collocation. Sinclair even developed terminologies for collocation in this sense:

> we may use the term **node** to refer to an item whose collocations we are studying, and we may then define a **span** as the number of lexical items on each side of a node that we consider relevant to that node. Items in the environment set by the span we will call **collocates**. (Sinclair, 1966, p. 415)

In Sinclair's description, lexical items could be any words that are considered relevant to the node, yet not all words in the span should be considered as collocations of the node. For example, *buy* and *cheap* are considered as a pair of frequent collocates in Sinclair's study, but the two items are habitually separated, as in the following examples:

4-1 What they did want was a spacious, easy-to-fly machine that was **cheap** to **buy**.

4-2 The resulting sounds are superb, and the product is **cheap** to **buy** and easy to use.

4-3 Even if you **buy** a **cheap** carpet to start with it's not gonna last.

4-4 The poor could **buy** a **cheap** lunch by the time he was sixteen.

(British National Corpus, 2007b)

In examples 4-1 and 4-2, the word between *buy* and *cheap* is *to*; and in example 4-3 and 4-4, the item between *buy* and *cheap* is *a*. it is not difficult to deduce that neither *to* or *a* are relevant to *buy* and *cheap* regarding their semantics. Therefore, *a* and *to* are not seen as collocates of *buy* and *cheap*.

Additionally, there is another main definition of collocation, which follows phraseological tradition. In general, collocation in this tradition is seen as a type of word combination (Nesselhauf, 2005, p. 14) with a collocation being defined as a phrase in which one of at least three criteria should be met – semantic non-compositionality, structural non-modifiability or structural non-substitutability (Manning & Schütze, 1999, p. 184). Specifically, the explanations for the three criteria are as follows (Manning & Schütze, 1999; Nesselhauf, 2005, p. 16):

Non-compositionality: there are two conditions. Firstly, the meaning of a collocation is completely different from the meanings of its compositional words (e.g. the idiom *kick the bucket*); secondly, there is a connotation or new aspect of meaning of a collocation which cannot be deduced from the compositional words

(e.g. the meanings of *white* in the phrases *white wine, white hair* and *white woman* are slightly different).

Non-modifiability: the compositional words in a collocation cannot be grammatically changed, no components can be omitted or new words cannot be added. This criterion especially applies to frozen expression like idioms. For example, *ugly* cannot be added before *frog* in the idiom *to get a frog in one's throat*, though *frog* goes with *ugly* in other contexts frequently; also *people* cannot be changed into singular form in the idiom *people as poor as church mice*.

Non-substitutability: the components in a collocation cannot be substituted by other words even if they are synonyms. For example, *do* cannot be replaced by *make or give* in the collocation *do a favour* even if in this context *do, make and give* represent similar actions.

The three criteria above reveal the definition of the phraseological collocation which means words that can "occur in a common grammatical unit and with a particular order" (Manning & Schütze, 1999, p. 185). Manning and Schütze (1999) suggested using terms like *association* and *co-occurrence* for the phenomenon that words have strong associations and tend to occur in the same context, this approach being related to the definition of coherence collocation. They also argued that if time and human resource are available, the quality of collocation analysis by manual work will be higher than computational analysis (Manning & Schütze, 1999, p. 185).

Within the broad scope of phraseological collocation, there are still several terms generated for the same definition, such as 'lexical bundle' and 'word cluster' in corpus linguistics; 'formulaic expression' in language education; and 'multiword unit' and 'n-gram' in natural language processing (Xiao, 2015, p. 107). However, as this current study is cohesion-oriented and discourse-based, this phraseological definition of collocation is excluded from the analysis of collocation as it falls outside the scope of this study.

Considering the three main different definitions of collocation mentioned above, i.e. coherence collocation, neighbourhood collocation and phraseological collocation, there are two key problems: firstly, the term collocation itself is not a fixed term for a single phenomenon; and secondly, the notion of collocation in relation to cohesion is also not precisely defined and still remains "notoriously difficult" (Bartsch, 2004, p. 65).

In order to effectively analyse collocation in the corpora used in the current research, this thesis needs to differentiate collocation in cohesion from collocation in other research areas. As a consequence of this requirement, an operational definition of collocation is derived for application in this study, which is mainly based on the definition of coherence collocation mentioned above. The reason is that, among the three main definitions of collocation, coherence collocation allows the collocates to cross clause boundaries and co-occur in larger stretches of text, which makes it possible for collocations to contribute to textual cohesion. This can be seen in the example below from Halliday:

4-5 A little fat man of Bombay
Was **smoking** one very hot day.
But a bird called a snipe
Flew away with his **pipe**,
Which vexed the fat man of Bombay.

(Halliday & Matthiessen, 2014, p. 649)

In example 4-5, *smoking* and *pipe* are in different clauses but they still create a collocational relation as they appear in the same activity – a man is *smoking* with his *pipe* (this activity-related relation is one type of collocational relation in this study, which will be mentioned in section 4.4.3). Based on this coherence collocation defined by Halliday and Hasan (1976), as well as adopting ideas from other two definitions, at this stage, a basic operational definition of collocation can be proposed for this study, i.e. the lexical relation between lexical items beyond the clause in which the lexical items tend to co-occur in similar contexts and semantically associate with each other.

As discussed above, there are at least three main definitions of collocation, and several research approaches have also been developed tied to these different definitions. Three main approaches will be discussed in order to support the process of generating a method of analysis for this study.

4.3 Previous approaches to researching collocation

In this section, three main research approaches will be introduced corresponding to the three definitions of collocation mentioned above: the psychological approach, the Firthian approach and the phraseological approach.

4.3.1 The psychological approach

The key basis of this approach is the emphasis on the strong connection between lexical items. This connection is related to a single kind of meaning of a lexical item, which is 'collocative meaning'. This type of meaning "consists of the associations a word acquires on account of the meanings of words which tend to occur in its environment (Leech, 1981, p. 17). This definition, with regard to the 'collocative meaning' of a lexical item and its associates, has been accepted by researchers adopting the psychological approach (Partington, 1998, p.16). Aitchison (2012) also stated that "word meaning is probably learned by noting the words which come alongside" (p. 10). In most cases, learners of a language are exposed to a lexical item associated with other lexical items. They try to interprete the meaning of the target lexical item by understanding the co-text which is made up of the target lexical item and its associative items. In this sense, meaning is regarded as a function in context (Partington, 1998, p. 16). Because of the strong tendency to occur together with other particular items, one lexical item will probably also be stored in the mental lexicon in the manner of a member of collocational pairs or groups, and this is verified by word association tests. Aitchison (2012) points out that collocation is the second most common response to stimulus lexical items in a test (p. 86).

However, there is a big gap between native speakers (henceforth NSs) and NNSs regarding the tightness of the associations between lexical items. For NSs, it is easy to find the associates with a particular lexical item in the mental lexicon; while for NNSs, the associations are much looser, which makes it more difficult

to navigate a closer associate with the particular item. Men (2018) provides an example to explain the difference: when asked to express the idea that a coffee is rich and dense, NSs can easily say *strong coffee*, but NNSs tend to struggle to choose the adjective between *strong* and *powerful*. This is because for NNSs, the association between *strong* and *coffee* is not as strong as that for NSs who have a life-time exposure to their mother language (Men, 2018, p. 16). It is this strong association between lexical items that helps predict the associates of a lexical item in texts, which then contributes to the creation of cohesion.

4.3.2 The Firthian approach

As the approach which is introduced below was developed by Firth himself and other researchers who follow this Firthian tradition, this approach is called 'the Firthian approach' or 'the frequency-based approach' (Nesselhauf, 2004). Related research in this tradition will also be discussed below.

4.3.2.1 The importance of text

Text is a constituent in a context of situation, which contributes to the statement of meaning. The reason is that text is the meaning being put into use. As mentioned in section 4.2 above, collocations are "actual words in habitual company" (Firth, 1968, p. 182). This aspect of Firth's definition of collocation is in line with the neighbourhood collocation definition. Firth also mentions that the collocational use of each word form is independent from the use of their lemmas. For example, the "collocations of *light* (noun, singular form) separate it from *lights* (n.s.) and *light* (adj.) from *lighter* and *lightest*. Then there are the specific contrastive collocations for *light / dark* and *light /heavy*" (Firth, 1968, p.180)

Firth makes this statement of the important role of text repeatedly in his works by claiming that words "stare you in the face from the text" (Firth, 1968, p. 195) and "the text […] can be said to have a physiognomy" (Firth, 1957, xii), not the words themselves.

4.3.2.2 Firth's methodology: restricted language analysis

Firth's idea is to treat combinations of words as linguistic units (Léon, 2007, p. 7). He used finite sets of texts to analyse idiosyncratic collocations without having

the intention to generalise the collocations. Focusing on the aspect of stylistics, he found that habitual collocations are not as significant as idiosyncratic collocations. As this study is discourse-based and the samples are also finite sets of texts in one register, the focus in this study is also on idiosyncratic collocations .

As Léon (2007) points out, ‘meaning by collocation’ is closely related to another key theoretical principle of Firth’s study of collocation which is called the “polysystemic approach of meaning” (p. 2). ‘Polysystemic’ in Firth’s sense refers to “multilevel and to ‘multistructural’, and is associated with restricted language” (Léon, 2007, p. 2). Firth claims that “linguistic analysis must be polysystemic. For any given language there is no coherent system which can handle and state all the facts” (Firth, 1968, p. 24), and “any given or selected restrcted language, i.e. the language under description is, from the present point of view, multi-structural and polysystemic” (Firth, 1968, p. 200). He suggests that a restricted language limits and circumscribes the field of linguistic investigation and is sufficient to state coherent grammatical structures. In particular, a restricted language has its own micro-grammar and micro-glossary (Firth, 1968, p. 106).

Collocation in the scope of restricted language is characteristic and contributes to refining the boundary of the field of linguistic investigation. Therefore, it is worthwhile to study collocation within restricted language for practice goals, such as foreign language teaching, dictionaries and translation. For example, restricted language is used in specific situations like industry, politics and so on. Each field of restricted language has its own vocabulary, grammar and style. The field of language teaching should also be divided into restricted fields and introduce related linguistic knowledge.

This restricted language approach is more practical and doable than the approach adopted by structuralists which views language as one system whole (Léon, 2007, p. 3). Here, Firth stressed the central role of text, again as he pointed out that the restricted language is exemplified by text in corpus (Firth, 1968, p. 112).

However, Firth does not mention the method of collecting texts to form a corpus, instead suggesting that linguistic theories should be put into test by language in use. Although researchers who follow the Firthian tradition, such as Sinclair (1991), have conducted a series of systematic computational studies of collocation with large corpora, Firth himself always focused on combinations of words, not a single word as a node collocated with other words within a particular span.

Based on Firth's general principle for the treatment of collocation and the previous arguments given in relation to cohesive collocation, this study also discards using computers as an analytical tool and focuses on idiosyncratic collocation and restricted language with manual analysis adopted.

4.3.2.3 Post-Firthian studies on collocation by British researchers

Following Firth's idea that collocation is a syntagmatic relation between lexical items in texts, substantial studies have focused on lexical items' co-occurrences in language in use, and further definitions of collocation have been proposed. There have been two main directions of research: one direction has examined the recognition and differentiation of lexical meaning by using collocation in a broader sense, this research encompassing other notions such as "colligation, semantic prosody and semantic preference" (Men, 2018, p. 17); the other direction investigates collocation in a narrower sense through frequency-based studies.

Text-oriented direction

Within the text-oriented research focus, collocation is seen as the patterning of language which describes "the occurrences of two or more words" within a short stretch of language (Sinclair, 1966; 1991, p. 170). This view of collocation is considered at the textual level, as the defining criterion is the occurrence of words in a stretch of text, while the levels of strength of the association between lexical items is not important in this case. Furthermore, this text-oriented research focus also confines the notion of collocation: the medium is *text* and the form is *two or more words co-occurring*. As can be seen from section 4.2 above, this direction follows the neighbourhood collocation tradition. Further, as mentioned above,

Sinclair (1966) has already defined the boundary of a short stretch of language – a **node** with a span referring to the number of collocates on each side of the node. Hoey (2000) went a step further through providing a technical definition of collocation in this sense: “words which occur within a few (six) words on either side of the [node] in naturally occurring spoken or written text” (p. 228). Therefore, in his definition, collocation is more related to lexicography with its short stretch in text. However, Hoey also emphasises studying collocation at the textual level, which is line with the coherence collocation.

Frequency-based direction

Another focus of collocation is “the relationship a lexical item has with items that appear with greater than random probability in its (textual) context” (Hoey, 1991, p. 5). This criterion for defining collocation refers to frequency of lexical collocates and concentrates on the strong association between collocates. “The higher the probability is, the more likely for a word combination to be a collocation” (Men, 2018, p. 19). However, this criterion itself cannot determine meaningful collocations in language in use (Men, 2018, p. 19) as some words can co-occur at high frequencies even if they do not have any semantic bonds. For example, the words *he* and *night* have 2132 frequencies of co-occurrences in the British National Corpus (2007).

To avoid such cases with the pure frequency-based approach of collocation, another criterion has been added, ‘grammatical well-formedness’, which means if two lexical items have a very distant grammatical relationship, they cannot be counted as collocates even though they occur within a short span (Kjellmer, 1994, p. xv). For example, *but too* is not collocational relation but *in April* is (Men, 2018, p. 19), as the former pair does not have a close grammatical relationship between the two lexical items, while in the latter pair, *in* and *April* form a typical ‘preposition + noun’ grammatical pattern. Collocation in this sense can be recognised easily, as based on the criterion of ‘grammatical well-formedness’ between two lexical items, the collocational relation between these two items is further identified by the frequency of co-occurrence.

At the same time, it is clear that significant collocations in this direction are identified based on statistical measures rather than context-specific meanings. However, the focus of this thesis study is meaningful collocations in language in use. Therefore, the definition and criterion in this direction are considered not suitable for the current study, but provide an approach to examining the significance of the collocates discovered in this study based on these two criteria, i.e. grammatical well-formedness between two lexical items and the frequency of co-occurrence between the two items. That is to say, if two lexical items have a relatively close grammatical relationship and higher frequency of co-occurrence in texts, the collocational relationship between these items is of greater significance.

Computer-based studies of collocation

Halliday (1966) added the paradigmatic dimension to collocation besides the syntagmatic dimension suggested by Firth. He gave the comparative example of *strong* and *powerful* to instantiate the two dimensions. As table 9 below shows, although both *strong* and *powerful* can collocate with *argument*, their meanings and usages are still different in general. For example, in contexts such as when collocating with *car*, *strong car* is not acceptable while *powerful car* is; by contrast, when the collocate is *tea*, *strong tea* is acceptable while *powerful tea* is not. From the comparison of the collocational usages between *strong* and *powerful*, it can be deduced that the paradigmatic relation of *powerful* and *strong* is not constant but based on which collocate they have the syntagmatic relation with. Halliday (1966) also argued that the linguistic system is inherently probabilistic, which is the basis for computational corpora studies on collocation, focusing on the words' mutual expectancy and tendency to co-occur in the vicinity of each other (Léon, 2007, p. 13).

Table 9 Collocates for *strong* and *powerful* (adapted from Halliday, 1966, p. 150)

	strong	powerful
argument	☑	☑
car	☒	☑

tea	☑	☒

Based on Halliday's development of Firth's notion of collocation, Sinclair (1966) developed the study of collocation in computational linguistics. As already mentioned in section 4.2, he proposed technical terms such as *node, span* and *collocates*. For example, in the stretch of text *He went back to the house. When he opened the door, the dog barked*, the node is *house*, and the words *went, back, to, the, when, he, opened* and *the* are all considered in the span as potential collocates.

Furthermore, collocation was divided into two types: 'significant and casual collocations'. It needs to be noted that sometimes 'collocation' was used by Sinclair (1991) to only refer to 'significant collocation' (p. 115) which was regarded as 'regular collocation' between two lexical items with greater co-occurrences than the respective frequencies of the separate two items (Susan & Sinclair, 1974). In the above example, *dog* and *bark* may form a significant collocation while *the* and *house* probably do not. The reason is that *the* is a frequent word by itself, while *barked* is not frequent by itself.

Sinclair's main contribution has been to bring in the use of corpora to study collocation. However, because Sinclair studied neighbourhood collocational relations via a computational approach (Xiao, 2015, p. 107), while the present study focuses on coherence collocational relations via a manual approach, this thesis study does not adopt his computational method of collocation research, but uses his corpus-based approach to investigating coherence collocational relations in two student corpora.

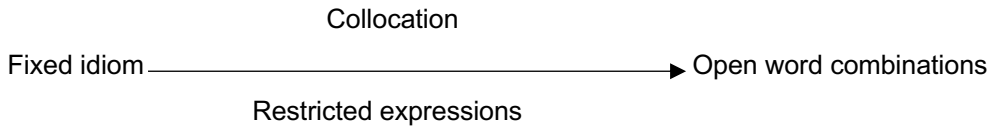
4.3.3 The phraseological approach

The phraseological approach follows the phraseological definition of collocation mentioned in section 4.2. In contrast to the frequency-based approach, the phraseological approach requires syntactic structure as one criterion for collocation formation (Nesselhauf, 2005, p. 17). The concern of this approach is the design of the evaluation criteria for collocation identification and the differentiation of collocation from other types of word combination, such as free

combination and fixed idioms. This approach was suggested by Russian phraseologists who developed different categories of word combination. For example, Vinogradov (1947, cited in Cowie, 1998, pp. 4-5) categorised word combination into three types. The first type is 'phraseological fusions' (or idioms, e.g. *spill the beans*) which are structurally fixed and semantically opaque; the second one is 'phraseological unities' (e.g. *blow off steam*) which extend the originally neutral meaning of words metaphorically; and the third type is phraseological combinations (also called 'restricted collocation' by Cowie (1981, p. 229), e.g. *meet the demand*) (Men, 2018, p. 20) which consist of two open-class words, one with the basic meaning and the other with figurative sense [22]. Here in this example, the meaning of *meet* is context-based, or 'phraseologically bound' in Vinogradov's term (Cowie, 1998, p. 5). This view is firstly similar to the notion of 'collocative meaning' in the psychological approach, as in both approaches, the meaning of a word in the collocation is related to the meaning of its collocate; and secondly, is related to the present study in the sense of analysing collocations based on the context, not on prefabricated phrases.

Another category Vinogradov did not cover is 'free word combination' or 'open collocation' (Cowie, 1981, p. 226). That is the main difference between Russian phraseologists' and Cowie's division. The former focused on the continuum from the fixed idiom end to collocation, while the latter concentrated on distinguishing restricted expressions from open word combinations (see figure 9 below).

Figure 9 Continuum from the fixed idiom end to open word combinations



As the object of this current study is second language learners' written work, the focus of this study is the difference between 'restricted expressions' and 'open

[22] "A figurative sense is a meaning that is derived from a primary sense by analogy (for example, personification), association (for example, metonymy), or similarity (for example, metaphor and simile)" (SIL International, 2003).

collocations'. The reason for this focus is that L2 learners may find it challenging to identify at what point they should combine words according to general grammatical rules and their basic meanings, or use words together as a whole expression (Howarth, 1996, p. 31). The reason for this potential challenge is that collocation is in the middle of the continuum from fixed idiom to free combination, and there is no clear boundary among the three types of word combination. Previous studies provided three criteria to identify collocation as a type of word combination which is different from the other two types of word combination: semantic transparency, specialised senses of one element and commutability/substitutability (Men, 2018, p. 21).

4.3.3.1 Semantic transparency

This criterion is applied for the differentiation between idioms and non-idiomatic expressions (i.e. collocation and free combinations) (Men, 2018, p. 21). It refers to whether the meaning of the whole combination is made up of the meanings of the individual lexical items. For idioms, the expression's whole meaning is different from the meanings of individual component items. For example, *kick the bucket* as an idiom means *to die* (OED Online, 2019). This meaning cannot be deduced just from the meaning of *kick* or *bucket*. However, for collocation and free combination, the whole meaning is not opaque and can be understood as the combination of the meanings of individual items. For example, the whole meaning of the collocation *commit a crime* and the free combination *control the crime* is easy to deduce from the combination of meanings of their individual lexical items – *commit* + *crime* and *control* + *crime* respectively (Men, 2018, p. 21). Therefore, this criterion is an effective means of distinguishing collocations from idioms.

4.3.3.2 Specialised senses of one element

This criterion is used for distinguishing collocations from free combinations. Collocation is one type of lexical combination. For collocation, specialised meanings must be included in either of the lexical item members. Take the structure of 'verb + noun' as an example. The semantic specialisation of the verb has several types: figurative senses (as *pay* in *pay one's respects*, *adopt* in *adopt a policy*), technical senses (as *obtain* in *obtain a warrant*) and de-lexical senses

(as *make* in *make a decision*) (Men, 2018, p. 19). On the other hand, for free combinations, the lexical members of a free combination have literal meanings (e.g. *bake bread, cut cheese*) (Men, 2018, p. 21). However, one thing should be noticed here: even in the range of 'restricted collocation', there are different levels of collocability. Examples are given below regarding the structure of 'verb + noun':

4-6 figurative: *assume a role*

4-7 de-lexical: *give emphasis to*

4-8 technical: *obtain a warrant*

(Howarth, 1998, pp. 169-170)

Although in examples 4-6 and 4-7, the verbs *assume* and *give* are used in their specialised meanings, in example 4-8, *obtain* is used in its original meaning — *get*, while the whole meaning of the phrase *obtain a warrant* is used in any technical texts. Therefore, this classification of semantic specialisation complicates the whole criterial system. Also, this criterion rules out other potential collocational relations between lexical items, such as *commit a crime*. The whole meaning of this collocation is based on the literal meanings of individual lexical members, i.e. the literal meanings of *commit* and *crime* respectively.

As the present study adopts a discourse-based approach to the identification of collocational relations in texts, the meanings of lexical items in a collocational relation depend on the contexts in which these items occur, which does not require the inclusion of specialised meanings used in either of these lexical items. Furthermore, as mentioned above, this criterion may exclude some potential collocational relations, while this thesis aims to investigate all potential collocational relations in the current corpora. Based on the two reasons made above, this criterion, therefore, is not applicable to the identification of collocation in this thesis (Men, 2018, pp. 21-22).

4.3.3.3 Commutability/Substitutability

For free combinations, each lexical element can be replaced by others without changing the sense of the whole phrase, while for collocation, there are more restrictions for substitution. This difference is related to their commutability with other elements (Aisenstadt, 1979; Cowie, 1992; Howarth, 1996, 1998). Commutability is syntactic flexibility of lexical items (Gries, 2013, p. 138). Aisenstadt (1979) provided two examples to explain the concept of 'restricted commutability' and tried to delimit collocations based on the 'restricted commutability' of the lexical items in a collocational pair:

4-9 **shrug** one's shoulders
shrug something off
shrug something away
shrug **one's shoulders**
square **one's shoulders**
hunch **one's shoulders**

4-10 make a **decision**
take a **decision**
have a **look**
give a **look**
take a **look**.

Aisenstadt (1979, p. 73)

In example 4-9, *shrug* and *shoulders* are the elements in collocations which cannot be replaced by other substitutes. In example 4-10, there is 'restricted commutability' among the verbs, i.e. *make* and *take*, in the collocations with *decision*, or *have*, *give* and *take* with *look*.

However, there are at least two reasons why commutability is not an effective criterion for collocation identification. Firstly, it depends on "the conceivability of a human mind" (Men, 2018, p. 22). For example, Aisenstadt did not make the concept of 'restricted commutability' clear when he used the example 4-9 above to explain the arbitrary restrictions on commutability. He assumed that the commutability of *shrug* is restricted to the elements like *one's shoulders, sth. off*,

and *sth. away*, and it is probably the case that *shrug* has limited collocates; however, for *one's shoulders*, there are much wider options for collocates, such as *straighten one's shoulders, wash one's shoulders, look at one's shoulders, rub one's shoulders, scratch one's shoulders* and so on (Nesselhauf, 2005, p. 27). Secondly, 'restricted commutability' can also happen in free combinations. For example, in the free combination *wash the glass*, if *wash* is replaced by another verb like *clean*, the original sense has been slightly changed; this is also the case for substituting the noun *cup* for *glass* (Men, 2018, p. 23). Therefore, there must be other factors that decide the 'restricted commutability' of collocations to differentiate them from free combinations.

There is a notion called "the given sense" in Cowie's (1992, p. 5) commutation tests for the demarcation of restricted collocation. This means that, for example, in *shrug one's shoulders*, substituting verbs should follow the meaning of the verb *shrug* which is 'the given sense' of this collocation. In this collocation, probably *shrug* is the only verb. The test itself is used when the verbs are assessed regarding whether they are in the same sense collocating with a 'given noun' (e.g. *shoulders*). For example, verbs are commutable in *abandon/give up a cherished principle,* while in *run a deficit,* there are no other commutable verbs to replace *run* (Men, 2018, p. 23).

A comprehensive classification of collocations on the basis of commutability was established by Howarth (1996, p. 102) in his categorisation of 'verb + noun' collocations from the most free to the most restricted (i.e. from Level 1 to Level 5), as summarised in table 10.

Table 10 Howarth's categorisation of collocations into five levels of restrictedness (Howarth, 1996, p. 102; Men, 2018, p. 23)

	verb	noun	examples
L1	some restriction	free substitution	*adopt/accept/agree* to a *proposal/suggestion*, etc.
L2	some substitution	some substitution	*introduce/table/bring forward* a *bill/an amendment*

L3	some substitution	complete restriction	*pay/take heed*
L4	complete restriction	some substitution	*give* the *appearance/impression*
L5	complete restriction	complete restriction	*curry favour*

The scale from level 1 to level 5 is a continuum with one end involving only some restriction on one element and the other end involving complete restrictions on both elements. The restrictions are based on the number of synonyms that one or two elements can have. For instance, at L1, only verbs have restrictions regarding the limited number of synonyms that they take, while at L5, both the verbs and nouns have no substitutions, which means they have complete restrictions.

Nevertheless, the introduction of synonyms makes the differentiation between collocations and free combinations more complex, as both commutability and synonyms rely on the conceivability of a human mind (Men, 2018, p. 24). For example, the combination *pay heed* has been regarded as a collocation at L3 in which *heed* has no other substitutes. However, according to the Oxford English Dictionary, *heed* has at least one synonym *attention* and *pay attention* is also appropriate in English (OED Online, 2019g). This indicates that making the judgment of the number of synonyms is subjective, not based on facts but more on personal feelings and knowledge (Men, 2015, p. 24). Even in Cowie's commutation test for verbs, there is high possibility that no synonyms can be found to substitute for the given verbs. For example, for the free combination *drink one's tea* (Cowie, 1994, p. 3169), it is hard to find many synonyms for both the verb *drink* and the noun *tea* respectively (Nesselhauf, 2005, p. 28). It can been seen that commutability is not a suitable criterion to differentiate collocations from free combinations.

In this section, collocation has been firstly introduced from the psychological perspective in which collocation is regarded as representing strong psycholinguistic lexical associations. Collocation in this sense focuses on the 'collocative meaning' of a word which is related to the meanings of the word's collocates, and this emphasises the strong semantic association between the

word and its collocates. Such a strong tendency of words to occur together results in the words being stored with their collocates in the mental lexicon.

The text-oriented direction of the Firthian approach concentrates on linear co-occurrence of lexical items, and ignores the syntactic and semantic aspects of collocations (Greenbaum, 1970, p. 10) which are more important in studying collocation in relation to its cohesive function. Furthermore, the definition of *span* confines the collocates within a relatively short stretch of texts (e.g. four words), which is not applicable to the analysis of collocates in longer segments of texts. For instance, in the following examples 4-11 and 4-12, the collocation *collect stamps* would not be counted as representing collocation in the text-oriented sense of collocation, as the component words occur beyond the span of four words on each side of the collocated items.

4-11 They **collect** many things, but chiefly **stamps**.

4-12 They **collect** many things, though their chief interest is in collecting coins. We, however, are only interested in **stamps**.

(Greenbaum, 1970, p. 11)

The frequency-based direction of the Firthian approach focuses on the identification of significant collocations via statistics of frequencies, which ignores the phraseological sense of collocations, such as *collect stamps* in examples 4-11 and 4-12 above. In addition, the Firthian approach, in general, studies collocation as a pure linguistic phenomenon without the demarcation between collocation and other types of word combinations (e.g. free word combination and idioms). These drawbacks of this approach are also the disadvantages of using the computer-based method to investigate collocations. In order to take a rich detailed and thorough look at collocation, as already stated, this study determined to use manual analysis.

To solve the problem of demarcation between collocation and other word combinations, the phraseological approach provides some defining criteria. However, there is still not a clear borderline between free combinations and

collocations (Men, 2018, pp. 23-25). What can be utilised from this approach is the idea that the collocation lies in the continuum between fixed idioms and open word combinations. Idioms which have fixed lexical meanings and grammatical structures are excluded in this current study, as the componential words in idioms cannot be separated into different clauses in texts, and cannot contribute to cohesion beyond the clause.

Furthermore, the concept of 'restricted collocation' is enlightening in regard to judging whether the relations between lexical items are collocational. Although the criteria above in general are not satisfactory in certain respects, the criteria of semantic transparency and commutability, for example, are still useful when differentiating 'restricted collocation' from some types of free word combination. For example, when at least one of the elements in a relation has a phraseologically bound meaning, this relation is probably 'restricted collocation', such as *adopt* in *adopt a policy* (Men, 2018, p. 21). As for commutability, if there are no other substitutes having the given sense, then the relation between the original lexical items is also 'restricted collocation', such as in the previously given example, *shrug one's shoulders* (Men, 2018, p. 23). Although some delexical verbs can also be used in collocations, such as *give* in the collocation *give evidence* meaning *provide* (Men, 2018, p. 22), delexical verbs (e.g. *do, give, have, make, take* and *get*) are not included as collocations in this study, as they are considered too general in meaning and too frequent in texts to contribute to cohesion in texts.

As can be seen from the above review, each approach to collocation identification in texts has its pros and cons. This study adopts an integrated approach to analysing collocation based on a range of aspects derived from the approaches above, where they contribute to the cohesive function of collocation beyond the clause. However, only taking an integrated approach is not enough to identify collocation in real world practical analysis. A classification of collocational relations is necessary to ensure that the analysis is both operational and replicable.

4.4 The classification of collocation

In the scope of 'coherence collocation', only limited studies have been conducted to develop detailed frameworks for collocational relations relating to lexical cohesion analysis. The reason for this is that in this definition, collocation is regarded as the most problematic category of lexical cohesion as "[this] category has been notoriously difficult to define, so much so that it has often been excluded from analyses" (Tanskanen, 2006, p. 60). The aim of the present study is to investigate lexical cohesion in detail in Chinese postgraduates' writing on two MA programmes. Therefore, as one of the two main categories of lexical cohesion (another one is reiteration according to Halliday and Hasan's (1976) model), collocation is of necessity included within the investigation.

Having considered the need for developing the preliminary definition of collocation proposed in section 4.2 into a more operational definition of collocation for the current study, and the lack of a comprehensive framework of collocation relations which can be readily used in the current corpora, the following sections now examine three related classification models of collocation originating in three studies, in order to fulfil such need.

4.4.1 Halliday and Hasan's model of collocation

Inspired by Firth's (1957) concept of collocation as 'mutual expectancies', Halliday and Hasan (1976) defined collocation in cohesion as "lexical items that regularly co-occur" (p. 284). Based on this loose definition, Halliday and Hasan have classified collocation into three categories:

Table 11 Categories of collocation in Halliday and Hasan's model (adapted from Halliday & Hasan, 1976, p. 285)

categories	examples
opposites	complementaries, e.g. boy – girl; antonyms, e.g. like – hate; converses, e.g. order – obey
ordered and unordered sets	ordered: north – south unordered: basement – roof

Non-systematic semantic relation	blade – sharp

It can be seen from the table that the first two categories are more systematically semantic-related while the last category is more intricate to define and identify in analysis as it has no clear criterion to follow but only the vague rule that the lexical items are semantically associated in similar environments.

Therefore, to avoid the ambiguities involved in defining and identifying the last category, in later work of Halliday's (1985) study on cohesion, he moved some categories previously found under collocation to categories under synonyms which includes synonymy "proper" (*sound – noise*), superordinates (*blackbirds – birds*), hyponymy (*tree – oak*), meronymy (*tree – trunk*), co-hyponymy (*oak – pine*), co-meronymy (*trunk – branch*) and even antonymy (Tanskanen, 2006, p. 36). This modification by Halliday involves the first two categories of collocation in Halliday and Hasan's model (i.e. opposites as well as ordered and unordered sets) being moved from collocation to synonyms as two new categories (i.e. antonymy and meronymy) in Halliday's new model. As a consequence, the scope of collocation becomes smaller compared with Halliday and Hasan's original framework. The definition of collocation was also modified by Halliday to refer to "the relationship of the lexical items which depend on the association between them" (Tanskanen, 2006, p. 36). Halliday further pointed out the cohesive function of collocation as "one of the factors on which we build our expectations of what is to come next" (Halliday, 1985, pp. 312–313). This current study will combine this definition with the previous definition in section 4.2, and generate a new definition as follows: collocation is seen as:

the lexical relation between lexical items beyond the clause, which are expected to occur in similar environments and depend on the semantic association between them.

However, there remains a need to develop some categories of collocation to make the analysis as replicable and objective as possible in a similar manner to which Tanskanen (2006) enacted in her analysis of several types of discourse.

4.4.2 Martin's framework of collocation

Martin (1992) followed the definition of 'mutual expectancy' discussed above (see section 4.3.2) but pointed out that the range of 'mutual expectancy' needed to be clarified. He opposed the idea provided by Sinclair (1988) that it is acceptable to treat the whole text as 'mutually expectant'. The reason for his disapproval is that if this is acceptable, then collocation can no longer be an effective criterion for lexical cohesive analysis anymore. Therefore, it was necessary to set the boundaries of 'mutual expectant' relationships.

Martin (1992) borrowed Halliday's notion of 'field' and redefined it as "sets of activity sequences oriented to some global institutional purpose" (p. 292). Several examples were used to explain this adapted definition, such as "linguistics, tennis, cooking, […] politics, religion and so on" (Martin, 1992, p. 292). A system with three aspects to break down 'activity sequences in a field' was also developed, and these comprise the three main semantic relations in his model:

Table 12 Martin's (1992, p. 292) system of activity sequences

aspects of activity sequences	examples
taxonomies of *actions, people, places, thing* and *qualities*	part/whole relations among *game-set-match*
configurations of actions with people, places, things and qualities, and of people, places and things with qualities	agent process medium structure: player-serve-ball
activity sequences of these configurations	player serve-opponent return-player volley

According to Martin's (1992) description, the network for lexical relations in his model of collocation is organised around the three aspects of "the contextual structure of activity sequences" (p. 293) mentioned in table 12; and the lexical relations are divided into three main categories. The first category is called 'taxonomically-based relations' which are related to the categories of reiteration in Halliday and Hasan's model, and about which there is a long tradition of research. As for the connection with collocation, according to the proposed concept of 'activity sequences in a field', Martin (1992) modified Halliday and

Hasan's collocation category and put forward two new terms – 'nuclear and activity sequence relations', which comprise the other two categories of lexical relations in his study.

What is worthy of particular mention in Martin's study is that 'taxonomic and activity sequence relations' overlap in some respects. For example, in a match, the two activities "player serves" and "opponent returns" are sequenced (Martin, 1992). Therefore, *player* and *opponent* demonstrate the 'activity sequence relation'. At the same time, *player* and *opponent* can act as antonyms in a semantic sense. There thus exist two relations within one pair of lexical items, a feature which substantially complicates analysis. Furthermore, the 'activity sequence relation' potentially connects almost every lexical item in texts, while potentially beneficial in identifying the reality of complexity of lexical relations in texts, also increases the difficulty for textual analysis (Tanskanen, 2006). Therefore, due to the complexity of determining sequential relations, in this study, 'activity sequence relations' are excluded. Instead, 'nuclear relations' will be the main collocation category here from Martin's model.

4.4.2.1 Nuclear relations – ideational collocation

In Martin's model, 'ideation' is part of the ideational system which is one of three metafunctions in Halliday's model of Systemic Functional Linguistics. 'Ideation' has been defined as being:

> … concerned with how our experience is construed in discourse. It focuses on sequences of activities, the people and things involved in them, and their associated places and qualities, and on how these elements are built up and related to each other as a text unfolds. (Martin & Rose, 2007, p. 73)

'Ideation' refers to more detailed interpretation of possible relations among semantic units realised through lexical items in Hasan's (1984) framework of lexical cohesion (including repetition, synonym, antonymy, hyponymy and meronymy); additionally, 'nuclear relations' are developed to capture the semantic relations under the heading of collocation in Halliday and Hasan's (1976) model. These 'nuclear relations' are those between the semantic units typically

realised through nominal groups, verbal groups and clauses, and are based on Halliday's notion of 'logico-semantic relations' of expansion which comprise elaboration, extension and enhancement (Martin, 1992, pp. 309-310). Furthermore, nuclear relations "reflect the ways in which actions, people, places, things and qualities configure as activities in activity sequences" (Martin, 1992, p. 309), which can be demonstrated by the example in table 13 below from the field of tennis with the structure of 'Process Medium':

Table 13 Activities in the field of tennis (Martin, 1992, p. 309)

Serve + ace
Smash + overhead
Put away + volley
Hit + winner
Net + passing shot
Intercept + volley
Drop + shot
Lob + return
Volley + winner

However, it is also pointed out that in the cohesive sense of collocation, 'nuclear relations' cannot be only explained by grammatical structure, such as Process Medium in the table above (Martin, 1992). For instance, the lexical items *serve* and *ace* from table 13 were shown to appear in the grammatical structure of 'Process Medium' within one clause, like *Ben serves aces*; these lexical items can also occur in different clauses, like *Ben serves … That's his fifth ace of the match*; in addition, they can also co-occur in other configurations, such as *Ben's* ***serve*** *produced very few* ***aces*** *today*. The three different usages of *serve* and *ace* demonstrate that 'cohesive collocation' (or coherence collocation) is different from 'neighbourhood collocation' which needs the collocates to be relatively close to each other. It also can be seen that verbs (processes) have fixed objects (media), which can be one type of collocation. As the focus of Martin's (1992) study is on unpacking the semantic relations involved in the umbrella of collocation (p. 309), nuclear relations cannot be explained thoroughly by

grammatical structure. In order to explain nuclear relations at a more abstract level, Halliday's (1985) general logico-semantic relations of expansion to clause, as well as nominal group and verbal group meanings were applied in Martin's study in which three types of logico-semantic relations, identified above, were discussed. These relations are briefly introduced below (Martin, 1992, p. 310):

Elaboration (=): the second clause adds more information about one of the elements in the first clause by restating, clarifying or refining the element. For example,

4-13 That clock doesn't go; it's not working. (restating)

4-14 She wasn't a show dog; I didn't buy her as a show dog. (clarifying)

4-15 Each argument was fatal to the other: both could not be true. (refining)

(Halliday 1985, p. 203)

Extension (+): the second clause adds new elements to the first clause. The added element can be an addition, replacement or an alternative. For example,

4-16 I breed the poultry and my husband looks after the garden. (addition)

4-17 I said you looked like an egg, sir; and some eggs are very pretty, you know. (addition)

(Halliday 1985, p. 207)

Enhancement (*): the second clause enhances the meaning of the first clause. The enhancement approach can be adding information about the time, place, manner, cause or condition. For example,

4-18 It's the Cheshire cat: now I shall have somebody to talk to. (time)

4-19 The three soldiers wandered about for a minute or two, and then quietly marched after the others. (manner)

(Halliday, 1985, p. 211)

With the aid of the three types of relations in clauses, nominal groups and verbal groups, ‘nuclear relations’ can be more systematically classified and identified as one type of collocation relation.

However, nuclear relations were not applied beyond the clause in Martin’s study, and further his analytical strategy for collocation itself would be more complicated to replicate in the present study. Therefore, it was decided to adopt only his concept of ‘nuclear relations’ for this study, in a similar manner to that adopted by Tanskanen (2006) in which a single category of collocation was developed based on the notion of ‘nuclear relations’.

4.4.3 Tanskanen’s framework of collocation

The basis for Tanskanen’s analysis is the imprecise definition of collocation given by Halliday and Hasan (1976), which stated that collocations occur when two items are related to similar things or events in a similar environment (Tanskanen, 2006, pp. 60-61). In order to make the definition of collocation more useful for conducting lexical cohesive analysis, Tanskanen (2006) classified collocational relationships into three categories:

1. Ordered sets
2. Activity-related collocations
3. Elaborative collocations

4.4.3.1 Ordered sets

The ordered sets were treated as the clearest and most systematic category among these three categories of collocation (Tanskanen, 2006, p. 61). Examples are colours, numbers, months, days of the week and so on (mentioned in section 2.2.4). This category is similar to Halliday and Hasan’s (1976) category of unordered/ordered set (see section 4.4.1). However, collocations in this category are not frequent in texts.

Examples 4-20 and 4-21 instantiate this category:

4-20 The working people of *today* are the pensioners of **tomorrow**; the single people of **today** were the children of **yesterday** and are the parents of **tomorrow**.

4-21 RG: . . .So, like, the term starts in **September** and runs through till **January**, when we have Spring festival, which is the Chinese New Year. Now schools and universities will close for three weeks, and that is a particularly cold time of the year in the North. And then the term starts again, finishing at **the end of June**.

(Tanskanen, 2006, p. 61)

This study will not include analysis of this category in collocation, as these relations can be placed into other categories under reiteration, such as hyperonymy, hyponymy, and meronymy (see sections 2.3.3, 2.3.4 and 2.3.5, chapter 2). For example, the collocational pairs in the two examples above (e.g. *tomorrow – today, September – January*) are regarded as cohesive pairs in the category of meronymy in this study, because *tomorrow* and *today* are members of the category of time, and similarly *September* and *January* are the members of the category of months. Both of the two pairs represent part-whole relations of abstract entities, which are included in the category of meronymy in the present study. Furthermore, in order to avoid the vagueness and complex of the category of collocation, this study also narrows down the boundary of collocation, in a similar manner to that implemented by Halliday in his study (1985).

4.4.3.2 Activity-related collocation

In definition, this category includes semantically non-systematic relations. It is based only on associations between items and thereby resists simple systematic classifications and definitions. Therefore, there is no construction of clear cut rules or models which can guide the readers to identify which items are related regarding this category of collocation and which are not.

Based on previous studies, there are some tendencies of associations which can be summarised as rules for researchers to understand and classify these complex relations. The first one is Martin's (1992) nuclear and activity sequence relations which have already been discussed (see section 4.4.2). As also already

mentioned, Tanskanen simply focused solely on the ‘nuclear relations’. Examples are provided to instantiate this category as follows:

4-22 C: well I expect you don’t need **cyphers** during if by that you mean people who e: people who can **decode** yeah

4-23 . . .it means of course that they will have the utmost difficulty in paying for their **meals** in the refectories and that means that the refectories go into deficit if they can’t afford to **eat** here. . .

4-24 HK: . . .and of course this meant that there was no alcohol, there was no **driving** with a member of the opposite sex. This was for Sudanese people, but, of course, you, as a foreigner, could not be seen in **the same car** with a Sudanese man.

(Tanskanen, 2006, p. 62)

In these examples, the pairs of collocates (bolded) are related to each other, because they participate in the same activities: *decode cyphers*, *eat meals* and *drive a car* respectively. Since such pairs, i.e. cyphers – decode, meals – eat, driving – car, have relations based on activities, such relations between the pairs were named “activity-related collocation” (Tanskanen, 2006, p. 62). Compared with the category of the ordered sets, the category of activity-related relations is more challenging to recognise in texts, as the collocates in this category are only related based on activities which are judged and interpreted by personal knowledge without following fixed membership of a category, such as *days of a week* in the ordered sets.

4.3.3.3 Elaborative collocations

The final category of collocation in Tanskanen’s framework includes all the remaining collocational relations aside from the relations in the two categories above. The lexical items in this category also have associations but cannot be defined concisely and recognised easily. The associations between the lexical items are based solely on the possibility of the two lexical items “[elaborating or expanding] on the same topic” (Tanskanen, 2006, p. 63). In line with this

elaborating function of the lexical items related to the topic, this category of collocation is named “elaborative collocation” (Tanskanen, 2006, p. 62).

To explain this function in detail, another two concepts were introduced in Tanskanen’s description of collocation: ‘frame’ and ‘trigger’. The former is cognitive and the latter textual. Frame was defined as “knowledge structures evoked by lexical items” (Tanskanen, 2006, p. 63). The first lexical item evokes a frame and the meanings of the following items can be interpreted within this frame. For example, in the sample texts provided by Fillmore and Baker (2001, p. 3) below, the lexical item *arraign* appears in the beginning of a text. Then this item evokes the arraignment frame. If other items like *magistrate* and *charges* appear later in the same text, these items will be decoded based on the arraignment frame. As Fillmore (1985) described, these lexical items are “lexical representatives of some single coherent schematization of experience or knowledge” (p. 223). This schematization of experience or framework of knowledge is the frame.

4-25 Washington (CNN) —— Alleged White House gunman Robert Pickett was **arraigned** Wednesday at a federal court in Washington and ordered held without bond.

4-26 A federal **magistrate** informed Pickett of the *charges* against him - assaulting a federal officer with a deadly weapon, which carries a maximum of ten years in prison.

4-27 The **magistrate** set a preliminary hearing for next Tuesday and ordered Pickett held without bond.

(Fillmore & Baker, 2001, p. 3)

Although *frame* contributes to the creation of coherence, this concept is subject to people’s conceptuality, which means it cannot be applied to the surface textual analysis which requires the inclusion of salient lexical elements at the surface level. To compensate for this deficiency, another concept ‘trigger’ was adopted as it can be used at the surface level and also helps recognise the association

between two lexical items. Tanskanen (2006) first adopted the definition of "trigger" provided by Jordan (1984) who introduced 're-entry' techniques for technical writers. A trigger was defined by Jordan as a repetition of the previous topic (or item), which can be used to clarify the association between an item and its re-entry (Tanskanen, 2006, p. 38). The 're-entry' techniques were divided into three types: basic re-entry, associated re-entry and perspective re-entry. The first type is in line with some categories of cohesion, such as repetition, substitution and synonymy. The second and the third types are related to the notion 'trigger', as occasionally the 'trigger' is needed to make the association between two lexical items clearer. An example below demonstrates this enhancing function of trigger in texts.

4-28 **The System 90 Users Group** was established in October of this year. **Membership** of *the Group* is open to all organizations.

(Tanskanen, 2006, p. 38)

In example 4-28, *the Group* acts as the trigger to enhance the association between *the system 90 Users Group* and *membership*.

Nevertheless, Jordan (1984) also pointed out the lack of necessity for writers/speakers to include the 'trigger' in every associated re-entry (p. 54). For textual analysis, it is helpful to use the 'trigger' as a device to help recognise the connections between one lexical item and its re-entry. The only difference between these two types of re-entry techniques are the length of the lexical items: the 'associated re-entry' refers to the nominal groups, while the 'perspective re-entry' is connected with longer stretches in texts (Jordan, 1984). Therefore, it is suggested to combine these two types into a larger type for convenience.

Additionally, another definition of 'trigger' provided by Hawkins (1978) was also mentioned in Tanskanen's discussion of elaborative collocations. The trigger in Hawkins' study was used particularly to refer to the associations created by lexical items (Tanskanen, 2006, p. 64). In his description of trigger, the first lexical item is regarded as the 'trigger' and the following triggered lexical items or "the first-mention definite descriptions" are 'the associates' (Hawkins, 1978, p. 123).

Hawkins acknowledged that there are no universal parameters which can decide these associates, however he did provide some general rules for determining the associates. The essential rule is that the speaker/writer and the hearer/reader should share "knowledge of the generic relationship" between the trigger and the associates (Tanskanen, 2006, p. 65) (i.e. the frames). The example below demonstrates this point:

> 4-29 . . . at the beginning of the Michaelmas term 1955, Sylvia's first year at **Cambridge**. I had walked into **the Mill Lane lecture room** a few minutes early. . .
>
> (Tanskanen, 2006, p. 62)

Obviously, in example 4-29, *the Mill Lane lecture room* is part of the University of Cambridge. Only if the readers share this knowledge with the writer, can they recognise the collocation between these two lexical items (*Cambridge – the Mill Lane lecture room*).

Another helpful rule is that "the trigger must conjure up a set of [associates] which are generally known to be part of some larger object or situation" (Hawkins, 2015, pp. 123-124), i.e. the associates are part of or members of a broader set of objects or situations, like the collocational relationship above – *the Mill Lane lecture room* is part of the *Cambridge* University. This indication relates hyponyms and meronyms with the collocation category, which is in line with Halliday and Hasan's (1976) framework of collocation. However, in this study, the concepts of 'frame' and 'trigger' are considered to encompass a wider range of items than the categories of hyponymy and meronymy.

As 'triggers' can help establish the relationship between lexical items on the surface of texts, it is necessary to include the "trigger-test" (Tanskanen, 2006, p. 64) when analysing collocations, and therefore, this test is also applied in the current analysis of collocation. For this test, Hawkins (1978) and Jordan (1984) provided different approaches. For Hawkins, the first-mention definite article 'the' is a good signal as the starting point of recognising the elaborative collocation; while Jordan suggested that the re-introduction of the first item by certain

methods is a sign of the collocational relation. Example 4-30 (adapted from examples 4-29 and 4-31) can be used to exemplify the application of these two approaches.

4-30 . . . at the beginning of the Michaelmas term 1955, Sylvia's first year at *Cambridge*. I had walked into *Cambridge's Mill Lane lecture room* a few minutes early. . .

(Tanskanen, 2006, p.63)

In example 4-30, it is possible to include the trigger *Cambridge* in the second sentence. The re-introduction of *Cambridge* makes it connected to the next lexical item *Mill Lane lecture room* and then creates the collocational relation.

4-31 A: yes the reason is you know the disgusting curmudgeonliness of school caretakers **the evening class** I went to we all agreed that it should start at quarter to eight so that we could make it instead of half past seven and we did this happily for four years and in the fifth year the school caretaker went on strike and said he wasn't going to have any more **classes** that finished after half past nine and everybody just had to knuckle under which was very annoying indeed
b: well this one starts at half past seven
A: particularly as **the lecturer** came down from Hampstead down to Wimbledon and he didn't like driving all that way through the rush hour

(Tanskanen, 2006, p. 64)

The first trigger in example 4-31 is *the evening class* while the second trigger which is more general is *classes.* The associate is *the lecturer*. What is interesting here is the use of the definite article *the*. As Hawkins (1978) mentioned, the function of the trigger is to create the possibility of "first-mention definite descriptions" (p. 123). In this example, *the evening class* and *class* make it possible to mention *the lecturer* in proceeding text, as it is shared knowledge that once there is *class*, there is generally a *lecturer*. The function of the definite article *the* is to instruct the reader to deduce the shared knowledge and which frame should be evoked. In this case, as Speaker A knows confidently that Speaker b shares the knowledge that classes normally have lecturers, Speaker A can state *the lecture* without being afraid of being misunderstood in this conversation.

Therefore, the key point of adopting the concept of the 'trigger' and the 'associate' is to ensure that the readers can identify the elaborative collocational relations between the trigger and its associate. In this study, the concepts of both 'frame' and 'trigger' are used to help the researcher recognise the collocational relations in the corpora. Although the analysis of these categories of collocation cannot be as replicable as other categories of lexical cohesion and cannot avoid intersubjectivity, it is nevertheless necessary to include them, since they comprise an important device for creating cohesion in texts (Tanskanen, 2006, p. 69).

4.5 Summary

In order to develop an operational definition and framework of the collocation category in this thesis study, this chapter has been concerned with reviewing previous definitions of collocation, introducing the main approaches to collocation and classifying collocational categories. Three main forms of collocation have been discussed: coherence collocation, neighbourhood collocation and word-combination collocation. These definitions are related to three different approaches to collocation: the psychological approach, which views collocation as psychological associations in the mental lexicon; the Firthian approach which defines collocation as lexical items in syntagmatic relations in texts; and the phraseological approach which tries to demarcate collocation from other types of word combinations.

As mentioned in section 4.2, the key difference between the lexicographical collocation and coherence collocation is the proximity of the items. In lexicography, collocation refers to adjacent items while in cohesion, collocation considers lexical items beyond the clause (Tanskanen, 2006, pp. 33-34). Since the aim of this study is to investigate the cohesive function of collocation beyond the clause, the notion of coherence collocation predominates in this study. The operational definition of collocation for this study is developed as follows:

Collocation comprises the semantic relations between lexical items beyond the clause which are generally known to be part of some larger objects or situations and which frequently co-occur.

Corresponding to the definition of coherence collocation, the study also adopts the psychological approach which emphasises the strong associations between lexical items. However, it needs to be pointed out that the psychological approach does not identify the distance between two lexical items which are semantically associated, while this study identifies the distance between two lexical items by focusing on the associations between lexical items which occur in different clauses.

As for the classification of collocation used in the analysis of collocational relations, this study develops a slightly refined classification of collocation based on the works by Tanskanen (2006) and Martin (1992), which comprises the categories of activity-related collocation and elaboration collocation. The related concepts of ‘frame’ and ‘trigger’ mentioned in section 4.3.3.3 are also adopted to help with the identification of elaborative collocational relations.

Chapter 5 Chinese students and their use of lexical cohesion in academic writing

5.1 Introduction

This chapter focuses on two elements: a description and broad characterisation of Chinese students as research subjects in this thesis, and a review of the previous studies related to the use of lexical cohesion in Chinese students' academic writing. Both of these areas of focus serve to further underpin the research in this thesis.

Section 5.2 starts with exploring the nature and characteristics of Chinese students, focusing on the similarities of identifying Chinese students as a unified, identifiable group. Section 5.3 reports studies on the features of lexical cohesive devices used by Chinese students in their academic writing. This section concentrates on two detailed elements of lexical cohesive device use: firstly errors regarding the usage of lexical cohesive devices, and secondly the correlation between the use of lexical devices and the writing quality of students' work.

5.2 Common characteristics of Chinese students

As mentioned in chapter 1, the definition of Chinese students in the present study is those Asian students whose L1 is Chinese. That is to say, this research concerns the writing of 'Chinese students' as a category in a very broad sense, a category which covers various linguistic, ethnic and national groupings. Therefore, it is necessary to examine the characteristics of this group in detail at first to substantiate whether this definition of Chinese students is indeed reasonable for the present study, since, as stated in the introductory chapter, this study aims at exploring lexical cohesion in Chinese students' academic writing.

5.2.1 Shared Mandarin writing system

The first common characteristics of Chinese students as a whole group is that those speaking Chinese, in spoken form, a language with many dialects, as mother tongue, share the same writing system. Although a range of the

mentioned dialects of Chinese are spoken in East Asia[23], Chinese speakers share a universal standard written system (simplified or traditional), which enables literacy to be a uniting force for all Chinese native speakers.

To be more specific, it is only Mandarin (also known as "Modern Standard Chinese" or "Putonghua") that has been consistently recognised as a "language" by 'Chinese people' because of its connection with the standard ideographic writing form while others are categorised as "dialects" of Chinese (Gao, 2000). The Mandarin written system allows communication between people with mutually unintelligible dialects (Hu, 2002, pp. 4-5), such as the 56 peoples of the PRC and Chinese speakers outside the PRC.

Scarcella (1984) mentioned that the first language may have a significant impact on EFL learners' ability to use cohesive devices in academic writing. Furthermore, the influence of Chinese (mainly written Mandarin) on the use of cohesion in English academic writing has been discussed by Jin (2001) who pointed out that the main reason for difficulties in the use of these cohesive devices lies in the difference between English and Chinese. English is a subject-prominent language while Chinese is mainly a topic-prominent language. An example is given below to demonstrate the difference:

nà kēshù shùyè dà wǒ bù xǐhuān
那 棵 树 树 叶 大， 我 不 喜 欢

That tree tree leaves big, I don't like

The leaves of that tree are big, and I don't like them.

(Jin, 2001, p. 5)

[23] In the PRC, there are basically seven big dialect groups: Beifang (the syntactic and lexical basis for Mandarin), Wu (a variety of which is Shanghainese), Xiang, Kejia (Hakka), Gan, Min (Hokkien), and Yue (Cantonese). Within these dialect groups, there are more sub-varieties (Hu, 2002, p. 4). In total, there are 129 ethnic dialects in the PRC (Jie & Keong, 2014, p. 496).

As demonstrated in this example, the Chinese version has a double subject (那棵树 树叶 and 我 nà kēshùshùyè wǒ), which lacks co-referentiality and overt causality linkage markers; while in the English version, the two clauses above with two different subjects (*the leaves of that tree* and *I*), are linked by the conjunction *and*, and the pronoun *them* is used as a reference device to refer back to *the leaves of that tree*. From this comparison, it is noticeable that Chinese relies on notional connectivity rather than on formal connections, which are regarded as incohesive by western people who tend to use formal connecting devices, i.e. cohesive devices. Since Chinese students share the same mother tongue which is Chinese, all of them will be influenced by Chinese to some extent regarding the use of cohesion in their English academic writing. According to the result of the interviews conducted in Jin's study, the majority of Chinese graduate students who participated in the interviews stated that

> According to their own writing experiences, there is not much difference between English and Chinese prose in terms of overall organization, but there is a marked difference in terms of the use of cohesive ties. […] [R]ather than relying on connective words for cohesion, Chinese learners would look for notional or logical connectivity for interclausal connection.
>
> (Jin, 2001, p. 5)

It can be deduced from these interviews that Chinese graduate students are aware that there is difference between English and Chinese regarding the use of cohesive devices, especially connective words such as conjunctions.

5.2.2 Shared learning culture and language learning method

The second commonly shared characteristic among Chinese students is the culture of learning and techniques adopted in learning Chinese, which is likely to heavily influence students' English learning. "'Chinese' as a defining term descriptively characterises a range of ethnic backgrounds [while] sharing a relatively homogeneous linguistic and cultural heritage" (Jin & Cortazzi, 2006, p. 9). Jin and Cortazzi (2006, p. 9) used Chinese 'cultures' to imply the diversities

in terms of social and individual identities that worldwide Chinese people embrace. A culture of learning is one of the factors that contribute to the creation of social and educational identities for Chinese people (Jin & Cortazzi, 2002). This factor is well reflected when Chinese students learn Chinese Literacy, because as a typical symbol for representing Chinese identity, Chinese literacy engenders a complex learning process to acquire its written characters.

For one thing, the learning of written characters consists of repeated practice through several stages which includes "demonstration, modelling, tracing, repeated copying, and ultimately active memorisation of the precise movement, direction and order of strokes" (Jin & Cortazzi, 2006, p. 9) of each Chinese character.

Furthermore, the features of Chinese written characters also increase the difficulties of acquiring the orthography of these characters. The first feature is that the written Chinese system has roughly 56,000 Chinese characters (Xu, 1995, cited in Li, 2014, p. 2), which increases the burden of memorising these characters (Li, 2014, p. 1). The minimum requirement for L1 Chinese students in Hong Kong and Mainland China to be literate is to acquire 2500–3000 characters (Li, 2000, cited in Li, 2014, p.2). Therefore, in primary schools and junior middle schools, Chinese students spend hours in practicing characters until they can commit these characters to memory (Everson, Chang, & Ross, 2016, p. 1). Another feature is that Chinese, as a tone language, has many homophones which may confuse learners and also force them to learn a large number of strokes[24] these being the only bases to distinguish each character (Butcher, 1995).

As mentioned above, such detailed practice of Chinese calligraphy and the great challenge involved in the memorisation of Chinese characters have a profound influence on Chinese learners' impression of language learning. Chinese students adopt the learning method of imitating and repeating models provided

[24] Strokes are the primary "building blocks" of Chinese characters (Everson, Chang & Ross, 2016, p. 2)(Everson et al., 2016).

by teachers and textbooks, and believe that mastery of basic forms of written language is the foundation for being creative and artistic.

It was further suggested by Alexander (2001) that learning Chinese shapes the way Chinese students learn English. “Chinese literacy practices seem to encourage Chinese students to approach the learning of English with a similar attention to specific detail and a similar respect for the authority of the teacher” (p.1, cited in Leedham, p.21, 2011). This spirit of paying attention to details in learning Chinese literacy unites all Chinese students in learning other languages, such as students’ focus on extracting and memorising new words in texts. This spirit of learning languages is typically encouraged in the Grammar Translation Method (henceforth GTM) adopted by Chinese teachers in English language teaching (henceforth ELT) (Leedham, 2011, p. 21).

Although ELT in East Asia has been influenced by other approaches, such as Audiolingualism, Communicative Language Teaching and Task-Based Learning (Littlewood, 2007), the GTM was the earliest and most widely adopted teaching method in China. This method has persisted throughout the history of ELT in the PRC and Taiwan, and, alongside the learning method influences already mentioned has itself influenced Chinese students’ perception of English learning (Chang, 2011, p. 13; Hu, 2002, p. 28). The key elements of this method are firstly learning grammar rules, and then practicing these rules in exercises (Stern, 1983). As a result, language is studied at the sentence level (Hu, 2002, p. 28), which may lead to Chinese students lacking of awareness of the unity and cohesiveness of the whole text. However, it is also admitted that there is no explicit evidence in Chang’s (2011) study and Hu’s study (2002) showing that students in the PRC and Taiwan lack awareness of the cohesiveness in their texts because of their focus on studying grammar at the sentence level.

5.2.3 Shared cultural heritage

The third commonality across Chinese students considered here is the shared background of culture heritage, mainly from Confucianism. Despite the difficulty in generalising the cultural behaviour of such a complex and extensive entity as the Chinese, there are still some deeply rooted cultural assumptions in this

complicated society. These apparently stable assumptions have a significant influence on Chinese models of teaching and learning.

There are two main cultural features deriving from Confucianism, which might affect Chinese students' academic writing styles which may be linked to the result of the lack of using cohesive devices in their writing. Firstly, traditional Chinese education emphasises the importance of maintaining a hierarchical but harmonious relation between teachers and students. Students are expected to respect and not to challenge teachers. A maxim said by the Confucius is that:

being a teacher for only one day entitles one to lifelong respect from the student that befits his father.

yī rì wéishī　　zhōngshēnwéifù
一日 为 师，　　终 身 为 父

(Hu, 2002, p. 34)

This maxim implies that there is a solid and permanent relationship between teachers and students; and students need to show their reverence to teachers. That is to say, since Chinese teachers have preferred to use this GTM in ELT, Chinese students have followed their teachers' instruction and used the GTM in their English learning. As mentioned above, the GTM is not applicable to learning the use of cohesion, which may lead to Chinese students' lacking of knowledge regarding the use of lexical cohesive devices in their English writing.

For example, in Mainland China and Taiwan, which are frequently categorised in the group of "Confucian Heritage Cultures" (CHCs) (Leedham, 2011, p. 22), the conventional teaching activity in the GTM is grammar drilling (Zhilong, 2018, p. 119). Furthermore, In language learning, generally speaking, "[c]ompetence in the organisation of written discourse develops late and […] appropriate instruction has an impact on this competence" (Mohan & Lo, 1985, p. 522). As Chinese teachers have emphasised on grammar drilling activities and ignored the importance of teaching academic writing style and discourse organisation, students have accepted this method of learning English at the grammar level,

which is likely to have influenced their lack of knowledge in regard to cohesion which is a central role in discourse organisation of academic writing.

Another element of their cultural heritage is that Chinese people have high tolerance of uncertainty (Hofstede, Hofstede, & Minkov, 2010, p. 218). This feature has been attributed to Chinese society as being collectivist-oriented, which means the main function of this kind of society in communication is to maintain social cohesion and harmony (Bloch & Chi 1995, cited in Liu & Jiao, 2014). Implicit messages are frequently seen in Chinese students' academic writing. As Zhang (2018) has pointed out, "Chinese academic writing, as a significant element of Asian culture, has the feature of the implicit message of the meaning rather than using linking words to demonstrate logic and cohesion" (p. 120). However, Zhang's research only conducted analysis of previous studies, which cannot provide any empirical evidence for the previous statement regarding Chinese writers' lack of cohesive devices in their academic writing.

The convention of including implicit information in Chinese writing is the inverse of western writing convention, representing a key difference arising from these two cultures. Kaplan (1966) has claimed that western cultures and Oriental cultures (mainly Chinese culture) lead to different thought patterns. Western people favour a linear thought pattern while Chinese people choose a nonlinear or even circular fashion. Accordingly, the difference, reflected in writing style, is that western writing style tends to be linear and hypotactic, whereas Chinese writing style prefers a nonlinear and paratactic[25] one. Regarding the use of cohesion in writing, western students depend on textual cohesive devices for creating coherent and linear texts, while Chinese students follow notional connections, which means the reduction of cohesive pairs in text, such as zero anaphora and ellipsis (Jin, 2001, p. 3).

Such cultural influences, including the hierarchical teacher-student relationship and high tolerance of uncertainty, may firstly prevent Chinese teachers and

[25] "In a paratactic language, connective elements are often optional or unnecessary while the opposite is true in a hypotactic language" (Yu, 1993, p. 1).

students from accepting Western academic writing style which is more argumentative (Vyncke, 2012; Wingate, 2012) and connects all relevant inferences (Kurland, 2000); and secondly, result in their lack of knowledge and understanding of the contribution of cohesiveness to high quality English academic writing.

However, the research mentioned above does not have sufficient empirical evidence for the connection between Chinese culture and its influence on Chinese students' use of cohesion as well as their approaches to academic writing. Furthermore, models such as those of Hofstede (Hofstede, Hofstede, & Minkov, 2010) and Kaplan (1966) have been extensively criticized (Helal, 2013; Kramsch, 2004; McSweeney, 2002; Schmitz & Weber, 2014), which indicates that the validity of their models is open to question. The application of their models in the current study is without exception. These two points suggest that there are also other possible alternative interpretations regarding the use of cohesion in Chinese students' writing. For example, all non-expert writers, including Chinese students, might have the same or similar characteristics in English academic writing as which is a skill that is only acquired after long-term training and practice, regardless of native/non-native speaker status.

5.3 Research on the use of lexical cohesion in Chinese students' writing

As mentioned in chapter one, only a small number of studies have focused on lexical cohesion in Chinese students' writing in depth. This section discusses the studies focusing on aspects of this area, in particular, the dominant usage of lexical cohesion in Chinese students' writing; the lexical cohesion errors appearing in their writing; and the correlation between the use of lexical cohesion and writing quality of Chinese students' work.

5.3.1 The predominance of lexical cohesion in Chinese students' writing

Writing strategies are categorised into rhetorical strategies, metacognitive strategies, cognitive strategies, and social/affective strategies (Mu & Carrington, 2007; Riazi, 1997; Wenden, 1991). Rhetorical strategies refer to strategies that writers use to organise and translate their ideas following writing conventions appropriate to native speakers (Mu & Carrington, 2007; Wenden, 1991).

Cohesion strategy belongs to the category of rhetorical strategy. In Mu and Carrington's study (2007), cohesive analysis was used to represent the exploration of the use of rhetorical strategy in three Chinese postgraduate students' written drafts of papers or proposals at an Australian university. While this is a small study, the result of this study points to a higher frequency of the use of lexical cohesive devices than that of the use of grammatical cohesive devices in Chinese students' writing (Mu & Carrington, 2007, p. 6). This conclusion has also been evidenced by several other studies (e.g. Zhang, 2000; Liu & Braine, 2005; Yang et al., 2018). For example, in Zhang's (2000) study, 107 expository essays written by second-year English-major Chinese undergraduate students at two Chinese universities were examined regarding the use of cohesive devices in these essays. Based on the percentage of cohesive ties of each cohesive category used in these essay samples, this study found that Chinese undergraduates employed more lexical cohesive devices (71.7%) than other grammatical cohesive devices, such as conjunction (17.5%) and reference (10.8%) (Zhang, 2000, p. 71). A similar result can be found in Liu and Braine's (2005) study which focused on the use of cohesive devices in 50 argumentative essays produced by non-English-major Chinese undergraduates. An interesting point of this study is that the student subjects were given explicit instruction regarding the use of cohesive devices in English academic writing classes before they created the essay samples, which has not been mentioned in Zhang's study above. Nevertheless, Liu and Braine's study also suggests that lexical cohesive devices were used more frequently by Chinese undergraduates than other grammatical cohesive devices. The percentage of each cohesive category used in the student samples is shown as follows: lexical cohesion (55.6%), the reference devices (29.8%) and the conjunction devices (14.6%) (Liu & Braine, 2005, pp. 627-628).

However, since lexical items involve both meaning and usage, compared with grammatical items which mainly deal with their usage, though these grammatical items have comparatively constant meanings when used in different contexts, it is more complicated for Chinese students as EFL learners to grasp the appropriate use of lexical cohesive devices in their academic writing (Liu & Braine,

2005, p. 633). Therefore, there are some typical errors identified in Chinese students' academic writing regarding the use of lexical cohesive devices.

5.3.2 The features of lexical cohesive devices used in Chinese students' academic writing

As found in many related studies (e.g. Connor, 1984; Khalil, 1989; Li & Thompson, 1981; Jin, 2001; Haris Fatimah & Yunus, 2014), repeating the same word is a common feature among EFL writers. Chinese EFL learners are no exception. For example, Ong (2011) investigated cohesive errors in expository compositions of Chinese students in a Singaporean university and identified that "the EFL learners had a strong tendency to repeat the same words in their writing" (p. 58). This claim is supported by several other researchers, e.g. Feng (2003, cited in Ong 2011) and Zhang (2000) who also reported that repetition of the same words was the predominant type of lexical features in Chinese EFL writing.

It has been pointed out that there are several negative influences that the redundant repetition can have on text composition. Firstly, the repetition of the same words retards the flow of text (Haris Fatimah & Yunus, 2014, p. 847) and makes text reading dull, monotonous and uninteresting. At the same time, if this is a general feature for Chinese students' English academic writing, this might create problems in terms of elaborating and extending their ideas something which can be achieved through the use of other lexical cohesive devices. However, in certain discourses, such as legal documents and laboratory reports, repetition is an efficient cohesive device to establish greater clarity (Bouchoux, 2017) and repetition of technical terms might be highly appropriate in these discourses. Although elegant variation" (or repetition), referring to using other devices, such as synonyms and more general words, other than repeating the same word in a text, is encouraged in both spoken and written contexts (Cook, 1989, p. 19), whether the use of repetitive lexical items is appropriate depends on the discourses in which the academic written texts are generated. It is not fair to judge the quality of a piece of academic written work only based on the variety of lexical cohesive devices used in that text.

Although the use of repetitive devices is discourse-oriented, possible reasons for the overuse of repetitions in Chinese students' academic writing have been discussed in previous studies. As Chen (2007) suggests, one possible explanation for the overuse of exact (simple) repetition in Chinese college EFL writing is a lack of knowledge of the concept of lexical cohesion in college classes. Chinese students are unlikely to have received any instructions or information about lexical cohesion, or how to use various lexical cohesive devices in their academic writing to create the cohesiveness of texts. This view is shared by Liu and Braine (2005). As a Chinese national writing test marker, Liu found that most non-English major Chinese undergraduates had no knowledge at all about the use of cohesive devices in academic writing.

Another related reason for this overuse of repetition lies in students' restricted use of lexical items (Liu & Braine, 2005). Since students often might have a limited vocabulary repertoire, it is easier and safer to keep repeating the same lexical item through the whole text, rather than risking trying to use other substitute words. A deeper reason behind students' lack of vocabulary knowledge is undoubtedly their relatively low exposure to English outside class and the limited time that they spend in practicing English writing.

A third reason might be the influence of Chinese in English composition. An example of the use of the word "car" in Chinese students' writing was made by Chen (2007) to elucidate this point. If Chinese students use Chinese at the idea-generating stage for composition, a picture of the object "car" is connected with the Chinese word for "car" which will be translated into the English word for "car" at the text-generating stage (Wang & Wen, 2002). The reason is that Chinese students often memorise English words with a vague meaning, rather than differentiating the meanings of an English word at different levels. In this case, the English prototype word for all vehicles is *car*, which is the first choice for Chinese students during the Chinese-English translation stage. Once this stage is complete, the students' attention is distracted by other stages which need to be the focus of attention during composition, such as content and organisation. In the subsequent writing process, "whenever there is a need for an English word referring to an object which looks like a car" (Chen, 2007, p. 51), the word *car* will

be used repeatedly. However, such Chinese-English translation strategy is not an effective writing strategy regarding the creation of cohesiveness in texts, which means that in fact, various lexical cohesive devices need to be taught to Chinese students to help improve their academic writing skills. As Chanyoo (2018) suggests, "teachers need to teach and encourage […] EFL students to use accurate devices for more efficient textual compositions" (p. 996), which leads to the subject of another feature in Chinese students' writing – misuse of collocation.

Misuse of collocation points again to the learners' lack of vocabulary and their inability to use alternative cohesive devices. The misuse happens under two conditions: first, learners make wrong choices of lexical items, which results in collocational cohesive errors; second, these errors occurring within a phrase itself. For example, in the following sentence from a student sample:

> 5-1 Through education, we can recaeve[26] good things, learn how to be a human being. How to deal with every possible situation, and the good mind must be **planted** and growing in our hears little by little.
>
> (Ong, 2011, p. 55)

It is noticeable in example 5-1 that the writer probably attempted to form a cohesive chain among *good mind*, *planted* and *growing*. However, a proper collocate for *good mind* is the lexical item *cultivated* instead of "*planted*" in this example. This is a typical instance of selecting a wrong lexical collocate within a phrase itself.

More examples of other collocational errors are provided in Liu and Braine's study (2005) regarding Chinese undergraduates' writing (the correct version is in parenthesis):

> 5-2 First, it makes students **interest** in the computer (interested).

> 5-3 It means not only the students but also the teachers should **face** computer games

[26] This misspelling of *receive* is from Ong's (2011) study.

in a proper attitude (take a proper attitude toward computer games).

5-4 Compare **to** its wasting time, it has much more advantages (compared with those disadvantages such as wasting time, it has many more advantages).

(Liu & Braine, 2005, p. 633)

As can been seen from the three examples above, the misuse of collocations include fixed phrases, verbs, nouns and prepositions. It needs to be pointed out that the errors in collocational pairs were identified within the clausal boundary, which is not the research area in this thesis study, which focuses instead on lexical cohesive pairs beyond the clausal level. However, these examples still reveal potential problems that Chinese students may also have across clauses in their writing. One reason for these problems may probably be that the Chinese traditional method of teaching vocabulary is to isolate the lexis from its context and focus only on its semantic meaning (Zhang, 2000).

The third lexical cohesive feature is the overuse of general nouns[27], which was found in Wu's (2010) study. Although her study focused on Chinese students' oral English competence, it still has meaningful implications for the present study, because as a main sub-category in the present study, general nouns are an importance focus of attention. Wu's (2010) study suggests that the overuse of general nouns has a negative influence on the expression of meaning in English. Two examples are given to demonstrate this point:

5-5 and I think the spare time is too difficult to spend for me; I just don't know what I should do in my spare time.

5-6 But I think because the college provides less class everyday, it provides our students more time to do their own business and take activities. Also, they can do their interesting things.

(Wu, 2010, p. 100)

[27] In this study, general nouns is regarded as one sub-category of signalling nouns. See chapter 3 for more details.

Example 5-5 is from what was referred to in her study as a high quality discourse sample, while example 5-6 is from a low quality one. In the first example, it is easy for listeners to infer that the referent of the general noun *do* is *spend spare time*. This interpretation between *do* and the previous referent phrase was simple and smooth, which suggests that competent Chinese students are aware of the appropriate use of general nouns and specific lexical items in discourse. In contrast, in example 5-6, the item *things* is also a general noun, however, it was much harder to deduce that whether its referent was *business* or *activities* in this specific discourse. Such a complicated or uncertain interpreting process may create obstacles for the listeners in understanding the information that the speaker is trying to convey. The comparison between these two examples suggests that the overuse of general nouns expressing vague meanings will reduce the sense of connectivity and cohesiveness between lexical items, which will likely increase the difficulties for listeners' comprehension of the whole discourse. A systematic way of using general words was suggested by Scarcella (1984) in order to avoid ambiguity. This involves use of explicit lexical items at first to introduce a referent; and once the referent is clear to the reader, then a less explicit item, e.g. a general word, is used to support the variety of lexical cohesive devices in writing and to clarify the centrality of the referent to the ongoing development of the text (Scarcella, 1984, pp. 90-91).

5.3.3 The correlation between English proficiency and the use of lexical cohesion

Conclusions in the literature as to the existence or not of correlation between the number of lexical cohesive devices used in Chinese students' writing and their writing quality have been inconsistent. Some studies have identified no significant correlation between English proficiency or score of written work and the use of lexical cohesive devices in Chinese students' writing.

Chen's study (2007) of Chinese college EFL writing suggests that the use of lexical cohesive devices in Chinese writers' assignments is not related to the students' language proficiency. An ANOVA test was used to assess the difference between two groups of Chinese students with lower and higher English proficiencies regarding mean frequencies of lexical cohesive devices identified in

60 150-200 word samples. Students were allowed 35 to 40 minutes to write two samples, one narration and one argumentation, on individual computers and submit them to the main computer in the language lab. The result of the test shows that the difference was not significant. However, the definition of the tested group is somewhat problematic. 30 English major undergraduates from a Chinese university were the tested subjects, in which 15 participants were the first-year students while another 15 were the third-year students. Their language proficiency levels were measured by the students' year of study at university. Therefore, the 15 first-year students were categorised in the group of lower English proficiency level, whereas the 15 third-year students formed the higher level group. Although it was mentioned in the study that during the sample selection process, top and poor students in each group were removed according to their academic records, this would still seem rather an unconvincing means to distinguish the students' English proficiency levels, being only based on their length of formal English training time at university. In theory, a third-year student could be far less proficient in English writing than a first-year student. It would have been more appropriate to categorise students' samples into different groups, representing higher and lower English proficiencies based on evaluation scores of the samples because the scores of the students' recent written assignments would have a comparatively true reflection on their current English proficiency levels. This strategy has been used in several other studies (e.g. Liu & Braine, 2005; Chanyoo, 2018). The exact level of students' English proficiency in each group was not demonstrated in Chen's study. The above selection criterion of research subjects and lack of information about the participants' English proficiency cast some doubts over the reliability of the results in the study. However, regarding the explanations about the results, Chen's study provides some interesting thoughts.

Firstly, Chen's study found that Chinese students with lower English levels tend to use more simple repetitions in their writing, a result supported by other related studies (e.g. Scarcella, 1984; Jin, 2000, see below).

Secondly, several reasons were put forward in Chen's study (2007, p. 52) to explain why there was no significant difference between the two student groups

with different English proficiencies. First of all, the gap of lexical knowledge between the two groups may be not as big as might be expected based on the teaching syllabus. The time allocated to teaching lexis was limited in the studied Chinese colleges. Students were expected to study vocabulary on their own initiatives. This might have resulted in students' lacking lexical knowledge even after three years of English study at university, this lack of study being unhelpful for developing students' ability to use lexical cohesive devices. Therefore, Chen's study could not confirm the students' level of vocabulary knowledge as it did not investigate the time that students spent in expanding their vocabulary repertoire.

Another possible explanation mentioned in Chen's study was that students in both groups studied already knew much vocabulary in isolation, but had not used these items in real contexts, e.g. using lexical items as lexical cohesive devices. McCarthy (1991) commented that

> an awareness of the usefulness of learning synonyms or hyponyms for text-repeating purposes may not always be psychologically present among learners. It is likely that vocabulary learning has been taken as word studying separated from actual use of only associated with receptive skills.
>
> (McCarthy, 1991, p. 68)

That is to say, students might have learned vocabulary out of context and have never psychologically realised the need to use the lexis in real contexts, let alone using them as lexical cohesive devices in academic writing.

The third reason put forward for the lack of correlation in Chen's study lies in the influence of the 'process approach' adopted in teaching academic writing at these Chinese colleges. Different from the traditional 'product approach' which emphasises linguistic knowledge, such as the appropriate use of vocabulary, syntax and cohesive devices (Pincas, 1982), the 'process approach' focuses on linguistic skills, like planning and drafting (Badger & White, 2000, p. 154). In such teaching contexts, while the teachers had realised that there might be problems of redundant repetition in students' writing, "few of them would take time to deal

with it in the stage of vocabulary instruction or in writing class" (Chen, 2007, p. 52).

Considering the possible three reasons given above, it is not difficult to conclude that the teaching method and learning environment in these colleges could have had a negative impact on students' knowledge of vocabulary and perception of lexical cohesion learning.

Another study conducted by Scarcella (1984) supports Chen's findings. In this study, twenty Taiwanese newly-enrolled students in graduate programs at an American university were part of the research subject group. Analysis of the Taiwanese students' writing shows that there is no significant correlation between the students' various English proficiency level and the use of repetitions, synonyms, superordinates and general words in their expository essays. However, Scarcella (1984) also points out that since general words and superordinates were rare in these samples, further investigation would be required before drawing any firm conclusions concerning the frequencies of the lexical cohesive ties in these two lexical cohesive categories. Furthermore, the collocation category was excluded in his study, which was a limitation of his study. However, what is noteworthy in his study is the finding that more repetitive cohesive ties are used in the beginner group compared with that in the advanced group.

By contrast with the two studies mentioned above, other studies have shown that there is a positive correlation between writing quality of students' works and the use of lexical cohesion in their writing. For example, as mentioned above, Liu and Braine (2005) studied the number of lexical cohesive devices in English argumentative essays written by Chinese students. Their study followed Halliday and Hasan's (1976) model of lexical cohesion in their corpus analysis, in which the samples were taken from a final 150-200-word exam argumentative writing produced in 30 minutes. Each sample was marked by two raters according specific marking criteria which state that "main ideas stated clearly and accurately; well organized and perfectly coherent; very effective choice of words;" (Liu & Braine, 2005, p. 627), and the mark was used as the criterion for determining the

writing quality of each sample. In terms of the use of lexical cohesive devices in the samples, superordinates and general words were extremely infrequent while repetitions comprised by far the highest frequency.

As mentioned above, what might be different in their study is that the participants were given explicit instructions on knowledge of lexical cohesive devices in advance of the examination during writing classes. Furthermore, as mentioned above, an important aspect in the marking criteria emphasised the feature of being "well organized and perfectly coherent" (Liu & Braine, 2005, p. 627). Therefore, it is perhaps not surprising that the essay scores were highly correlated with lexical cohesive devices. One other possible explanation for this result was also given in the study, which was that "compositions with high scores tended to be longer and involved more lexical items" (Liu & Braine, 2005, p. 631).

Another valuable point in Liu and Braine's study was the comparatively high quality of the participants' written samples. These Chinese participants were from Tsinghua University which is generally considered as a top university in China, signifying the likelihood that the students' English proficiency levels were far above the average level of Chinese undergraduates in the rest of the country. Furthermore, these participants had already achieved advanced levels of English and been trained for English writing for one term. What can be deduced from this study is that, regarding the use of lexical cohesive devices, the more general capacity for Chinese undergraduates in non-English majors to use lexical cohesive devices could be significantly different than the results in Liu and Braine's study.

Another study that supports the positive correlation between the use of lexical cohesive devices and writing quality was Jin's (2001) study. Six Chinese students who studied at a university in the United States were investigated in her study. The criteria used to determine their English proficiency level were their TOEFL scores and their placement in two English rhetoric and composition courses at the intermediate and advanced levels respectively at university. According to the two criteria, three participants were selected from the intermediate level course and the remaining three were from the advanced level course. 18 text samples

were collected in total, 3 from each student's assignments for the English rhetoric and composition courses. The problem of these classification criteria for determining English proficiency levels in Jin's study is similar to that of Chen's (2007) study. The standard of classification is based on the division of the students' abilities when they started their study at university rather than in accordance with the marks of their assignment samples which were used for the specific studies. However, the students' English levels were not positively correlated with their written assignment marks. A further difficulty is that, the samples were produced through the whole academic year. The students' English level might well vary during that period of time, which might influence the quality judgements of the samples, and accordingly have a negative impact on the study.

In terms of the categories of lexical cohesion examined in Jin's study (2001), there were only four categories included, namely same word (i.e. simple repetition in the present study), superordinate (hyperonymy, hyponymy and meronymy in this thesis), and general word (a subtype of signalling noun in this thesis). The categories of collocation and antonymy which are widely-recognised in previous studies (e.g. Halliday & Hasan 1976; Tanskanen 2006) were excluded in Jin's study. Although this strategy narrowed down the research scope of the study, it helped this study focus on investigating the specified categories in detail. On the other hand, similar to other studies (e.g. Chen 2007), Jin's study reported the frequencies of lexical cohesive devices across proficiency levels in each lexical cohesion category.

In addition to the similar finding that Chinese students favoured lexical repetition in their academic writing (e.g. Chen, 2007), Jin's study also revealed that there was no significant difference between the two groups of Chinese students with different English proficiencies regarding the use of superordinates and repetitions, while the advanced group used more general words and many more synonyms at a statistically significant level than the intermediate group did. Based on this result, Jin (2001) concluded that "the writing of the advanced group demonstrated a relatively higher level of sophistication than that of the intermediate group" (p. 22). However, such a claim seems unjustified as it was based on significant difference identified in only one category of lexical cohesion (i.e. synonymy).

The final study which will be mentioned here is Wu's (2010) study. Although her object of study was oral English rather than written English, some findings and discussions from this study regarding qualitative analysis of lexical cohesion are considered of relevance to this thesis. 10 samples were selected from 118 pieces of sound recordings of final oral examination produced by second year English-major undergraduates at a Chinese university. These samples were divided into two groups: five discourses with higher quality and five discourses with lower quality. The result of this study showed that there was significant difference between higher quality discourses (henceforth HQDs) and lower quality discourses (henceforth LQDs) in terms of the use of lexical cohesion, as in the quantitative analysis of the discourse samples, the number of mean ties per sentence employed by HQDs (2.24) is higher than that employed by LQDs (1.82) (Wu, 2010, p. 98).

When demonstrating the difference between high quality discourses and low quality discourses regarding the use of synonyms, antonyms and superordinates, Wu's study found that there was no significant quantitative difference between these two groups of discourses. However, she pointed out that the effectiveness of the usage of these three types of lexical cohesive device was positively related to the quality of discourses. Unfortunately, no further evidence was mentioned in her study to support this claim, though an example was provided to illustrate a more effective way of learning lexis regarding Chinese students' English study. The learning of the lexical item *body* should be connected to its related lexis, such as *head, hand, feet, eyes, mouth* and *face*. By building up a hierarchical network among lexical items, it was posited that students would be able to retrieve relevant words in their internal lexicon more easily (Wu, 2010). This strategy would be helpful for Chinese students to improve their efficacy of vocabulary learning, and also support their familiarisation with lexical cohesive relations between lexical items, such as hyponymy and hyperonymy.

In addition to the elements mentioned above, there are still gaps in the related research field, such as the selection of criteria for the determination of writing quality of students' assignments, and the use of each type of lexical cohesive

device in Chinese postgraduates' English academic writing. To contribute to the whole picture of the research, this thesis will investigate the characteristics of the use of lexical cohesion in Chinese postgraduates' academic writing, as Chinese postgraduates form an important segment of Chinese international students in the UK (British Council, 2017). At the same time, it is also important to bear in mind that cohesion is used as an assessment criterion in EFL or EAP writing classes (the context for previous studies mentioned above), while it might not be considered much in the context of marking MA assignments which are the targets of this study. This indicates that claims from these previous studies might not apply for the samples in the current study.

5.4. Conclusion

Three areas of commonality among Chinese students as a whole group mentioned in section 5.2 suggest the feasibility of adopting the thesis definition of Chinese students as people whose L1 is Chinese. This is firstly because their mother tongue influences every Chinese student when they use lexical cohesive devices in English academic writing. Secondly, the traditional language learning approach adopted from the learning process of written Chinese characters makes Chinese students focus on learning the target language at the grammar and sentence level, which may contribute to their lack of knowledge of writing cohesive compositions at the textual level. Furthermore, their shared cultural background – the traditional teacher-centred educational culture and high tolerance of ambiguity may also mean that Chinese students lack sufficient awareness of using lexical cohesion in English academic writing.

The three main features of using lexical cohesive devices in Chinese students' English academic writing mentioned in previous studies are dominant use of repetition, misuse of collocation and overuse of general nouns. The main reasons behind these features are Chinese students' small vocabulary repertoire, the traditional vocabulary-out-of-context learning method adopted, and the lack of knowledge regarding the use of lexical cohesive devices. It is suggested that teachers emphasise these three aspects in EAP classes to help students improve their appropriate use of lexical cohesive devices.

As for the correlation between the use of lexical cohesion in Chinese students' writing and the quality of their writing, as described in this chapter, the results in previous studies are contradictory. Some studies suggest that there is a positive correlation between these two elements, while the other researchers claim that there is no significant difference in different proficiency groups regarding the use of lexical cohesion. The result of this thesis will hopefully provide more evidence for this issue by investigating the use of lexical cohesion in Chinese postgraduates'' writing in the UK context.

Chapter 6 Methodology

6.1 Introduction

This chapter describes the methodology adopted in the present study. Firstly, Section 6.2 will describe the development of the corpora of Chinese postgraduates' writing constructed for this thesis research. Specifically, as mentioned in chapter 1, two small-sized corpora were created: the MA TESOL Module assignments Corpus [28] (henceforth MTMC) and the MA TESOL Dissertation Corpus (henceforth MTDC), supporting the two key goals of this research. MTMC was used for the purpose of comparing the use of lexical cohesive devices between various levels of written works classified based on marks awarded in module assignments. The second corpus – MTDC – was used to enable analysis of the lexical cohesion in functional sections of dissertations.

The present study used manual analysis with two analytical purposes: the quantitative analysis of the MTMC and the MTDC corpora aimed to support the identification of potential areas of linguistic challenge for Chinese students in terms of the usage of lexical cohesion in their academic writing, while the purpose of the qualitative analysis of the features of lexical cohesive devices in the corpora was to inform the identification of Chinese students' preferences for different devices in their academic writing, and shed light on potential practical implications for EAP pedagogy.

In order to identify the necessary size and number of samples required for analysis and the analytical strategies used in the full-scale analysis, and develop an operational framework for analysis of lexical cohesive devices, a pilot study was conducted which is discussed in section 6.3.

Section 6.4 will introduce the work of co-raters. Two co-raters were involved in the early stages of sample analysis. Their background and the training procedures which were implemented before conducting any analysis are

[28] The samples in MTMC are from a Syllabus Design and Assessment module for MA TESOL programme.

discussed. Furthermore, the results of the co-rater analysis and the test of inter-rater reliability are also presented and discussed.

Section 6.5 will focus on the data segmentation in the corpora for discourse analytical purposes. The two corpora are divided into different groups to facilitate analysis. The MTMC corpus is divided into four marking-scale groups (i.e. failed, pass, merit and distinction); and the MTDC corpus into five functional-section groups (i.e. introduction, literature review, methodology, findings and discussion, and conclusion). The selection criteria of the samples in the corpora are introduced. This is followed by the presentation and demonstration of the operational framework and the tagging rules applied to the samples in the analysis. Finally, the annotation system applied to the lexical cohesive devices identified in the sample analysis will be discussed.

Section 6.6 will detail and summarise the rationale for adopting manual analysis as the main approach in this research, and the evaluation of the data. Further, the operation process of manual analysis in the corpora will be introduced.

Section 6.7 will focus on the discussion of the statistical methods used in the lexical analysis of the two corpora and the rationale for the selection of the specific statistical tests used in the quantitative analysis of the samples.

Based on the results of the pilot study and the co-raters' evaluation, section 6.8 will demonstrate the developed analytical strategy for lexical cohesive identification and classification in this research, and section 6.9 will summarise the content of this chapter.

6.2 The two corpora in the present study

As mentioned in section 6.1 above, two small-sized corpora were constructed for this study: MTMC and MTDC. The information is summarised in table 14 below (other related information regarding these two corpora are provided in Appendix A):

Table 14 Summary information: Corpora MTMC and MTDC

corpus	number of text samples	word count (total)	mean	maximum	minimum
MTMC	52	17538	337	486	197
MTDC	45	19148	426	457	389

It can be seen that these two corpora are small in size and are specialised corpora collected from the particular academic programmes at a UK Top 10 University (i.e. MA TESOL and MA Applied Linguistics for TESOL) in the present study. There are three reasons for selecting these two programmes: firstly, the majority of students studying on these programmes are Chinese; secondly, these Chinese students' English proficiency level is relatively advanced as they achieved the minimum entry requirement which is 7 or above band score in the IELTS Test before starting their study on the masters programmes; and thirdly, these students have higher chances to learn the knowledge of cohesion through pre-sessional EAP courses or an optional module offered by the programmes – Discourse Analysis. Based on these reasons noted above, 97 samples were selected from an existing corpus comprising written texts produced by students taking these programmes. The rationale for collecting the samples from this corpus is based on the principles of convenient sampling, a type of nonprobability sampling techniques in which "members of the target population that meet certain practical criteria, such as easy accessibility, […], availability at a given time or the willingness to participate are included for the purpose of the study" (Etikan, Musa, & Alkassim, 2016). In this case, this selected corpus was established by the academic department which offers the two programmes, and samples in this corpus are from students who have signed consent forms to give permission to use their written texts for research purposes.

This thesis allocated the 97 samples into two corpora. MTMC is made up of 52 text excerpts from Chinese postgraduates' module assignment samples and MTDC consists of 45 text excerpts from 9 Chinese postgraduates' dissertations. The nature of the selected excerpts and the processes through which the particular excerpts were selected is described below. Overall, the corpora data yields a total word count of 36,686 words which were used for a manual analysis

of lexical cohesion in the relevant academic discipline. MTMC was assembled at an early stage for pilot study and rater work, while MTDC was compiled at a later stage for complementary research of the use of lexical cohesion in different functional sections of long and linear texts. For convenience of information retrieval, the samples were coded based on which marking-scale groups or functional-section groups they belong to in the two corpora respectively.

As mentioned in chapter 1, the main purpose for analysing the samples in the MTMC corpus was to investigate whether there is variation between marking-scale groups regarding the use of lexical cohesion. Since the nature of lexical cohesion is context sensitivity. It is important to choose samples produced in similar contexts and in the same genre for the comparison of the samples in different marking-scale groups. The genre of the module assignments in the MTMC corpus is expository writing because the samples were selected from the module called “Syllabus Design and Assessment”, and the contents of the samples are either descriptions of the learners for designing a syllabus or the design statements of assessment tools. In fact, samples share a range of lexical items for technical terms, such as *learners, syllabus, test, design statement, TLU (Target Language Unit)* and so on.

As for the nature of the MA dissertation samples, since dissertations are normally extended reports of an individual study, which can display the use of lexical cohesion during the flow of text, each dissertation was divided into five parts according the function of each part to examine whether there would be significant difference between the five parts regarding the use of lexical cohesion. The samples chosen for the current study were empirical-based dissertation texts for two reasons. Firstly, the majority of students on the master programmes have chosen this type for their dissertations (another type is the library-based dissertation) (Durham University, 2019). Secondly, the functional sections in empirical-based dissertation texts are the same, i.e. introduction, literature review, methodology, findings/discussion and conclusion, which made it convenient to allocate different parts of these dissertations into the five functional sections.

Since dissertations are long documents reporting individual studies, the nature of the five functional sections in an empirical-based dissertation text varies, which is demonstrated as follows (Durham University, 2019, pp. 21-22):

- The introduction section should prepare the readers for what they will be reading by introducing the topic and research questions being investigated, why these were selected, and why they are relevant and important to the field of English Language Teaching. It should also provide the structure of the dissertation.

- The literature review section presents a concise and critical overview of relevant previous studies, establishing the ground for the research questions and research methods of the dissertation, and introducing the key research themes and arguments.

- The methodology section describes the research methods and tools used in the dissertation, the subjects/participants involved in the research, and the rationale for selecting them.

- The findings/discussion section presents the findings of the research by comparing data, exploring similarities and differences, and trying to account for these similarities and differences with other research.

- The conclusion section summarises the key themes of the dissertation and draws out implications for practice or recommendations for further research in the area and discusses limitations of the dissertation.

The reason for testing the hypothesis that there would be variation in the use of lexical cohesive devices in assignments from different grading bands was based on the findings from previous studies discussed in section 5.3.3 above. These studies focused on investigating to what extend the use of lexical cohesion is influenced by NNS students' language proficiency (e.g. Chen, 2007; Jin, 2001). This study intended to follow this tradition to examine whether there would be

variation in the use of lexical cohesion in Chinese students' assignments from four grading bands which indicate various language proficiency of students.

However, the samples in previous studies were from EFL or EAP classes whose focus was English language skills, while the samples in the current study were from MA module assignments which were evaluated not only based on language skills but also other criteria, such as the connection between the contents and the topic, the critical review of prior research and so on. Therefore, the appropriateness and effectiveness regarding the use of lexical cohesion were not considered much when marking MA assignments. This indicated that the variety of language proficiency might not be reflected well by the division of grading bands.

The rationale for investigating whether there would be variation between sections in the dissertations was to examined whether writers make different choice of lexical cohesive devices for sections with different functions during the flow of an extended text. However, the fact that dissertation samples were taken from beginnings of texts in this study might reduce the likelihood of variation between functional-section groups as the function of beginnings of each section in dissertation samples is similar, i.e. introducing the contents of the sections briefly. Furthermore, other factors which were not considered in the current study could also influence the results of the comparison. For example, it is possible that topic, genre, text type, methodological approach and writer background and experience might also impact the use of lexical cohesion, as with the findings from previous studies (Chen, 2007; Jin, 2001).

6.3 Pilot study

There were three reasons for doing a pilot study for this research. Firstly, the context-sensitive nature of lexical cohesion requires development of a detailed framework for analysis of lexical cohesion for the present study; secondly, the specialised small-sized corpora engender a practical set of strategies for the analysis of lexical cohesive devices used in the samples; thirdly, the time limitation of the present project requires the selection of proper sample sizes for

the full-scale analysis. Based on these three reasons, the objectives of this pilot study are as follows:

1. To decide how large a sample size would be appropriate for realistic analysis as the manual analysis approach adopted in this research is time-consuming.

2. To develop a basic operational classification of lexical cohesion for this study.

3. To build up a coding system and specify analytical strategies which are needed in the full-scale analysis at a later stage.

To achieve these objectives, it was necessary to determine how to conduct a pilot study in general. A pilot study in Applied Linguistics refers to “a dress rehearsal” of the full data collection procedures (Dörnyei, 2007, p. 75). There are several possible purposes for this ‘dress rehearsal’ piloting, such as:

1. to try out research instruments (e.g. interview and questionnaire (Burns, 1999);

2. to assess the practicality of data collection procedures (Bryman, 2001);

3. to identify problems before doing the actual study (Mackey & Gass, 2005);

4. to enhance validity and reliability of the research instruments (Cohen, Manion, & Morrison, 2007).

In this study, because of the manual analytical approach adopted, the analysis process required substantially more human analytical work than computer-assisted analysis. Therefore, there were two pilot studies conducted prior to the full-scale analysis in order to adapt and fine-tune the research instrument (i.e. the lexical cohesive categories and the tagging system for analysing lexical cohesive pairs in tables) and the analytical strategies, for example, supporting

determination of the boundary of each analysed segment, and the necessary size of the corpus.

In pilot study 1, three approximately 3000-word samples were chosen from Chinese students taking the MA programmes noted above, and these samples were analysed adopting Halliday and Matthiessen's (2004, p. 578) discourse-based analytical procedures of text analysis for lexical cohesive devices:

- Build up tables of categories of lexical cohesion as column headers and divide a text into clauses;

- Put lexical cohesive devices into six categories (i.e. repetition, synonymy, identity, hyperonymy, meronymy and collocation);

- Use italics and bold fonts to represent presupposed and presupposing items[29] in a connected tie (a term which refers to a pair of lexical cohesive items (Halliday & Hasan, 1976, p.3)

Table 15 Sample analysis in pilot study 1

	rep[30]	syn.	ide.	sup.	mer.	col.
Vocabulary has always been a central **subject** for *language* teachers,	language = language	vocabulary – words	vocabulary – subject			
for **language**, as				vocabulary – language		

[29] In a cohesive pair, there are two lexical items. Presupposed item refers to the item that appears before the other item; the presupposing item refers to the item which appears after the presupposed item.
[30] Rep. = repetition; syn. = synonymy; ide. = identity; sup. = superordination; mer. = meronymy; and col. = collocation.

Thornbury states, “emerges as **words**”.						

Excerpt from a Chinese student’s assignment

In the example shown in table 15, *vocabulary* is the item presupposed and **subject** is the presupposing item. Both of the items constitute a tie in the category of identity.

Analysis of lexical cohesive device frequency in the 3000-word samples in pilot study 1 generated the data shown in table 16 below.

Table 16 Normalised frequencies of lexical cohesive pairs in six categories (per 3000 words) in three samples

	repetition	synonymy	identity	superordination	meronymy	collocation
sample 1	715	71	19	15	17	41
sample 2	754	14	7	9	4	33
sample 3	729	11	10	9	5	27

From the figures shown in this table, the numbers of cohesive pairs in sample 1 (distinction) are significantly higher than that in sample 2 (merit) and sample 3 (merit), except in the repetition category, where the figure in sample 1 is lower than the other two. This could be seen as providing a highly preliminary indication that there might be a positive correlation between the marking-scale of the assignment and the variety of lexical cohesive devices adopted in the assignment.

Importantly, the level of analytical work required for this pilot study demonstrated that, due to the high level of effort and time required for the analysis, a 3000-word document was too long for manual lexical cohesive analysis if significant numbers of texts were to be analysed. Due to the nature of lexical cohesion and the complex and time-consuming analysis involved in qualitative text analysis, therefore, the length of the samples used in this study was reduced to 300-word excerpts from students’ assignments allowing analysis of multiple excerpts from

different writers. Developing the pilot study, 10 further excerpts with varied marks were selected from 10 samples in the same corpus noted above, which are written assignments from a Syllabus Design and Assessment module for MA TESOL programme. These excerpt samples were analysed to examine the applicability of the size of 300-word texts at the following stage of this study. Table 17 shows the results of this analysis below:

Table 17 Nomalised frequencies (per 300 words) of lexical cohesion in 10 excerpts of students' assignment samples

	mark	repetition	synonymy	hyperonymy	hyponymy	meronymy	identity	collocation
sample 1	29%	32	2	0	9	2	2	10
sample 2	48%	34	2	3	3	2	2	2
sample 3	52%	40	2	0	5	3	2	1
sample 4	54%	20	4	2	4	2	2	5
sample 5	54%	33	5	1	0	3	1	3
sample 6	68%	35	6	2	5	3	1	3
sample 7	71%	23	3	1	0	0	2	3
sample 8	71%	31	2	2	1	1	3	0
sample 9	72%	29	4	0	1	2	1	7
sample 10	76%	18	3	1	1	4	3	1

As can be seen in table 17, the framework of lexical cohesive devices was developed from 6 categories involved in pilot study 1 (see table 15 above) to 7 categories in pilot study 2. The reason for this adjustment is that the superordination category included two lexical cohesive relations in pilot study 1 (i.e. part/kind – whole and whole – kind), which might cause confusion in terms of the identification of lexical cohesive relations between two lexical items . For clarity and convenience in further analysis of lexical cohesion in the present study, this original superordination category in pilot study 1 was divided into two categories in pilot study 2, i.e. hyperonymy (part/kind – whole relation) and hyponymy (whole – kind relation).

Through the analysis of samples included in pilot study 2, it could be confirmed that 300-word length was more feasible for manual analysis. For the analytical strategy, the use of bold and italics to indicate forms of cohesive pairs was found to be insufficient to clearly indicate the range of lexical cohesive categories and therefore a more complex indicative system was developed for further full-scale manual analysis, which is described below.

From both pilot studies 1 and 2, some patterns of usage of lexical cohesive devices between the different categories were clearly present: the repetition category had the highest number of lexical cohesive pairs (or 'ties' in Halliday and Hasan's term (1976)) of lexical cohesion, while in the other categories the frequencies of lexical cohesive pairs were varied between samples. The distribution of lexical pairs observed in the pilot studies was in line with the majority of previous studies (e.g. Tanskanen, 2006; Hoffman 2012). This feature of the distribution of lexical cohesive pairs in different categories in the pilot studies as well as in previous studies also hypothesised that the students tend to use more repetition devices in their assignments than other cohesive devices.

6.4 The work of co-raters

As this study adopted manual qualitative analysis for the identification of cohesive text features, subjectivity could not be avoided. To reduce the influence of research subjectivity and to provide further judgments as to whether the adopted classification of lexical cohesion were sufficient, accurate and capable of

implementation, two co-raters were recruited and involved at an early stage of the analytical process. The two co-raters had expertise in applied linguistics and education, and also had basic knowledge of linguistic cohesion through having finished their MA TESOL programme at the same university as the sample writers had and through having studied a discourse related module. They were familiar with the register and language in the samples, which helped them in understanding the content in the samples and enabled them to work more efficiently in identifying lexical cohesive text features.

Both of the co-raters were sent a copy of the introduction to analytical strategy sheet[31] in Chinese, as the researcher and the co-raters were native Chinese speakers. After reading the introduction, they were trained via email and WeChat (a Chinese social software like WhatsApp). Firstly, they read through the analytical strategy document and were asked questions where necessary to ensure thorough understanding of the analytical strategies. Then one sample was sent to both co-raters at the same time. After one week, the analysed samples were sent back to the researcher to enable assessment of the levels of inter-rater agreement in this sample analysis. This meant that three versions of lexical cohesive analyses of one sample could be compared using Fleiss' [32] generalisation of kappa to more than two raters. Despite several hours of joint discussion and one week allowed for individual analysis, satisfactory inter-rater agreement was not reached with the inter-rater reliability calculated as below 0.5: kappa ≈ 0.413 (Hallgren, 2012, p. 28). Zaiontz (2019) has also pointed out that there are no clear-cut rules which indicate to what extent that the levels of agreement are good or bad. However, in general, the set of criteria is:

less than 0 = no agreement

[31] See appendix B: both the Chinese version and English version.

[32] "The statistic kappa was introduced to measure nominal scale agreement between a fixed pair of raters" (Fleiss, 1971, p. 378). There are two typical kappa definition: Cohen's kappa and Fleiss' kappa. The former is for measuring the inter-rater reliability between two raters who evaluate each subject. The latter is used in the case of more than *m* raters ($m \geq 2$) involved and it is not necessary for each rater to evaluate each subject. "What is important is that each subject is evaluated *m* times" (Zaiontz, 2019). Therefore, as in this case, three raters evaluated each subject, Fleiss' kappa was appropriate to measure the inter-rater reliability. The original kappa is 0.413.

0-0.2 = poor

0.2-0.4 = fair

0.4-0.6 = moderate

0.6-0.8 = good

0.8 or higher = very good (adapted from Zaiontz, 2019)

“1” indicates perfect agreement while “0” indicates that any agreement is entirely by chance.

Table 18 Kappa[33] for measuring inter-rater reliability of sample MR1[34] by three raters

Fleiss's Kappa									
	total	repetition	synonymy	hyperonymy	hyponymy	meronymy	identity[35]	collocation.	none[36]
kappa	0.416	0.696	0.240	-0.008	1	-0.004	0.592	0.189	0.125

[33] The calculation of kappa is assisted by Excel add-in software *Real Statistics Using Excel* (Zaiontz, 2019).
[34] Sample MR1 refers to the sample which is the first sample in the merit marking-scale group and analysed by the researcher and two co-raters.
[35] *Identity* refers to the category of other relations with identity of reference
[36] *None* refers to that the rater did not recognise the lexical item with cohesive relations in text.

It can be seen from table 18 that there were three categories whose kappa scores were above 0.5 (i.e. repetition, hyponymy and identity); three categories were between 0 and 0.5 (i.e. synonymy, collocation and none); and two categories below 0 (i.e. meronymy and hyperonymy):

Table 19 Kappa for measuring inter-rater reliability in each category in three scale ranges

$k \geq 5$	repetition (k = 0.696) hyponymy (k = 1) identity (k = 0.592)
$0 \leq k < 5$	synonymy (k = 0.240) collocation (k = 0.189) none (k = 0.125)
$k < 0$	meronymy (k = -0.004) hyperonymy (k = -0.008)

It seemed like three co-raters had good agreement only on the category of repetition. In fact, this result was anticipated by the researcher because firstly, the other two co-raters, while familiar with the concept of lexical cohesion, did not have detailed and extensive knowledge about lexical cohesion, a complex phenomenon, even though the explanation of each category of lexical cohesion was elucidated in the analytical strategy sheet. Secondly, judgements of lexical cohesion tend to involve a degree of subjectivity, especially in the collocation category. Therefore, after the trial practice, more discussion was undertaken amongst the co-raters. The definition of each category was explained with more examples and the analytical strategy was revised. The analysed unit was changed to lexical item, not word, which meant several words could form one lexical item (e.g. *business English*). Such that, when analysing these items, co-raters were instructed to treat them as one unified lexical item when necessary. For example, *business English* could be regarded as one form of repetition rather than two separate repetitive items.

Even though one week was allowed for co-rater analysis, it is worth pointing out that both of the other raters had a range of other commitments, which meant that,

in reality, they only had limited time for this sample analysis work. In the second co-rating stage, due to this time limitation, and bearing in mind their other commitments, only 5 samples in total were sent to these two raters, 3 to co-rater A and 2 to co-rater B. As an important purpose of this co-rating stage was to examine the level of subjectivity influencing analytical results, the inter-rater agreement needed to be examined again after this second co-rating process. Table 20 shows the results of this second examination:

Table 20 Kappa for measuring inter-rater reliability in other 5 samples analysed by rater 1 and rater 2, and rater 1 and rater 3[37]

sample	sample A (rater 1&3)	sample B (1&3)	sample C (1&2)	sample D (1&2)	sample E (1&2)	average
kappa[38]	0.812	0.836	0.721	0.697	0.616	0.737

Table 20 indicates that after further discussion and training, the inter-rater agreement figures increased, achieving 0.74 on average overall. In this comparison, the inter-rater agreement figures between rater 1 and 3 were higher than that between rater 1 and 2. The reason was that in the category of repetition, rater 3 was more accurate than rater 2 who made some noticeable mistakes when identifying items in this category. It was understandable that rater 2 made mistakes in repetition analysis, as there were many more pairs in this category compared with the remaining categories. A possible speculation about this result is that when analysing the latter part of the text sample, raters would easily forget the lexical items in the previous part of the sample and sometimes could not recognise the repetitive pairs. This was also one of the reasons for those engaged in analysis of samples to check one sample through several times to avoid such simple mistakes.

Whether the application of inter-rater reliability in applied linguistics is sufficiently helpful has been questioned, as the data studied in this co-rater research area "never has an ultimate truth" (Hoek & Scholman, 2017, p. 2), which means the

[37] Rater 1 is the researcher, rater 2 is associate rater A and rater 3 is another associate rater B.
[38] Here kappa refers to Cohen's (1960).

judgment of the data always involves subjectivity. In fact, in many cases, linguistics researchers focus on “gradient phenomena where there are no right answers and where it is not uncommon for data to be ambiguous” (Hoek & Scholman, 2017, p. 2). For example, in relation to lexical cohesion research, a cohesive pair, *boy* – *girl*, can be collocational and antonymic at the same time, and researchers may decide to regard their relation as antonymic based on their semantic meanings, or treat this pair as collocational based on their contextual meanings. In this study, the work of co-raters can only help enhance the reliability of the result of such analytical method, but it cannot ensure that an interpretation is accurate. Furthermore, “a key assumption [of inter-rater reliability] is that the [raters] act independently, an assumption which isn’t easy to satisfy completely in the real world” (Zaiontz, 2019). For example, in this study, although the analysis work was conducted by raters individually, following this stage, if there were disagreements, the raters would negotiate and reach final agreement. In most cases, the key rater (the researcher) convinced the other two raters to agree on her own analysis result. The researcher inevitably became the authority during the after-analysis discussion due to her deeper knowledge of the area of lexical cohesion. Therefore, it was really difficult to satisfactorily conduct independent analysis in the real world context, and the co-rater method can only be seen as supporting the reliability of the outcome or analytical method to a small degree.

Additionally to this somewhat developed analytical reliability, this rater work also contributed to other aspects of this study. Based on this rater work, a further conclusion was reached for the full-scale analysis. It was decided that the cohesive relations between adjacent lexical items would not be counted in the analysis anymore, as the cohesive force between these items were already reinforced by the grammatical boundary between clauses. Besides, the manual analysis was extremely time-consuming. If the workload could be reduced, this would be more practical for the full-scale analysis.

6.5 Data segmentation

The following is a condensed overview of the main steps towards adapting the data of students’ module assignments and dissertations to the needs of the linguists.

As noted in section 6.2, the raw data was collected from two academic programmes. Firstly, ten students' assignments were collected from an existing corpus with the students' permissions, and used for a pilot study in 2016. Once the pilot study was complete, following similar procedures, other students' assignments were collected for co-rater analysis and full-scale analysis. Considering that several studies have already been conducted regarding lexical cohesion and assessment scores, a new research angle was added which was comparison between functional chapters in master dissertations. To conduct this study, dissertation samples needed to be collected. The following selection criteria were used:

1. The dissertations and assignments must be written by students whose first language is Chinese.

2. The dissertations must include all five conventional functional sections, i.e. introduction, literature review, methodology, findings and discussions, and conclusion.

3. The assignment corpus must have equal number of samples in four marking-scale groups: failed, pass, merit and distinction.

The analysed samples were chosen based on these criteria. Overall, based on these selection criteria, there were 9 dissertations and 52 module assignments included in this analysis. The dissertations were further divided into 45 texts according to the functions of the chapters in these dissertations. All the analysed samples are extracts from these selected dissertation and assignment texts, and the 9 texts, illustrated in the text extract (or excerpt[39]) analysis sections in chapters 7 and 8 respectively, are excerpts from these analysed samples.

[39] 'Extract' and 'excerpt' have similar meanings in this study. In order to differentiate texts taken from different corpora, 'extract' is used to refer to texts chosen from the MTMC corpus and 'excerpt' is used for texts taken from the MTDC corpus.

The contents of the samples were selected from the beginning of these dissertation and assignment texts. Because some of the raw data from module assignments was not extensive (nearly half of the collected texts were less than 500 words), for a more applicable comparison and practical analysis, the size of each sample was reduced to the range of 250 to 500 words. After this adaptation, each sample was represented in the format of tables, as shown in table 22 below with each sample/excerpt was segmented into clauses, i.e. each clause[40] was put into one cell. “The tables acted as basic data files and served as a springboard for all of the ensuing steps of the empirical study” (Hoffmann, 2012, p. 98).

As mentioned in the previous chapter, the code for each sample followed the marking-scale group or functional-section group which the sample belonged to, with some additional symbols used to support recognition of individual samples. For example, as some of the samples were initially collected as printed (hard) copies, “paper” or “P” was added after the document’s original name. as already stated in this chapter, for the MTMC corpus, the samples were classified based on their marks. Each table was labelled individually by coded identification tags. For the MTMC corpus, one tag included some or all of the following very basic but recognisable pieces of information:

1. Short form of score classification achieved (i.e. F(failed), P(pass), M(merit) or D(distinction))
2. if the sample was collected as printed (hard) copies, there would be added “P” (for “paper”) or “paper” for the samples in the pass group.

For example, the excerpt “F6P” means the sample was from the sixth excerpt on the list of samples from the failed group and collected as printed (hard) copies.

For MTDC, one tag includes three very basic but recognisable pieces of information:

[40] Here, clause is “defined through its particular syntactic function, i.e. the presence of subject and predicate” (Hoffmann, 2012, p. 73).

1. dissertation number
2. chapter number
3. functional section

For example, D2C1I refers to dissertation 2 chapter 1 in the introduction section. The purpose of these labels was to speed up and facilitate later identification and localisation of samples in the corpora as there were 97 samples in total.

Next in each corpus, each sample was segmented into individual data tables. Each table contained ten different columns instead of the seven in the pilot study (see table 15): one on the left side provided sample content, and other remaining nine columns on the right side presented nine lexical cohesive categories: repetition, synonymy, hyperonymy, hyponymy, meronymy, signalling nouns, identity, antonymy and collocation. The reasons for this further division of lexical cohesive categories are discussed in section 6.8 below.

To make the raw data more analysable, unnecessary or inappropriate contents as regards lexical cohesion analysis was deleted manually, e.g. in-text citations and raw quotations from other sources. The following colour-based analytical system was used for presenting the analysis in tables, as is shown in table 21 below.

Table 21 Highlighting system in the present manual lexical cohesion analysis

lexical cohesion category	example
bold is for repetition	**speaking – speaking**
orange is for synonymy	test – assessment
dark red is for hyperonymy	English – language
green is for hyponymy	language – English
red is for meronymy	English skills – listening
highlight grey is for signalling noun	literature on how to respond to the inappropriateness of arguments clearly in teacher feedback is scant,// and

	pedagogical suggestions for ESL writing teachers on how to deal with such issues could be a subject for further studies.
blue is for identity	test – element
purple is for antonymy	learning – teaching
highlight yellow is for collocation	lecture – student

Based on this highlighting system, the adapted version of lexical cohesion categories is shown in table 22 below, which also demonstrates how lexical cohesive relations are highlighted in the real analysis:

Table 22 Classification of lexical cohesive devices in excerpt D2C6C in the conclusion group

text	repetition	synonymy	hyperonymy	hyponymy	meronymy	signalling noun	identity	antonym	collocation
Chapter 6 **Pedagogical** Implications and conclusion									
6.1 **Teaching English plural morphemes** in a **different way** to **perceive** information through **nouns**									Pedagogical – teaching Plural morphemes – nouns
According to the discussion so far, we can see that there are mainly two reasons for **Chinese advanced learners** to make **errors** persistently on **English Plural Morphemes** in Spontaneous **Speaking** Situations.	English Plural Morphemes								

As for the examples used in the present study, "//" is used to indicate the separation of clauses, while the related lexical item in analysis is emboldened. A hyphen is used to connect two lexical items in a cohesive pair in both the analytical table and in the discussions in the main thesis text, with the exception of simple repetition pairs where only repetitive lexical item itself is included in the analytical table. The colour coding system allows each lexical cohesive category to be recognised in an efficient way.

If one item has more than one relation with other items, this first item will be coloured according to its relation with the item which occurs later in text in order to avoid misunderstanding of the identification of several lexical cohesive relations related to one lexical item. An extract from sample P6 in the pass group will demonstrate this point as well as other types of lexical cohesive relations identified in the text:

6-1 This is a **speaking**-oriented **assessment** tool// which is designed for the **high level English Speaking Contest** qualification trials – the **national** level. //The **test takers** are upper intermediate and advanced **English** learners. //It is within the scope of **high level**'s **English** majors' **spoken test**. //Take IELTS as the **English** proficiency criteria as reference, the **test takers** for this **assessment** should have scored 7 point (each section is no less than 6.5 point) and **higher**. //The **test takers** are Chinese students// who want to compete in the final round in the **national English speaking contest**. // (I) Details of Recourses **[Resources]**[41] // **Test developer[s]** and **Test** writers //Bachman and Palmer points out that **test developers** is one of the most important factors in human **recourses [resources]**.// They are responsible for the **test** specification, management, try-out, achievement and use. //They administrate and take control of the entire **test** process and make sure the **test** can be well carried out. //The **developers** in this **assessment** are not the daily teachers// who are familiar to the **test takers**.

[41] Grammatical typos which tie to lexical cohesion in the original sample text are corrected in "[]" where applicable.

In this extract, *students* (line 7) in *Chinese students* has two relations with two different lexical items, i.e. *students – teachers* as an antonymic relation and *Chinese students – test takers* as an identity relation. The reason for *students* being finally coloured in blue is that *test takers* appeared later in the text than *teachers*. Therefore, in line with the system applied, *students* is coloured according to its relation with *test takers* instead of *teachers*.

There are further other lexical cohesive relations involved in this extract, which are shown in table 23 as follows:

Table 23 Lexical cohesive pairs in the extract of sample P6

repetition	synonymy	hyponymy	identity	collocation
speaking	assessment – test	speaking-oriented assessment tool – English majors' spoken test	the national level – high level	assessment – test takers
English	learners – students	test – IELTS	upper intermediate and advanced English learners – Chinese students	upper intermediate and advanced English learners – English majors
high level				English – IELTS
English				
speaking – spoken				
test				
English				
test takers				
assessment				
high – higher				
test takers				
national				
English				
speaking				
contest				

After the segmentation and tagging of the data, the results of the lexical cohesion frequency analysis in the corpora were then noted and classified according to the lexical cohesive categories to which they belonged as shown in table 22 above. To support comparison, frequencies were normalised to occurrences per 1000 words. The procedure produced the average number of lexical cohesive pairs per 1000 words for the particular cohesive category used in this study. The corpus was now open to further analysis evaluation and interpretation, these being discussed in chapters 7 and 8.

6.6 Manual analysis and evaluation of the data

Compared with grammatical cohesion comprising clear-cut formal indicators (e.g. personal pronouns) which can be easily searched with concordance tools, lexical cohesive relations, as already discussed through the literature review chapters, are based on semantic grounds, which cannot (at least at the moment) be identified with the assistance of such concordance tools (Hoffman, 2012, p. 101). Thus manual analysis of lexical cohesion was conducted and this was time-consuming work.

Despite the existence of degrees of gradience between different approaches applied in corpus-based research, there are generally accepted as being three main approaches to corpus study (Bednarek, 2009) – small-scale corpus analysis, large-scale corpus analysis and manual analysis of individual texts. As the corpora in this study was relatively small-scale identified as being fewer than 100,000 words (Ghadessy & Gao, 2001), the approach in this study could only use small-scale corpus analysis and manual analysis. The benefit of this approach is that the researcher can become familiar with each sample and individual examples can be analysed in their micro- and macro contexts. Because of this familiarity, it was possible to conduct “more complex, richer interpretive, dynamic, and flexible analysis” than is the case with large-scale corpora (Flowerdew & Forest, 2015, p. 71). Specifically, every lexical item in each text was checked regarding its relation with other items in the same text (Tanskanen, 2006, p. 44). As this work was time-consuming, the corpora size had to be reduced to a workable level. The manual examination and analysis of the data

sets is now to set out in detail. First, a sample of the actual analysis is displayed below in table 24:

Table 24 Tables for the manual analysis of lexical cohesion (from sample F9 in the failed group)

text	repetition	synonymy	hyperonymy	hyponymy	meronymy	signalling nouns	identity	antonymy	collocation
Part One Target Group									
This **assessment** tool aims at a **group** of **business-major students**	group		business – major				target group – students		
who are joining an **English training** program,									students – training program
after which they are going to become **interpreters** and **bilingual sales agents** in an **international furniture fair**.							business-major students – interpreters and bilingual sales agents		English – interpreters business – sales agents

As it can be seen in table 24, the left-hand column of the analysis sheet specifies the clauses in the sample while the other columns demonstrate the lexical cohesive categories of the relations and their corresponding cohesive pairs found within the text. The analysed lexical items in column one (left) were coloured or highlighted in the analysis for the convenience of retrieval and analysis. Each pair recognised was allocated to its corresponding category column, and a hyphen was introduced between elements in that pair. The cohesive pairs in a clause were placed in the cells which were in the same row of that clause cell. For example, in table 24 above, because *interpreters* and *sales agents* were in the same clause, the collocation pairs *English – interpreters* and *business – sales agents* were placed in the same cell in the same column.

6.7 Statistical methods in the corpora

In order to investigate firstly the relation between the use of lexical cohesion and the marks achieved in Chinese students' writing, and secondly the relation between the use of lexical cohesion and the functional sections in dissertation texts, several statistical tests were conducted and implemented using the SPSS 24.0 programme in the comparative analysis of the two corpora. Specifically, four steps were taken at this statistical test stage :

Firstly, the analysis of variance (ANOVA) test was used to evaluate the variance of the use of lexical cohesive devices between different groups in two corpora respectively. As a parametric test, the ANOVA test assumes that the distribution of data on the dependent variable is normally distributed (Pallant, 2016, p. 255). In this case, a typical normal distribution test, the Kolmogorov-Smirnov (K-S) test, was chosen to assess the normal distribution of frequencies of cohesive pairs for the lexical cohesion variable in each corpus. A non-significant result of the K-S test (**Sig.** value of more than 0.05) indicates normality (Pallant, 2016, p. 63). Furthermore, as parametric tests also assume that "samples are obtained from populations of equal variances, a Levene's test was taken for equality of variances assessment between groups. This means that "the variability of scores for each of the groups is similar" (Pallant, 2016, p. 208). If a significance value of greater than 0.05 in the Levene's test, this suggests that variances for the groups are approximately equal.

Secondly, when meeting the two requirements of normal distribution and homogeneity of variance in the data, an one-way between-groups ANOVA test was conducted in the corpus to compare the mean scores of different marking-scale groups or functional-section groups regarding the use of lexical cohesion in each group. The choice of ANOVA test was made because the condition in which this test can be conducted was achieved by the data in the corpora: the independent variable, the marking-scale groups or the functional-section groups, has three or more groups; and the dependent variable, the lexical cohesive devices, is a continuous variable. If either of the requirements mentioned above was not achieved in the K-S test and in the Levene's test, which indicates that at least one of the assumptions was violated, a non-parametric alternative to the one-way ANOVA test, the Kruskal-Wallis (K-W) test, was adopted to compare the scores on the lexical cohesion variable for the marking-scale groups or the functional-section groups. In the K-W test, "[scores] are converted to ranks and the mean rank for each group is compared" (Pallant, 2016, p. 236).

Thirdly, as both the ANOVA test and the K-W test can only indicate whether there are significant differences in the mean scores on the dependent variable across different groups, if the significance value was less than or equal to 0.05, there was a significant difference somewhere among the mean scores on the dependent variable for the groups. In order to find out which group was different from which other group, post-hoc tests were taken (Pallant, 2016, p. 256). The recommended post-hoc test for the ANOVA test is the Turkey's HSD post hoc test (Hinton, Brownlow, McMurray, & Cozens, 2004, p. 169) while that for the K-W test is the follow-up Mann-Whitney U tests between pairs of groups (Pallant, 2016, p. 240).

Fourthly, as for the investigation of the correlation between the use of lexical cohesion and the marking scale groups in the MTMC corpus, Spearman's rho correlation test was used as the marking scale groups were regarded as an ordinal variable rather than measured on an interval scale (Hinton, Brownlow, McMurray, & Cozens, 2004, p. 300).

6.8 Analytical strategy

Based on the pilot studies and co-rater work, the following final analytical strategy was developed for the full-scale analysis:

1. The smallest unit of analysis was the lexical item.
2. Only two lexical items could form one cohesive pair.
3. The scope of cohesion was interclausal.
4. The nearest items (across the clause) form one cohesive pair.
5. One lexical item could have more than one cohesive relation.
6. Nine categories (as specified in table 24) were included in the analysis.

With regard to the boundary of the basic unit of analysis in the present study, the term 'lexical item' is used as the basic unit, which acts as a cover term including single words, phrasal verbs, idioms and other multi-word units. As Tanskanen (2006) pointed out, if putting lexical cohesion, a dynamic textual device, into the scope of a single orthographic English word as the basic unit of analysis, lexical cohesion would be analysed "within the limits of a highly conventionalised system, namely orthography" (p. 9). However, the lexical units involved in lexical cohesive relations, are not orthographically restricted. In fact, it was considered that the number of words in a lexical unit is irrelevant to the cohesive function of the lexical unit (item)[42] (Halliday, 1994, p. 311). Therefore, the first rule in this study was that the smallest unit of analysis was the lexical item rather than words in terms of an orthographical definition.

As this study involved a discourse-based approach, the identification of lexical cohesion was determined by contexts, which meant one pair of lexical items might be identified as cohesive in one context but not cohesive in another (Tanskanen, 2006, p. 10). With regard to the specific identification of cohesive pairs, the second rule stated above was that three related lexical items would be divided into two cohesive pairs. (Tanskanen, 2006, p. 50). In example 6-2 (from

[42] In Tanskanen's (2006) study, 'lexical unit' was the basic unit of analysis. However, in the present study, the term 'lexical item' was adopted to refer to the same unit of analysis, since the term 'lexical item', which originally came from Halliday and Hasan (1976), was seen as a more straightforward term, more amenable to understanding and clearly expressed the distinction between the notions of 'lexical item' and 'word'.

sample F9) below, three lexical items formed two cohesive pairs. In order to recognise the types of cohesive relation more easily, the highlighted lexical items in examples follow the highlighting code for different lexical cohesive relations.

6-2 In 3F, leading furniture manufacturers from mainland China, Hong Kong and Taiwan will take part in this exhibition, seeking for new contracts with byers [buyers] all over the world. // During exhibition time, which lasts for 14 days semi-annually, // about 80 interpreters and bilingual sales agent[s] are needed. // To meet the needs of trading companies and provide them with better interpretation services, the exhibition committee cooperates with business faculty in Sun Yat-sen University (SYEU), // which is one of the top universities in southern China, // to provide internship opportunities for their business-major students.

In this example of identity relation, *trading companies* included both *manufacturers* and *buyers*. Therefore, this co-referential cohesive pair was *manufacturers/buyers – trading companies*[43].

The third rule related to the scope of cohesion analysis. This study only analysed interclausal relations "as intra-clausal lexical cohesive relations [were] less important in creating texture when clause as a grammatical structure [was] already a good signal for cohesive effect" (Tanskanen, 2006, p. 50).

The fourth rule was that if the relation of two items has been already analysed once, even if these two items were repeated in a succeeding part of in the same text, this relation would not be counted as another pair of lexical cohesion.

Since the scope of cohesion in this study was interclausal relations. It was necessary to clarify the boundary of clause. As mentioned in section 6.5, the definition of clause was 'subject + predicate'. As a tradition from Halliday and Hasan (1976), Hoffman's (2012, p. 89) study also focused on interclausal

[43] Since manufacturers and buyers occurred in the same clause, they were not counted as co-hyponyms in this analysis.

cohesive relations rather than intraclausal ones. Three examples from his study are shown below in order to demonstrate the difference between intraclausal and interclausal levels:

6-3 Fiachna called, after spending most of the day (2pm – 7:30pm) at the local hospital, helping out **one of the little Irish girls** – **a friend of his daughter**.

6-4 **The owners**, **Claude and Zaina**, have not updated its classic 19th century bistro style interior,

6-5 Apparently Earth has **a natural soundtrack**: **a high-pitched series of chirps and whistles, that could, potentially, be heard by aliens** if they knew how to listen.

(Hoffmann, 2012, p. 89)

Examples 6-3 and 6-4 both include intraclausal types of synonymous relations[44]. In contrast, example 6-5 presents an interclausal relation – the two clauses being connected by a colon (Hoffmann, 2012, p. 89). To make the analysis more clear and convenient, as mentioned in section 6.6, clauses were allocated to table cells in the analysis in the present study, which also helps distinguish between intra- and inter-clausal cohesive relations.

The fifth rule was that if an item *a* could have similar relations with other two items *b* and *c* (*a* was geographically closer to *b* than to *c*), and items *b* and *c* happened to be in the same clause, only the pair that was formed by *a* and *b* would be counted as a lexical cohesive pair in this analysis. For example:

6-6 This assessment tool aims at a group of business-major students // who are joining an English training program,// after which they are going to become interpreters and bilingual sales agents in an international furniture fair.

(F9)

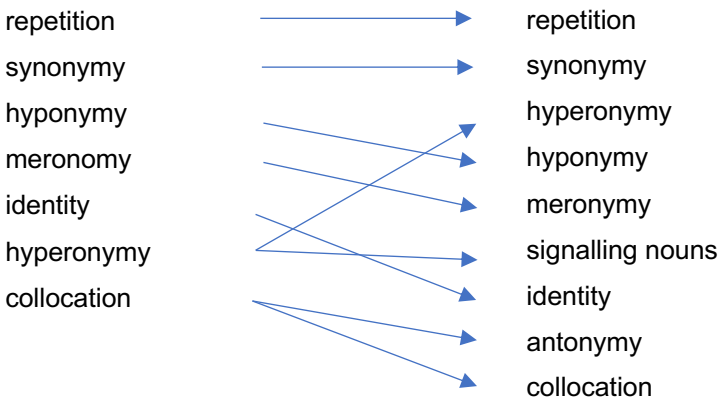
[44] They were regarded as synonymous relations in Hoffman's model while in the present study they were types of identity relations.

In this example of collocation relation, *business* was the item *a*, *sales agents* was *b* and *fair* was *c*. The three items were in the 'business frame' to form collocational relations. However, as the proximity between *business* and *sales agents* was closer than that between *business* and *fair*, only the *business – sales agents* pair was counted in this analysis. Martin adopted a similar strategy in which the latter item was seen as referring back to the nearest related item rather than all the preceding related items (Tanskanen, 2006, p. 46). The difference between Martin's approach and the present study was if the related items were nearer and occurred in the same clause, their relation would not be counted in this study, as in example 6-6 *sales agents – fair* . By contrast, in Martin's strategy, no matter whether the scope of cohesion was interclausal or intraclausal, the nearest related items formed a cohesive pair. Nevertheless, both strategies are considered efficient in terms of analysis. The point is to tailor the scope of cohesion in a manner which is suitable for a particular study.

The sixth rule was to allow one lexical item to have more than one relation with other items across clausal boundaries. The reason for this was related to the texture of text. Texture in cohesion is an extended meaning of texture "as the warp thread and weft of a piece of textile" (Renkema, 2009, p. 10, cited in Hoffman, 2012, p. 71). The extended meaning of texture in text combines two meanings: one is the structure of components in text, and another is the characteristics or quality of text. This texture was regarded as "the essence as a result of interwoven segments" in text (Renkema, 2009, p. 10, cited in Hoffman, 2012, p. 71), which can reasonably be seen as the target of discourse study. As the cohesive force created by the segments in text contributes to the creation of texture, the relation between the segments which are expressed by lexical items in text needs to be analysed in detail.

The seventh rule relates to designation of the lexical cohesive categories expanded from seven to nine (the original seven categories of lexical cohesion were based on Halliday and Hasan's model of lexical cohesion (1976)), as shown in figure 10 below:

Figure 10 Designation of lexical cohesive categories expanded from seven to nine



The reason for dividing the original hyperonymy category into hyperonymy and signalling nouns was that in the pilot studies and rater analysis, signalling nouns were salient in the samples and these were therefore seen as being worthy of separate investigation. For the separation between collocation and antonymy, on the one hand, this study followed the tradition after Halliday and Hasan's work (1976) that antonymy was separate from collocation; on the other hand, collocation in this study refers to non-systematic relations while antonymy was similar to synonymy in that both of them were seen as systematic relations.

The seven rules above are adaptions developed following previous trial study stages – pilot study and co-rater analysis. These basic rules formed the main analytical strategy in the full-scale analysis, aimed at supporting the efficacy of the manual text analysis.

6.9 Summary

This chapter has presented a description of the methodology used in the present study. The construction of the corpora used in this study has been described alongside their different purposes with MTMC developed as a means of focusing on the relation between the use of lexical cohesion and the quality of students' writing, and MTDC constructed to examine the distribution of lexical cohesive devices in the flow of dissertations. In order to determine the sample size, the operational framework of lexical cohesion and the coding system as well as the analytical strategy of the full-scale analysis of the present study, two pilot studies

were carried out. The detailed procedures involved in the pilot studies have been described in this chapter. In addition, in order to evaluate the feasibility of the analytical strategy of this study, two-staged rater work was also conducted. Levels of inter-rater reliability were examined based on the results of analyses accomplished by three individual raters.

Computer-based analysis and manual analysis were compared, and based on this comparison, this chapter provided a rationale for using manual analysis in the present study. The selection of statistical tests for the quantitative analysis of two corpora was also introduced and discussed. Finally, based on the pilot studies and co-rater analyses as well as theoretical principles, data segmentation and the final developed analytical strategy have been demonstrated with seven rules described which underpinned the analysis of lexical cohesion applied in this study.

Having examined the main methods and analytical strategy used in this research, this thesis now presents a detailed and in-depth examination of the lexical cohesive devices used in Chinese students' writing in the two corpora.

Chapter 7 Analyses of lexical cohesive devices in the MTMC corpus

7.1 Introduction

This chapter sets out to achieve two of the thesis objectives. The first objective was to compare the frequencies and features of different lexical cohesive devices in four marking-scale groups of samples (i.e. failed, pass, merit and distinction) in the MTMC[45] corpus in order to establish whether there are similarities and differences in lexical cohesion in these marking-scale groups. Arising from and to some degree overlapping with the first objective, the second objective was to use statistical tools to investigate the correlation between the use of lexical cohesion and marks achieved for Chinese students' assignments.

More specifically, this chapter first identifies the frequencies of lexical cohesive devices identified in MTMC presenting an analysis of the distribution of each lexical cohesive category in the samples. As quantitative analysis of MTMC only provides macro-level information about the use of lexical cohesive devices, qualitative analysis is also needed to take the investigation to the micro level (Tanskanen, 2006, p. 94). Thus, following the discussion of the normalised frequencies of lexical cohesive devices identified in MTMC in section 7.2, section 7.3 presents the results of qualitative analysis of four extracts selected from each of the four marking-scale groups.

An additional purpose of this detailed analysis is to describe how Chinese students use lexical cohesive devices to create lexical cohesion in their assignments, for which examples are identified in the extracts. The final section of the chapter summarises the overall findings and presents conclusions derived from comparison of the findings regarding different marking-scale groups.

[45] MA TESOL Module assignment Corpus

7.2 Results of the MTMC analysis

The frequency figures shown in the tables discussed in this chapter are, where relevant, normalised frequencies[46] of the cohesive pairs occurring in the texts, which enables direct comparison of figures between texts.

As mentioned in chapter 6, there are 52 samples in MTMC, 13 in each marking-scale group. Table 25 below shows the normalised number of lexical cohesive pairs in each category across marking scale groups in MTMC[47], including the maximum, minimum and mean numbers of pairs of the nine lexical cohesive categories found in each group. The topics of the samples naturally vary, but all lie within the scope of Applied Linguistics and TESOL.

What seems clear from table 25 is that the four marking-scale groups contain similar levels of mean numbers of lexical cohesive devices in all eight categories except in repetition. Specifically, three groups, failed, pass and distinction, have similar numbers of repetition pairs, while that in the merit group is much less. This difference in the repetition category has a significant influence on the overall comparison between the four groups, i.e. the merit group has fewer lexical cohesive pairs that the other three groups.

[46] The raw distributions of lexical cohesive devices in MTMC can be found in appendix D.
[47] The detailed normalised frequencies of lexical cohesive devices in MTMC can be found in appendix E.

Table 25 Normalised frequencies (per 1,000 words) of lexical cohesive categories in marking-scale groups

	failed			pass			merit			distinction		
	max.	min.	mean	max.	min.	mean	max.	min.	mean	max.	min.	mean
repetition	227	109	163	224	101	159	177	85	136	205	123	159
synonymy	29	3	11	27	4	14	25	3	10	20	4	13
hyperonymy	30	0	9	15	2	9	22	0	8	28	0	8
hyponymy	67	3	20	37	0	16	27	0	15	42	9	20
meronymy	15	0	5	20	0	6	11	0	4	30	0	8
signalling noun	21	0	8	14	4	10	20	0	7	30	0	8
identity	13	0	6	16	0	8	13	0	7	14	3	8
antonymy	15	0	6	12	4	7	20	0	5	14	0	6
collocation	45	16	28	45	8	24	50	11	25	40	9	23
total			256			254			218			253

It also can be seen from table 25 that repetition relations are far more frequent than the remaining relations, an observation which is in line with the results of the majority of previous studies (e.g. Chen, 2007; Jin, 2001), which show that Chinese students adopt repetitive devices as the main cohesive devices in their English academic writing. Furthermore, according to the minimum numbers in each group, apart from the categories of repetition, synonymy and collocation, there were no pairs identified in some other categories in the samples from different marking scale groups. This reveals that Chinese students' writing in these samples does not reflect variety of lexical cohesive devices no matter which marking scale category their assignments belong to. This finding is different from some other studies (e.g. Wu, 2010) which claim that advanced learners show more sophistication in the use of lexical cohesive device than lower level students do.

Another interesting point is that the frequencies of meronymic devices are the lowest in all four groups. The reason for this observation may lie in the view that meronyms are typically used in texts which describes entities, such as the guidebook for an exhibition centre referred to by Halliday and Matthiessen (2014, p. 648), while the samples in MTMC are from MA TESOL students' module assignments which describe the information about learners for a syllabus design or introduce the design statement of assessment tools. Another possible reason might be that meronymic relations might be rarely used in academic writing.

Regarding the levels of standard deviation (see table F1 in appendix F), the sequence (from the lowest to the highest) is: merit, distinction, pass and failed. This means in terms of using lexical cohesive devices, the variability of the samples in the failed group is the highest, while that of the samples in the merit group is the lowest. This is an interesting finding, however, considering the relatively small size of this corpus, further investigations need to be conducted before reaching a valid conclusion.

In order to investigate the relationship between the number of lexical cohesive devices and writing quality, after checking the normal distribution and homogeneity of variance of the data in the MTMC corpus through the K-W test

and the Levene's test (see tables F2 and F3 in appendix F), one-factor ANOVA test (see table F4 in appendix F) were conducted to determine whether there was significant difference in the frequency of lexical cohesive devices (total normalised number of devices per sample) between the four marking scale categories. The result of the ANOVA test ($F = 4.675$, $p = 0.006 < 0.05 = \alpha$) indicates that there is statistically significant difference between module assignments in four marking-scale groups regarding the frequency of lexical cohesion. Contrary to the findings claimed by some previous studies (e.g. Zhang, 2000; Chen, 2007), this finding was unexpected and might be partly due to the fact that, as noted above, the lexical cohesive pairs in the merit group appear much fewer than in the remaining three groups. This comparative result was further investigated using the Tukey's HSD test (see table F5 in appendix F). In terms of the mean number of lexical cohesive devices, there is no significant difference among the failed, pass and distinction groups, while the mean number in the merit group is significantly lower than the above three groups confirming the impression that this difference in the merit group influenced the result of the ANOVA test. It is unclear to understand why the merit group should contain so many fewer lexical cohesive devices, especially the repetitive devices. This may be an artefact of the small sample size used in the analysis, which was necessary due to the intricate nature of the analysis.

Another objective of the study was to investigate whether there was any correlation between the number of lexical cohesive pairs used in the samples and the quality of writing. Correlation was computed between the marking scales of assignment samples and the frequency of lexical cohesive pairs (total pairs per sample). The result of the Spearman's rho correlation test ($p = 0.29 > 0.05$, see table F6 in appendix F) suggests that there is no significant correlation between the use of lexical cohesion and the marks of the student assignments. This result is in line with results from previous studies (e.g. Zhang, 2000).

The ratio of overall average distribution in the categories of reiteration and collocation in this study is compared with the result of Tanskanen's (2006) study of academic articles, which is shown in table 26 below.

Table 26 Comparison of the frequencies (per 1,000 words) of reiteration and collocation pairs between the analysis of research articles in Tanskanen's study and the analysis of MA assignments in the present study

	Tanskanen's study	the present study
average number of reiteration pairs	90.25	219
average number of collocation pairs	15	25
ratio	6.01	8.76

Table 26 shows that the ratio of reiteration pairs to collocation pairs in the present study is higher than that of Tanskanen's study (8.76>6.01), which provides preliminary evidence that Chinese students may tend to use reiterative pairs rather than collocational pairs to create cohesiveness compared to academic authors of research articles. This comparison suggests the potential value of providing more knowledge of collocational pairs to Chinese students in order to support their use of collocational devices and contribute to cohesiveness in their academic writing (Tanskanen, 2006, p. 167). However, it also needs to bear in mind that research articles in Tanskanen's study might have different genres or text types from the MA assignments in the current study, and the use of lexical cohesion is context-sensitive, which may suggest the inappropriateness of comparing the results of these two studies.

Reducing textual features into numbers may not be regarded as representing a complete analysis. However, in the analysis of lexical cohesion, this method can shed light on the more general tendencies of uses in regard to the lexical cohesive devices (Tanskanen, 2006, p. 96) in each marking-scale group. After examining the macro result from the sample analysis in MTMC, the next section of this chapter presents detailed information about the use of lexical cohesion in the samples, in order to demonstrate the operation of the developed framework of lexical cohesion in the present study.

Specifically, one extract is selected from each marking scale group for detailed analysis. The selection criterion is whether the extract includes typical features of

lexical cohesive devices which are considered worthy of discussion in this chapter. The purpose of demonstrating these extract analyses is that the analyses of the four extracts enable elucidation of the features of different lexical cohesive devices identified in the four marking scale groups.

7.3 Fine analysis of text extracts in the MTMC corpus

Four extracts were selected from the beginning of four samples in the marking-scale groups respectively as this strategy was adopted in the selection of all sample in the corpora (see chapter 6, p. 183). The analysis of these extracts follows the analytical strategies described in chapter 6. Specifically, the identified lexical cohesive pairs in the extracts below are coded according to the highlight system (see table 21); each extract is divided into clauses; the text versions[48] of cohesive analysis is displayed for readers' convenience and comprehension in the analysis of each extract.

Specifically, the extracts are selected from four samples in four marking scale groups. The detailed information about each extract is shown in table 27 below:

Table 27 Information about four extracts in marking scale groups

text	source	word count
extract 1	sample F9 (failed group)	179
extract 2	sample P6 (pass group)	174
extract 3	sample M13P (merit group)	209
extract 4	sample D5P (distinction group)	211

Extracts 1, 2, 3 and 4 are from failed, pass, merit and distinction groups respectively, showing both reiteration and collocation relations at work in context. What needs to be emphasised here is that in the analysis below, only those lexical items which are related to other items included in the displayed extracts are highlighted or marked. The reason for this marking alternative is that the discussion of lexical cohesive relations between lexical items inside the extract

[48] In real analysis, the analytical table is the main framework for constructing the analysis. The tabulated versions of cohesive analysis for the extracts are displayed in appendix G.

and items outside the extracts would mislead the readers as some of these lexical items involved in the cohesive relations do not appear in the extracts. “Although this may slightly skew the picture of the use of cohesion in the texts, it is still the best and most reader-friendly alternative.” (Tanskanen, 2006, p. 96). Firstly, extract 1 will be analysed and discussed.

7.3.1 Analysis of extract 1 from the failed group

To begin with, the cohesive analysis of extract 1 is shown in a colour-coded text.

Extract 1

This assessment tool aims at a group of **business-major students** // who are joining an English training program,// after which they are going to become **interpreters** and **bilingual sales agents** in an **international furniture fair**. //**International** Famous **Furniture Fair** (Dongguan), commonly known as **3F**, is the biggest and most representative **furniture exhibition** in **China**.// In **3F**, leading **furniture** manufacturers from mainland **China**, Hong Kong and Taiwan will take part in this **exhibition**, seeking for new contracts with byers [buyers][49] all over the world.// During **exhibition** time, which lasts for 14 days semi-annually,// about 80 **interpreters and bilingual sales agent**[s] are **needed**.// To meet the **needs** of trading companies and **provide** them with better **interpretation** services, the **exhibition committee** cooperates with **business** faculty in Sun Yat-sen **University** (SYEU),// which is one of the top **universities** in southern **China**, to **provide internship** opportunities for their **business-major students**.// **Students** // who are interested in this **internship** can apply for it// and they are **required** to at first pass interviews and then language tests to ensure// they meet relative **requirements** of the **committee**.

[49] Grammatical typos in the original sample text are corrected in “[]” where applicable.

As mentioned in chapter 6 (see section 6.5), extract 1 comes from the first part of an assignment for a Syllabus Design and Assessment module, and thus provides an introduction to the topics of the assignment. Since the samples in MTMC are all from this module, the analysis of extract 1 can be regarded as a starting point for detailed investigation of features of lexical cohesive devices used in this corpus.

First of all, the analysis of extract 1 starts with collocational relations. There are 8 collocational relations in this extract, which are highlighted in yellow in the text above:

1. students – training program
2. English –interpreters
3. business – sales agents
4. fair – manufacturers
5. international – over the world
6. business – trading
7. internship – interviews
8. interviews/language tests – requirements of the committee

As mentioned in chapter 4, the collocation category is considered to have two sub-categories in the present study: activity-related collocation and elaborative collocation. Accordingly, the above eight collocational pairs can be allocated to either of the sub-categories. Before discussing their allocation to these categories, there is one more point which needs to be addressed at first, which is the definition of activity-related collocation. Chapter 4 has already stated that the idea of activity-related collocation derives from the work of Tanskanen (2006). In her definition of the activity-related collocation sub-category, *actions, people, places, things* and *qualities* configure as activities. *Activity* as a concept itself was not included in the elements which form the activity, and, therefore, was not regarded as part of activity-related collocation relations in Tanskanen's work. For example, in her example below, *the First World War* and *the British army* were regarded as an elaborative collocational pair:

B: I'd be interested to know in fact if they had made maps of the the area before the **First World War** broke out

a: surely surely they must have Napoleon I shouldn't be surprised

B: there must have been maps yes

a: there are certain French maps much earlier than that I mean like early nineteenth century like ours

B: oh yes I mean **the British army**

a: no I only meant that maps must have existed

B: oh maps must have existed certainly. . .

(Tanskanen, 2006, pp. 96-97)

However, in the present study, *activity* as a concept is included in this relation as one of the elements in the activity-related collocation sub-category, as the name of the category is activity-related collocation, it is judged as reasonable to include the term *activity* itself. Therefore, in Tanskenen's example above, *the First World War* is the *activity* and *the British Army* is one of the participants, i.e. the *people* element, which will be regarded as activity-related collocation in the present study. In extract 1, there are four similar activity-related collocation examples and one activity-related collocation pair in Tanskanen's sense:

7-1 *students* and *training program* form an activity-related collocation relation (students are the participants of the activity training program).

7-2 *business* and *sales agents* are related by activity-related collocation (sales agents act as the participants in business activities).

7-3 *fair* and *manufacturers* are related by activity-related collocation (manufacturers are the participants of the activity fair).

7-4 *business* and *trading* are elaborative collocation (activity *Business* with the action of the activity *trading*).

7-5 *English* and *interpreters* are related by elaborative collocation (*English* is the "thing" and *interpreters* are the "people" in the activity of language interpretation).

In collocational pairs 7-1, 7-2 and 7-3, two lexical elements are involved in the relation between *activity* and *person*, and in collocation example 7-4, the

elements have the relation between *activity* and *action*; while in example 7-5 the relation between the two lexical items *English* and *interpreters* is between *thing* and *person*.

Another three collocations in extract 1 are elaborative ones:

7-6 *International* and *over the world* are elaborative collocation (*over the world* elaborates the meaning of *international*).
7-7 *Internship* and *interviews* are elaborative collocation (*interviews* is used as an approach to selecting interns for *internship*).
7-8 *Interviews*/*languages tests* and *requirements of the committee* are elaborative collocation (*Interviews* and *language tests* include the contents of *requirements of the committee*).

Apart from the examples for collocations above, it can be seen from the table analysis (see table G1 in appendix G) that repetition has the highest number of cohesive pairs. Simple repetitions form the majority of repetition pairs. There are only three complex repetition pairs: *needed – needs; interpreters – interpretation and required – requirements*.

In terms of hyperonymy, two pairs display part-whole relations and one pair represents kind-whole relation:

7-9 Dongguan – mainland China (Dongguan is a city of Guangdong Province in mainland China)
7-10 Sun Yat-sen University – Universities (Sun Yat-sen University is one member of universities)
7-11 English – language (English is a kind of language)

By contrast, there is one hyponymic pair and there are two meronymic pairs:

7-12 *Furniture exhibition – 3F*[50] (*3F* is one kind of *Furniture exhibition*)

7-13 *China – mainland China* (*mainland China* is part of *China*)

7-14 *mainland China – southern China* (*southern China* is part of *Mainland China*)

Only one synonymous pair occurs: *assessment – test*, which is a typical synonymous pair in samples with the topic of *assessment*. There are two categories (i.e. signalling nouns and antonymy) without any pairs in the scope of this extract. These absences may be explained by the fact that firstly these two categories also have smaller number of pairs according to the raw frequencies of lexical cohesive pairs in the sample F9[51] itself (i.e. one pair of signalling nouns and two pairs of antonymy); secondly, the extract is only 176-word length, and therefore includes only a small number of pairs compared with the whole text sample (i.e. sample F9).

Overall, the numerical information in this extract demonstrates that, as with the overall sample counts in the fail category, the number of repetitive pairs is the highest and the number of collocational pairs is the second highest. However, several differences exist between extract 1 and the average level of frequencies of lexical cohesive pairs identified in the failed group regarding other categories, as shown in table 28 below:

Table 28 Comparison of lexical cohesion distribution (per 1,000 words) between the overall average figures in the failed group, sample F9 and extract 1

	rep[52].	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.
failed group	163	11	9	20	5	8	6	6	28
sample F9	181	6	19	16	10	3	13	6	44
extract 1	117	6	17	6	11	0	11	0	45

[50] International Famous Furniture Fair

[51] See appendix D.

[52] Rep. = repetition, syn. = synonymy, hyper. = hyperonymy, hypo. = hyponymy, mero. = meronymy, sig. = signalling nouns, ide. = identity, ant. = antonymy, and col. = collocation

Table 28 indicates that in terms of repetition, extract 1 is below the average level while in terms of collocation it is above average. Other categories also demonstrate substantial variability between the extract and average lexical cohesion category level. An interesting point which needs to be mentioned here is the surprisingly high frequency of meronymic pairs identified in extract 1, especially noticeable as the meronymy category comprises the lowest average number of pairs in the failed group (see table 25). The reason for this lies in the geographical description of different regions in China tied to the main topic of introducing a Chinese furniture exhibition in extract 1. This description forms several hierarchical relations regarding geographical locations, which leads to the identification of two meronymic pairs in this extract (i.e. *China – Mainland China* and *Mainland China – southern China*). Therefore, the topic under discussion may be a more important predictor of lexical cohesion type frequency than, for example, the mark achieved by the writer, which also reveals the context-sensitive nature of lexical cohesion.

Overall, the comparison above reveals that substantial differences exist between extract 1 and the average distribution in the failed group regarding the frequencies of lexical cohesive pairs, which indicates that, to some extent, may suggest that perhaps every extract or sample is unique in the failed group in terms of the use of different lexical cohesive devices.

7.3.2 Analysis of extract 2 from the pass group

As mentioned in table 27, extract 2 is from the pass group. Therefore, the result of extract 2 is compared with the average distribution of lexical cohesive pairs in the pass group. The related contents in the analysis of extract 2 are shown in the colour-coded text below.

Extract 2

This is a **speaking**-oriented **assessment** tool// which is designed for the **high level English Speaking Contest** qualification trials-- the **national** level. //The **test takers** are upper intermediate and advanced **English** learners. //It is within the scope of **high level**'s **English** majors' **spoken test**. //Take IELTS as the

English proficiency criteria as reference, the **test takers** for this **assessment** should have scored 7 point (each section is no less than 6.5 point) and **higher**. //The **test takers** are Chinese students// who want to compete in the final round in the **national English speaking contest**. // (I) Details of Recourses [Resources] // **Test developer[s]** and **Test** writers //Bachman and Palmer points out that // **test developers** is one of the most important factors in human **recourses [resources]**.// They are responsible for the **test** specification, management, try-out, achievement and use. //They administrate and take control of the entire **test** process and make sure // the **test** can be well carried out. //The **developers** in this **assessment** are not the daily teachers// who are familiar to the **test takers**.

Similar to extract 1, the most frequent cohesive pairs in extract 2 are still simple repetitive pairs. Only two of the 23 repetitive pairs are complex repetitive ones: *speaking – spoken* and *test developer – test developers*. This similarity reveals the dominant use of simple repetitive pairs in Chinese students' writing. Another similarity between extract 1 and 2 is the empty categories, i.e. in extract 2, there are two categories that have no pairs (meronymy and signalling nouns). In terms of the average figure for cohesive pairs in the pass group, meronymy has the smallest number of cohesive pairs, and in extract 2, there are no pairs in meronymy, which demonstrates the lower frequent use of meronymic pairs in students' written samples in the pass group. However, the average distribution of signalling nouns in the pass group is relatively higher compared to that of meronymy in this group (see table 25). In extract 2, there is also no pairs in the category of signalling nouns. This comparison result demonstrates substantial variability in individual samples, which is a salient and important feature in the sample analysis of the MTMC corpus.

As for other categories, there are two synonymous pairs: *assessment – test* and *learners – students*. As mentioned in extract 1, *assessment – test* is a frequent synonymous pair in the samples sharing the topic of assessment. Furthermore, *learners – students* is a frequent pair in all the samples, as their topics are in the English education domain. These repetitive uses of certain lexical cohesive items and pairs suggest that topic is an important factor which has a significant influence on the use of lexical cohesion in students' academic writing. In terms of

hyperonymy, one pair is counted: *speaking-oriented assessment tool – English majors' spoken test*. What is interesting here is the adding of information for helping identify this hyperonymic relation. Looking at the clause in which *English majors' spoken test* occurs:

It is within the scope of high level's English majors' spoken test (lines 3-4).

The phrase *within the scope of* indicates that *English majors' spoken test* is a more general category which includes sub-categories. In this case, the sub-category has already been mentioned at the beginning of the extract, which is *speaking-oriented assessment tool*. Therefore, this hyperonymic relation is a kind-whole relation, i.e. *speaking-oriented assessment* is a kind of *English majors' spoken test*. It is noted that there is a possible pedagogical application arising from this example, which might be used to demonstrate how to use hyperonimic devices in academic texts. Similar to the frequency of hyperonimic relations in this extract, in the hyponymy category, also one pair is also counted: *test – IELTS*. This pair is the contrary of the hyperonymic pair above because *test* is a superclass and *IELTS*[53] is a kind of *test*.

In terms of identity which is developed as a new category in the lexical cohesion model in this study (see chapter 2), this category has not been systematically included in previous models of lexical cohesion. The operational definition of identity presents challenges during the cohesive analysis. Extract 2 provides an opportunity to explain further this operational definition of identity with examples. There are three identity pairs in this extract:

7-15 *the national level – high level*
7-16 *upper intermediate and advanced English learners – Chinese students*
7-17 *Chinese students – test takers*

[53] The International English Language Testing System

As a co-referential relation, identity is extremely context sensitive, which is to say that this kind of relation is only identified between two lexical items when they appear in a specific context; otherwise, in a different context, such relation will no longer exist. Looking at the three examples mentioned above, in example 7-15, the contextual meaning of *high level* is based on its referent – *the national level*. The latter specifies the former with identity of reference. However, in examples 7-16 and 7-17, the presupposing items add characteristics or new identity to the presupposed items. Specifically, in the pair of *upper intermediate and advanced English learners – Chinese students, Chinese students* offers another identity to *Upper intermediate and advanced English learners,* both of which refer to the same group of people but carry different aspects of information about the relevant group. In another pair, *Chinese students – test takers, test takers* also gives another identity to *Chinese students*.

It can been seen from the explanation of the identity pairs above that the identity relation consists of two kinds. The first relation refers to the co-referential relation between two lexical items, in which one item specifies the meaning of another in a certain context. The second kind of relation is between two lexical items which contribute different identities or characteristics to the same referent. These examples may provide another pedagogical implication regarding the teaching of identity devices in text, which is to demonstrate these devices with ample examples.

As for antonymy, there are two pairs in extract 2:

7-18 *test takers – test developers/test writers*
7-19 *students – teachers*

Both of the pairs in examples 7-18 and 7-19 belong to the sub-category of converse antonymy (see section 2.3.6), as the items in these pairs are “nouns which express reciprocal social roles” (Hoffmann, 2012, p. 91). Specifically, *test takers* are the participants in the test, while *test developers* and *test writers* are the designers of the test. Therefore, these two lexical items form a converse

relation in the test domain in this text. The second pair *students – teachers* is highly frequent in the two corpora which are situated in the larger subject domain of English education. This indicates again the context-specific feature of lexical cohesion in general.

The final category examined in this extract is collocation. There are four collocational pairs in extract 2:

7-20 *assessment – test takers*
7-21 *upper intermediate and advanced English learners – English majors*
7-22 *English – IELTS*
7-23 *test developers – the test specification, management, try-out, achievement and use*

Assessment – test takers is an activity-related collocation relation, while the other three are elaborative collocational relations. Specifically, *Assessment* is the '*thing'* and *test takers* are the '*people'* who take the *assessment*. In terms of the elaborative collocation sub-category, *English-major students* are normally and specifically in this context *upper intermediate and advanced English learners*. The latter evokes the English major frame. As English is the language tested in IELTS test, English evokes the IELTS test frame. *Test developers* are responsible for several aspects of the test – the test specification, management, try-out, achievement and use. Therefore, *test developers* evokes the test frame.

The numerical information comparison between extract 2 and the average figure in the pass group is in the following table:

Table 29 Comparison of lexical cohesion distribution (per 1,000 words) between the overall average figures in the pass group, sample P6 and extract 2

	rep.	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.
pass group	159	14	9	16	6	10	8	7	24

sample P6	187	11	6	22	8	11	14	8	8
extract 2	121	11	6	6	0	0	17	11	23

As can be seen from this table, the highest and second highest frequencies of pairs are in the repetition and collocation categories respectively both in extract 2 and at the average level in the pass group. However, the repetitive pairs in extract 2 are fewer than the average distribution of repetitive pairs in the pass group, while the number of collocational pairs in extract 2 is similar to that at the average level in the pass group.

Providing additional evidence of high sample variability, there are substantial variations between extract 2 and the average distribution in the pass group regarding the number of pairs in other categories of lexical cohesion. Specifically, as demonstrated in table 29 above, there are more pairs in extract 2 than in the average distribution of the pass group regarding the categories of identity and antonymy. In terms of the antonymic pairs used in extract 2, the lexical items in the pair *students – teachers* are frequently used either in their singular or plural forms. Specifically, *student* (or *students*) and *teacher* (or *teachers*) co-occur in 6 samples (i.e. P1, P2, P3, P4, P6 and P7). However, the lexical items in the second antonymic pair in extract 2, *test takes – test developers/test writers*, only occur in sample P6. In terms of the difference of the normalised frequencies of antonymic pairs between extract 2 and the average distribution in the pass group, this result may be related to the dominant use of the first antonymic pair in several samples and the exclusive use of the second antonymic pair in sample P6. As regards the normalised frequencies of the identity pairs, the number of pairs in extract 2 is the highest among the three compared groups in table 29, while the number at the average level is the lowest, which indicates that the author of sample P6 used more identity pairs compared to the average use of such pairs in the pass group, and the majority of these pairs occur in the beginning part of sample P6, which are included in extract 2.

By contrast, in the remaining five categories, the average frequencies of cohesive pairs in the pass group are higher than those in extract 2 at various levels. The biggest gap lies in the categories of signalling nouns and hyponymy. It is also

worthy of mention that there are two categories which have no pairs in this extract: meronymy and signalling nouns. As already mentioned in section 7.2, meronymic relations have the lowest average frequencies in all four marking-scale groups. It is not surprising that there is no meronimic relations in extract 2. Furthermore, regarding the zero pairs in the category of signalling nouns, as the number of pairs in this category used in sample P6 is slightly higher than that at the average level, this suggests that the author of sample P6 tends to use signalling nouns in his/her academic writing in general, but did not use such devices in the beginning part of the sample. This case also happens in extract 1 which does not include the use of signalling nouns. This result may indicate that the use of signalling nouns in the beginning parts of samples in the failed and pass groups is not frequent. This indication is possible as signalling nouns are used when being provided certain contexts, while the function of the beginning parts of texts is generally to set the broad contexts for the whole texts.

The overall comparison indicates that to some extent, every extract or sample is also unique in the pass group regarding the use of lexical cohesive devices, which is substantiated by the standard deviation (henceforth SD) data in table F1 (see appendix F), i.e. the SD value in each marking-scale group is 33.497 (failed), 31.811 (pass), 25.287 (merit) and 30.285 (distinction) respectively. As the SD values of the failed and pass groups are higher than those of the merit and distinction groups, the sample variabilities in the former two groups are higher than the latter two groups. Furthermore, the SD figure in the failed group is higher than that in the pass group (i.e. 33.497 > 31.811). That is to say, the samples in the failed group show more variability than those in the pass group. This comparison between the failed group and the pass group in general is substantiated by the comparison between extract 1 and extract 2 regarding their varieties with the average distributions in their corresponding marking scale groups, i.e. extract 2 shows more consistency with its corresponding pass group compared to the distribution of lexical cohesive pairs in extract 1 and the average figures in the failed group.

7.3.3 Analysis of extract 3 from the merit group

Extract 3 is from the merit group, and thus, the frequencies of lexical cohesive pairs identified in extract 3 is compared with the average distribution of such pairs in the merit group. In line with the analysis of extracts 1 and 2, the content of extract 3 and the cohesive analysis is shown in the colour-coded text below.

Extract 3

1.1 learner profile and job background// the syllabus is **designed** for a group of adults // who are the workers of the **Canton** Exhibition **Centre**.// This **centre** is located in **Guangzhou**,// where **China** Import and Export Fair, also called Canton **Fair**, is held twice a year in **Spring** and Autumn.// It is **China**'s largest **trade fair** of the highest **level**, with the most complete varieties and the largest attendance and business turnover,// which attracts thousands of millions of people both home and abroad, including foreign **trade** companies, factories, scientific research institutions, **foreign** invested enterprises and so on.// Recently, the administrative **level** of this **centre** launched an investigation to identify the areas for further development, and found that// the **English** communicative competence of the staff was unsatisfactory and always led to misunderstandings, inconvenience and complaints from the customers.// As a result, in order to fully meet such challenges brought by the coming spring-term **fair** as the surging number of **foreign** clients and specific fair-related needs and issues, this **centre** decided to introduce an **English** training program to their sixteen **staff**//, more specifically, among whom, five are receptionists of the front desk, // four are ushers,// three are **workers** of **customer** service,// and four are the personal [personnel] of function rooms.

Apart from the repetition category which has only simple repetitions and is thus not further discussed in this section, the remaining categories are considered worthy of further elucidation, as lexical cohesive pairs in these categories represent typical examples regarding the use of various lexical cohesive devices in texts more generally. Also, these pairs exemplify the use of sub-types of each category with examples identified in extract 3, and these examples might contribute to the teaching of these lexical cohesive devices for EAP class.

Considering firstly the category of synonymy, there are three pairs in this extract, which provides typical examples regarding the two main types of synonymous relation in the present study, i.e. the synonymous relation in traditional sense and the near-synonymous relation (see section 2.3.2):

7-24 *workers – staff*
7-25 *abroad – foreign*
7-26 *customers – clients*

Workers – staff and *customers – clients* are examples of synonymy in the traditional sense, i.e. having similar meanings and belonging to the same word class. Another pair *abroad – foreign* is an example of what previous researchers (e.g. Halliday & Hasan, 1976; Tanskanen 2006) called 'near synonyms' because the lexical items in this pair have similar meanings but belong to different word classes (*abroad* as an adverb and *foreign* as an adjective while both of them express the meaning of being in another country).

In terms of hyperonymy, there are also three pairs identified in extract 3:

7-27 *workers – centre*
7-28 *Guangzhou – China*
7-29 *Canton Fair – trade fair*

The first two are part-whole relations while the third is a kind-whole relation. However, even within the part-whole relations, *workers – centre* and *Guangzhou – China* are clearly different, as the former is a member-organisation relation while the latter is a geographical part-whole relationship. As mentioned above, In the hyponymy category, the pair *staff – receptionist/ushers/ workers/personnel* is an example of different kinds of job positions in an organisation, but the items refer to people with different job positions rather than job positions themselves. That is why the superordinate is given here as *staff* instead of *the centre*. Regarding the meronymy category, there are three pairs in extract 3:

7-30 *centre – administrative level*

7-31 *centre – front desk/customer service/function rooms*

7-32 *autumn – spring*

The first two pairs, *centre – administrative level and centre – front desk/customer service/function rooms,* share the same superordinate *centre*, but the contextual meaning of *centre* is different in the two cases. The first *centre* is the organisation which includes different departments, whereas the second *centre* is a building in which there are different functional sections. What should also be noted here is the inclusion of the third pair *autumn – spring*. In other researchers' models, this pair will be included in the category of 'ordered set'[54] (Tanskanen 2006) or 'closed set' (Hoey, 1994). The reason for including this pair under the meronymy category in the present study is that the lexical items, *autumn* and *spring*, are parts of a more general system – four seasons.

As for the only two signalling noun pairs in the three extracts so far, the first signalling noun pair *areas – the English communicative competence of the staff* is a cataphoric one, the SN *areas* appearing before the latter segment that this SN refers to. The second pair, *people – customers*, is related to the use of GNs. *People*, as a GN, is used as a cataphor, whose contextual meaning is specified by the lexical item, *customers*, appearing in the subsequent clause. As mentioned in chapter 3, SNs have "a variable, pragmatic meaning which depends on contextual lexicalisation" (Jiang & Hyland, 2017, pp. 2-3). In this example, the segment *the English communicative competence of the staff* specifies the pragmatic meaning of the SN *areas*.

With regard to the identity category, there are four identity pairs in extract 3:

7-33 *learner – adults*

7-34 *adults – workers*

7-35 *China – home*

[54] "The category includes members of ordered sets of lexical items, for example, colours, numbers, months, days of the week and the like" (Tanskanen, 2006, p. 61)

7-36 *foreign…enterprises – customers*

In fact, the first two form a chain of lexical cohesion as the lexical items in these two pairs refer to the same group of people. One reason for excluding the discussion of chains (i.e. sequence of 'cohesive ties' in Halliday and Hasan's term, or 'cohesive pairs' in this study) in this thesis is its complexity in analysis. Only a very small-sized sample analysis can include lexical cohesive chains as an analytical object (e.g. Tanskanen, 2006) due to the complexity of such analysis. In the present case, identifying lexical cohesive pairs is enough to fulfill the aim of this study which is to investigate Chinese students' usage of lexical cohesive devices in their written work, instead of focusing on how to make a text more cohesive at the textual level.

Returning to the identity category itself, all the latter items in the pairs here state another identity of the previous items. For the first two pairs *learner – adults* and *adults – workers*, the items refer to the same group of people who have several social roles or characteristics: *learners, adults* and *workers*. For the pair *China – home*, in extract 3, *China* is regarded as the *home* for Chinese people as the furniture exhibition is held in *China*, and thus, Chinese people are the host of this exhibition, whereas people from other countries are the foreigners. That is to say, in this context, *China* has another identity which is *home* to the Chinese people; at the same time, *home* is specified as a country, *China*. With regards to the fourth pair *foreign…enterprises – customers*, *foreign trade companies, factories, scientific research institutions* and *foreign invested enterprises* are connected by sharing one identity – *customers*. This identity pair is a good example to show the extremely context-sensitivity of identity relations, as the contextual meaning of *customers* can only be interpreted as *foreign trade companies, factories, scientific research institutions* and *foreign invested enterprises* in extract 3, but cannot be directly generalised to other contexts.

In the antonymy category, *home – foreign* and *staff – clients* are identified as antonymic relations in extract 3. The first pair is a complementary antonymic relation because the compoments are binary contrasts without any gradeability.

The lexical items *staff* and *clients* in the second pair are converse antonyms as they are reciprocal social roles (Hoffmann, 2012, p. 91).

As for the collocation category, there are only four pairs in extract 3:

7-37 *job – workers*
7-38 *Import and Export Fair – business turnover*
7-39 *business – trade*
7-40 *complaints – clients*

For detailed explanation, *business* is the activity and *trade* is the action, which form an activity-related collocation pair; *Import and Export Fair* is the place for business, and *business turnover* is the outcome of the activity business, both items existing in the business frame. Therefore, *Import and Export Fair* and *business turnover* form an elaborative collocation. In fact, both of the pairs are located within the business frame. The reason for dividing them into two sub-categories of collocation was based on the elements which form the activity-related collocation: *actions, people, places, things, qualities* and *activity* itself. In the pair, *business – trade*, the two lexical items in this pair are two elements in the activity-related collocation relation. However, in the pair, *Import and Export Fair – business turnover*, only the former item is the element – *place* – in such a relation. Therefore, the second pair is regarded as elaborative collocation.

For pairs 7-39 and 7-40, they are also included in the job frame and regarded as elaborative collocation. In the pairs of *job – workers* and *complaints – clients*, people who are doing *jobs* are called *workers;* and *workers* receive *complaints* from their *clients.* These are therefore items which elaborate on the same topic which is 'doing a job', in this case specifically, doing a service job at an exhibition centre. It can be seen from this extract that collocation pairs in a text are sometimes connected by several broad frames.

The quantitative data, according to table 30 below, shows a further illustration of variability among individual samples in the merit group and provides possible

evidence for some findings in previous sections. It is noticeable that extract 3 has fewer repetition pairs than the average number in the merit group, which is in line with the comparison between previous two extracts and their corresponding failed and pass groups regarding the use of repetition pairs. However, the use of repetition pairs in extract 3 is less frequently compared with those in extracts 1 and 2, which perhaps substantiates the result in table 25 above that the use of repetition in the merit group is less frequent compared with that in the pass and failed groups.

In the categories of synonymy, hyperonymy, meronymy and identity, extract 3 surpasses the average number in the merit group; however, compared with the use of devices in these categories in sample M13P, the average number of synonymous pairs is higher, which indicates that the student author of sample M13P does not use synonymous pairs constantly in this sample.

while in other categories (i.e. hyponymy, signalling nouns, antonymy and collocation), the average number is higher than extract 3. It is notable that for the merit text in extract 3, the signalling category has pairs whose number is very close to the average number in the merit group, which is different from extracts 1 and 2, as there are no pairs in the category of signalling nouns in extracts 1 and 2. As mentioned above, in general, signalling nouns are used after their contexts are already set in the texts. That is to say, the majority use of signalling nouns is in anaphoric relations rather than in cataphoric relations. However, the two cases of signalling nouns used in extract 3 are in the cataphoric relations, which means the signalling nouns occur before the lexical specification parts in the texts. These cases are not typical in the use of signalling nouns, which provide potential special examples for EAP teaching of signalling nouns used as cataphors.

Overall, the frequencies in extract 3 follows the general tendency in all marking scale groups that repetition is the biggest category. Furthermore, the pairs in other categories have various conditions compared with the average number in the merit group, which also evidences variabilities in individual samples in the merit group. What is interesting in this extract is that, compared with extract 1 and

extract 2, all nine lexical cohesive categories have pairs identified in extract 3 as there are two special cataphoric uses of signalling nouns in extract 3.

Table 30 Comparison of lexical cohesion distribution (per 1,000 words) between the overall average figures in the merit group, sample M13P and extract 3

	rep.	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.
merit group	136	10	8	15	4	7	7	5	25
sample M13P	85	8	12	27	12	8	12	4	50
extract 3	72	14	14	5	14	5	19	10	19

7.3.4 Analysis of extract 4 from the distinction group

Finally, the cohesive analysis of extract 4 is presented. As extract 4 is from the distinction group, the frequencies of cohesive pairs in this extract is compared with the average distribution of the cohesive pairs in the distinction group. Firstly, the cohesive analysis of extract 4 is demonstrated in the colour-coded text below.

Extract 4

a) the **learners**// The **learners** are 40 **students**// who are of first year in an **elite** senior **high school** in **China**.// They just **experienced** the **entrance examination**// and their overall performances were **good**,// which means that most of them have solid foundation in **English**.// Due to the **exam**-oriented **educational** **atmosphere** in **China**, they used to pay more emphasis on the written **English** other than **oral** **part**.// For them, **English** **learning** is by learning the explicit input from teacher and drill the **knowledge** through different types of **exam** paper.// On the other hand, although majority of them have **good** command of vocabulary and grammar,// they have little **confidence**// while **speaking** and their **oral** expressions were not so proficient owing to lack of time and opportunities to **practice**.// While the **oral** **part** is included in **the** **National** **College** **Entrance** **Examination**,// and they realize the importance of **speaking** **English** and also the deficiency of their speaking English,// so **part** of their focuses begin to switch to the **speaking** part.// Exposed to the idea that once you **enter** an **elite** **high** **school**, you've already been on the way to an **elite** university, the learners believe that their school will

provide them with good education and help them to make progress according to their needs.

Similar to other extracts, simple repetition contributes the highest number of pairs to cohesion in extract 4 as there are only five complex repetitive pairs which are derivational repetitions:

7-41 *examination – exam*
7-42 *learners – learning*
7-43 *entrance – enter*
7-44 *learning – learners*
7-45 *educational – education*

As for synonymy, there are four pairs:

7-46 *oral – speaking*
7-47 *have good command of – proficient*
7-48 *emphasis – focuses*
7-49 *college – university*

Only the last pair *college – university* can be regarded as a synonymous relation, while the remaining pairs are near synonymous relations, because both *college* and *university* are nouns, whereas the lexical items in other three pairs belong to different word classes or have varied lengths. Specifically, pair 7-46 is near synonymous as *oral* is an adjective while *speaking* is a noun. This is also the case with pair 7-48 in which *emphasis* is a noun while *focuses* is a verb. Pair 7-47 is also regarded as a near synonymous relation but the reason is slightly different in this case, because *have good command of* is a phrase while *proficient* is an adjective. This example demonstrates the essence of a cohesive lexical item which is irrelevant to the number of words in one item, but only considers the cohesive force between items. The synonymous pairs shown above provide good examples of using both two subtypes of synonyms (i.e. synonyms in the traditional sense and near synonyms) as cohesive devices in Chinese students'

academic written assignments, which suggests that synonyms in lexical cohesion could be taught with these two sub-types as examples.

For superordinate-subordinate relations, extract 4 only has two pairs which are hyponymic and meronymic respectively. The pair *senior high school – college* belongs to the hyponymy category as both of components are the members of the school system as the superordinate. The pair *written – speaking* is counted as a meronymic relation in this study as the two items in this pair are parts of language skills. Here, the meronymic relation is more abstract than traditional meronymic relations which refer to the constructional parts of an entity, e.g. roof and wall of a house (Hoffmann, 2012, p. 90). However, in the academic discourse of Language Education, most pairs in meronymy are abstract concepts rather than concrete entities. As a special case of superordinate-subordinate relations, the category of signalling nouns in the extract has only one pair *English – part* (which is repeated in later segment of the extract and the repeated pair is not included in the analysis).

In the identity category, there are two pairs:

7-50 *learners – students*
7-51 *China – national*

The pair *learners – students* is usually counted as a synonymous pair. However, in this extract, they are included in the identity category. The original text in which these items occur is:

> *the **learners**. // The learners are 40 **students** who are of first year in an elite senior high school in China.*

In this case, *students* not only refers back to *learners* but also specifies the identity of *learners* in this particular context. Therefore, component elements form an identity relation rather than a synonymous relation. This case reveals the context-sensitivity of lexical cohesion from another perspective: the cohesive relation of two lexical items depends on the specific context in which the items

co-occur, and this relation vary in different contexts. This example supports the pedagogical implication that the teaching of lexical cohesive devices should be within specific contexts. Otherwise, it is difficult for students to understand the context-sensivity of lexical cohesion in use. In pair 7-51, although *China* is a noun and *national* is an adjective, the component elements still form an identity relation, as *China* is an antecedent clue for the interpretation of the contextual identity of *national* in this text. This example reveals again the principle of cohesive analysis in this study that lexical items from different word classes still can be cohesive pairs.

There are only two pairs in the antonymy category:

7-52 *students – teacher*
7-53 *proficient – deficiency*

Pair 7-52 is a typical antonymic pair in the corpus, while pair 7-53 is regarded as comprising near antonyms, as the lexical items *proficient* (adjective) and *deficiency* (noun) are from different word classes and their meanings are not directly opposite. There are two missing parts in the connection between *proficient* and *deficiency*: *good* and *deficient*. The complete connection chain should be *proficient – good – deficient – deficiency*. In this chain, *proficient* means "skilled and experienced" (Cambridge Dictionary, 2019), which includes the meaning of *good*; *deficient* means "not good enough" (Cambridge Dictionary, 2019b), which is obviously the opposite of *good*; and finally, *deficency* is the noun form of *deficient*. This example shows that sometimes the cohesive relation between two lexical items is not obvious or straightforward, and needs some extra information to identify the implicit cohesive force between the two items. As for the EAP implication regarding the teaching of lexical cohesion, teachers should explain explicitly with examples (such as pair 7-53 above) that some antonymic relations are not straightforward but can still create lexical cohesiveness in texts.

The last category is collocation which has 8 pairs in extract 4:

7-54 students – first year
7-55 students – senior high school
7-56 students – examination
7-57 China – entrance examination
7-58 performance were good – solid foundation in English
7-59 students – educational
7-60 English learning – vocabulary and grammar
7-61 little confidence – not so proficient

It can be seen that four pairs exist within the school frame: 7-54, 7-55, 7-56 and 7-59. As this study does not include specific discussion about cohesive chains and the cohesive relations are across clauses, the school chain is separated into four individual pairs in which the item *students* acts as a connector to unify other items (i.e. *first year, senior high school, examination* and *educational*) in the school frame. As for the other 4 pairs, *entrance examination* adds more information to the educational system in *China*; the *solid foundation in English* is the deduced result of the statement that the students' *performance were good* in English; *vocabulary and grammar* specifies the content of *English learning*; and the idea that the students' oral English is *not so proficient* can explain why the students have *little confidence* in speaking. It can be seen that in these four pairs, the latter lexical items elaborate the former items in various ways. Therefore, all of them form elaborative collocational relations.

With regard to the qualitative data in extract 4, the number of repetitive pairs in this extract is similar to those in extracts 1 and 2, while still the number is smaller than the average number in the distinction group, which is shown in table 31 below.

Table 31 Comparison of lexical cohesion distribution (per 1,000 words) between the overall average figures in the distinction group, sample D5P and extract 4

	rep.	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.

distinction group	159	13	8	20	8	8	8	6	23
sample D5P	150	20	5	10	2	2	10	7	20
extract 4	104	19	0	9	5	5	9	9	47

As shown in this table, in the categories of hyperonymy, hyponymy, meronymy and signalling nouns, numbers of lexical cohesive pairs in extract 4 are lower than the average number in the distinction group. Further there is no pairs in the hyperonymy category. This is a noteworthy result as it indicates that the student author does not prefer to use superordinate – subordinate cohesive devices in the assignment, which is further evidenced by the numbers in sample D5P. this result in extract 5 is different from the previous four extracts, which indicates that personal preference may act as a key determinant of the use of different lexical cohesive devices. In the other three categories (i.e. synonymy, antonymy and collocation), the frequency in extract 4 is higher than the average distribution, especially in the collocation category in which the number of extract 4 is more than twice as much as is in the average number of the distinction group. This may lie in the use of a long cohesive chain of the school frame in extract 4. Four collocations pairs are included in this chain, i.e. *students – first year, students – senior high school, students – examination* and *students – educational*. In general, the school frame is involved frequently in the samples in the MTMC corpus because it relates to the topics involved in the module where the samples were selected. However, forming such a long collocation chain within a short text, as the case in extract 4, is not usual. It is not surprising that this long chain used in extract 4 contributes to the overall higher number of collocational pairs used in this extract. In the identity category, extract 4 shares a similar figure with the average level in this group, which is supported by the frequency of identity pairs used in sample D5P. Overall, extract 4 follows the tendency in this group to have much more repetitive pairs and the second highest number of pairs is in collocation category.

7.4 Conclusions

This chapter has firstly presented the quantitative analysis of lexical cohesive devices used in the MTMC corpus. The result of the one-way ANOVA test regarding the comparison of lexical cohesive elements between samples in

different marking scale categories shows that there is significant difference between the number of lexical cohesive pairs and marks achieved in MTMC. This conclusion is similar to several previous studies concerning Chinese EFL learners' academic writing (e.g. Liu & Braine, 2005). By contrast, the result of the Spearman correlation test illustrates that there is no significant correlation between the use of lexical cohesion and the marks achieved in MTMC, which is also in line with other previous studies (e.g. Zhang, 2000). As discussed in chapter 5, the results shown in existing research are contradictory. The reason may lie in the fact that different measurements and analytical strategies have been adopted in various contexts, and lexical cohesion is context sensitive, which indicates that the use of lexical cohesive devices varies in contexts.

The figures of standard deviation of four marking-scale groups reveal the overall variation of samples in the four marking scale groups. Specifically, the samples in the merit group show the lowest variation level while the samples in the other three groups achieve higher variation levels, in which the samples in the failed group show the highest variations to the average number in this group. Further investigation is needed in order to give possible explanations.

In addition, the quantitative data[55] in the analyses of the four extracts when compared with the average figures for their corresponding marking-scale groups above implies that each sample has a relatively individual usage of lexical cohesive devices, though all samples use much more repetitive devices than other forms of lexical cohesive devices. There are several factors underpinning this result, such as personal preference, topic relevance and students' knowledge of lexical cohesion.

Regarding the overall comparison among the four marking scale groups, shown in table 25, similar figures are found in each group, though there are about 30 repetitive pairs/1000 words fewer in the merit group compared to the other three groups. The reason for this probably lies in the smaller number of repetitive pairs identified in two samples in the merit group, i.e. samples M12P and M13P. Only

[55] See tables 28, 29, 30 and 31.

these two samples have less than 100 repetitive pairs/1000 words with all the 52 samples in MTMC (see appendix E). Furthermore, compared to the use of repetitive and collocational pairs, other lexical cohesive devices are significantly fewer in MTMC. For example, signalling nouns are not frequently used in general. The reason for this observation is partly that signalling nouns which appeared in intraclausal cohesive pairs were not included in the analysis of MTMC; but this observed result may also have its origins in the students' lack of knowledge about signalling nouns. If this is the case, it would suggest that other lexical cohesive devices, such as signalling nouns, need to be taught explicitly to Chinese EFL learners in EAP writing classes.

As for the analysis in the 4 extracts, some interesting findings have pedagogical applications for teaching lexical cohesive devices in EAP classes. Although Chinese students in the present study tend to use simple repetitive pairs in their module assignments in all four marking scale groups, they also favour using collocational devices in their writing. This is a positive sign in regard to using various devices in the creation of lexical cohesion, because in previous studies, misuse and lack of use of collocation devices have been identified as the main problems in ESL students' writing (Liu, 2000). In terms of other lexical cohesive devices, the quantitative data in the extract analyses indicates that these devices are not as frequently used as the repetitive devices and collocation ones. However, the exemplary use of the devices from the remaining lexical cohesive categories in the extracts provides valuable pedagogical implications for EAP classes. Some implications are summarised as follows:

When teaching the use of identity devices in text, it is important for teachers to stress the two sub-types of the identity category with examples, as the cohesive relation between two lexical items in the identity category is highly contextual sensitive.

As for the teaching of synonyms in EAP classes, both two subtypes of synonyms (i.e. synonyms in the traditional sense and near synonyms) as cohesive devices should be mentioned as Chinese students may not be familiar with the concept

of near-synonyms, let alone consciously using this type of synonyms as cohesive devices in their writing.

This is also the case with teaching antonyms. It is suggested that teachers should give explicit instructions on the point that there are some relations which cannot be directly interepreted as antonymic relations straightforward, but still act as cohesive devices for the contribution to the whole cohesiveness in texts.

The overall conclusion from the discussion above indicates that different lexical cohesive devices should be taught to all levels of Chinese students to help them select appropriate devices for creating a more cohesive text. Further implications for EAP pedagogy will be discussed in chapter 9.

Chapter 8 Analyses of lexical cohesive devices in the MTDC corpus

8.1 Introduction

This chapter presents the quantitative and qualitative analyses of the MTDC corpus to achieve two objectives: firstly to explore the relation between the use of lexical cohesion and the flow of text through the different functional sections of dissertation samples in this corpus; and investigate the specific features of various lexical cohesive devices used in the MTDC corpus. More specifically, section 8.2 will present a quantitative analysis of the use of lexical cohesive devices in MTDC[56], i.e. investigating the frequencies of various lexical cohesive devices. This is an examination at the macro-level (Tanskanen, 2006, p. 94). The normalised frequencies of lexical cohesive devices used in the five functional sections of dissertation samples will be presented in this section.

Next, section 8.3 will discuss the detailed quantitative and qualitative results of five excerpts selected from samples in the functional sections (i.e. introduction, literature review, methodology, findings and discussion, and conclusion) and compare these results with the average figures of lexical cohesive devices used in each functional-section group. This analysis supports two research purposes. The first purpose is to describe and characterise lexical cohesive devices used in MTDC; and the second purpose is to evaluate variations between the functional sections regarding the distribution of lexical cohesive devices. Finally, section 8.4 summarises the findings in this chapter and presents an overall discussion of the results generated in the chapter.

8.2 Results of quantitative analysis in MTDC

As described in chapter 6, there are 45 text excerpts from functional sections of dissertation samples overall in the MTDC corpus, 9 in each functional section group. The word count slightly varies in each sample, ranging from 389 to 457 words (see appendix H). The reasons for the variability in excerpt word count are that when choosing the excerpts from each dissertation chapter, either it was

[56] MA TESOL Dissertation Corpus

important not to separate clauses in one sentence, or one functional section in a dissertation was simply shorter than its counterparts in other dissertations. For example, the conclusion section of Dissertation 14 (i.e. D14C6C) is much shorter at 389 words than the remaining conclusion sections in other dissertations, where the average number of words among the 52 excerpt samples in MTDC is 426.

Table 32 below shows the normalised number of pairs across dissertation functional sections in MTDC[57], including the maximum, minimum and mean numbers of pairs of nine lexical cohesive categories identified in each section. It is not surprising to see that the repetition category is the biggest one among all nine categories, which is in line with the results of the MTMC corpus. In the remaining categories of lexical cohesion, the pairs in each category in terms of five functional sections vary (see figure 13 below).

[57] The detailed normalised frequencies of lexical cohesive devices in MTDC are shown in appendix I.

Table 32 Normalised frequencies (per 1,000 word) of lexical cohesion categories in functional sections

	introduction			literature review			methodology			finding and discussion			conclusion		
	max.	min.	mean	max.	min.	mean	max.	min.	mean	max.	min.	mean	max.	min.	mean
repetition	216	97	161	200	134	172	198	130	171	184	137	162	177	119	151
synonymy	36	7	21	24	7	14	36	9	18	37	5	18	27	5	16
hyperonymy	10	2	6	9	0	4	18	0	7	7	0	4	5	0	3
hyponymy	17	0	9	35	2	16	11	0	6	20	0	7	15	2	6
meronymy	10	0	3	29	0	5	22	0	8	15	0	7	10	0	3
signalling nouns	19	5	8	20	2	12	19	7	11	17	5	10	20	3	10
identity	26	2	10	31	2	12	14	5	9	23	0	9	21	2	7
antonymy	23	2	9	11	0	6	16	2	6	14	2	7	22	0	9
collocation	21	5	11	20	4	13	31	7	16	24	2	12	36	10	20
total			239			253			253			235			225

Based on appendix I, the range of frequencies in each lexical cohesive category are shown in the table below:

Table 33 Range of lexical cohesive pairs in MTDC

category	rep.[58]	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.
number range	97 – 216	5 – 37	0 – 18	0 – 35	0 – 29	2 – 20	0 – 31	0 – 23	2 – 36

Compared with MTMC in which only three categories have the numbers of minimum pairs above zero, there are four categories in MTDC. The reason for this may lie in the fact that the sizes of samples in MTDC (426 in average per sample) are larger than those in MTMC (337 in average per sample[59]). Another finding here is that in the meronymy category, several samples in all functional section categories do not have any pairs (see appendix I). This result is similar to other previous studies (e.g. Hoffman, 2012), as this category is used more frequently in describing parts of an entity (Halliday & Matthiessen, 2014, p. 648) rather than in elucidating abstract relations in text, which is the main feature of the samples in this study for both corpora.

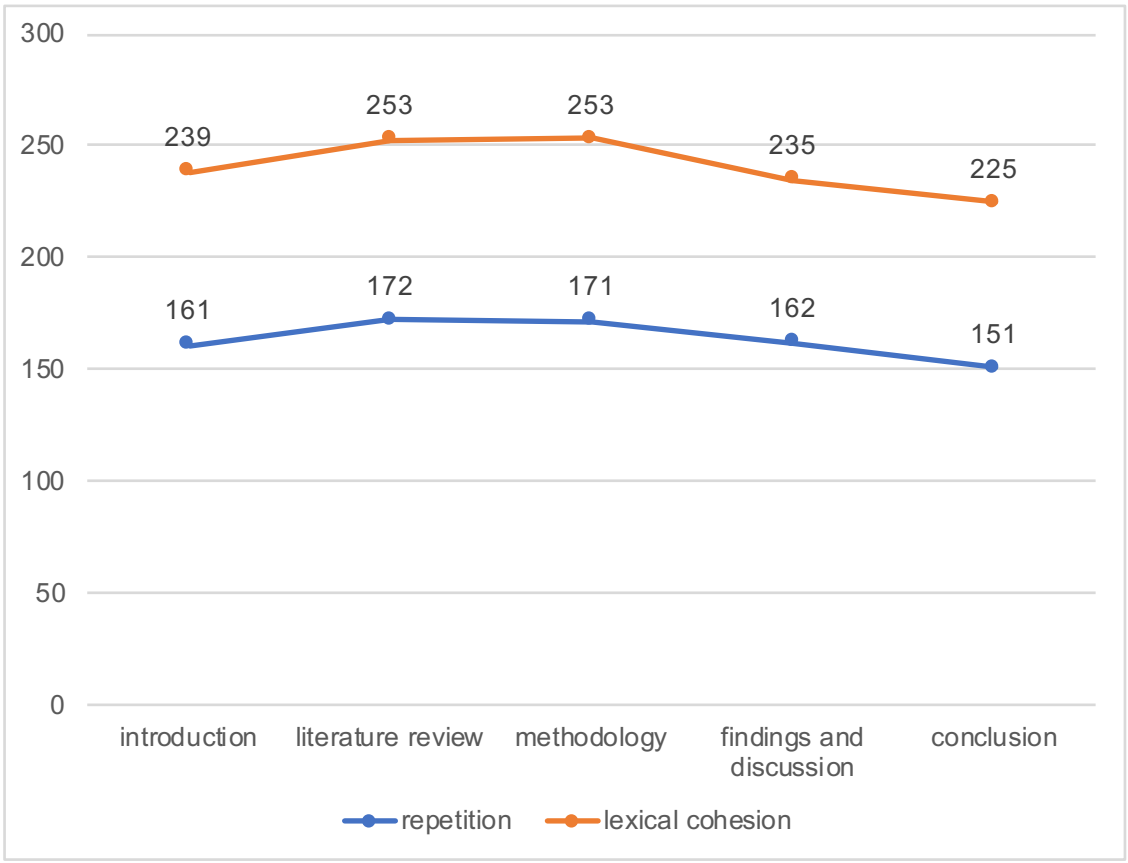
The Levene's test ($p = 0.001 < 0.05$) (see table J3 in appendix J) indicated that the functional section groups do not have equal variance on the use of lexical cohesion. As discussed in chapter 6 (see section 6.7), in this case, the K-W test (see table J4 in appendix J) was conducted for the comparison among functional-section groups regarding the frequencies of lexical cohesive devices. The result of this test (Chi square = 7.706, $p=0.103 > 0.05$, df = 4) shows that there is no significant difference between functional sections regarding the use of lexical cohesive devices. However, some interesting points were found in the comparison of lexical cohesive pairs used in functional section groups.

[58] Rep. = repetition; syn. = synonymy; hyper. = hyperonymy; hypo. = hyponymy; mer.= meronymy; sig. = signalling nouns; ide. = identity; ant. = antonymy; and col. = collocation

[59] The size of samples in MTDC is generally larger than that in MTMC because some raw data in the latter corpus are less than 500 words (see section 6.5, p. 183) while in the former corpus each collected dissertation text is more than 10,000 word length.

As shown in figure 11 below, there are more pairs in the literature review and methodology sections than in the remaining three sections. Comparing the tendencies of cohesive pairs in all nine categories and repetition pairs across functional sections, a noteworthy point is that these tendencies are similar, which reveals that the overall tendencies of lexical cohesive pairs used in functional sections are greatly influenced by the tendencies of repetitive pairs used in these sections. The frequencies of repetition pairs in each sample across the functional sections are shown in table 34 and related comparisons of the use of repetition pairs among the functional sections are demonstrated in table 35 and figure 12 below.

Figure 11 Comparison between normalised frequencies (per 1,000 word) of lexical cohesive pairs in nine categories and repetition pairs in five functional sections



As shown in table 35, the maximum and minimum values of the frequencies of repetition pairs are in the introduction section. This indicates the individual

variability in this section may be higher than that in the other functional sections, which is substantiated by the highest standard deviation value (38.873) of the introduction section in table 35 and the larger size of the box in the boxplot in figure 12. This higher individual variability in the introduction section does not contribute to the overall frequencies of repetition pairs in this section, although the highest number of repetition pairs occurs in sample D4C1I.

As regards the reason for the higher frequencies of repetition pairs in the literature review and methodology sections, it can be deduced from the information shown in figure 12. The pictorial representation of the distribution of frequencies of repetition pairs shown in the boxplot below reveals that the median values (represented by the line across the inside of the box) of the literature review section and the methodology section are higher than the remaining functional sections. Furthermore, as the length of each box contains 50% of cases in each group, it can be noted from the boxplot that the majority of samples in the literature review section and the methodology section have higher frequencies of repetition pairs than the other three sections, which contributes to the overall higher frequencies of repetitional pairs in the former two sections. The reason for this result may lie in the function of each section in dissertation texts.

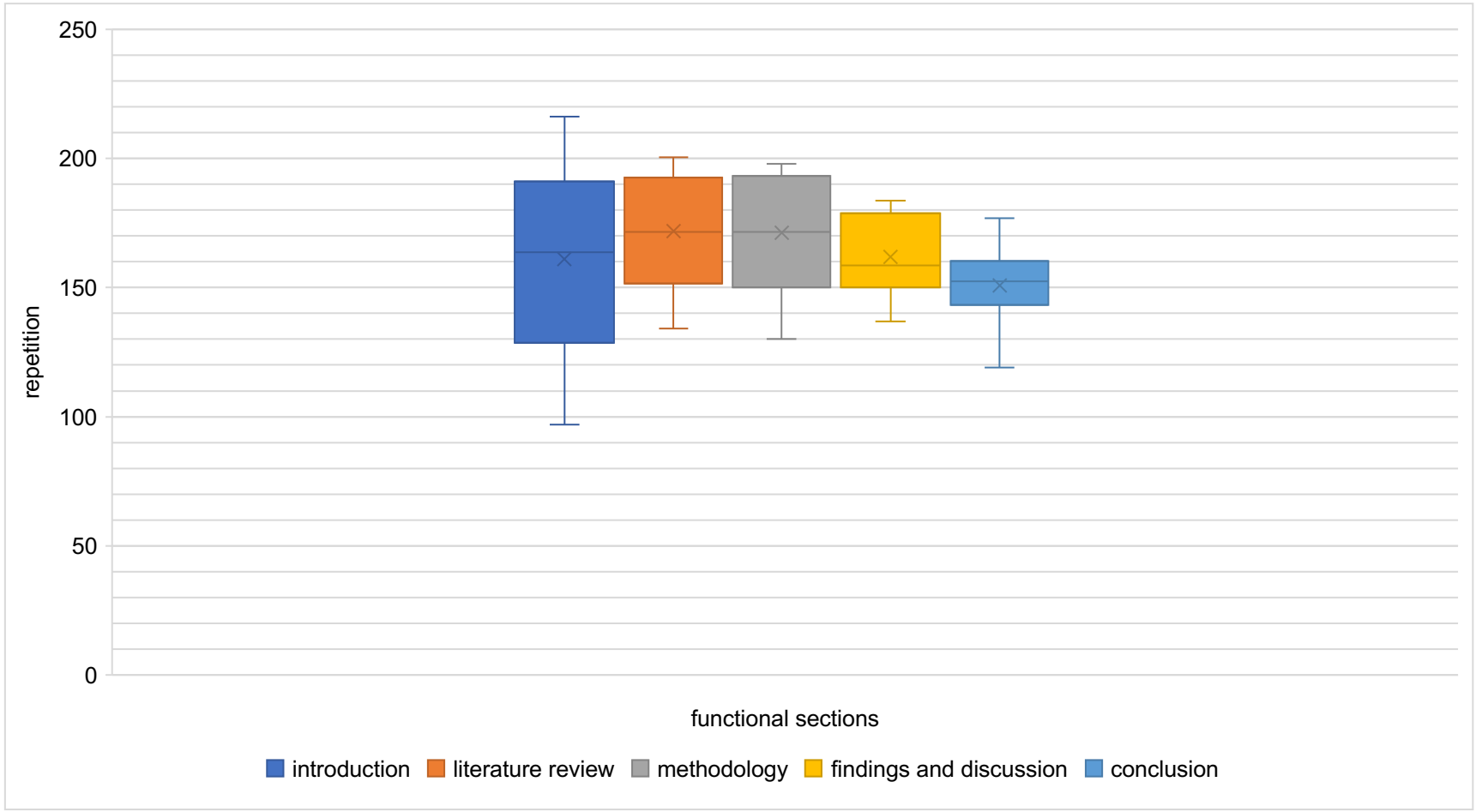
Table 34 Normalised frequencies (per 1,000 words) of repetition pairs in each sample across the functional sections

introduction		literature review		methodology		findings and discussion		conclusion	
D2C1I	186	D2C2L	166	D2C4M	160	D2C5F	159	D2C6C	177
D4C1I	216	D4C2L	193	D4C3M	140	D4C4F	179	D4C5C	158
D6C1I	190	D6C2L	148	D6C3M	168	D6C4F	151	D6C5C	119
D7C1I	192	D7C2L	187	D7C3M	198	D7C4F	163	D7C5C	148
D8C1I	164	D8C2L	200	D8C3M	188	D8C4F	184	D8C5C	152
D11C1I	117	D11C2L	134	D11C3M	130	D11C4F	149	D11C5C	138
D12C1I	141	D12C2L	192	D12C3M	195	D12C4F	158	D12C6C	153
D13C1I	147	D13C2L	171	D13C3M	191	D13C4F	178	D13C5&6&7C	150
D14C1I	97	D14C2L	156	D14C3M	172	D14C4F	137	D14C6C	162

Table 35 Standard deviation of the normalised frequencies (per 1,000 words) of repetition pairs in the functional sections

	N	minimum	maximum	mean	std. deviation
introduction	9	97.00	216.00	161.111	38.873
literature review	9	134.00	200.00	171.889	22.811
methodology	9	130.00	198.00	171.333	24.408
findings and discussion	9	137.00	184.00	162.000	15.692
conclusion	9	119.00	177.00	150.778	16.022

Figure 12 Comparison of the distribution of normalised frequencies (per 1,000 words) of repetition pairs in functional sections



As for the comparison of the frequencies of pairs used in other lexical cohesive categories (apart from repetition), as shown in figure 13 below, larger variations occur in the categories of synonymy, hyponymy and collocation as the gap between the maximum and the minimum values in these categories is larger than 5. In the collocation category, the highest figure (20) is in the conclusion section while the lowest figure (11) is in the introduction section, and their difference is 9, which indicates that students tend to use more collocational pairs in the final part of their dissertation texts rather than in the beginning part. This result may lie in the fact that collocational relations involve the elaboration on the same activities or topics, however, new topics and activities are introduced briefly in the introduction section, elaborated during the following functional sections, and finally summarised in the conclusion section. This may suggest that core lexical items involved in the elaboration of topics and activities have higher chances to co-occur in the conclusion section rather than in the introduction section. Therefore, as with repetition pairs, the use of collocational pairs is also influenced by the function of each section in dissertation texts.

As for the variation in the hyponymy category, it is noticeable that the literature review section has the highest frequencies of hyponymic pairs. The comparison of frequencies of hyponymic pairs in each sample across the functional sections is shown in table 36 below.

As can be seen in table 36, three samples in the literature review section have the highest numbers of hyponymic pairs across the 45 samples (i.e. 35 in D2C2L, 23 in D4C2L and 18 in D11C2L), which make the main contribution to the highest frequencies of hyponymic pairs in the literature review section while the other two samples have comparatively lower frequencies (i.e. 2 in D4C2L and 7 in D14C2L).

Table 37 further demonstrates that compared with the other four functional sections, the standard deviation in the literature review section is the highest, which indicates that there is higher variability between individual samples in this section. When taking a further look at the contexts for each hyponymic pair identified in the these samples, this study found out that the use of hyponymic pairs is highly related to the topics discussed in the contexts in which these pairs

occur. For example, as shown in table 38 below, in sample D2C2L, two cohesive chains with six hyponymic pairs each and an independent hyponymic pair were identified within successive clauses: six pairs form a cohesive chain of ‘abbreviations’; six pairs form another chain of ‘words’, and the independent pair is *symbols – [](()*. By contrast, in the same literature review category, sample D4C2L only has the lowest frequencies of hyponymic pairs with only one pair recognised, *second language – English* (see table 39 below). The comparison of the use of hyponymic pairs between the samples D2C2L and D4C2L in the same functional section category indicates individual variabilities between samples, which may be caused by topic preference. Certain context-based topics need elaboration in texts, such as the topics of ‘abbreviations’ and ‘words’ in sample D2C2L, while other topics in general sense may not require further explanation with more lexical items, such as the topic of ‘English as a second language’ in sample D4C2L.

Figure 13 Normalised frequencies (per 1,000 words) of lexical cohesive pairs in other categories (apart from repetition) in functional sections

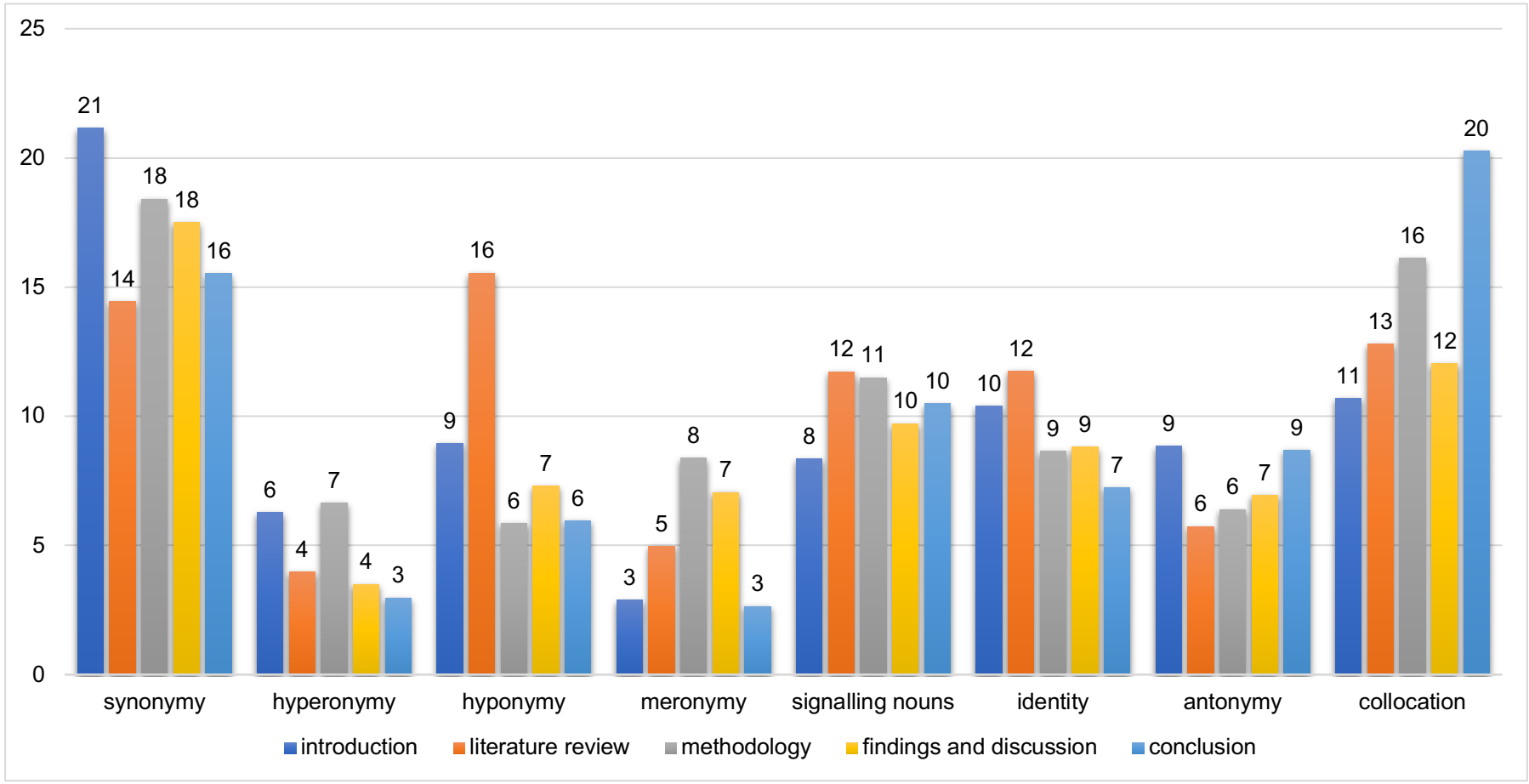


Table 36 Normalised frequencies (per 1,000 words) of hyponymic pairs in each sample across the functional sections

introduction		literature review		methodology		findings and discussion		conclusion	
D2C1I	14	D2C2L	35	D2C4M	9	D2C5F	11	D2C6C	5
D4C1I	7	D4C2L	2	D4C3M	11	D4C4F	13	D4C5C	5
D6C1I	12	D6C2L	12	D6C3M	2	D6C4F	7	D6C5C	5
D7C1I	5	D7C2L	16	D7C3M	9	D7C4F	0	D7C5C	10
D8C1I	9	D8C2L	23	D8C3M	0	D8C4F	15	D8C5C	2
D11C1I	17	D11C2L	18	D11C3M	5	D11C4F	0	D11C5C	2
D12C1I	0	D12C2L	11	D12C3M	7	D12C4F	13	D12C6C	7
D13C1I	12	D13C2L	16	D13C3M	2	D13C4F	2	D13C5&6&7C	15
D14C1I	5	D14C2L	7	D14C3M	7	D14C4F	2	D14C6C	3

Table 37 Standard deviation of the normalised frequencies (per 1,000 words) of hyponymic pairs in the functional sections

	N	minimum	maximum	mean	std. deviation
introduction	9	.00	16.99	8.946	5.362
literature review	9	2.19	35.48	15.539	9.661
methodology	9	.00	11.29	5.852	3.757
findings and discussion	9	.00	14.89	7.060	5.987
conclusion	9	2.38	14.74	5.956	4.084

Table 38 Part of hyponymic pairs in the sample D2C2L

Since there will be a large number of Chinese characters occurring in this chapter, it is a good idea to clarify some symbols and abbreviations of words before the departure.	
I will use "[ ]" to mark the phonetic pronunciation of the characters and "( )" to show their meanings.	symbols – []/()
The abbreviation "PL." stands for "plural";	abbreviations – PL; words – plural
"MW." **means** "measure word";	PL. – MW; plural – measure word
"PT." **means** "past tense";	MW. – PT; measure word – past tense
"NP." represents "**noun** phrase";	PT. – NP; past tense – noun phrase
"pred." **stands for** "predicative";	NP. – pred; noun phrase – predicative
and "arg." **means** "argumental".	pred. – arg; predicative – argumental

Table 39 Hyponymic pairs in the sample D4C2L

Teacher feedback has been the focus in the research on second language writing.
A large amount of research has focuses on investigating the usefulness and effectiveness of teacher feedback.
A bulk of surveys of students' feedback preferences indicate that ESL students greatly value teacher feedback and consistently rate it more highly than alternative forms, such as peer.
The reason for this trend is partly
because learners believe that teachers 'possessed a better knowledge of English language and therefore provide more helpful feedback than peer learners'.

8.3 Fine analysis of text excerpts in the MTDC corpus

In order to be compatible with the extract analysis conducted for the MTMC corpus in chapter 7, five texts below were selected from the introduction sections of five samples in the MTDC corpus. The analysis of these texts follows the analytical strategies discussed in chapter 6. Specifically, excerpts 1, 2, 3, 4 and 5 are from the five functional-section groups respectively. In order to explore the use of lexical cohesion in the functional sections of texts, each excerpt was analysed in an analytical table (see appendix K), and will be illustrated using the colour-coded text framework described in chapter 6, in the following sub-sections. Table 40 demonstrates the detailed source for these excerpts as follows:

Table 40 Details for selected excerpts

name	word count	functional section	source
excerpt 1	188	introduction	Dissertation 2 Chapter 1 (D2C1I)
excerpt 2	181	literature review	Dissertation 4 Chapter 2 (D4C2L)
excerpt 3	172	methodology	Dissertation 6 Chapter 3 (D6C3M)
excerpt 4	172	findings and discussion	Dissertation 7 Chapter 4 (D7C4F)
excerpt 5	173	conclusion	Dissertation 8 Chapter 5 (D8C5C)

8.3.1 Analysis of excerpt 1 from the introduction section

Firstly, excerpt 1 from the introduction group will be analysed and discussed. The text version of cohesive analysis for excerpt 1 is shown below.

Excerpt 1

Chapter 1 introduction// 1.1 Research **Background**// English Noun **plural morpheme** is one of the few **grammatical morphemes**// which **Chinese-speaking** learners started to **learn** from a young age.// Actually the early **learning** of this is not only true for **Chinese-speaking L2 learners**.// The study done by Dulay and Burt with Spanish-**speaking** children also indicated that// **plural "-s"** came first in the common order of acquisition for certain **structures** in **L2 acquisition**.// Krashen, Sferlazza, Feldman, and Fathman also did a picture questioning test among **speakers** from different language **background** and found that// the most **frequent grammatical item** was **plural "-s"**.// These **studies** show that **plural morpheme** is a **grammatical item**// which has been

taught to **learners** for a long time and been used with a high **frequency**.// Surprisingly, it is such a **grammatical structure**// which is widely believed to be easy to **acquire**// that causes persistent trouble to **Chinese learners** especially in spontaneous **speaking** situations.// This makes me think that// the reason may be laid in the linguistic differences between **English** and **Chinese** in terms of **plural** formation and the different ways of psychological processing between English and Chinese **speakers**.

The following qualitative analysis of excerpt 1 will detail the use of lexical cohesive devices in each category. In terms of the category of repetition, there are simple repetitions (e.g. *speaking*[60]), as well as complex repetitions with derivational variations (e.g. *acquire* – *acquisition*) and grammatical changes (*structures* – *structure*). This result is not unusual throughout the analysis of all the samples in MTDC.

There are four collocational relations in excerpt 1:

8-1 your age – children (elaborative collocation, henceforth EC)
8-2 L2 learners – acquisition (EC)
8-3 Language – linguistic (EC)
8-4 L2 acquisition – psychological processing (EC)

All four pairs are elaborative collocations related to the same topic, 'second language acquisition'. *Young age* is a characteristic of *Children*, both of which items elaborate on the topic of children. *Acquisition* is what *L2 learners* do when learning second language. These two items are closely connected with the topic, 'second language acquisition'. *Linguistic* here is connected with *language*, referring to some aspects of *language* regarding the difference between English and Chinese as two languages. *Psychological processing* is one stage in *L2 acquisition* process. Therefore, they also refer to the same topic. The use of these four pairs demonstrates the co-occurrence nature of collocation within a short span of text, contributing to the overall lexical cohesiveness of excerpt 1.

[60] The related lexical items in cohesive pairs will be italicised in discussion paragraph.

However, there are overall 5 collocational pairs (the fifth pair is *processing – encoded*) in the sample D2C1L (excerpt 1 is from this sample), all of which elaborate on the same topic, second language acquisition, although the fifth pair represents an activity-related collocation relation. This may indicate that topic is an important factor which influences the frequencies of collocational pairs used in students' writing.

As for the antonymy category, there are two pairs:

8-5 learn – taught
8-6 easy – trouble

Learn and *taught* are converse antonyms, as they are reciprocal procedure verbs which refer to the same action but from different perspectives of students and teachers respectively. *Easy* and *trouble* are not strictly antonyms, because they belong to different word classes with *easy* being an adjective and *trouble* a noun. They form a pair of near-antonymic relation. Furthermore, they are contrary antonyms as they are gradable and allow for comparative relations. For example, *plural "-s"* is believed to be **easier** to acquire than other *grammatical structures*; or other *grammatical structures* are believed to be **more troublesome** to acquire than *plural "-s"*.

There is only one pair in the identity category in this excerpt, which is *English noun plural morpheme – plural "-s"*. Both items refer to the same grammatical structure, but the former is the name of the structure *plural "-s"* while the latter is the written form of this structure. The two lexical items emphasise various aspects of the same subject.

There is also one lexical cohesive pair in the category of SNs: *the study done by Dulay and Burt /*[61] *a picture questioning test among speakers from different language background – these studies.* The SN in this pair is *studies* in the phrase

[61] "/" separates two segments which are referred to in a cohesive pair but occur in different places in a text.

these studies which refers back to two individual segments in the same text, which are *the study done by Dulay and Burt* and *a picture questioning test among speakers from different language background*. The reason for including the determiner *these* when discussing the use of SNs in this case is to indicate that *studies* here is used as an anaphoric cohesive device to replace previous contents. As mentioned in chapter 3, the function of SNs in texts is to make the texts more succinct and cohesive, i.e. "there should be no verbosity and [redundant] repetitions" (Beijing University, 1973, pp. 104-105, cited in Mohan & Lo, 1985, p. 520). In this case, the SN phrase *these studies* is an economic way of encapsulating two chunks of segments without distorting necessary information.

As for superordinate relations, there are three hyperonymic relations and three hyponymic relations:

Hyperonymy:

8-7 English – L2
8-8 Spanish – language
8-9 Plural "-s" – grammatical item

English here is one type of *second language* (i.e. *L2*); *Spanish* is a type of *language*; and *plural "-s"* is a type of *grammatical item*.

Hyponymy:

8-10 English – Chinese
8-11 Chinese – Spanish
8-12 structures – plural "-s"

English, Chinese and *Spanish* are three kinds of languages. Therefore, they form two co-hyponymic relations. *Plural "-s"* is a type of *grammatical structure*, which is thus a subordinate of *structure*. It is noticeable that the above lexical items in the pairs of hyperonymy and hyponymy are similar or the same. The only

difference is just that their sequences in the text are changed. For example, if *plural "-s"* appears before *grammatical item* (or *structure*) in the text, the relation between *plural "-s"* and *grammatical item* (or *structure*) is hyperonymic; if *grammatical item* (or *structure*) occur before *plural "-s"*, then the relation between them is hyponymic. This fact shows that the high frequencies of hyperonymic and hyponymic pairs used in excerpt 1 may not indicate the diverse use of lexical cohesive pairs in these two categories in this excerpt.

As for the synonymy category, there are three pairs:

8-13 research – study
8-14 the most frequent – high frequency
8-15 structure – formation

Pairs 8-13 and 8-15 are synonymous relations while pair 8-14 is a near-synonymic relation. The reason for counting pair 8-14 as near-synonymy is that firstly, the items in this pair do not belong to the same word class: *frequent* is an adjective while *frequency* is a noun; secondly, the meanings of these two items are not exactly the same: *the most frequent* is an expression of superlative form of *frequent*, which means in this excerpt that *plural "-s"* appears more frequently than any other grammatical items in the picture questioning test. However, *high frequency* in this context means *plural morpheme* occurs very frequently without an obvious comparative meaning. Although these two lexical items are slightly different in the exact contextual and semantic meanings, they still elaborate on the topic that *plural "-s"* (or *plural morpheme*) is frequently used in text. Therefore, these two items still form a near-synonymic pair in this excerpt.

8.3.2 Analysis of excerpt 2 from the literature review section

The colour-coded text version of the detailed analysis of excerpt 2 is as follows:

Excerpt 2

1. **Literature** review// 1.1 Definition// The **literature** has showed that// the effect of **teacher feedback** on **students**' improvement is not explicit and does not achieve its full potential.// Albeit with the uncertainty of validity of **feedback**, it is

undeniable that// **teacher feedback** is indispensable in **student** learning, especially in a process-approached class.// Therefore, it is **necessary** to look at// what **feedback** is and //what should **teacher feedback** focus on.// Kluger & DeNisi define **feedback** interventions as 'actions taken by an external agent to **provide information** regarding some aspects of one's task **performance'**.// This argument is further explored by **Ramaprasad** that// **feedback** should **provide information** about how to **narrow** the **gap** between **students'** current **performance** and the reference level,// otherwise **teacher** commentary is less useful in helping **students** improve their skills.// The difference between the two interpretations is// **Ramaprasad** emphasizes the **necessity** of the **information** on how to alter the **gap**.// When **narrowing** the research scope down to writing **feedback**, Keh's regards feedback as 'input from a reader to a writer with the effect of **providing information** to the writer for revision'.

The discussion of the analysis of excerpt 2 will begin with the use of repetition pairs in this excerpt. There are 20 pairs in total in the repetition category in which one repetitive pair is with derivational change (*necessary* – *necessity*) while the remaining pairs are simple repetitions.

There are three synonymous pairs:

8-16 indispensable – necessary
8-17 regarding – about
8-18 feedback – commentary

Pair 8-16 is synonymous as both of the lexical items are adjectives and express the similar meaning that 'something' is highly important. In the case of indispensable, 'something' refers to teacher feedback which is important in student learning. In the case of necessary, "something" denotes that examining some aspects of teacher feedback is important because of its centrality in student learning. Here, indispensable and necessary are used in succeeding clauses, which indicates that the author probably intended to avoid simple repetitions which can make a text dull to read (Ong, 2011, p. 55). In terms of pair 8-17, in the Oxford English Dictionary, the explanation for regarding is "about" (OED

Online, 2019l), and both regarding and about in this pair are used based on their semantic meaning which is “connected with”. Therefore, these two items also form a synonymous relation. As for the third pair, the meanings of feedback and commentary are slightly different. Feedback refers to “opinions about something” (Cambridge Dictionary, 2019) which is more general, while commentary means a set of written remarks on a text (Cambridge Dictionary, 2019) which is more specific. However, in this context, both items point to the same referent which is teachers’ opinions about students’ performance. It is thus suitable to regard them as near-synonyms.

With regards to the use of SNs in this excerpt, there are three pairs:

8-19 feedback interventions…performance – this argument
8-20 feedback…performance/feedback…skills – the two interpretations
8-21 the difference – Ramaprasad…gap

In pair 8-19, the SN argum*ent* is accompanied by the determiner *this*, which indicates that *argument* is used as an anaphoric device here to refer back to the previous clause *feedback interventions as ‘actions taken by an external agent to provide information regarding some aspects of one’s task performance’*. The same is the case with pair 8-20, in which the lexical item *the two interpretations* refers back to two previous clauses which are:

a. *feedback interventions as ‘actions taken by an external agent to provide information regarding some aspects of one’s task performance’.*

b. *feedback should provide information about how to narrow the gap between students’ current performance and the reference level, otherwise teacher commentary is less useful in helping students improve their skills.*

By contrast, the SN *difference* is used as a cataphoric device in pair 8-21, which signals the following segment *Ramaprasad emphasizes the necessity of the information on how to alter the gap*. As mentioned above, the encapsulating function of SNs as lexical cohesive devices operates by using condensed lexical

items to replace bigger chunks of information in order to reduce the length of the texts while increasing the cohesiveness of the whole texts. The three SN pairs above demonstrate this function successfully via two opposite referential devices – anaphora and cataphora.

There is only one pair in the identity category in excerpt 2, which is *teacher – external agent*. The co-referential relation between these two lexical items is highly contextual, as *teacher* is exclusively regarded as *external agent* in this context because the teacher takes action to provide feedback for students. In excerpt 2, *Kluger & DeNisi's* definition of *feedback interventions* elaborates on the definition of *external agent* in a general sense rather than denoting *external agent* as a specific role in the process of providing feedback. However, being put into the relevant context in excerpt 2, *external agent* refers to *teachers*. Therefore, the relation of identity of referent between these two items is incontrovertible in this context.

As for collocational relations in this excerpt, there are three pairs:

8-22 *students – teacher*
8-23 *students – class*
8-24 *literature – research*

The first two pairs exist under the same frame of school. *Students* and *teacher* are two roles in school, and *class* is the period of time when *students* and *teacher* need to spend at school. In these two pairs, *students* is regarded as the trigger to initiate this frame because *students* appears before *teacher* and *class* in excerpt 2. Since *teacher* and *class* are included in the same clause, which cannot be counted as cohesive pairs in this study, *student* is identified as the preceding item in both pairs 8-22 and 8-23. The third pair elaborates on the topic of *research*. *Literature review* is part of doing *research*. Therefore, *literature* is used at first to trigger the *research* frame, which makes the use of *research* more appropriate in the following text.

8.3.3 Analysis of excerpt 3 from the methodology section

As with sections 8.3.1 and 8.3.2, the analysis of excerpt 3 will explore and describe the use of lexical cohesive devices identified in this excerpt. The colour-coded text framework version of excerpt 3 is shown below.

Excerpt 3

3. Methodology// This study aims to **find** out **reasons**// why **CLT** cannot be **implemented** in English discipline class of Yunnan Nationalities **University**,// and if possible, **find** some **solutions** and **give** some feasible **suggestions** to solve these problems.// Till now, there has been a lot scholars and **researchers**// who **researched** the constraints and challenges of **CLT implementation** in **China**.// Most of these **researche[r]s** talk about impediments generally in the whole **Chinese context**,// thus this **dissertation** can be a test to prove// whether findings in the **university** match the previous statements.// Since the concept of **CLT** was initially spread to **China** in 1970s,// why it cannot be **implemented** in this **context** after over 40 years.// Thus, **research** questions can be:// how many **reasons** are due to teachers?// How many are due to students?// Do grammar based exams matter a lot?// Or is it because lacking authentic language environment?// After finishing this **dissertation**, **reasons** for **CLT** fails [failing] to be **implemented** and **solutions** proposed are listed.// At the same time, some **suggestions** can be **given**.

As with the discussion of the previous two excerpts, the use of lexical cohesive devices in different categories will be discussed separately. The first category discussed here is repetition. There are 20 pairs in excerpt 3, in which 8 pairs are complex repetitions. Among the 8 pairs, *research*-stemmed repetitions form three pairs: *researchers – researched – researchers – research*, while *China* and *Chinese* form two pairs. It is often the case that several pairs share the same stem, which indicates that this group of lexical items is central for the proposition of this text. The remaining complex repetitive pairs are *implemented – implementation* and *give – given*. Except the pair *give – given* which is with grammatical changes, other seven pairs are derivational variants.

As regards the synonymy category, there are four pairs in excerpt 3:

8-25 *scholars – researchers*
8-26 *constraints – impediments*
8-27 *study – research*
8-28 *cannot be – Failing to be*

Pair 8-28 is a special case, in which *cannot be* is a modal verb structure while *failing to be* is an intransitive verb structure. It is important to include the grammatical structure following the verbs in order to express the whole meaning of the phrase in which the verbs are included. In this case, *cannot* and *failing* are not synonyms per se. Only when they are followed by the '+ *to* infinitive' structure in this text can they express the same meaning which is 'not being able to do something'. It is also interesting to note that these two lexical items are used for the same purpose of expressing the idea that it is unsuccessful to implement CLT in China. This indicates that the author was aware of using synonyms as an alternative cohesive device to avoid simple repetitions. This can be seen as a positive evidence to challenge the idea that Chinese students tend to overuse simple repetitions in their academic writing (e.g. Ong, 2011). The remaining three pairs above are clearly identifiable as synonyms as the related lexical items are in the same word classes as well as sharing similar meanings in the context in excerpt 3.

Regarding superordinate relations, only one pair is identified in hyperonymy, which is *Yunnan – China*. This pair refers to part-whole relation as *Yunnan* is a province of *China*. Until now, according to the analyses of the three excerpts above, it is not surprising to figure out that superordinate relations do not frequently occur in Chinese students' dissertations. The reason is probably that Chinese students have not been taught such cohesive devices explicitly in EAP classes, and thus are not aware of using these devices consciously in their own academic writing.

There are four pairs of SNs identified in excerpt 3, which are:

8-29 *reasons – why CLT cannot be implemented in English discipline class of Yunnan Nationalities University*

8-30 *CLT cannot be implemented in English discipline class of Yunnan Nationalities University – these problems*

8-31 *impediments generally in the whole Chinese context – the previous statements*

8-32 *research questions – how many reasons are due to teachers? How many are due to students? Do grammar based exams matter a lot? Or is it because lacking authentic language environment?*

What is also worthy of mention here regarding the first two pairs is that the lexical items involved occur in neighbour clauses, and the segments replaced by these two SNs *reasons* and *problems* are almost the same apart from the fact that the first segment includes an extra word *why*. The formulae below can demonstrate their relations more clearly:

a) *reasons* = *Why* + the remaining part of the first segment;
b) the remaining part of the first segment = the second segment
c) the second segment = *these problems*

It also can be seen from the three formulae that if the SNs are anaphoric, normally they are preceded by a determiner, such as *this, these* or *that*. Therefore, the first pair clearly represents a cataphoric relation (which has no determiner) while the second one is anaphoric (which has a determiner *these*). As for the remaining two pairs, pair 8-31 represents an anaphoric relation between its two lexical items, while the lexical items in pair 8-32 form a cataphoric relation.

In the identity category, there is only one pair: *this study – this dissertation*. The lexical items in this pair denote two identities of the same referent: *this study* reveals the activity nature of the referent in question, while *this dissertation* demonstrates which document genre this referent belongs to.

There is also only one pair showing antonymy in excerpt 3, which is the classical antonymic pair in this study, *teachers – students*. As this pair has been discussed before (see pair 8-22 in section 8.3.2), there will be no further explanation here.

Finally, there are two pairs of collocational relations in this excerpt:

8-33 methodology – study
8-34 class – teachers

These two pairs represent ECs, which are very similar to the collocation pairs in excerpt 2 (see pairs 8-22, 8-23 and 8-24), as both pairs 8-33 and 8-34 are related to two frames: school and research. In excerpt 3, the first pair regards *methodology* as one of the stages of conducting a *research*, and *study* has similar meaning to *research*. Therefore, *methodology* and *study* form an elaborative collocational relation through elaboration of the research topic. The second pair starts with the item *class* which refers to a typical activity happening at school. Therefore, *class* triggers the school frame. Teachers are the participants in this activity and in excerpt 3 this lexical item, *teachers*, is used to expand the school frame in the subsequent text.

8.3.4 Analysis of excerpt 4 from the findings and discussion section

As with the discussion about the three excerpts above, the qualitative analysis of excerpt 4 will be shown regarding its use of lexical cohesion, following the demonstration of the colour-coded text version of this excerpt.

Excerpt 4

1 Findings// **1.1 Law seminars'** influences on **spoken English//** According to the information from the **interviewed law students**, it is suggested that// the problem-based **law seminar** has some positive effects on their **spoken English**, but not remarkably advantageous.// The general consensus they have reached is that// through seminar learning, their confidence to **speak English** has been noticeably promoted,// and they are able to express themselves clearer and more logical.// Basically, **law students** can choose four **modules** in their LLM programme,// but **some modules** are lecture-based, not containing **seminars**// and **some modules** have a mixed arrangement.// The **law students**// I have **interviewed**, […] have at least two **seminars** in a week,// and some of them may have three or four times.// **Some seminars** are 2 **hours** as an independent

session,// but **some seminars** are preceded by a one-**hour lecture**, only lasting for I hour.// The **size** of **seminars** varies:// **some seminars** are quite small, only consisting of 6-8 people,// but the normal **size** is around 20, with the maximum number of 40 **students**.

The analysis of excerpt 4 starts with the repetition category. There are 19 repetitive pairs, in which only two pairs are complex repetitions:

8-35 law seminars – law seminar
8-36 spoken – speak

Pair 8-35 represents repetition with grammatical changes while pair 8-36 denotes repetition with derivational variations.

As for the remaining categories, there is only one synonymous pair which is *influences* – *effects*. The two items in this pair are regarded as synonyms as they belong to the same word class of nouns. In this excerpt, *influences* is used to express the general impact of law seminars on spoken English, while *effects* is adopted to express a good impact on law seminars as it is preceded by an adjective *positive*. Therefore, both items elaborate on the same topic. It seems likely that the author replaced *influence* with *effect* in the following text to avoid simple repetitions, which is in line with the use of synonyms in excerpt 2.

For the superordinate relations, there is one pair in hyponymy and four in meronymy, which is not usual compared with other excerpts. The hyponymic pair is *seminar* – *lecture*, in which *seminar* and *lecture* are co-hyponyms of a more general category of module teaching approaches. The four meronymic pairs are all about different numbers:

8-37 four – two
8-38 two – three/four
8-39 2 – one
8-40 6-8 – 20

The first two pairs are related to the number of modules or seminars; the third pair describes the number of hours; and the fourth pair is about the number of students in the seminar. Such high frequencies of meronymic relations are not typical in the MTDC corpus. In fact, the category of numbers is used to be included in the sub-category of ordered set in collocation (Tanskanen, 2006, p. 61). The reason for treating these numbers as meronyms in the present study is that different numbers are part of the number 'as a whole category'. The example above indicates that compared with meronymic relations used to describe entities in expository texts, meronymic devices can also be used to refer to abstract relations in other texts (e.g. excerpt 4). Therefore, based on the analysis of samples in the current corpora, this study expands the definition of meronymic relations from part-whole relations of 'concrete entities' into part-whole relations of 'abstract categories'.

There are two pairs identified in excerpt 4 in the SN category:

8-41 *the general consensus* – through seminar learning, their confidence to speak English has been noticeably promoted,// and they are able to express themselves clearer and more logical
8-42 law students – *people*

Pair 8-41 is a typical use of the SN *consensus* as a cataphor to encapsulate a chunk of information in the succeeding text. Pair 8-42 is more related to the function of GNs which is to point out a more general superordinate of a specific referent. In this case, *people* is the more general superordinate of the group of *law students*.

In the collocation category, there are six pairs used in this excerpt:

8-43 seminars – students
8-44 students – learning
8-45 learning – modules
8-46 seminar – LLM programme
8-47 size – small

8-48 size – the maximum number of 40 students

In these collocational pairs, only pair 8-44 is an activity-related collocation while the remaining five pairs denote elaborative collocational relations. Specifically, pair 8-44 represents an activity-related collocation because *students* are the participants of the activity of *learning*. As for the remaining collocation pairs, in pair 8-43, *seminars* are occasions when *students* discuss something, which triggers the frame of seminar, and makes the occurrence of *students* more appropriate later in the text. In pair 8-45, *learning* and *modules* expand on the frame of university, because *learning* is one of students' main activities at university, and *module* refers to the course at university. In pair 8-46, *seminar* is one of the teaching approaches in the modules of the *LLM programme* which is the frame in this relation. Pairs 8-47 and 8-48 are within the same frame of *size*: *small* elaborates on the general feature of *size* while the multi-word lexical item *the maximum number of 40 students* describes a certain feature of *size*. These pairs above demonstrate the use of two types of collocational relations in several aspects, providing exemplars for the teaching of such relations in EAP classes.

8.3.5 Analysis of excerpt 5 from the conclusion section

The final part of this section focuses on the examination of the lexical cohesive analysis of excerpt 5. The colour-coded text version is as follows:

Excerpt 5

Conclusion// This **concluding section** will first cover relevant **limitations**, **considering** their **impact** on the **conclusions//** which can be **drawn** (5.1)//. Then, being aware of these **limitations**, a summary of the results will **follow** (5.2).// **Following** this, implications will be **drawn**, specifically for pedagogy (5.3).// The final **section** will open up possibilities for future research (5.4).// 5.1 Limitations// There are several **limitations** that must [be] taken into consideration in this **study**,// which have an **impact** on **conclusions**, and sometimes present options for further investigation.// First, there are **limitations** concerning the used **corpora**.// The Yale University lectures may not be representative of lectures at **other Universities**,// so the findings may not be applicable in **other** countries and cultures, or even other **Universities** in the USA.// Similarly, the TED talk

corpus was compiled by the **most viewed talks** and was **limited** to the three main **tags**.// **Talks** outside of these **tags**// (which **includes** the […] six **most viewed**)// and **talks** not selected due to the practical **considerations** of this **study** were not **included** in the **corpus**.

As with the analysis of previous 4 excerpts, the discussion of the analysis of excerpt 5 firstly focuses on the repetition category. There are 16 repetitive pairs, in which 7 pairs are complex repetitions:

8-49 conclusion – concluding
8-50 follow – following (verb)
8-51 concluding – conclusions
8-52 limitations – limited
8-53 considering – considerations
8-54 includes – included

Pairs 8-49, 8-51, 8-52 and 8-53 are repetitions with derivational variations while pairs 8-50 and 8-54 are with grammatical changes. Compared with previous excerpts, even the total number of repetitive pairs are less, the percentage of complex repetitive pairs in this excerpt is higher than other excerpts, which reveals the diversity of repetitive pairs used in this excerpt.

As for the remaining categories, four synonymous pairs are identified in this excerpt, which are:

8-55 future research – further investigation
8-56 considering – concerning
8-57 results – findings
8-58 cover – includes

All four pairs are synonymous in the traditional sense. The first pair is interesting to discuss as the words in each item are also synonymous respectively: *future* – *further* and *research* – *investigation*. Therefore, the two whole lexical items form a synonymous relation.

There is only one meronymic pair among all superordinate relations, which is *the used corpora – The Yale University lectures/the TED talk corpus*. From the use of *corpora* and *corpus*, it can be seen that the *corpora* included more than one *corpus*. In this pair, *the used corpora* comprise two corpora: *The Yale University lectures/the TED talk corpus*.

There is also one pair of SNs: *conclusion – this concluding section*. Here the SN is *section*. Furthermore, the use of the determiner *this* and the adjective *concluding* define the boundary of *section* in this context, which refers back to *conclusion* as one section of a dissertation.

Finally, there are four collocational pairs used in excerpt 5:

8-59 results – research
8-60 research – findings
8-61 the YALE University – the USA
8-62 corpora – compiled

All four pairs are ECs. Pairs 8-59 and 8-60 are related to the same frame of *research* as *results* and *findings* are two stages of conducting a *research*. The third pair expands on the USA frame as the Yale University locates in the USA. The final pair shares the corpus frame, in which the elaborative relation lies in the fact that the data needs to be *compiled* to form the *corpora*.

8.4 Conclusion

This chapter has firstly presented the quantitative analysis of the frequencies of lexical cohesive devices used in the samples across the functional section groups in the MTDC corpus; and demonstrated the fine analysis of 5 excerpts selected from 5 samples in the functional section groups.

According to the result of the non-parametric K-S test, overall, there is no significant difference between the five functional sections of dissertations in MTDC regarding the use of lexical cohesion. As was seen in figure 11, more pairs

are identified in the literature review and methodology sections, whereas the smallest number of pairs is found in the conclusion section. As mentioned in section 8.2, the distributions of repetitive pairs and overall pairs are similar through the different functional sections while, by contrast there is some variability in other categories. These observations indicate that the high frequency of repetitive pairs in the functional section groups has a significant influence on the frequency counts for the overall pairs in these groups.

As for the use of cohesive pairs in the remaining categories, there is no specific pattern or tendency in general, however, two noticeable points are worthy of mention. Firstly, in the hyponymy category, there are substantially more pairs in the literature review section than in other sections (see table 37). The reason for this is that certain topics, addressed in sample D2C2L, need more hyponymic pairs for further elaboration, and these pairs contribute significantly to the overall high frequencies of hyponymic pairs in the literature review section. Secondly, in the collocation category, the conclusion section has more collocational pairs than the remaining four sections, especially compare with the introduction section. This result may relate to different functions which these sections have in dissertation texts. Specifically, the conclusion section may have higher possibility of summarising core lexical items involved in the same topics or activities, which enables the co-occurrence of these items within a shorter span of text.

Based on this detailed comparison between each section in terms of the use of lexical cohesion, it can be noted that the use of lexical cohesive devices in the remaining lexical cohesive categories (except the repetition category) fluctuates significantly during the flow of text because of different factors, i.e. topic preference, writer's choice and students' knowledge of lexical cohesion.

In terms of the analyses based on the five excerpts, there are some noteworthy findings as well. For example, there are several numbers (e.g. *2, 4* and *one*) used in excerpt 4, and every two numbers form one cohesive pair. This particular use of 'numbers' to form meronymic pairs was only identified in this excerpt, which leads to the surprisingly high frequency of meronymic pairs in excerpt 4 compared

with other 4 excerpts[62]. This result indicates that subject matter can influence the types of lexical cohesion observed in texts. Thus, the observation of lexical cohesion is to some degree down to the context or topic, rather than individual writer's choice.

In the categories of synonymy and antonymy, there are a few near-synonymous or near-antonymic pairs used in the excerpts. Table 41 below summarises the near-synonymous or near-antonymic pairs appearing in these excerpts:

Table 41 Near-synonymous and near-antonymic pairs in 5 excerpts

	near synonymy	near antonymy
excerpt 1	the most frequent – high frequency	easy – trouble
excerpt 2	feedback – commentary	
excerpt 3	cannot be – failing to be	

It can be drawn from the samples in this table that the features for being near-synonymous or near-antonymic pairs are twofold: firstly, the lexical items in a pair belong to different word classes; secondly, the semantic meanings of the lexical items are somehow slightly different while their contextual meanings are relatively similar. The identification of these pairs substantiates the context-sensitivity of lexical cohesion which, in relation to pedagogy, clearly need to be taught with examples in specific contexts rather than being presented as abstract, contextless devices in classes and textbooks.

Similarly, as a new category developed in this study, the identity category is also context-sensitive, as is demonstrated by the excerpt analysis. For example, in the segment from excerpt 2 below, the identity pair *teacher – external agent* refers to the same subject in this particular context though in general *external agent* has a broader definition.

> teacher feedback is indispensable in student learning, especially in a process-approached class … Kluger & DeNisi define feedback interventions as 'actions

[62] The normalised frequencies of meronymic pairs in the excerpts are: 0 in excerpts 1, 2 and 3; 23 in excerpt 4 and 3 in excerpt 5.

> taken by an external agent to provide information regarding some aspects of one's task performance'. (from excerpt 2)

The encapsulating function of SNs is also revealed by the examples found in these excerpts. For example, the pair from excerpt 1 below illustrates this function:

> The study done by Dulay and Burt / a picture questioning test among speakers from different language background – these studies

In this pair, *these studies* refers back to two longer segments in excerpt 1, but compresses these information into only two words.

As for sub-categories in collocation, there are many more elaborative collocational pairs than activity-related collocational ones. In fact, only one activity-related collocational pair was found in the excerpt analysis, *students – learning*. The other 18 pairs are all elaborative collocational pairs. This result is similar to Tanskanen's (2006, p. 135) results in her analysis of academic articles, which shows that elaborative collocation is the dominant sub-category in collocation.

The qualitative analysis also shows students' use of lexical cohesive devices other than repetition to avoid redundant repetition in their writing. For example, in excerpt 2, the student used *indispensable* and *necessary* as synonyms in adjacent clauses:

> teacher feedback is indispensable in student learning, especially in a process-approached class.// Therefore, it is necessary to look at// what feedback is and //what should teacher feedback focus on.

From the analysis so far it can be seen that more repetitive pairs are used in the literature review and methodology sections compared to the remaining three functional sections. The low levels in certain categories are similar to the frequencies involved in MTDC. These common low frequencies of pairs used in these lexical cohesive categories are a finding which is worth emphasising in the

EAP pedagogy. However, there are still some evidences in the excerpt analysis, which show students' attempts to replace repetitive devices with other diver lexical cohesive devices.

Chapter 9 Discussion and conclusion

9.1 Introduction

This thesis has presented an examination of the use of lexical cohesion in Chinese postgraduates' academic writing. Two corpora, the first constructed from module assignments, and the second from dissertation texts, were analysed in order to explore the relationship between the use of lexical cohesion and the quality of Chinese students' written works; the usage of lexical cohesive devices in the flow of text; and linguistic features of lexical cohesive devices used in the student samples in the corpora.

Based on the quantitative and qualitative analysis of lexical cohesive devices used in the corpora in this thesis, this final chapter will firstly review the key findings of the study providing further discussion of these results (section 9.2). This will be followed by the discussion of the limitation of this study (section 9.3). Potential pedagogical implications for the teaching of lexical cohesion in EAP classes for Chinese students and further for EFL learners in general will be discussed in section 9.4. Noting the limitations for this study, suggestions will be made in section 9.5 for future research with regard to both the further investigation of lexical cohesion in general and the study of the use of lexical cohesion in Chinese students' writing in particular, as well as in relation to the development of research in learning the effective use of lexical cohesive devices. This will be followed by the concluding remarks of this thesis.

9.2 Key findings of this study

The following are considered to be the key achievements and findings in the current study. These achievements and findings are discussed in detail the in the subsequent sub-sections below:

1. a new analytical subcategory of identity was generated to support the operational analytical framework for lexical cohesion in the present study;

2. the complex SN-like phenomenon was evaluated and summarised as a subcategory of SNs from the perspective of lexical cohesion;

3. the tailored definition of collocation was developed for the present study, mainly based on the exploration of three main definitions of collocation and their corresponding approaches to collocation study;

4. the linguistic and cultural reasons for Chinese students' features regarding their use of lexical cohesion in English academic writing were evaluated from the existing studies;

5. the result of the ANOVA test conducted in the MTMC corpus substantiated the claim by previous studies that there is significant difference between Chinese students' written works with different marks regarding the use of lexical cohesion; and the result of the Turkey post hoc test reveals this difference lies in the fewer lexical cohesive pairs used in the merit group compared with the other three marking-scale groups, which might be caused by the small size of the sample in the merit group.

6. the quantitative analysis of MTDC suggested that there is no significant difference between functional sections in the flow of text within dissertations in terms of the use of lexical cohesive devices;

7. the qualitative analysis of both corpora identified several new grammatical and semantic features of different lexical cohesive devices.

9.2.1 Framework of lexical cohesion

While based on Halliday and Hasan's (1976) seminal model of lexical cohesion, this study adapted and developed this model in a number of ways in order to provide an enhanced and more fine-tuned approach to analysing lexical cohesion. Firstly, a new sub-category called 'other relations with identity of reference' was added to their original model. Secondly, the category of general nouns in Halliday and Hasan's model was developed into a broader definition of signalling nouns. Further, the definition of collocation in their model was modified to be more

specific and the sub-category of collocation was narrowed down in order to be more operationally utilisable in the analysis. Through the use of such a fine-tuned approach, it was possible to generate a number of interesting findings.

Firstly, identity was established as a new sub-category of reiteration. This identity sub-category was built on the identification of lexical cohesive pairs which had not been identified in previous studies. The identity sub-category[63] included lexical cohesive pairs in which the two lexical items form a co-referential relation from two perspectives: one is identity of reference and the other one denotes different identities of the same referent. This development of the identity sub-category will hopefully add a new element to the whole picture of developing models of lexical cohesion, and shed light on those co-referential relations in lexical cohesion which have not been focused on before.

Compared with previous models, the expanded definition of SNs adopted in this study and applied to the analysis of the two thesis corpora, also allowed the exploration of various functions of SNs as a type of lexical cohesive device. Within this expanded definition of SNs, firstly, groups of nouns with different terms describing the same SN-like phenomenon were included in the cover term 'signalling nouns', and they were regarded as sub-categories of SNs from a perspective of lexical cohesion. Specifically, SNs in this expanded definition adopted in this study comprise three main sub-categories: GNs from Halliday and Hasan's (1976) model, shell nouns in Schmid's (2000) study and DBSNs in Flowerdew and Forest's (2015) research. However, these sub-categories overlap. Therefore, this study could not establish clear-cut boundaries between these sub-categories. So, rather than counting all lexical cohesive devices in these sub-categories, this thesis focused on the contributions of these SNs to lexical cohesion. The achievement of this study regarding SNs is to expand the category of GNs into a broader area of SNs, and to examine and apply the broad SN-like phenomenon to analysis from the lexical cohesion perspective. This expansion is seen as having the potential to positively contribute to the connection between

[63] For the convenience in the real analysis, sub-categories in the reiteration category were referred to as 'categories', e.g. 'the identity category' and 'the category of SNs'.

the study of lexical cohesion and that of SN-like phenomenon by exploring the functions of SNs as lexical cohesive devices.

In comparison with SNs, the case with collocation in this study is slightly more complicated because the development in the analytical framework regarding the investigation of collocation were twofold: the definition of collocation was developed to be more precise, and the number of sub-categories of collocation was reduced based on the classification of collocation from previous studies (e.g. Halliday & Hasan, 1976; Tanskanen, 2006).

It was noticed that the nature of collocation is likely to be influenced by approaches to collocation and the proximity of the two lexical items in a collocational pair. Since this study focused on lexical cohesion as semantic relations across the clause, the psychological approach was chosen to collocation-related research as this approach emphasises semantic relations between collocates rather than the close proximity of the collocates. Based on this approach, this study developed a new definition and a modified classification of collocation which seemed of benefit in the identification of collocational relations in the current analysis of the corpora.

Firstly, this new definition of collocation made it clear that for this study the proximity between two lexical items in a collocational pair was beyond the clause, which clearly placed boundaries on the distance between two collocates. Secondly, this definition of collocation specified that the meaning of two items "co-occurring in similar environment" is that two items are involved in the same activity or expanding the same topic, something which had not been discussed precisely in previous studies. Finally, the modified classification of collocation used in this study, i.e. activity-related collocation and elaborative collocation, focused only on non-systematic relations between lexical items, which made the analysis of collocation more easy to implement and simplified the process of identification of collocational pairs; and the adoption of two notions 'frame' and 'trigger' also helped in the identification of collocational relations.

Overall, the operationalisation of collocation in lexical cohesion analysis is a necessary step towards the depiction of the complete picture of the classification system of lexical cohesion, as collocation has been regarded as a notoriously problematic category or a 'ragbag' (Hoey, 1991) in lexical cohesion in previous studies. The present study developed a more clear-cut definition and structural division of this complex category, which will hopefully contribute to the lexical cohesive analysis in further research.

9.2.2 Chinese students' use of lexical cohesion

The backgrounds of the subjects in the corpora are not completely homogeneous, i.e. two samples were written by Taiwanese students while the remaining samples were produced by the students from the PRC. Therefore, there was a necessity to substantiate that there exist commonalities among this group of subjects in the present study. The first strategy adopted in this research was to expand the definition of 'Chinese students' to Asian students whose L1 is Chinese. The second step was to further summarise shared characteristics which might have a negative influence on Chinese students' use of lexical cohesion in English academic writing.

Based on the review of previous studies, as presented in chapter 5 of this thesis, three homogeneities were identified: shared writing system; mutual learning culture and language learning method; and the same cultural heritage regarding hierarchical teacher-student relationship and high tolerance of uncertainty (Hofstede, n.d.). The present study has presented some evidence which might support the claims made by previous studies that some of these characteristics have influenced Chinese students' ability to use lexical cohesive devices (e.g. Jin, 2001; Zhang, 2018).

Specifically, as Jin (2001) pointed out, the entrenched habit of depending on the notional connectivity used in Chinese may contribute to Chinese students' preference of the use of lexical cohesive devices in English academic writing. According to the data in the present study, the tested Chinese postgraduates' written samples demonstrated a smaller number of the use of other lexical cohesive devices compared with the dominant use of repetitive devices. This

feature is consistent with the majority of prior studies (e.g. Ong 2011; Zhang 2000), which suggests the possibility of redundant repetitions as a feature for Chinese postgraduates' writing or that certain types of lexical cohesive devices might be fairly rare in academic writing while it might be appropriate to use repetitions of technical terms in the sample texts.

The grammar-based teaching method in China may result in Chinese students' focus on grammar and their ignoring important aspects of lexis, which itself may be connected to Chinese students' reported small vocabulary repertoire (Wu, 2010), an observation which is also supported in the present study as an error regarding the use of synonyms was found in sample M3P in this study and is shown below:

> Though they might [be] proficient in reading tests and memorize [memorizing] thousands of English words, they still have little chance to practice and improve their speaking skills. Since most of their classes are lectured, they never have chances to speak English in class. … Therefore, these students are not as sufficient in their speaking skills as reading or writing.

The author of this sample probably intended to use a synonym to replace the adjective *proficient* in the previous clause, but used an inappropriate adjective *sufficient*. The meaning of *sufficient* is that the amount of entities or people is enough for a particular purpose (OED Online, 2019), this adjective being inappropriate to directly modify the concept of people, i.e. *students*, in this context. Furthermore, *proficient* is used in this context to express the idea that students are good at reading tests and memorising English words. A more accurate synonym for *proficient* would be *capable* or *skilful*, rather than *sufficient*.

9.2.3 Findings from the analyses of the corpora

The key findings from analysis of the corpora will be presented in two parts: the findings from the quantitative analysis of the corpora followed by the results from the qualitative analysis of the corpora.

9.2.3.1 Quantitative findings in the corpora

Regarding one of the main research goals in this thesis, investigation of possible relationships between the use of lexical cohesion and writing quality, as determined by assessment grades, the result of this thesis (see chapter 7, page 205) has shown that there is significant difference between the use of lexical cohesive devices and students' assignments on different marking scales. This is in line with some studies (e.g. Liu & Braine 2005) discussed in chapter 5 which conducted research at the undergraduates' level. Comparatively, this study provides further empirical evidence for such conclusions at the postgraduates' level.

Another main research goal was to investigate whether there was any relationship between the use of lexical cohesion and the flow of text as seen through different dissertation functional sections. In order to operationalise the notion of the flow of text, as mentioned above, this thesis used the sequence of functional sections in dissertations to investigate the distribution of lexical cohesion through the dissertation. According to the analysis of the dissertation corpus, MTDC, there was no statistically significant difference identified regarding the frequencies of lexical cohesive devices used in MTDC in the sequence of functional sections in dissertations (see chapter 8). Specifically, the overall lexical cohesive pairs are more frequent in literature review and methodology sections than other sections. This distribution is similar to that of repetitive pairs distributed in each section, which indicates that the overall distribution of lexical cohesive pairs is significantly influenced by that of repetitive pairs in each section, while lexical cohesive pairs in other categories have much less impact on the overall levels of lexical cohesive pairs. A possible reason lies in the different functions of functional sections in dissertation texts.

As mentioned in chapter 1, the research angle of examining the connection between lexical cohesion and such flow of text is relatively innovative. Only one related study was identified investigating the relation between the use of several types of lexical cohesive devices and sections (i.e. introduction, methodology, results, and discussion/conclusion) of research articles (Wang & Zhang, 2019). This study suggests that research article writers are prone to use more lexical

cohesive devices in the introduction and discussion/conclusion part. By contrast, the result of the present study indicates that Chinese students, as writers of master dissertations, tend to make heavy use of lexical cohesive devices in the literature review and methodology sections. The reason for this difference may lie in the different genres investigated in these two studies. As research articles have shorter lengths compared with dissertation texts, the introduction part of research articles has a more broader function by combining the introduction and the literature review sections as one part; while dissertation texts regard the introduction part and the literature review part as two separated sections. Therefore, the result of Wang and Zhang's study on research articles and that of this thesis are not comparable.

Overall, the quantitative analysis of MTDC in this study suggests that there is value in exploring the use of lexical cohesion in the flow of text. Although the results here showed that in general there was no significant difference between the five sections of dissertations in terms of the use of lexical cohesion, the results also identify the existence of notable differences between individual dissertation samples. This opens a new scope of the investigation of lexical cohesion in dissertation texts. This thesis provides a possible explanation for this individual variability that the use of lexical cohesive devices in texts is influenced by the topics selected in the texts. For example, in terms of the use of hyponymic pairs in sample D2C2L (see chapter 8, page 246), as the topic of 'abbreviations of words' in this sample require a list of different abbreviations and words for this topic, a significant number of hyponymic devices were used for this purpose.

9.2.3.2 Qualitative findings from the corpora

As for the grammatical patterns found in the qualitative analysis of the corpora, this study has identified several novel features regarding the use of SNs which have not been discussed in previous studies.

The first novel feature identified the partial specification as one type of lexical specification of SNs, which relates to the 'borderline pattern' of DBSNs in Flowerdew and Forest's (2015) term, the 'comparative specifics'. Two examples in terms of the use of DBSNs in this pattern were illustrated (see examples 3-27

and 3-28 in section 3.3.4.3), in which the lexical specifications of the DBSNs are by analogy and exemplification respectively. However, an example of SNs found in the present study regarding this grammatical pattern demonstrates that the lexical specification of the SN in that example is by partial specification, which is shown below:

9-1 The size of seminars varies: some seminars are quite small, only consisting of 6-8 people, but the normal size is around 20, with the maximum number of 40 students.

(D7C4F)

The grammatical structure of the GN *people* in this example is:

comparative specifics: SN + referring item (Noun phrase) (partial specification)

As a typical GN, *people* is highly unspecific in isolation. The use of *40 students* as the lexical specification of *people* specifies a certain sub-group of students in the group of *people* as a whole concept.

The second novel feature is concerned with the relational process verbs used with SNs to replace the verb *be* in the grammatical structure 'SN + verb/verb phrase + nominalisation' (Flowerdew & Forest, 2015, p. 21). In this structure, the nominalisation part is the lexical specification of SNs, which is realised by 'identifying clauses', "clauses in which the subject and complement are equated and presented as having a shared identity" (Flowerdew & Forest, 2015, p. 19). Therefore, the verbs or verb phrases used to connect the SNs and the identifying clauses are named 'relational process verbs'. One example is given as follows:

9-2 The reasons emerge from the data collected by interview that // activities organized by their university are attractive and they can actively participate in them and are not afraid to communicate with foreigners in English, …

(D11C5C)

In example 9-2, the verb phrase *emerge from* is used to relate the SN reasons to the following highlighted identifying clause. This verb phrase was not mentioned in Flowerdew and Forest's study regarding this grammatical structure with identifying clauses. The reason may lie in their claim that such relational process verbs in this structure are relatively rare (Flowerdew & Forest, 2015, p. 20). On the other hand, the use of these relational process verbs (e.g. *lie in*, *consists of*) contributes to the creation of lexical cohesive force between SNs and their identifying clauses, and therefore surely deserves more attention in both related research and EAP pedagogy.

The semantic features of lexical cohesion found in this study focus on two components. The first component lies in the types of modifiers co-occurring with different lexical cohesive devices, acting as indicators of the co-referential function of the lexical items which can be immediately found in the co-text, and confining the contextual meaning of those items.

As mentioned in section 3.3.1, there are several pre-modifiers that SNs typically appear with, such as demonstratives, quantifiers and ordinals, adjectives as well as indefinite and definite articles. The findings in the present study suggest that these pre-modifiers not only occur with SNs, but are also used with other reiteration devices to contribute to the creation of cohesive force between lexical items in texts. Furthermore, what is of further interest in this study is the use of other lexical items as modifiers or indicators of co-referential relations. Examples are now given to demonstrate how the use of different modifiers and indicators in several lexical cohesive relations helps convey different lexical cohesive relations in texts, which is of benefit for the teaching of these abstract lexical cohesive relations with examples.

Synonymy

9-3 It seems that the "conceptualizer" stage is to do with psychological domain // while the "formulator" phase is the one within linguistic field.

(D2C1I)

In example 9-3, the premodifiers are two adjectives: *psychological* and *linguistic*. *While* is used as a conjunction to connect the two clauses, which indicates the comparative status between these two adjectives, and that the nouns after these adjectives are suggested to have similar meanings. This is exactly how the nouns are used in this example, i.e. *domain* and *field* are used as synonyms to refer to specific areas.

9-4 1.1 Research Background// […] The emphasis of this study will be put on the "formulator" phase.

(D2C1I)

In example 9-4, the demonstrative *this* is used before *study* to specify that referent of study is *research*, the term *that* is used in the title of a dissertation section.

Hyperonymy

9-5 According to British Council annual report, there were 2.5 million IELTS tests, an increase of over half a million from last year. // […] In Australia, it is now the only test accepted by universities and one of the most widely accepted language qualifications in UK universities.

(D4C1I)

In example 9-5, quantifier *2.5 million* indicates the popularity of *IELTS tests*. The comparative adjective phrase *one of the most widely accepted* also implies the same popular status of *IELTS tests*, and the use of *one* suggests that *IELTS tests* is one type of *language qualifications*.

Hyponymy

9-6 Students may need to be prepared for various modes of online education, which range from the big, obtaining bachelor's degrees, master's degrees, and PhDs through distance learning, to the small, infotainment channels on YouTube.

(D8C1I)

In example 9-6, the prepositional phrase *various modes of* indicates *online education* as a category which includes members; and the verb phrase *range from* signal that the contents in the following clause are the member of this category, i.e. *obtaining bachelor's degrees, master's degrees, and PhDs through distance learning* and *infotainment channels on YouTube.*

Meronymy

9-7 In the academic year 2003/04 the United States welcomed just over half a million international students to study in its higher education institutions. // This number accelerated year on year to just under one million in 2014/15…

(D8C1I)

In example 9-7, the noun phrase *the academic year*, acting as the superordinate for the two lexical items *2003/04* and *2014/15* appeared before the first lexical item *2003/04*, which helps to clarify the meronymic relation between these two lexical items.

Signalling nouns

9-8 There are some cases showing that if CLT is implemented successfully, both students' class performance and communicative competence have been enhanced. // For example, Ni says CLT is implemented successfully in her class and the students' overall communicative competence has improved compared with other class whose students are taught in traditional way.

(D6C1I)

In example 9-8, the determiner *some* is used to indicate the amount of the SN *cases*; and the prepositional phrase *for example* indicates the lexical specification of *cases* is by exemplification.

Identity

9-9 Accordingly, Hong Kong is “a racially homogeneous, largely monolingual society” // […] However, as a former colony of Britain, its English Language Teaching system mainly modelled on Britain’s comprehensive education system.

(D12C1I)

In example 9-9, the use of the preposition *as* and the indefinite article *a* before the lexical item *former colony of Britain* indicates that this lexical item is one identity of *Hong Kong*.

Antonymy

9-10 Supplementary education, also known as “shadow education” or “private **tutoring**” has been expanded rapidly in the globe since this century. // According to Stevenson et al., supplementary **tutoring** is defined as “a set of educational activities // that occur outside formal schooling and are designed to enhance the student’s formal school career.”

(D12C1I)

In example 9-10, the preposition *outside* indicates the exclusion status of *supplementary tutoring* from formal schooling, and it can interpret from the co-text that *private tutoring* refers to *supplementary tutoring*. Therefore, *private tutoring* and *formal schooling* are mutually exclusive, which helps to interpret the contradictory meaning between *private* and *formal*.

It can be seen from the examples above that different types of lexical items can act as indicators of certain lexical cohesive relations in specific contexts, which suggests an effective way of contributing to the creation of lexical cohesion. Teaching such indicators with their corresponding lexical cohesive devices in examples will help students learn how to use different lexical cohesive relations more efficiently to create lexical cohesiveness in their English academic writing.

The second component of semantic features found in the present study is related to the uses of SNs identified in the corpora. As mentioned in chapter 3, in order to identify SNs in this study, a semantic classification of SNs was developed (see table 8) based on previous studies. The findings in this study substantiate the

operationalisation of this classification of SNs, which will be shown in the table below followed by examples from each semantic category of SNs. The demonstration of these examples also provide information about the use of different types of SNs to create lexical cohesiveness in texts as previous studies (e.g. Flowerdew & Forest, 2015; Schmid, 2000) have not studied SNs from the perspective of lexical cohesion in depth, while as a type of lexical cohesive device SNs are of importance as they provide an economical way of creating cohesiveness in texts, which need to be acquired by students to improve their English academic writing skills.

Table 42 Semantic classification of SNs and examples in the present study (see table 8, p. 108)

name	signalling nouns
factual nouns	reason (9[64]), fact (7), problem (4), aspects (4), aspect (3), thing (3), parts (2), picture, phenomenon, difference, popularity, advantage, advantages, value, feature, truth, consequence, findings, resources, disadvantages, , causes, components, reasons, barrier, feature, example, result, evidence, part, importance, aspect
mental nouns	purpose (6), Issue (3), topic (2), perception, point, picture, conflict, mental, interpretation, key, ideas, principle, idea, issues, attitudes, impression
linguistic nouns	definition, argument (3), questions (3), section (3), part (2), parts (2), explanation (2), structure, term, chapter, conclusion, question, overview, statement, knowledge,
eventive nouns	effort, activities, choice, practice, part, measures, process, activity
modal nouns	trend, requirement, unpredictability, task, part, roles
circumstantial nouns	case (5), way (3), context (3), circumstance, cases, fields, systems, tool, environment, situation, areas
first-order entity nouns	people (7), human, individual (2), individuals (2), part (2), population, group (3), person

[64] Numbers in the brackets after SNs refer to the frequencies of the SNs in the present corpora.

Factual nouns

9-11 There is a similar picture in the United Kingdom, which saw just under half a million non-domicile students in higher education in 2014/15, accounting for 19% of the student population.

(D8C1I)

In example 9-11, the SN is *picture* which encapsulates the content in the following *wh*-clause. As the content in this clause refers to a fact in terms of the increasing number of international students in the UK, the use of *picture* in this example is factual. However, *picture* was put into the category of idea noun (i.e. mental nouns in the present study) in Flowerdew and Forest's study (2015, p. 140). Their example is as follows:

9-12 [towards the end of a lecture] What I have tried to give you in the handouts and in the slides I have gone through is a picture globally of the disease as of the end of nineteen ninety seven.

(Flowerdew & Forest, 2015, p. 77)

In example 9-12, the use of the DBSN *picture* refers to the lecturer's idea of the disease in late 1997. Therefore, *picture* was regarded as an idea noun in this context.

Examples 9-11 and 9-12 demonstrate the 'multicategory membership' characteristic of SN-like phenomenon, which reveals that the semantic categories in which the identified SNs belong to in this study can be different from the categories of the same SNs in other studies.

Mental nouns

9-13 Understanding communities is of particular importance in EAP // because one perspective in EAP is to see students as vying for entry to academic communities, all with their own unique cultures, norms, and practices. // This idea dates back to Becher who conceptualised academia as divided up into various "tribes".

(D8C3M)

In example 9-13, *idea* is used as a mental noun to provide a 'conceptual shell' for the previous highlighted content which is the product of mental process.

Linguistic nouns

As discussed in chapter 3, linguistic nouns in this study include several sub-categories: linguistic nouns can represent illocutionary acts semantically, project the content of verbal activities and signal the product of the verbal activities. Accordingly, there are three examples given below to demonstrate the three sub-categories of contents that linguistic nouns denote in the present study.

9-14 For example, in S4's (Participant4) writing, she wrote 'The older people have more work pressures'. // This argument sounds weak, but the teacher did not indicate any inappropriateness.

(D4C4F)

In example 9-14, *argument* is used as a linguistic noun to refer to the content of the verbal activity, i.e. the sentence that *participant 4* wrote.

9-15 This will frame the following discussion of lecture discourse research, followed by research into TED talks (2.3). // As this study takes a pedagogical perspective it is necessary that section 2.3 includes sources from teaching blogs and media outlets alongside research papers.

(D8C2L)

In example 9-15, *section* is used as a linguistic noun to signal the product of verbal activities, i.e. the product of part of a dissertation regarding *research into TED talks*.

9-16 it can approximately make a conclusion that textbooks in China do not contain much implicature.

(D14C3M)

In example 9-16, *conclusion* is used as a linguistic noun to represent the illocutionary act of concluding utterances in a text, which reveals the writer's intention to use the utterance *that textbooks in china do not contain much implicature* as a conclusion to conclude previous segments of the same text, and further build up the cohesive relation between this utterance and other segments.

Eventive nouns

9-17 However, students' perceptions and attitudes toward these two kinds of feedback should be considered // […] and they are the main beneficiaries of these activities.

(D13C1I)

In example 9-17, *activities* is used as an eventive noun to refer back to *these two kinds of feedback*.

Modal nouns

This category of SNs is also divided into three sub-categories: modality (i.e. the judgement of possibility), modulation (i.e. obligation) and dynamic modality (i.e. possibility under circumstances) (see section 3.5).

Modality

9-18 Similarly, lexical density only hints at difficulty in terms of comprehensibility, [and] many other factors can also be considered, such as grammatical complexity, conceptual complexity, familiarity of vocabulary, and so on. // This unpredictability in terms of difficulty is exacerbated by the individual listener who has their own strengths and weaknesses.

(D8C5C)

As *modality* concerns "judgments about the possibility, probability or certainty that something is or is not the case" (Schmid, 2000, p. 235), the SNs in this group represent different degrees of modality. In example 9-18, the use of

unpredictability demonstrates that the author's judgment of the outcome expressed by previous highlighted content, which indicate the *difficulty* is unpredictable.

Modulation

9-19 Thus, this test is designed to evidence the students' capability to use verb forms and 'he' and 'she' correctly in a simulated real-world conversation with peers. // Moreover, by situating students into a collaborative discussion, the task also purposes an evaluation of the extent to which the students can use the target language to communicate and co-construct conversation regarding familiar topics.

(D1)

As Schmid (2000) described, the modulation use of *task* as a modal SN refers to activities that "are supposed, meant, or even more or less required, but not absolutely obliged or forced to do" (p. 248), which indicates some resistible forces portrayed by the SN *task*. This feature of *task* as a modal SN is demonstrated in example 9-19. The referent of *task* in this example is *this test* which is designed to examine students' capability to use specific linguistic forms. That is to say, students are required to take *this test* as a *task*, which indicates a mild obligation imposed on the students.

Dynamic modality

9-20 A bulk of surveys of students' feedback preferences indicate that ESL students greatly value teacher feedback and consistently rate it more highly than alternative forms, such as peer. // The reason for this trend is partly because learners believe that teachers 'possessed a better knowledge of English language and therefore provide more helpful feedback than peer learners'.

(D4C2L)

In example 9-20, *trend* is used as a dynamic modal noun to denote the possibility of the highlighted future event that *ESL students greatly value teacher feedback and consistently rate it more highly than alternative forms, such as peer*. The use

of *trend* also indicates the source of the probability of this event or "the circumstances under which this event is likely to happen" (Schmid, 2000, p. 259), i.e. the author of sample D4C2, is the protagonist of this event, even though the author, who provides the circumstance for this event, is backgrounded in the grammatical pattern '*this* + tendency + *is'*.

Circumstantial nouns

Compared with Schmid's (2000) work as well as Flowerdew and Forest's study (2015), there is a new circumstantial noun identified in this study, which is shown in the following example:

9-21 A questionnaire survey of pupils from fresh secondary school 2015/2016 graduates helps reflect the students' recognition of supplementary education, // while a semi-structured interview with former and current tutors in the fields helps illustrate a clear view from different side.

(D12C3M)

In example 9-21, the noun *fields* is used to refer back to the education system *supplementary education*. As supplementary education is a unified concept, *fields* should be changed into *field* for grammatical uniformity. The circumstantial use of *field* in this example reveals the meaning of *field* which refers to an area of activity (OED Online, 2019), i.e. *supplementary education* in this context.

First-order entity nouns

This category was developed for this study based on Lyon's tripartite taxonomy of experiences (1977, p. 442) and Halliday and Hasan's GNs (see section 3.5.4), mainly addressing the use of nouns denoting people. One example is demonstrated below:

9-22 The participants are 16 Mandarin-speaking students aged 22 to 26 currently studying on MA TESOL program at Durham University. // […] They are grouped into two test groups, 8 people in Group A and another 8 people in Group B.

(D2C4M)

In example 9-22, the GN *people* is used twice. Each use of *people* is modified by the quantifier 8 to refer back to the previous lexical specification content – *16 Mandarin-speaking students*. Although *people* is not an abstract noun like *event* or *idea*, the contextual meaning of *people* is still dependent on its corresponding lexical specification, as is shown in this example. In this way, the lexical cohesive relation between the GN *people* and its lexical specification can be established.

Overall, the discussion above regarding grammatical and semantic features of lexical cohesive devices provide examples identified from the corpora regarding the teaching of effective use of typical modifiers and grammatical patterns with different cohesive devices in texts, as well as the use of SNs based on their semantic characteristics. As the nature of lexical cohesion is context sensitivity, different types of lexical cohesive device can be used in various contexts based on the topics in the contexts, which is substantiated by the individual variability of the samples in this thesis. This suggests that it is necessary to teach the use of lexical cohesive devices with examples rather than explaining the meaning of these relations in isolation.

9.3 Limitations of the study

9.3.1 Limited focus of lexical cohesion

The present study only describes a subset of the linguistic features of lexical cohesion, i.e. the lexical cohesive pairs in texts. Cohesive chains, another important feature of lexical cohesion, which have also been investigated in previous studies, are not included in the present study. The reason for this is that each chain consists of several lexical cohesive pairs, and mapping these cohesive pairs onto different chains requires more labour work, which is not allowed considering the time limitation of this thesis research. Researchers have suggested that it is the long cohesive chains that differentiate the poor and good qualities of writing (e.g. Neuner, 1987), because longer chains which include a greater variety of words and maturity of word choice contribute to good writing (Yang, 1989, p. 235). However, it has also been pointed out that lexical cohesive pairs themselves constitute lexical cohesive chains (Halliday & Hasan, 1976).

Furthermore, the present study has focused on the lexical cohesive relations across clauses, which to a certain degree, already expands the scope of connectedness and interrelatedness between segments in a text from the intra-clausal level to the inter-clausal level. It can be asserted that understanding lexical cohesive pairs across clauses leads to a better insight into the concept of texture and of how the meanings of segments intertwine with each other in a text. However, It is certainly agreed that study of lexical cohesive chains would be a useful focus for future research.

9.3.2 Issues of corpora size and diversity

"[O]ne of the key requirements for a corpus to be considered valid in terms of making broad generalisations about text characteristics is that it must be representative of those texts to which those generalisations are targeted" (Nathan, 2010). While it can be argued that although the corpora of MTMC and MTDC in the present study may be representative of relevant MA TESOL postgraduates' module assignments and dissertations at the disciplinary level, within the scope of the broader picture of students' assignments in EFL and EAP writing, the level and diversity of texts in this study must be considered small and institutionally narrow in origin. Nevertheless, as mentioned in chapter 5, several departments and disciplines have been studied at both undergraduate and postgraduate levels in previous studies. The purpose of the investigation in the present study is to contribute additional information to the picture of lexical cohesion research on the writing of Chinese students in the specific discipline at postgraduate level. In addition, claims based on previous research with Chinese students in EFL classrooms might not apply to Chinese postgraduates in MA programmes at a UK university, and although cohesion is used as an assessment criterion in writing classes, it is probably not considered much when marking MA assignments.

The main reason for the small size of the corpora used in this study is related to the adoption of manual analysis as the main method of this study. Use of much larger corpora would have meant that the analytical workload would be unrealistic for the researcher. Although the accuracy of the annotation of the corpus has been enhanced and is considered of greater validity and reliability due to the use

of such a manual method, it is impossible to rule out issues of intersubjectivity in the coding and analysis of the corpora, issues which might also be problematic with a fully automated approach. The size of the two corpora is probably not as large as other more automated corpus-based studies. Nevertheless, the present study has done both quantitative and thorough qualitative analysis of samples in context which is also important for the study of lexical cohesion which is context-sensitive.

As for the diversity of the subjects in the corpora, the main focus of this study is on Chinese students, and the majority of Chinese students in the UK universities are Chinese postgraduates (see chapter 5). All of the samples analysed were written by postgraduates whose first language is Chinese, among which, only two were written by Taiwanese (see appendix D) while the remaining samples were produced by students from the PRC. The reason for the regions where the Chinese students in the present study come from is determined by the ethnographic and demographic backgrounds of the potential participants who have taken the MA TESOL programme in previous years.

As for the sampling and grouping of the corpora, since the use of lexical cohesion is context-sensitive, there are also other possible approaches to investigate this issue, some of which have already been adopted in previous studies, such as dividing up assignments according to text types, comparing texts across disciplines or genres or between groups of speakers, or carrying out a longitudinal study with a cohort over a year.

9.4 Implications for EAP teaching

9.4.1 Reason for teaching lexical cohesion in EAP classes

This study focused on EAP teaching because previous studies have showed that Chinese students lack knowledge of lexical cohesion in their English academic writing (e.g. Zhang, 2018) and there is a “shortage of teaching materials on cohesion—particularly lexical cohesion—in second language (L2) writing” (Johnson, 2017). Specifically, the first reason for including lexical cohesion in EAP teaching is that lexical cohesion is a significant feature of texture, which should be taught to students to help improve their text-creating and decoding

abilities. Another reason for the importance of teaching lexical cohesion is that cohesion is often included in the assessment criteria for evaluating the quality of students' written assignments (Yang, 1989, p. 235). As an important type of cohesion, lexical cohesion requires more attention in EAP classes as "most [TESOL] writing textbooks have either totally neglected the subject or merely mentioned it in passing" (Liu, 2000, p. 28) . Furthermore, "teaching sub-types of lexical cohesion to foreign language learners will improve the quality of their reading and writing" (Gholami & Alizadeh, 2017, p. 307). Based on the reasons noted above, the purpose of the present study was to analyse Chinese-speaking postgraduates' assignments in terms of lexical cohesion in an effort to explore significantly different patterns of lexical cohesion in texts, and to identify and recommend materials and methods to help Chinese students improve their writing skills. The following recommendations for the teaching of lexical cohesion in EAP classes are made based on the results of the use of lexical cohesive devices in the corpora used for the present study. Although the focus was Chinese students, other EFL learners and practitioners in EAP classes can also adapt the implications to suit their own needs and levels.

9.4.2 Teaching lexical cohesion with examples in EAP pedagogy

9.4.2.1 Teaching lexical cohesive devices other than repetition

The results of this study and those of several other previous studies (e.g. Jin, 2001) suggest that Chinese students employ lexical repetition as a main lexical cohesive device, while they do not substantially employ other more sophisticated lexical cohesive devices in their writing. This is the case regardless of their English proficiency levels. Therefore, the first implication for EAP pedagogy is to place more emphasis on teaching more sophisticated lexical cohesive devices in classes to help Chinese students grasp the whole picture of lexical cohesive devices which can be used in their English academic writing, rather than simply focusing on repetition to create lexical cohesiveness in texts.

Specifically, the first type of sophisticated lexical cohesive devices which needs to be explicitly taught is signalling nouns as previous studies suggest that general nouns have been taught in EAP classes (Nga, 2012), but not signalling nouns defined in the present study with several sub-categories. As mentioned in chapter

7, signalling nouns are not frequently used in general in MTMC. Part of the reason for this may lie in students' lack of knowledge about signalling nouns. However, this lexical cohesive device contributes to the cohesion in the whole text, as its encapsulating function in summarising complicated chunks of information contributes significantly to the overall cohesiveness and succinctness of texts.

The second category of lexical cohesive devices which deserves more attention is collocation. In chapter 7, the present study conducted a comparison between Chinese students' module assignments and academic authors' research articles regarding the use of reiteration and collocation devices (see table 26). The result shows that compared with professional researchers, Chinese students tend to use more reiteration devices and collocational pairs in their writing, which suggests that Chinese students may have a good knowledge of the use of collocation as cohesive devices, or the use of collocational pairs is topic-based and certain topics in students' samples in the MTMC corpus need more collocational pairs for further elaboration.

Furthermore, the quantitative analysis in MTDC (see chapter 8) suggests that not all samples used every kind of lexical cohesive device in all nine lexical cohesive categories. The absence of cohesive pairs in the categories of hyperonymy, hyponymy, meronymy, identity and antonymy occurred in some samples from different functional-section groups in the dissertations (see appendix I). This result might be seen as suggesting that these absent categories need to be the focus of some explicit instructions in EAP courses so that Chinese students can learn how to use them as cohesive devices effectively in their academic writing.

In addition to pointing out the necessity to emphasise the teaching of particular types of lexical cohesion in EAP classes, the next section will demonstrate how to combine the teaching of lexical cohesive devices with relevant grammatical patterns using examples identified in the analysis in the MTMC and MTDC corpora in the present study.

9.4.2.2 Teaching lexical cohesive devices with grammatical structures using examples

Since lexical cohesion is highly context-sensitive (Tanskanen, 2006, p. 174), it is considered essential to teach lexical cohesive devices with their co-occurring grammatical structures and to present and demonstrate the use of these devices with examples. The examples gathered in MTMC and MTDC will be used here to support suggestions for EAP pedagogy. The examples will be illustrated in accordance with the lexical cohesive categories which they belong to.

Repetition

The repetition category has been divided into two types in the present study: simple repetitions (i.e. the same lexical item), and complex repetitions with derivational variations or grammatical changes (see section 2.3.1). The findings in chapters 7 and 8 suggest that Chinese students tend to use simple repetitions rather than complex repetitions, which points to the necessity of introducing the use of complex repetitions to students, in order to raise their awareness of avoiding overusing simple repetitions which may make texts seem uninteresting and might suggest the writer has a small vocabulary repertoire. Especially when the lexical item is central for the proposition of the text, it is suggested to that writers use derivatives of the item to make the text more interesting and colourful while maintaining the preciseness of the meaning of the lexical item. An example will demonstrate the use of such complex repetitions below:

> 9-23 there has been a lot scholars and **researchers**// who **researched** …// Most of these **researche[r]s** …// Thus, **research** questions …
>
> (excerpt 3, chapter 8)

In example 9-23, *research*-stemmed repetitions form a repetitive chain with three cohesive pairs: *researchers – researched – researchers – research*. As the three pairs share the same word stem – *research*, they may potentially be interpreted as being lexically cohesive by readers. On the other hand, these *research*-stemmed lexical items are slightly different in forms, which helps to reduce the monotony of the text for more comfortable reading experiences. This would be a useful example, alongside others to teach the notion of complex repetition and its

realisation. What is missing in teaching EAP as far as the researcher can see is a substantive focus on this. On-course activities, such as group work on discussing different derivational variations or grammatical changes of the lexical item which needs to be replaced in an example (e.g. example 9-23 above), or underlining or identifying complex repetition forms of a given item in examples.

Signalling nouns

As mentioned in chapter 3, the analysis of signalling nouns includes its surrounding grammatical structures or modifiers. The reason for this inclusion is that the surrounding elements contribute to confining the contextual meaning and signifying the co-referential function of signalling nouns, and therefore guiding the readers to decode the complex information that the signalling nouns encapsulate. Several examples, derived from the current research are shown here to illustrate the use of signalling nouns in texts.

9-24 From the information which is attained from the interviews, the overview of participants[’] perspective on peer feedback is that// five of the eight students (Students A, B, C, D, E and Student G) felt that this activity was helpful.

(D13C4F)

9-25 There were a couple of previous studies investigated the ways how plural formation works in English and Chinese, with the representative of Quirk et al. in English and Yang in Chinese. Based on their effort, we know that English and Chinese means of marking number are basically different.

(D2C1I)

9-26 With its development, the popular and dominant ELT methods in China are grammar translation and audio lingual,// [and] they are 2 ways that make great contribution to language teaching.

(D6C1I)

In example 9-24, there are two pairs of SNs. The first SN is *activity* in the grammatical structure of ‘*this* + SN’. The whole structure refers back to *peer feedback* in the text. Furthermore, the use of the determiner *this* indicates that *activity* is used as an anaphor to replace the previous more specific nominal

phrase *peer feedback*. The second SN is *overview* in the structure of '*the* + noun + *of*'. The referent of *overview* is the whole succeeding clause *five of the eight students (Students A, B, C, D, E and Student G) felt that this activity was helpful.* The usage of the SN *overview* in this example is a demonstration par excellence of the encapsulating function of SNs by summarising a complicated stretch of text into a smaller nominal phrase, which shows a more sophisticated way of expressing the lexical cohesiveness between segments in the same text than the use of repetitive devices.

Similarly, in example 9-25, the determiner *their* is the indicator of the anaphorically-referential function of the SN *effort*. As for example 9-26, what is worth mentioning is the use of another pre-modifier, cardinal number *2*, prior to the SN *ways*. Compared with the use of the determiner *their* in example 9-25, *2* has another function of describing a specific characteristic of the SN *ways*.

It is noticeable in the three examples that in general, when used as anaphorical cohesive devices, the SNs are normally preceded by determiners, such as *their* and *these*, or other modifiers, such as the cardinal numbers. Another feature regarding using SNs as lexical cohesive devices is that the contextual meaning of the SNs is dependent on their referents, while the SNs express more general meanings of the referents. This feature is demonstrated explicitly when GNs[65] are used in texts, which can be illustrated in the example below:

9-27 The law students […] have at least two seminars in a week […]// The size of seminars varies:// some seminars are quite small, only consisting of 6-8 people …

(excerpt 4)

In example 9-27, *people* is a GN which refers back to *the law students*. The use of *people* as a GN in this case provides a valuable pedagogical implication regarding the teaching of GNs. As a type of SNs, the contextual meaning of general nouns is also highly dependent on the referents. As in example 9-27, the

[65] General nouns is a type of signalling nouns in the present study (see chapter 3).

contextual meaning of *people* is provided by the prior use of *the law students*. This example suggests that GNs should be used when it is clear for the readers to decode the contextual meaning of the GNs. Otherwise, the overuse of GNs may result in the vagueness of the text and disconnections between segments in texts, which may further create difficulties for readers' comprehension of the overall text (Wu, 2010).

From the examples above, it can be seen that the use of SNs in texts is context-based. Activities related to the teaching of this lexical cohesive device need to involve exercises with examples. For example, students are asked to choose the appropriate SN from a list of potential SNs for different sentences, explain to the class why this SN is chosen for this particular sentence, and the teacher as well as other students give feedback during class.

Identity

As for the teaching of the use of identity devices in texts, there are two points which arise from this research: the use of two types of the identity category, and the highly context-sensitive feature of the identity relations. The two points will be elucidated with four examples below.

Firstly, the two types of identity relations are shown in examples 9-28 and 9-29 identified in this research, these examples serve as exemplars of this type through which in order to give EAP teachers advice on how to teach identity cohesive devices in academic writing.

9-28 For example, if only grade one is analysed, perhaps the higher grades start to introduce implicature// because designers suppose these freshmen do not acquire sufficient pragmalinguistic knowledge …

(D14C3M)

9-29 This means that more university applicants will choose IELTS examination to attain a place in the course.// If students want to apply for a Tier4 general student visa, their IELTS overall score has to reach the band 5.5 and above…

(D4C1I)

Example 9-28 includes an identity pair *grade one* – *these freshmen*, in which *grade one* refers to the whole group of students who attend the first level of classes at school. This interpretation of *grade one* is supported by the anaphorical use of *these freshmen* in the succeeding clause, as the meaning of *freshmen* is first-year students (OED Online, 2019f), and the determiner *these* indicates that *freshmen* is used as an anaphor to refer back to *grade one*. Therefore, *grade one* and *these freshmen* form a co-referential relation which is the first type of the identity category. This pair can be used as a good example for teachers to explain the use of co-referential relations between identity devices.

As for example 9-29, the use of *applicants* and *students* forms the second type of identity relation which refers to the relation between two lexical items expressing different identities to the same referent in the same text. Particularly, *applicants* emphasises one identity of the group of referents who request to study in a course; while *students* imply another identity of these referents who have been already learning at school. This example provides a good demonstration of using two lexical items to express different identities of one referent, which not only creates the cohesiveness in the text, but also adds more information to the referent in an economic way. It is suggested for teachers to introduce this cohesive relation in writing for students to learn how to create lexical cohesion in texts in a more effective way.

Furthermore, the second point regarding the identity category will be discussed with another two examples as follows:

9-30 This assessment tool takes the form of a speaking test, devised as a progress test based on the syllabus// which aims to entrench the automatic cognitive processing to encode time information into verbs and gender information into third personal pronouns for Chinese learners of English.// Moreover, … the task also purposes [proposes] an evaluation of the extent// to which the students can use the target language to communicate and co-construct conversation regarding familiar topics.

(D1)

9-31 These new employees have all achieved satisfied scores in the English test// and they are regarded as upper-intermediate English language users.

(M2P)

In example 9-30, the lexical items *English* and *the target language* create an identity pair, in which the former specifies the contextual meaning of the latter. That is to say, *English* is regarded as *the target language* for Chinese learners in this context. This co-referential relation between *English* and *the target language* generate based on this specific context, which exactly demonstrates the highly context-sensitive feature of the identity relations.

Example 9-31 is another case to elucidate this feature. In this example, *these new employees* and *upper-intermediate English language users* are cohesively connected as they provide two identities to the same group of referents. The first lexical item provides the employment status of the referents, while the second one reveals the English proficiency level of these referents. The cohesive relation between these two lexical items can only be deduced in the context of example 9-31, because the referents which the items denote are a specific group of people in this context.

Overall, these complex situated examples above demonstrate the context-sensitive characteristic of the identity category in specific textual contexts, and therefore, can provide suitable cases for EAP teachers regarding the use of types of identity device in classes. For example, after introducing the types of identity devices with exemplars (e.g. examples 9-28, 9-29, 9-30 and 9-31 above) in class, students are asked to write a paragragh of a given topic and include the use of these types of identity device. During their composition, the teacher walks around the class to provide advice and answer questions from students.

Synonymy and antonymy

The reason for combining the EAP implications for teaching the categories of synonymy and antonymy is related to the division of the two categories in use. Both categories can be divided into two types in terms of their use in texts.

Specifically, synonymy is divided into synonymy in the traditional sense and near-synonymy; and similarly, antonymy is divided into antonymy in the traditional sense and near-antonymy. The rationale for their divisions is the same, which is based on two criteria in the present study (see chapter 2): whether the lexical items in a synonymous or antonymic pair belong to the same word class; or whether the meanings of the lexical items can be interpreted straightforwardly as being similar or opposite. The examples from the present study are shown below for explaining the usage of different types of synonymous and antonymic relations, and further contributing to implications for EAP pedagogy.

Firstly, the two types of synonymy will be discussed through examples 9-32, 9-33 and 9-34.

9-32 a) the learners// The course will be taken by 10 to 12 Chinese young adults (late teens or early twenties) as part of their preparation for meeting the level of English language proficiency required for admission to UK universities.// All of the students wish to achieve band score at least 5.5 in IELTS speaking …

(M12P)

9-33 The second part reflects the approach of task-based analysis, which helps to ensure the course to possess a high degree of real-life relevance … // … they can answer the questions according to the occasions where they cooperate with Chinese clients. This could also be conducive to figure out the tasks they are likely to carry out with foreigners.

(D6)

9-34 As an ending of the course, the department of human resource (HR) in the company wants to find out the participants' learning outcome.// Therefore, an assessment tool is required to design not only for evaluating achievement relevant to the short-term program, but also for deciding the final list of employees.

(P5)

Example 9-32 demonstrates the synonymous relation in the traditional sense between *learners* and *students*. The lexical item *learners* is replaced by its

synonym *students* in the following clause. It is not difficult to interpret their synonymous relation from this text, as *learners* and *students* have similar word meanings: *learners* represent people who are learning something (OED Online, 2019j), and *students* denote people who are learning at school (OED Online, 2019m). Furthermore, the two items share the same referent, i.e. *10 to 12 Chinese young adults*, in this example. These two points made above make it clear that *learners* and *students* form a synonymous relation in example 9-32.

By contrast, examples 9-33 and 9-34 demonstrate near-synonymous relations from two perspectives: example 9-33 elucidates the near-synonymous relation from the perspective of two lexical items belonging to different word classes; while in example 9-34, the two near-synonyms have similar meanings which cannot be interpreted directly. Specifically, the synonymous relation in example 9-33 is between *helps* and *conducive*. Although both items express the meaning of providing something good to make other things happen (OED Online, 2019h; 2019d), *helps* and *conducive* belong to different word classes, i.e. *help* is a noun while *conducive* is an adjective. Therefore, according the first criterion for dividing the synonymous relations mentioned above, *helps* and *conducive* are regarded as near-synonyms in example 9-33. As for example 9-34, *outcome* and *achievement* form a near-synonymous relation, because they have the same referent, i.e. the participants' learning results in this example; and their meanings are similar in this context. *Outcomes* refers to the participants' learning result of the course, and *achievement* is interpreted as the good result of the participants' learning in the course. Both of the items denote the participants' learning results, however, *achievement* adds a positive feature to the results while *outcomes* expresses a more neutral meaning. Therefore, in example 9-34, *outcomes* and *achievement* are regarded as near-synonyms.

According to the analysis of the three examples above, it is noticeable that the use of the two types of synonymy is not restricted to the word meaning of the lexical items and the word classes which the items belong to, but depends on the specific context in which the items occur. It is important for teachers to bear in mind that as lexical cohesive devices, synonyms and near-synonyms should be

taught with examples which emphasise the contextual meanings of the synonymous devices.

In terms of the teaching of antonymy, in addition to the introduction of the two types of antonymy mentioned above, this category can also be categorised into four sub-categories according to the opposite relations between two lexical items in a antonymic pair. As discussed in chapter 2, there are four kinds of opposite relations categorised in the present study, which are shown below. However, only the first three have pairs identified in the analysis of the current corpora. Some examples found in the corpora will be displayed to demonstrate the use of the first three types of opposite relations in Chinese students' academic writing, in order to provide advice to EAP teachers for the teaching of antonyms used as lexical cohesive devices.

The first sub-category of antonymy is complementary antonymy. The key of this relation between two antonyms is that there is binary contrast between the two antonyms, which cannot be gradable. Two examples are shown below:

9-35 The students may even find it difficult to notice the **errors** when speaking English. Thus the perceived needs of students are, […] explicit teaching of the thinking difference which is accountable for the **errors,** a raised awareness of the **thinking processes** and the thinking habits in speaking English to encode time information in verbs and gender information in the third person pronouns. Therefore, the syllabus is designed to achieve the goal of the "entrenchment" of a chain of **thinking processes** that can generate correctly-tensed and gender-referred speaking English.

(M1P)

9-36 There are totally 22 students in class,// among which 16 are girls// and 6 are boys.

(D9P)

In example 9-35, *errors* and *correctly* are regarded as a near-antonymic pair, as *errors* is a noun while *correctly* is an adjective, and they express contradictory meanings in this context. Specifically, *errors* refers to the inappropriate contents

in students' spoken English, whereas *correctly* in *correctly-tensed* denotes the appropriate verb tense use in students' speaking English, which in general also refers to the contents in students' spoken English. That is to say, *errors* refers to the inappropriate contents in spoken English, while *correctly* is related to the appropriate contents in spoken English. Therefore, *errors* and *correctly* are contradictory in meanings in example 9-35. The analysis of this pair demonstrates the feature of antonymy as lexical cohesive devices that two lexical items from different word classes can form near-antonymic relations when they express contradictory meanings in a specific context.

Example 9-36 has an antonymic pair, *girls* – *boys*. This pair is much more straightforward than the pair in example 9-35 regarding the contradictory meanings of the lexical items. *Girls* and *boys* are semantically regarded as contradictory in general, while *errors* and *correctly* are more abstract in word meanings, which need to be interpreted in specific contexts to decode their contradictory relation. Overall, it can be seen that examples 9-35 and 9-36 demonstrate how to use complementary antonyms in academic writing; furthermore, example 9-35 also shows the use of near-antonyms in a text. The use of the complementary antonyms in the two examples above provide the models for examples used in the EAP pedagogy regarding teaching such complementary cohesive devices.

The second sub-category of antonymy is contrary antonymy, which refers to relations between gradable antonyms. Examples 9-37 and 9-38 demonstrate the use of gradable antonyms in academic writing.

9-37 But even where performance test materials appear to be very realistic compared to traditional paper-and-pencil tests, it is clear that the test performance does not exist for its own sake. However, it is necessary to have a procedure that is fair to all candidates, and elicits a scorable performance, even if this means involving the candidates in somewhat artificial behaviour.

(D2)

9-38 After the investigation, it shows that these textbooks barely contain implicature contents, while they emphasize on grammar and lexis, pragmalinguistics, capturing explicit information, and neglect authentic conversations and contexts.// … they are not proved to be practical and effective to cultivate students' competences to understand and produce utterances in [at] the implicit level.

(D14C6C)

In example 9-37, *realistic* and *artificial* constitute a contrary pair, as the former is related to the reality or authenticity, while the latter means not authentic. The interesting point is the use of the surrounding lexical items which indicates the comparable sense between *realistic* and *artificial*, such as *very, compared to* and *somewhat*. These indicators make it clearer that *realistic* and *artificial* are gradable in terms of their contrary meanings, which is a good suggestion for teaching gradable antonyms that the use of such indicators in the co-texts of these gradable antonyms helps readers identify the gradability of the antonymic pairs in texts.

The third sub-category is converse antonymy, which typically exists in two kinds of situations: used with procedural verbs and nouns which express reciprocal social roles (see section 2.3.6). In the first situation, the actions that are expressed by the verbs are involved in a unidimensional movement which can be seen from two perspectives: that of the source and that of the goal (Murphy, 2003). In the second situation, one social role cannot exist without the other. The two examples below illustrate the two situations respectively.

9-39 It represents a particular realization of communicative language teaching.// … they can enjoy the activities and create more active learning atmosphere with satisfactory effect.

(M7P)

9-40 However, the uneven representation of values could be a feature of the data set and the individual speakers, … // … This unpredictability in terms of difficulty is exacerbated by the individual listener who has their own strengths and weaknesses.

(D8C5C)

In example 9-39, *teaching* and *learning* form a near-antonymic converse pair. On the one hand, the two items are near antonyms because *teaching* is a noun while *learning* is an adjective. On the other hand, the items express the same procedure from teachers' perspective (the source) and students' perspective (the goal) respectively. Therefore, *teaching* and *learning* are converse antonyms in this context. In example 9-40, *speakers* and *listener* are two social roles which are interdependent. As Crystal (2008) commented, there is symmetry of dependence in the reciprocal social role. The object of the *speakers* who speak to is the listener, and at the same time the object of the *listener* who pay attention to is the contents that the *speakers* say. Both *speakers* and *listener* cannot exist without each other. Overall, the two examples show the interdependence between the two lexical items in a converse relation in two typical situations, which gives practical suggestions to teacher regarding designing appropriate examples for teaching the use of converse antonyms.

As for the fourth sub-category, directional antonymy (see section 2.3.6), since there is no such relations found in the present study, there is no further discussion here for EAP implications.

Overall, it can be seen from the examples above that three types of antonyms can be used as cohesive devices in texts. As for the EAP pedagogical implication in terms of teaching these types of antonyms as cohesive devices, teachers can illustrate each type of antonyms with several exemplars such as the examples mentioned above at first, and design matching activities where learners need to match listed six antonymic relationships with six individual sentences in which different antonyms are used. Through such matching activities, students can have a better understanding of the use of different types of antonymic device in lexical cohesive relations.

Superordinate relations: Hyperonymy, hyponymy and meronymy

As mentioned in chapter 8 (see section 8.3.2), hyperonymy, hyponymy and meronymy are included under the umbrella term 'superordinate relations'. Six

examples are selected from the present study to further illustrate the use of the three types of superordinate relations.

The first relation discussed here is the hyperonymic relation which is divided into two types.

9-41 This dissertation is a study focusing on the exploration of law students' past English learning experiences and spoken English issues in seminars through their reflection on seminar learning in the LLM programme. … Ethnographic methods offer this study a holistic approach to … systematically document[ing] the influences of students' background on their seminar learning in a rich, contextualised detail with the aim of suggesting proper measures to deal with language issues in law seminars.

(D7C1I)

9-42 Chapter One Introduction//… With regards to peer feedback in this dissertation, it refers to the activity where students read each other's essay and then express not only negative criticism but also supportive and appreciated evaluation.

(D13C1)

Example 9-41 refers to a kind-whole relation, *English – language*, while example 9-42 denotes a part-whole relation, *Chapter One Introduction – this dissertation*.

The relation between hyponymy and hyperonymy is that a hyperonym consists of several types of hyponyms. Examples 9-43 and 9-44 below demonstrate this hyponymic relation.

9-43 because essays and reports are usually regarded as the most popular assessment forms in academic courses in western countries.// They may also have exams but compared with short exam answers,// it will cost more time like several weeks to write essays as assignments for course work.

(D11C1I)

9-44 Native speakers unconsciously talk in the implied way, while non-native speakers face this language barrier.// Most international students find it difficult

to comprehend native English speakers' hidden meanings as they come to the United Kingdom.

(D14C1I)

It can be seen from example 9-43 that *assessment forms* and *exams* form a hyponymic relation, as the more general item *assessment forms* appear before the more specific item *exams*. A co-hyponymic pair is also identified within the same category of *assessment forms*: co-hyponyms, *exams* and *essays*, are two kinds of *assessment forms*, and therefore are co-hyponymic.

The relation between meronymy and hyperonymy is that a hyperonym is made up of several parts which are regarded as meronyms. Examples 9-45 and 9-46 will explain the use of meronymic devices in texts:

9-45 Supplementary education, also known as "shadow education" or "private tutoring" has been expanded rapidly in the globe since this century. This phenomenon has first been developed in East Asia and has become externally visible throughout Asia as well as in other world regions in the present days.

(D12C1I)

9-46 We can see this through an exquisite job done by Quirk et al. , which categorized number classes of nouns mainly into four groups// … Nouns in group (A) are occurring only in singular form, which include (Aa) mass nouns such as gold, music, (Ab) abstract adjective heads like the unreal, and (Ac) some proper nouns like Henry, the Thames.// On the contrary, nouns in group (B) are occurring only in plural forms, which can be distinguished into five subgroups…

(D2C2L)

In example 9-45, *the globe* is the whole entity while *East Asia, Asia* and *other world regions* are parts of *the globe*. Therefore, the meronymic pair is the *globe – East Asia/Asia/other world regions*. As the three parts of the globe (i.e. *East Asia, Asia* and *other world regions*.) occur in the same clause, they will not be counted as co-meronyms in this study. As for example 9-46, *group (A)* and *group (B)* are two parts of the four groups of 'number classes of nouns' in sample D2C2L. As these two items occur in two clauses, *group (A)* and *group (B)* are regarded

as co-meronyms in this thesis. It is noticeable that the hyperonym of *group (A)* and *group (B)* also appear in the surrounding clause, which is the *groups of number classes of nouns*. This indicates that the hyperonym and its meronyms co-occur in certain contexts, which gives the EAP implication for teaching meronyms that providing the hyperonym of the co-meronyms can contribute to readers' correct interpretation of the co-meronymic relation between two lexical items in texts.

As mentioned in chapter 8, superordinate cohesive devices are not frequently used by Chinese students in their academic writing, and one reason may be that they have not been taught explicitly how to use the three types of superordinate devices above. It is recommendable for teachers to introduce these cohesive devices in EAP classes with examples as the six ones above. Fun activities can be conducted regarding the teaching of the three types of superordinate relation, such as crosswords of a set of hyperonyms with their corresponding hyponyms and meronyms or matching hyperonyms with the correct hyponyms or meronyms in a group of sentences.

Collocation

In the present study, collocation has been divided into two sub-categories: activity-related collocation and elaborative collocation (see chapter 4). The elaborative collocational pairs have been identified more frequently than the activity-related collocational pairs in the present study. The reason for this observation may lie in the loose definition of elaborative collocation which only requires two lexical items elaborating or expanding on the same topic, while the activity-related collocation entails the lexical items to be elements of the same activity. Three examples are selected from this study to show the use of the two sub-categories of collocation below.

9-47 it is common that researchers prefer questionnaires rather than interviews,// since perhaps the former tools can be used to attain information from a large number of participants …

(D13C3M)

9-48 Chapter Two Literature Review// This chapter will look at relevant research based on theoretical research as well as the investigation related to peer feedback from students' perspective, including Asian students, Chinese and Japanese and European learners from Spain.

(D13C2L)

9-49 The selection of the participants is established on the principle that efforts should be spent on finding a representative group of sample regardless of a small-scale research to seek context-specific answers.// Thus, the participants in this research design included a Chinese IELTS writing teacher and his five Chinese students who have prepared themselves for participating in the IELTS examination and who have taken IELTS examinations several times.

(D4C3M)

Example 9-47 shows an activity-related collocational relation between *questionnaires* and *participants*. The justification for this interpretation is that in the activity of filling in the *questionnaires*, *participants* are the 'people' who take the action, and *questionnaires* is the 'thing' which is the object of the action. That is to say, *participants* and *questionnaires* are two of the elements (i.e. 'people' and 'thing', see section 3.4.2) in the 'activity', which, therefore, form an activity-related collocational pair.

Examples 9-48 and 9-49 demonstrate the elaborative collocation relations. In example 9-48, *literature* and *research* form an elaborative collocational pair because *literature* triggers the *research* frame, which indicates the occurrence of *research* in the succeeding clause. In this example, *literature* refers to the information relating to the subject *peer feedback*, and *research* denotes the detailed study of the same subject from two different angles which are the theoretical research and students' perspectives. The content of *research* provides the information which is included in the *literature*. Therefore, both *research* and *literature* are interpreted as elaborating on the topic of *research* on *peer feedback*.

In example 9-49, the elaborative collocational relation between *small-scale* and *a Chinese IELTS writing teacher and his five Chinese students* is highly context-

sensitive, because the latter bigger chunk specifies and elaborates on the former only based on this context. Both of the two items expand on the topic that is related to the feature of the participants' size in example 9-49.

Generally, the examples above demonstrate how activity-related collocations and elaborative collocations can be used in academic writing. It is recommended for the teachers to use such examples for the teaching of collocation, as these types of collocational relations are activity-centred or topic-centred, which suggests a greater difficulty of explaining their use out of context only with word meanings of the lexical items.

9.4.2.3 Teaching lexical cohesion systematically with interactive activities

Another EAP implication is based on the relationship between teachers and students as well as the context-sensitive nature of lexical cohesion. As mentioned in chapter 5, the traditional relationship between Chinese teachers and students in class is hierarchical and teacher-centred. The predominance of grammar drilling activities which has been emphasised by Chinese teachers is seen here as contributing to the lack of teaching of lexical cohesion in classes. An important reason for what is considered such a negative influence is that learning the use of lexical cohesive devices in academic writing cannot occur in a similar manner to memorising fixed grammatical rules or structures via repeated exercises, but needs to incorporate developing understanding of the variability between different lexical cohesive devices in use through the use of a range of examples. It is of significant importance for teachers to communicate the use of lexical cohesive devices with students during the teaching and practice of such devices in students' assignments or in-class activities, provide prompt feedbacks for the students and help them avoid potential cohesive errors, such as overusing or misusing lexical cohesive devices.

Some suggested interactive activities based on the present study could include the following procedures:

1. First of all, teachers introduce broad conceptual knowledge of lexical cohesion and point out that lexical cohesion contributes to the creation of the texture in a text.

2. Then teachers guide students to learn different types of lexical cohesive devices via reading academic texts, such as research articles or other academic related texts dependent on students' English proficiency levels.

3. Furthermore, students are asked to have group discussions about the features of lexical cohesive devices used in the authentic texts.

4. In addition, based on the analysis of lexical cohesion in the texts, teachers ask students to summarise the lexical cohesive devices identified in the texts orally, in order to test the level of students' understanding of the use of lexical cohesive devices in texts. Then, teachers give oral feedbacks and answer students' questions.

5. Finally, teachers conduct some fun and interesting activities in order to let students practice lexical cohesive devices in examples, such as finding the close repetitions and filling the blanks in sentences with appropriate synonyms or antonyms.

Such procedures for in-class activities firstly aim to modify the hierarchical relation between Chinese teachers and students, and encourage them to communicate more frequently during class; and secondly help the teachers move away from the grammar drilling method at the sentential level and focus on teaching academic writing at the discourse level instead, which is likely to benefit students' learning of the usage of lexical cohesive devices in their academic writing.

9.4.3 Summary

This chapter has identified several reasons for teaching lexical cohesion in EAP classes as well as providing examples of types of lexical cohesion which might be used themselves or adapted in the classroom in the teaching of the different

forms of lexical cohesion. It has also discussed some approaches to teaching lexical cohesion which might be used with Chinese and indeed other students.

Specifically, four key conclusions have been presented and discussed. Thus, firstly regarding the high levels of repetitive lexical devices used in Chinese students' writing, the suggesion of teaching other lexical cohesive devices has been proposed as the first implication arising out of this work. Secondly, tied to the context-sensitive nature of lexical cohesion, the second implication is the advice that teachers should give explicit instructions on different types of cohesive devices using appropriate examples, such as those identified in the present study which are already elucidated in relation to their individual category of lexical cohesion.

The third implication combines the traditional hierarchical relation between Chinese teachers and students as well as the context-sensitive feature of lexical cohesion. Interactive activities are recommended for teachers to conduct in EAP writing classes for teaching lexical cohesion. One purpose is to encourage the students to express their own responses to issues relating to lexical cohesion but such an approach will also help promote the communication between the teachers and students, which will be helpful for the students to understand the use of lexical cohesive devices correctly and effectively. Another purpose of such interactive activites is to emphasise the teaching of academic writing at the discoursal level and discard the traditional grammar drilling activities, something which is also beneficial and perhaps necessary for the learning of lexical cohesion which is related to the texture of the whole text instead of sentential features.

9.5 Future research and concluding remarks

The thesis study has investigated the use of lexical cohesion in Chinese postgraduates' writing in UK academic settings using a specific framework of lexical cohesion developed for this study, and using both quantitative and qualitative analysis implemented through manual analysis of postgraduate academic writing samples. There were two research points of focus in this study: the match between scores of students' module assignments and lexical cohesion;

and the difference of lexical cohesion used in different functional sections of the dissertation samples.

The research results obtained from the two foci in the MTMC and MTDC corpora are seen here as a useful starting point in terms of lexical cohesion in student written texts. For future research, it is suggested that further in-depth analyses of the variability between individual Chinese students' samples in terms of lexical cohesion should be conducted using a larger size of corpus. this would of course involve intricate and detailed analysis and may require the work of several researchers if substantive conclusions are to be generated.

Further, the study of lexical cohesive chains needs to be adopted in future research although this itself will require detailed and complex investigate. For reasons already mentioned in this chapter, the present study did not include the study of lexical cohesive chains which from a further level or form of connectivity between lexical cohesive pairs in texts.

In order to answer the questions such as how EFL students can improve their ability of using lexical cohesion in academic writing over prolonged periods, another direction for future study should focus on a particular group of students for a diachronic research, comparing their usage of lexical cohesive devices in their written work and their perception of lexical cohesion in texts over a multi-year-based period of time. It is suggested that the methodology of such research should combine interviews with the targeted students and text analysis of lexical cohesion in their written work, as the former contribute to the explanation of the research result from the student writers' perspective while the latter focuses on the result from the researcher's angle.

While there is need for further work in this area, the results and implications in the present study have hopefully shed some more light on the nature of Chinese students' academic writing in terms of the use of lexical cohesion, and will be seen as having contributed to the ongoing study of EFL students' writing as a whole.

Specifically, it is expected that the identification of relevant lexical cohesive features presented will support the teaching and learning of lexical cohesive devices used in academic writing by all EAP writers, especially Chinese EFL learners. Although lexical cohesive features identified in this thesis are seen as evidential and informative in regard to Chinese students' use of such lexical cohesive devices, the pedagogical suggestions taken from these features are still likely to require adaptation to suit other teaching contexts, not just in terms of different types of language learners but also in terms of disciplinary variations.

In situations like the one in the present study in which Chinese students were all engaged in one-year master programmes, their experience of English academic writing practice is highly limited. Therefore, pedagogical encounters with lexical cohesive patterns supporting cohesiveness of texts, should be introduced and discussed in cohesion-focused teaching which aims at developing writing strategies for producing high quality of texts. Learning how to use lexical cohesive devices in EAP classes is considered to be a key area in terms of supporting EFL learners' success in academic writing and therefore in terms of supporting their performance on their academic programmes.

Appendices

Appendix A

Table A1 Details of corpora MTMC and MTDC

The details of MTMC

text	word count
1. F1P	239
2. F2	432
3. F3P	247
4. F4P	233
5. F5P	252
6. F6P	381
7. F7	359
8. F8	406
9. F9	315
10. F10P	300
11. F11P	266
12. F12	353
13. F13	335
14. P1	486
15. P2	395
16. P3	456
17. P4	357
18. P5	439
19. P6	358
20. P7	347
21. P8	456
22. P9	387
23. P10Paper	244
24. P11Paper	331
25. P12Paper	243
26. P13Paper	208
27. M1P	197
28. M2P	228

The details of MTDC

text	word count
1. D2C1I	419
2. D4C1I	421
3. D6C1I	421
4. D7C1I	427
5. D8C1I	434
6. D11C1I	412
7. D12C1I	441
8. D13C1I	428
9. D14C1I	433
10. D2C2L	451
11. D4C2L	457
12. D6C2L	427
13. D7C2L	428
14. D8C2L	439
15. D11C2L	447
16. D12C2L	447
17. D13C2L	449
18. D14C2L	450
19. D2C4M	443
20. D4C3M	443
21. D6C3M	435
22. D7C3M	450
23. D8C3M	431
24. D11C3M	423
25. D12C3M	441
26. D13C3M	439
27. D14C3M	408
28. D2C5F	435

29. M3P	374
30. M4	435
31. M5	315
32. M6	330
33. M7P	374
34. M8P	301
35. M9P	250
36. M10P	261
37. M11P	224
38. M12P	350
39. M13P	260
40. D1	449
41. D2	385
42. D3	455
43. D4P	361
44. D5P	407
45. D6	377
46. D7	373
47. D8P	212
48. D9P	213
49. D10P	258
50. D11	435
51. D12	449
52. D13	440

∑ 17538

29. D4C4F	391
30. D6C4F	423
31. D7C4F	418
32. D8C4F	403
33. D11C4F	403
34. D12C4F	400
35. D13C4F	437
36. D14C4F	402
37. D2C6C	424
38. D4C5C	404
39. D6C5C	420
40. D7C5C	412
41. D8C5C	420
42. D11C5C	405
43. D12C6C	411
44. D13 5&6&7C	407
45. D14C6C	389

∑ 19148

∑ 36686

Table A2 Topics and briefs of samples in the MTMC corpus

Assignments	Topics/Briefs
F1P	Descriptions about learners for a syllabus: Chinese undergraduates, their learning needs and admission requirements
F2	Plan for evaluating the qualities of usefulness of a language test: discussion about the concept of reliability, construct validity and authenticity of a test
F3P	Descriptions about learners for designing a syllabus: Chinese learners with mixed backgrounds and their learning needs
F4P	Descriptions about learners for designing a curriculum: Chinese learners at pre-intermediate or intermediate level and their learning needs
F5P	Descriptions about learners for designing a syllabus: background information about Chinese learners at a summer academic writing course
F6P	Design statement of a low-stakes written test: purpose of test, description of characteristics of test takers, definition of constructs and consideration of qualities of usefulness
F7	Design statement of a low-stakes competence-based listening test: purpose of test, description of the TLU domain and description of characteristics of test takers
F8	Description of test takers who are intermediate level students and whose purpose of taking the test is to get a higher level of speaking and listening English competence so that to get a standard certification of interpretation.
F9	Description of test takers who are business-major students, joining an English training program to become interpreters and bilingual sales agents in an international furniture fair.
F10P	Descriptions about learners for designing a syllabus: background information about Taiwanese learners at a high school English class with multitudinous needs

F11P	Descriptions about learners for designing a syllabus: background information about Taiwanese learners between the age of about 10 to 13 with boys and girls mixed together and their learning needs
F12	Design statement of a relatively high-stakes screening test: purpose of test, description of the target language use (TLU) domain
F13	Design statement of a preparation test for high school entrance examination: purpose of test, inference, stakes, specific decisions, description of TLU domain and task types
P1	Description of the target group of an assessment which is designed to test the English speaking skills of second-year medical students from a University in China after taking their English speaking training course
P2	Design statement of a formative speaking test for a class of 24 students in the second grade of senior high school: education background introduction, purpose of the test and characteristics of test takers
P3	Descriptions about learners for designing a syllabus: senior college students in Shanghai, China, who study Human Resource in college
P4	Design statement of a listening test designed for twenty-five Grade Three students, who are aging from 8 to 10 years old in Suzhou, China: test purposes, description of TLU domain and task types
P5	Design statement of a high-stake test designed for evaluating 20 prospective airline stewardesses' achievement relevant to the short-term spoken English training program and for deciding the final list of employees of an airline company: test purposes and make-up of the test
P6	Design statement of a speaking-oriented assessment tool for Chinese students who want to compete in the final round in

	the national English speaking contest: description of test developer and test writers
P7	Design statement of a classroom achievement test for primary English learners who are tested as 'pioneers' to be taught based on a new teaching syllabus which focuses on communication in daily life: purpose of the test, description of the TLU domain and task types
P8	Design statement of a test for a Chinese company which specialised in International trade around the world to examine the employees' English proficiency: target group, purpose of the test and description of tasks in the TLU domain
P9	Descriptions about learners for designing a syllabus: data and goals for a group of employees of Research and Development Department in a Chinese branch company of a multinational company
P10Paper	Design statement of a test to assess the academic writing in-sessional courses for international students: test purpose, description of TLU domain and task types, description of characteristics of test takers and definition of constructs
P11Paper	Descriptions about learners for designing a syllabus: all 15 learners come from the same company in China and are going to have a British tour in April. Their purpose in attending a 15 days short-term travelling English training is to learn some basic expression which will be used in their trip
P12Paper	Descriptions about learners for designing a syllabus: Chinese adults are applying master or bachelor degree in abroad university but now are still working or studying. They attend a night class for IELTS-writing training purposes
P13Paper	Descriptions about learners for designing a syllabus: fourteen junior college students who are going to be volunteers for the London Olympic Games of 2012 attend a language training course to improve spoken English skills

M1P	Descriptions about learners for designing a syllabus: Chinese teenage learners whose English reaches an intermediate level need explicit teaching of the thinking difference which is accountable for errors in spoken English
M2P	Descriptions about learners for designing a syllabus: a group of new employees of a multinational business company in China attend a two-week business English training course for their work
M3P	Descriptions about learners for designing a syllabus: a group of university students in China attend an English summer camp in China to improve their English, especially speaking skills
M4	Design statement of a low-stakes speaking test for Chinese students in their first year in a private senior secondary school, who have one-hour weekly English speaking course
M5	Design statement of a screening test for candidates who is going to be hired by an international hotel chain: purposes of test
M6	Design statement of an oral test to screen the job applicants in terms of their ability of spoken English: purpose of the test and characteristics of the test takers
M7P	Descriptions about learners for designing a syllabus: part-time adult EFL learners who take night classes to improve their communication and English skills in business
M8P	Descriptions about learners for designing a syllabus: 30 soldiers who work for Chinese navy forces take a ten-day training course to learn basic spoken English for their military exchanges with British navy forces
M9P	Descriptions about learners for designing a syllabus: a group of senior high school graduates attend an “English Speaking Summer Camp” in China to develop their oral English for giving presentations and having seminars at university

M10P	Descriptions about learners for designing a syllabus: a group of Chinese junior high school students who are going to have a 20-day summer camp in the UK during the summer vocation, attend a course to improve their English communicative abilities
M11P	Descriptions about learners for designing a syllabus: Chinese undergraduates attend a ten-week optional in-sessional British Parliamentary (BP) Debate training course to improve their linguistic and cognitive skills
M12P	Descriptions about learners for designing a syllabus: 10 to 12 Chinese young adults take a two-month summer speaking course in a private English training school in China as part of their preparation for meeting the level of English language proficiency required for admission to UK universities
M13P	Descriptions about learners for designing a syllabus: a group of adults who are the workers of the Canton Exhibition Centre attend an English training program to improve oral English skills for communicating with foreign customers in China Import and Export Fair
D1	Design statement of a progress speaking test to evidence Chinese students' capability to use verb forms and 'he' and 'she' correctly in a simulated real-world conversation with peers: purpose of the test and characteristics of the test takers
D2	Design statement of 50 minutes computer-based oral English test in a language laboratory
D3	Design statement of a progress or mid-term achievement test to measure language and skills progress of junior three students (aged 14~16) studying in a Chinese international school in relation to the syllabus they are following
D4P	Descriptions about learners for designing a syllabus: a total number of 25 students who come from multi-disciplinary background take an optional English course running in a

	foreign studies university in mainland China, aiming to improve their speaking and listening skills
D5P	Descriptions about learners for designing a syllabus: 40 students who are of first year in an elite senior high school in China take English classes, aiming to build confidence in spoken English and practice their conversational and communicative skills for the National College Entrance Examination
D6	Design statement of an assessment to examine the oral English skills of ESP (English for Specific Purposes) adult learners who are employees in a major company in the Chinese chemicals supplies market and try to improve their oral English to communicate with their foreign customers
D7	Design statement of an assessment to examine to what extent Chinese undergraduates experience foreign language speaking anxiety and possible reasons and influences about it
D8P	Descriptions about learners for designing a syllabus: 26 young learners (aged 14 or 15) who have just finished their first-year study in a foreign language junior high school in China and will attend an exchange programme in a secondary school in the UK for 12 months, aiming to reach the intermediate level in English after a three-week intensive English course
D9P	Descriptions about learners for designing a syllabus: 22 Business Administration major students at their second school year in a joint-funded university in Shanghai attend a continuation Business English course for one year, aiming to improve Business English learning through completing different tasks in 13 thematic units
D10P	Descriptions about learners for designing a syllabus: 20 learners whose ages range from 13 to 18, take the speaking module of a summer course of the Listening and Speaking Department of Beijing Neworiental School, aiming at fluency and the improvement of confidence in speaking English

D11	Design statement of an tape-based oral English test for selection of candidates into the final round of interview for jobs as English-speaking tour guides in a local travel agency in China: purpose of the test, description of TLU domain and characteristics of the test takers
D12	Design statement of a syllabus-based read and write test to evaluate the learning outcomes of the students in their first year English study for helping them better prepare second-year study in a Chinese primary school: characteristics of the test takers, purpose of the test and implementation of the test
D13	Descriptions about learners for designing a syllabus: Chinese learners aged from 10 to 11 take an English summer camp training course, aiming to increase their interests, make up for their poor exposure to English in schools and have more chances to use English instead of learning English rules blindly

Table A3 Titles of samples in the MTDC corpus

Dissertations	Titles
D2	Investigating the Reasons for Chinese Students' Incorrect Use of Plural Morphemes in Spontaneous Speaking Situations: a Crosslinguistic Difference in Marking Number of Nouns
D4	What good practices should be included in teacher feedback on IELTS writing task 2 for Chinese learners?
D6	An investigation of the reasons CLT cannot be implemented in Yunnan Nationalities University
D7	The LLM seminar and the Non Native English Speaker: an exploration of student language learning background, their expectations, and the implications for their learning experience
D8	How suitable are TED Talks for Academic Listening? An investigation of Academic Vocabulary List representation, Speed, and Lexical Density in TED Talks and Lecture Discourse

D11	Exploring Difficulties Chinese Students Who Learn Public Health in Durham University Have in Academic Writing and the Reasons Why They Encounter Such Difficulties
D12	Efficacy of Supplementary Education in Hong Kong to Assist Secondary School Students Acquire ESL and its Effect to the English Education Environment
D13	Chinese Students' Perspectives on Peer Feedback in EAP Writing Classes
D14	An investigation of the extent to which the English textbooks used in Shanghai senior schools promote implicit meaning and offer opportunities for modification

Appendix B

分析规则和方法 (analytical strategies)[66]

- 关系标记对照表 (coding system for categories of lexical cohesion)

lexical cohesive relation	representation	example
repetition 重复	加粗字体	**toe – toe**
synonym 近义	橙色 orange	test – assessment
hyponym 下义（种类）	绿色 green	language – English
meronym 下义（部分）	红色 red	English skills – listening
identity 下定义	蓝色 blue	test – element
superordination 上义	褐色 dark red	English – language
collocation 搭配（包括反义）	黄色高亮 highlight yellow	learning – teaching

lexical cohesive relation	definition
repetition 重复	同一单词或词组的重复使用，包括相同形式，语法变换形式，派生词形式和词性变换形式，除去一词多义 The repetition of a lexical item (including Identical form, form with grammatical change, derivational change or word class change; excluding polysemy)
synonym 近义	单词或词组意思相同或相近，不仅限于同一词性 the meanings of two lexical items are somehow synonymous (no word class limited)
hyponym 下义（种类）	后面出现的词或词组是前面出现的词或词组所指类别的一种 the second item is either a subclass or another class at the same level of classification of the first one
meronym 下义（部分）	后面出现的词或词组是前面出现的词或词组的一个部分 the second item is either a part or another part at the same level of classification of the first one
identity 下定义	后面出现的词或词组和前面出现的词或词组有相同指称对象，后者是前者的另一种身份或表现前者的某种特征 the lexical items in a cohesive pair refer to the same subject, in which the latter shows another identity or characteristics of the former item
superordination 上义	后面出现的词或词组是前面出现的词或词组的上一级或更广泛的类别 the second item is either a superclass of the first one

[66] Mandarin is included here because co-raters, who are Chinese native speakers, needed this for sufficient understanding of the analytical strategies.

collocation 搭配（包括反义）	后面出现的词或词组与前面出现的词或词组经常在相似情境下一起使用（包括反义词和互补词） Two lexical items tend to occur together in similar environment (including antonyms and complementaries)

- 除了重复之外的其他关系当中涉及到的两个 lexical item 都要在分析表中列出，方便日后检查。

  Except repetition, the two lexical items of a cohesive pair in the remaining categories need to be displayed in the analysing table for future examination.

- 遇到拿捏不定的同义词可用 word 的同义词选项作为参考。

  When coming across uncertain synonymous items, use the “synonyms” function in Word for reference.

- collocation 当中词汇搭配判断主观性较强，认为像是经常在语境中同时出现的搭配就列出来，后期我会在 BNC 语料库当中查找出现频率进行进一步判断和取舍。

  The items in collocation are relatively subjective. It should be identified based on rater’s personal judgement of their tendency of co-occurrence in similar environment. The researcher will re-examine with recourse to the BNC Corpus.

Appendix C

The training procedure for co-raters

1. Thanks for agreeing to participate in this rater work for lexical cohesion analysis. This work is not paid in any form but your name will be mentioned and your contribution acknowledged, should you so wish when the thesis is published.

2. I'm going to tell you what is lexical cohesion and how you can use this basic knowledge to analyse the students' assignments in MA TESOL and MA Applied Linguistics for TESOL programmes. The purpose of this rater work is to check whether the current framework of lexical cohesion and analytical strategy will work for the full-scale analysis. So after you finish your analysis, I will compare your results of analysis with mine using a statistical tool to check the level of inter-rater agreement reliability.

3. If the level of inter-rater agreement reliability is low, it means we have disagreement to a certain extent. Then we need to discuss the existing disagreement as well as the lexical cohesive model fit and the understanding of this model to see whether we can reach agreement afterwards.

4. As the purpose of this rater work is to check the practicality of the current model to see whether it needs to be revised or developed, I will start by introducing the model to you in the following steps:

    a. Send the electronic document of model explanation via WeChat to the you.
    b. Run through the model and ask if you have any questions about the model.
    c. Send the analytical strategy document to you via WeChat and explain how the samples should be analysed using this model.

5. After the introduction of this model, I will provide an illustrative sample analysis R1[67] and demonstrate the analysing strategy with this analysed sample.

6. Then I will send one sample R2 to you and another co-rater, and ask you to finish the analysis work by one week. During this period, you can ask any questions related to the analysis.

7. After the first three-rater analysis stage, the analyses will be compared via the statistical tool. Further discussion will be held between three of us.

8. More samples will be sent to you for two-rater analysis stage, after which the analyses will also be checked via this statistical tool.

[67] R1 = sample 1 for rater work 1; R2 = sample 2 for rater work

Appendix D

Table D1 MA TESOL Module assignment corpus with raw distribution of lexical cohesion categories by marking scales

No.	text	rep.	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.	wc[68]
1.	F1P	43	7	0	16	0	1	1	2	5	239
2.	F2	68	4	0	2	0	9	3	3	7	432
3.	F3P	27	2	3	12	2	1	2	3	11	247
4.	F4P	27	4	7	5	0	1	1	0	9	233
5.	F5P	35	1	6	6	1	1	0	0	6	252
6.	F6P	70	6	0	8	2	3	3	2	7	381
7.	F7	60	5	2	6	0	3	1	0	7	359
8.	F8	65	4	1	4	6	8	2	2	7	406
9.	F9	57	2	6	5	3	1	4	2	14	315
10.	F10P[69]	37	5	3	1	2	3	2	3	8	300
11.	F11P	41	2	1	2	2	3	1	4	8	266
12.	F12	80	2	1	1	4	0	1	2	12	353
13.	F13	74	1	1	4	0	3	3	3	9	335
14.	P1	76	10	5	4	5	5	3	2	7	486
15.	P2	62	5	5	8	1	5	4	4	8	395

[68] Word count
[69] F10P and F11P are written by Taiwanese while the other samples are written by students from the PRC.

16.	P3	102	2	1	17	3	5	6	4	14	456
17.	P4	63	3	3	4	2	3	4	3	8	357
18.	P5	68	12	5	0	1	6	3	2	4	439
19.	P6	67	4	2	8	3	4	5	3	3	358
20.	P7	70	4	1	1	0	5	0	3	9	347
21.	P8	79	4	1	5	0	5	1	4	12	456
22.	P9	59	3	1	7	1	5	4	3	12	387
23.	P10Paper	32	3	3	3	5	2	1	1	11	244
24.	P11Paper	35	5	5	12	5	3	2	4	11	331
25.	P12Paper	35	5	3	4	1	1	4	1	6	243
26.	P13Paper	21	4	3	3	0	2	2	1	5	208
27.	M1P	24	4	2	5	0	3	2	4	7	197
28.	M2P	37	2	5	0	1	1	3	0	6	228
29.	M3P	65	6	0	9	0	1	4	2	6	374
30.	M4	77	3	1	2	2	5	1	4	10	435
31.	M5	44	8	4	2	2	6	0	1	4	315
32.	M6	52	1	2	4	1	2	1	1	4	330
33.	M7P	62	2	4	5	0	2	2	1	4	374
34.	M8P	41	2	1	2	2	1	4	0	8	301
35.	M9P	31	3	1	6	2	0	1	1	10	250
36.	M10P	32	2	2	7	2	0	3	1	3	261
37.	M11P	24	1	2	6	1	1	1	2	7	224
38.	M12P	33	3	1	1	4	3	1	2	10	350

39.	M13P	22	2	3	7	0	2	3	1	13	260
40.	D1	71	5	0	5	2	4	2	3	18	449
41.	D2	63	7	0	10	0	1	1	2	9	385
42.	D3	67	5	5	8	4	2	4	5	10	455
43.	D4P	48	4	1	10	4	2	4	5	8	361
44.	D5P	61	8	2	4	2	1	4	3	8	407
45.	D6	53	7	2	8	4	3	4	0	9	377
46.	D7	60	6	3	6	0	4	3	2	6	373
47.	D8P	26	1	2	9	2	1	2	1	6	212
48.	D9P	34	4	6	2	4	0	2	2	8	213
49.	D10P	40	1	3	4	5	2	1	0	7	258
50.	D11	86	5	5	13	2	6	6	1	6	435
51.	D12	77	4	2	7	6	2	4	4	8	449
52.	D13	90	8	2	6	2	13	3	3	4	440
total		2773	213	130	296	103	156	129	112	419	17538
average		53	4	3	6	2	3	2	2	8	337

Appendix E

Table E1 MA TESOL Module assignment corpus with normalised distribution of lexical cohesion categories (occurrences/1000 words) linked to assignment marks[70]

text	rep.	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.
F1P	180	29	0	67	0	4	4	8	21
F2	157	9	0	5	0	21	7	7	16
F3P	109	8	12	49	8	4	8	12	45
F4P	116	17	30	21	0	4	4	0	39
F5P	139	4	24	24	4	4	0	0	24
F6P	184	16	0	21	5	8	8	5	18
F7	167	14	6	17	0	8	3	0	19
F8	160	10	2	10	15	20	5	5	17
F9	181	6	19	16	10	3	13	6	44
F10P	123	17	10	3	7	10	7	10	27
F11P	154	8	4	8	8	11	4	15	30
F12	227	6	3	3	11	0	3	6	34
F13	221	3	3	12	0	9	9	9	27
P1	156	21	10	8	10	10	6	4	14
P2	157	13	13	20	3	13	10	10	20

[70] Red highlighted numbers represent the maximum value in its lexical cohesive category while green highlighted numbers denote the minimum in that category.

P3	224	4	2	37	7	11	13	9	31
P4	176	8	8	11	6	8	11	8	22
P5	155	27	11	0	2	14	7	5	9
P6	187	11	6	22	8	11	14	8	8
P7	202	12	3	3	0	14	0	9	26
P8	173	9	2	11	0	11	2	9	26
P9	152	8	3	18	3	13	10	8	31
P10Paper	131	12	12	12	20	8	4	4	45
P11Paper	106	15	15	36	15	9	6	12	33
P12Paper	144	21	12	16	4	4	16	4	25
P13Paper	101	19	14	14	0	10	10	5	24
M1P	122	20	10	25	0	20	5	20	36
M2P	162	9	22	0	4	4	13	0	26
M3P	174	16	0	24	0	3	11	5	16
M4	177	7	2	5	5	11	2	9	23
M5	140	25	13	6	6	19	0	3	13
M6	158	3	6	12	3	6	3	3	12
M7P	166	5	11	13	0	5	5	3	11
M8P	136	7	3	7	7	3	13	0	27
M9P	124	12	4	24	8	0	4	4	40
M10P	123	8	8	27	8	0	11	4	11
M11P	107	4	9	27	4	4	4	9	31
M12P	94	9	3	3	11	9	3	6	29

M13P	85	8	12	27	0	8	12	4	50
D1	158	11	0	11	9	9	4	7	40
D2	164	18	0	26	3	3	3	5	23
D3	147	11	11	18	4	4	9	11	22
D4P	133	11	3	28	6	6	11	14	22
D5P	150	20	5	10	2	2	10	7	20
D6	141	19	5	21	8	8	11	0	24
D7	161	16	8	16	11	11	8	5	16
D8P	123	5	9	42	5	5	9	5	28
D9P	160	19	28	9	0	0	9	9	38
D10P	155	4	12	16	8	8	4	0	27
D11	198	11	11	30	14	14	14	2	14
D12	171	9	4	16	4	4	9	9	18
D13	205	18	5	14	30	30	7	7	9
Average	154	12	8	18	6	8	7	6	25

Appendix F

Statistical result of MTMC

Table F1 Descriptive figures of total lexical cohesive devices used in marking-scale groups

marking scale categories	N	mean	std. deviation
failed	13	255.769	33.497
pass	13	254.077	31.811
merit	13	217.923	25.287
distinction	13	253.000	30.285
total	52	245.192	33.497

Table F2 Result of one-sample Kolmogorov-Smirnov test of lexical cohesion devices in marking scale groups

		lexical cohesion
N		52
normal parameters[a,b]	mean	245.1923
	std. deviation	33.49746
most extreme differences	absolute	.135
	positive	.135
	negative	-.063
test statistic		.135
asymp. sig. (2-tailed)		.019[c]

a. Test distribution is normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

Table F3 Result of Levene's test of lexical cohesive devices in marking scale groups

Levene statistic	df1	df2	sig.
.543	3	48	.655

Table F4 Result of ANOVA test of lexical cohesive devices in marking scale groups

	sum of squares	df	mean square	F	sig.
between Groups	12939.923	3	4313.308	4.675	.006
within Groups	44286.154	48	922.628		
total	57226.077	51			

Table F5 Result of Tukey's HSD test of comparison of lexical cohesive devices between marking scale groups

marking scale	N	subset for alpha = 0.05	
		1	2
merit	13	217.9231[a]	
distinction	13		253.0000
pass	13		254.0769
failed	13		255.7692
sig.		1.000	.996

a. Means for groups in homogeneous subsets are displayed.

Table F6 Result of Spearman's rho correlation test of lexical cohesive devices in marking scale groups

			marking scale	lexical cohesion
Spearman's rho	marking scale	correlation coefficient	1.000	-.150
		sig. (2-tailed)	.	.290
		N	52	52
	lexical cohesion	correlation coefficient	-.150	1.000
		sig. (2-tailed)	.290	.
		N	52	52

Appendix G

Table G1 Lexical cohesive pairs in extract 1

text	rep.	syn.	hyper.	hypo.	mer.	sig.	ide.	ant.	col.
This assessment tool aims at a group of **business-major students**									
who are joining an English training program,									Students – training program
after which they are going to become **interpreters** and **bilingual sales agents** in an **international furniture fair**.							Business-major students – interpreters and bilingual sales agents		English – Interpreters Business – sales agents
International Famous **Furniture Fair** (Dongguan), commonly known as **3F**, is the biggest and most representative **furniture exhibition** in **China**.	International Furniture fair		Dongguan – mainland China	Furniture exhibition – 3F					

In 3F, leading **furniture** manufacturers from mainland **China**, Hong Kong and Taiwan will take part in this **exhibition**, seeking for new contracts with byers [buyers] all over the world.	3F Furniture China Exhibition				China – Mainland China				Fair – manufacturers International – over the world
During **exhibition** time, which lasts for 14 days semi-annually,	Exhibition								
about 80 **interpreters and bilingual sales agent**[s] are **needed**.	Interpreters Bilingual sales agents								
To meet the **needs** of trading companies and **provide** them with better **interpretation** services, the **exhibition committee** cooperates with **business** faculty in Sun Yat-sen **University** (SYEU),	Needed – needs Interpreters – interpretation Exhibition Business						Manufacturer/buyers – trading companies		Business – trading
which is one of the top **universities** in southern	University – universities		Sun		Mainland China –				

China, to **provide internship** opportunities for their **business-major students**.	China Provide Business-major students		Yat-sen University – Universities		southern China				
Students who are interested in this **internship** can apply for it	Students Internship								
and they are **required** to at first pass interviews and then language tests to ensure	Required – requirements	Assessment – tests	English – language						Internship – interviews
they meet relative **requirements** of the **committee**.	Committee								Interviews /language tests – requirements of the committee
Raw frequency	21	1	3	1	2	0	2	0	8
Normalised frequency	117	6	17	6	11	0	11	0	45

Table G2 Lexical cohesive pairs in extract 2

text	rep.	syn.	hyper.	hypo.	mer.	sig.	ide.	ant.	col.
This is a **speaking**-oriented **assessment** tool	Speaking								
which is designed for the **high level English Speaking Contest** qualification trials-- the **national** level.	English								Assessment – test takers
The **test takers** are upper intermediate and advanced **English** learners.		Assessment – test							
It is within the scope of **high level**'s **English** majors' **spoken** **test**.	High level English Speaking – spoken Test		Speaking-oriented assessment – English majors' spoken test				The national level – high level		Upper … learners – English majors
Take IELTS as the **English** proficiency criteria as reference, the **test takers** for this **assessment** should have scored 7 point (each	English Test takers Assessment High – higher			Test – IELTS					English – IELTS

section is no less than 6.5 point) and **higher**.									
The **test takers** are Chinese students	Test takers	Learners – students					Upper … learners – Chinese students		
who want to compete in the final round in the **national English speaking contest**.	National English speaking contest								
(I) Details of Recourses [Resources]									
Test developer[s] and **Test** writers	Test							Test takers – test … writers	
Bachman and Palmer points out that **test developers** is one of the most important factors in human **recourses [resources]**.	Test developer – test developers Resources								
They are responsible for the **test** specification,	Test								Test developers – the … use

management, try-out, achievement and use.									
They administrate and take control of the entire **test** process and make sure the test can be well carried out.	Test								
The **developers** in this **assessment** are not the daily teachers	Developers Assessment							Students – teachers	
who are familiar to the **test takers**.	Test takers						Chinese students – test takers		
Raw frequency	21	2	1	1	0	0	3	2	4
Normalised frequency	121	11	6	6	0	0	17	11	23

Table G3 Lexical cohesive pairs in extract 3

text	rep.	syn.	hyper.	hypo.	mer.	sig.	ide.	ant.	col.
1.1 **learner** profile and job background									
the syllabus is **designed** for a group of adults							Learner – adults		
who are the workers of the **Canton** Exhibition **Centre**.							Adults – workers		Job – workers
This **centre** is located in **Guangzhou**,	Centre Canton – Guangzhou		Workers – centre						
where **China** Import and Export Fair, also called **Canton Fair**, is held twice a year in **Spring** and Autumn.	Guangzhou – Canton		Guangzhou – China						
It is **China's** largest **trade fair** of the highest **level**, with the most complete varieties and the largest attendance and business turnover,	China Fair		Canton Fair – trade fair						Import and Export Fair – business turnover
which attracts thousands of millions of people both home and abroad, including **foreign**							China – home		Business – trade

trade companies, factories, scientific research institutions, **foreign** invested enterprises and so on.									
Recently, the administrative **level** of this **centre** launched an investigation to **identify** the areas for further development, and found that	Level Centre				Centre – administrative level				
the English communicative competence of the **staff** was unsatisfactory and always led to **misunderstandings**, inconvenience and complaints from the **customers**.		Workers – staff				Areas – English … staff	Foreign… Enterprises – customers		
As a result, in order to fully meet such challenges brought by the coming **spring**-term **fair** as the surging number of **foreign** clients and specific **fair**-related **needs** and issues, this **centre** decided to introduce an **English** training program to their sixteen **staff**,	Spring Fair Foreign Centre English Staff Workers Customer	Abroad – foreign Customers – clients			Autumn – spring			Home – foreign Staff – clients	Complaints – clients

more specifically, among whom, five are receptionists of the front desk, four are ushers, three are **workers** of **customer** service, and four are the personal [personnel] of function rooms.				Staff – receptionist /ushers/ workers/ personnel	Centre – front desk/ customer service/ function rooms				
Raw frequency	15	3	3	1	3	1	4	2	4
Normalised frequency	72	14	14	5	14	5	19	10	19

Table G4 Lexical cohesive pairs in extract 4

text	rep.	syn.	hyper.	hypo.	mer.	sig.	ide.	ant.	col.
a) the **learners**									
The **learners** are 40 **students**	Learners						Learners – students		
who are of first year in an **elite** senior **high school** in **China**.									Students – first year Students – senior high school
They just **experienced** the **entrance examination**									Students – examination China – entrance examination
and their overall performances were **good**,									
which means that most of them have solid foundation in **English**.									Performance were good – solid … English
Due to the **exam**-oriented **educational** atmosphere in **china**, they used to pay more emphasis on the	Examination – exam China English					English –part			Students – educational

written **English** other than **oral part**.									
For them, **English learning** is by **learning** the explicit input from teacher and drill the **knowledge** through different types of **exam** paper.	English Learners – learning Exam							Students – teacher	
On the other hand, although majority of them have **good** command of vocabulary and grammar,	Good Oral								English learning – vocabulary and grammar
they have little **confidence**									
while **speaking** and their **oral** expressions were not so proficient owing to lack of time and opportunities to **practice**.	Oral	Oral – speaking Have…of – proficient			Written – speaking				Little confidence – not so proficient
While the **oral part** is included in **the National College Entrance Examination**,	Oral part Entrance examination			Senior high school – college			China – national		
and they realize the importance of **speaking**	Speaking English							Proficient –	

English and also the deficiency of their speaking English,								deficiency	
so **part** of their focuses begin to switch to the **speaking** part.	Part Speaking	Emphasis – focuses							
Exposed to the idea that once you **enter** an **elite high school**,	Entrance – enter Elite High school								
you've already been on the way to an **elite** university,	Elite	College – university		High school – university					
the **learners** believe that	Learning – learners								
their **school** will provide them with good **education** and help them to make progress according to their needs,	School Educational – Education								Learners – school Learners – education
Raw frequency	22	4	0	2	1	1	2	2	10
Normalised frequency	104	19	0	9	5	5	9	9	47

Appendix H

Table H1 MA TESOL Dissertation Corpus with raw distribution of lexical cohesion categories in functional sections

No.	sample	rep.[71]	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.	wc[72]
1.	D2C1I[73]	78	10	3	6	4	2	11	3	2	419
2.	D4C1I	91	15	3	3	4	3	3	1	9	421
3.	D6C1I	80	11	3	5	0	6	2	6	3	421
4.	D7C1I	82	10	2	2	0	8	1	2	7	427
5.	D8C1I	71	3	4	4	1	3	4	4	6	434
6.	D11C1I	48	4	4	7	0	2	3	3	4	412
7.	D12C1I	62	7	3	0	2	3	9	3	4	441
8.	D13C1I	63	10	1	5	0	3	5	2	3	428
9.	D14C1I	42	11	1	2	0	2	2	10	3	433
10.	D2C2L	75	5	3	16	13	3	14	3	8	451
11.	D4C2L	88	10	1	1	0	7	4	2	4	457
12.	D6C2L	63	3	1	5	2	6	5	3	7	427
13.	D7C2L	80	5	1	7	0	2	5	1	7	428

[71] Rep. = repetition; syn. = synonymy; hype. = hyperonymy; hypo. = hyponymy; mer. = meronymy; sig. = signalling nouns; ide. = identity of reference; ant. = antonymy; and col. = collocation. They are categories or sub-categories of lexical cohesion.
[72] Word count
[73] D2C1I is abbreviation for dissertation 2 chapter 1 in the introduction section. The same rule applies to other abbreviations for all of the sample names in this chapter.

14.	D8C2L	88	3	0	10	2	4	3	0	2	439
15.	D11C2L	60	8	4	8	0	9	5	3	7	447
16.	D12C2L	86	7	2	5	0	1	8	4	9	447
17.	D13C2L	77	11	2	7	1	6	1	2	5	449
18.	D14C2L	70	6	2	3	2	9	2	5	2	450
19.	D2C4M	71	4	3	4	6	5	6	2	4	443
20.	D4C3M	62	9	8	5	4	3	3	1	9	443
21.	D6C3M	73	4	3	1	7	5	6	2	7	435
22.	D7C3M	89	5	1	4	10	4	3	2	9	450
23.	D8C3M	81	4	0	0	1	8	2	1	5	431
24.	D11C3M	55	12	3	2	1	4	2	2	13	423
25.	D12C3M	86	12	3	3	0	6	5	7	8	441
26.	D13C3M	84	16	2	1	0	7	4	5	3	439
27.	D14C3M	70	6	3	3	4	3	3	3	5	408
28.	D2C5F	69	4	1	4	5	2	10	1	8	435
29.	D4C4F	70	5	0	0	5	4	3	2	1	391
30.	D6C4F	64	2	1	2	3	5	2	5	7	423
31.	D7C4F	68	9	1	3	0	2	4	3	10	418
32.	D8C4F	74	3	1	8	6	2	0	1	1	403
33.	D11C4F	60	8	2	5	0	4	4	2	6	403
34.	D12C4F	63	6	2	2	5	4	4	3	6	400
35.	D13C4F	78	13	2	1	1	6	3	6	2	437
36.	D14C4F	55	15	3	2	1	7	3	3	4	402

37.	D2C6C	75	9	1	2	4	7	9	3	5	424
38.	D4C5C	64	7	1	2	1	2	1	9	5	404
39.	D6C5C	50	3	2	2	0	5	2	4	13	420
40.	D7C5C	61	3	0	4	0	3	3	3	15	412
41.	D8C5C	64	2	1	1	4	7	4	4	6	420
42.	D11C5C	56	8	1	1	0	8	1	3	4	405
43.	D12C6C	63	6	2	3	0	3	1	2	11	411
44.	D13C5&6&7C	61	11	2	6	1	3	4	0	10	407
45.	D14C6C	63	8	1	1	0	1	2	4	6	389
	total	3133	333	90	168	100	199	181	140	275	19148
	average	70	7	2	4	2	4	4	3	6	426

Appendix I

Table I1 MA TESOL Dissertation Corpus with normalised distribution (per 1,000 words) of lexical cohesion categories in functional sections

sample	rep.	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.
D2C1I	186	24	7	14	10	5	26	7	5
D4C1I	216	36	7	7	10	7	7	2	21
D6C1I	190	26	7	12	0	14	5	14	7
D7C1I	192	23	5	5	0	19	2	5	16
D8C1I	164	7	9	9	2	7	9	9	14
D11C1I	117	10	10	17	0	5	7	7	10
D12C1I	141	16	7	0	5	7	20	7	9
D13C1I	147	23	2	12	0	7	12	5	7
D14C1I	97	25	2	5	0	5	5	23	7
D2C2L	166	11	7	35	29	7	31	7	18
D4C2L	193	22	2	2	0	15	9	4	9
D6C2L	148	7	2	12	5	14	12	7	16
D7C2L	187	12	2	16	0	5	12	2	16
D8C2L	200	7	0	23	5	9	7	0	5
D11C2L	134	18	9	18	0	20	11	7	16
D12C2L	192	16	4	11	0	2	18	9	20
D13C2L	171	24	4	16	2	13	2	4	11

D14C2L	156	13	4	7	4	20	4	11	4
D2C4M	160	9	7	9	14	11	14	5	9
D4C3M	140	20	18	11	9	7	7	2	20
D6C3M	168	9	7	2	16	11	14	5	16
D7C3M	198	11	2	9	22	9	7	4	20
D8C3M	188	9	0	0	2	19	5	2	12
D11C3M	130	28	7	5	2	9	5	5	31
D12C3M	195	27	7	7	0	14	11	16	18
D13C3M	191	36	5	2	0	16	9	11	7
D14C3M	172	15	7	7	10	7	7	7	12
D2C5F	159	9	2	9	11	5	23	2	18
D4C4F	179	13	0	0	13	10	8	5	3
D6C4F	151	5	2	5	7	12	5	12	17
D7C4F	163	22	2	7	0	5	10	7	24
D8C4F	184	7	2	20	15	5	0	2	2
D11C4F	149	20	5	12	0	10	10	5	15
D12C4F	158	15	5	5	13	10	10	8	15
D13C4F	178	30	5	2	2	14	7	14	5
D14C4F	137	37	7	5	2	17	7	7	10
D2C6C	177	21	2	5	9	17	21	7	12
D4C5C	158	17	2	5	2	5	2	22	12
D6C5C	119	7	5	5	0	12	5	10	31
D7C5C	148	7	0	10	0	7	7	7	36

D8C5C	152	5	2	2	10	17	10	10	14
D11C5C	138	20	2	2	0	20	2	7	10
D12C6C	153	15	5	7	0	7	2	5	27
D13C5&6&7C	150	27	5	15	2	7	10	0	25
D14C6C	162	21	3	3	0	3	5	10	15
total	7354	783	210	392	233	466	422	329	647
average	163	17	5	9	5	10	9	7	14

Appendix J

Statistical results of MTDC

Table J1 Descriptive figures of total lexical cohesive devices used in functional-section groups

functional section categories	N	mean	std. deviation
introduction	9	238.549	49.171
literature review	9	252.874	27.114
methodology	9	253.217	25.265
findings and discussion	9	234.753	11.291
conclusion	9	224.698	22.297
total	45	240.818	30.427

Table J2 Result of One-sample Kolmogorov-Smirnov test of lexical cohesive devices in functional sections

		lexical cohesion
N		45
normal parameters[a,b]	mean	240.818
	std. deviation	30.427
most extreme differences	absolute	.123
	positive	.123
	negative	-.103

test statistic	.123
asymp. sig. (2-tailed)	.087[c]

a. Test distribution is normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

Table J3 Result of Levene's test of lexical cohesive devices in functional sections

test of homogeneity of variances			
lexical cohesion			
Levene Statistic	df1	df2	Sig.
5.427	4	40	.001

Table J4 Result of Kruskal-Wallis test of lexical cohesive devices in functional sections

total N	45
test statistic	7.706
degree of freedom	4
asymptotic Sig. (2-sided test)	.103

Appendix K

Table K1 Lexical cohesive pairs in excerpt 1 in MTDC

text	rep.	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.
Chapter 1 introduction									
1.1 Research **Background**									
English Noun **plural morpheme** is one of the few **grammatical morphemes**									
which **Chinese-speaking** learners started to **learn** from a young age.				English – Chinese					
Actually the early **learning** of this is not only true for **Chinese-speaking L2 learners**.	Learn – learning Chinese-speaking		English – L2						
The study done by Dulay and Burt with Spanish-**speaking** children also indicated that	Speaking	Research – study		Chinese – Spanish					Young age – children
plural "-s" came first in the common order of acquisition	Background						English noun plural		L2 learners – acquisition

for certain **structures** in **L2 acquisition**.							morpheme – plural "-s"		
Krashen, Sferlazza, Feldman, and Fathman also did a picture questioning test among **speakers** from different language **background** and found that	Speaking – speakers		Spanish – language	Structures – plural "-s"					
the most **frequent grammatical item** was **plural "-s"**.	Grammatical Plural "-s"		Plural "-s" – grammatical item						
These **studies** show that **plural morpheme** is a **grammatical item**	Plural morpheme Grammatical item					The study… children/a picture … background – these studies			
which has been taught to **learners** for a long time and been used with a high **frequency**.	Learners Frequent – frequency	The most frequent – high frequency						Learn – taught	
Surprisingly, it is such a grammatical **structure**	Grammatical Structures – structure								

which is widely believed to be easy to **acquire**	Acquire – acquisition								
that causes persistent trouble to **Chinese learners** especially in spontaneous **speaking** situations.	Chinese Learners Speakers – speaking							Easy – trouble	
This makes me think that									
the reason may be laid in the linguistic differences between **English** and **Chinese** in terms of **plural** formation and the different ways of psychological processing between **English** and **Chinese speakers**.	English Chinese Plural Speaking – speakers	Structure – formation							Language – linguistic L2 acquisition – psychological processing
Raw	21	3	3	3	0	1	1	2	4
Normalised	112	16	16	16	0	5	5	11	21

Table K2 Lexical cohesive pairs in excerpt 2 in MTDC

text	rep.	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.
1. **Literature** review									
1.1 Definition									
The **literature** has showed that	Literature								
the effect of **teacher feedback** on students' improvement is not explicit and does not achieve its full potential.									
Albeit with the uncertainty of validity of **feedback**, it is undeniable that	Feedback								
teacher feedback is indispensable in **student** learning, especially in a process-approached class.	teacher feedback students – student								Students – teacher Students – class
Therefore, it is **necessary** to look at		Indispensable – necessary							
what **feedback** is and	Feedback								
what should **teacher feedback** focus on.	Teacher feedback								
Kluger & DeNisi define **feedback** interventions as 'actions taken by an external agent to **provide information**	Feedback						Teacher – external agent		

regarding some aspects of one's task **performance'**.									
This argument is further explored by **Ramaprasad** that						Feedback … performance – this argument			
feedback should **provide information** about how to **narrow** the **gap** between **students'** current **performance** and the reference level,	Feedback Provide information Students Performance	Regarding – about							
otherwise **teacher** commentary is less useful in helping **students** improve their skills.	Teacher Students	Feedback – commentary							
The difference between the two interpretations is						Feedback…performance/feedback…skills – the two interpretations			
Ramaprasad emphasizes the **necessity** of the **information** on how to alter the **gap**.	Ramaprasad Necessary – necessity Information Gap					The difference – Ramaprasad …gap			

When **narrowing** the research scope down to writing **feedback**, Keh's regards feedback as 'input from a reader to a writer with the effect of **providing information** to the writer for revision'.	Narrow – narrowing Feedback Provide information – providing information								Literature – research
Raw	20	3	0	0	0	3	1	0	3
Normalised	110	17	0	0	0	17	6	0	17

Table K3 Lexical cohesive pairs in excerpt 3 in MTDC

text	rep.	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.
3. Methodology									
This study aims to **find** out **reasons**									Methodology – study
why **CLT** cannot be **implemented** in English discipline class of Yunnan Nationalities **University**,						Reasons – why… University			
and if possible, **find** some **solutions** and **give** some feasible **suggestions** to solve these problems.	Find					CLT… University – problems			
Till now, there has been a lot scholars and **researchers**									
who **researched** the constraints and challenges of **CLT implementation** in **China**.	Researchers – researched CLT Implemented - implementation		Yunnan – China						
Most of these **researche[r]s** talk about impediments generally in the whole **Chinese context**,	Researched – researchers China – Chinese	Scholars – researchers							

		Constraints – impediments							
thus this **dissertation** can be a test to prove							This study – this dissertation		
whether findings in the **university** match the previous statements.	University – university						Impediments …context – statements		
Since the concept of **CLT** was initially spread to **China** in 1970s,	CLT Chinese – China								
why it cannot be **implemented** in this **context** after over 40 years.	Implementation – implemented Context								
Thus, research questions can be:	Researchers – research	Study – research							
how many **reasons** are due to teachers?	Reasons								Class – teachers
How many are due to students?								Teachers – students	
Do grammar based exams matter a lot?//									

Or is it because lacking authentic language environment?							Research questions – how… environment		
After finishing this **dissertation**, **reasons** for **CLT** fails [failing] to be **implemented** and **solutions** proposed are listed.	Dissertation Reasons CLT Implemented Solutions	Cannot be – Failing to be							
At the same time, some **suggestions** can be **given**.	Suggestions Give – given								
Raw	20	4	1	0	0	4	1	1	2
Normalised	116	23	6	0	0	23	6	6	12

Table K4 Lexical cohesive pairs in excerpt 4 in MTDC

text	rep.	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.
Findings									
Law seminars' influences on spoken English									
According to the information from the interviewed **law** students, it is suggested that//	Law								Seminars – students
the problem-based law seminar has some positive effects on their spoken English, but not remarkably advantageous.	Law seminars – law seminar Spoken English	Influences – effects							
The general consensus they have reached is that									
through **seminar** learning, their confidence to **speak English** has been noticeably promoted,//	Spoken – speak English								Students – learning
and they are able to express themselves clearer and more logical.//						The general consensus – through…logical			
Basically, **law students** can choose four **modules** in their LLM programme	Law students								Learning – modules

									Seminar – LLM programme
but **some modules** are lecture-based, not containing **seminars**	Modules Seminars			Seminar – lecture					LLM programme
and **some modules** have a mixed arrangement.	Some modules								
The **law students** I have **interviewed**, […] have at least two **seminars** in a week,	Law student Interviewed Seminars				Four – two				
and some of them may have three or four times.//					Two – three/four				
Some seminars are 2 **hours** as an independent session,	Some Seminars				Four – 2				
but **some seminars** are preceded by a one-**hour lecture**, only lasting for I hour.	Some seminars Lecture				2 – one				
The **size** of **seminars** varies:	Seminars								
some seminars are quite small, only consisting of 6-8 people,	Some seminars				One – 6-8	Law students – people			Size – small
but the normal **size** is around 20, with the maximum number of 40 **students**.	Size Students				6-8 – 20				Size – the … students
Raw	19	1	0	1	4	2	0	0	6
Normalised	110	6	0	6	23	12	0	0	35

Table K5 Lexical cohesive pairs in excerpt 5 in MTDC

text	rep.	syn.	hype.	hypo.	mer.	sig.	ide.	ant.	col.
Conclusion									
This **concluding section** will first cover relevant **limitations**, **considering** their **impact** on the **conclusions**	Conclusion – concluding					Conclusion – this concluding section			
which can be **drawn** (5.1).									
Then, being aware of these **limitations**, a summary of the results will **follow** (5.2).	Limitations								
Following this, implications will be **drawn**, specifically for pedagogy (5.3).	Follow – following								
The final **section** will open up possibilities for future research (5.4).	Section								Results – research
5.1 **limitations**									
There are several **limitations** that must [be]									

taken into consideration in this **study**,									
which have an **impact** on **conclusions**, and sometimes present options for further investigation.	Concluding – conclusions	Future research – further investigation							
First, there are **limitations** concerning the used **corpora**.	Limitations	Considering – concerning							
The Yale University lectures may not be representative of lectures at **other Universities**									
so the findings may not be applicable in **other** countries and cultures, or even other **Universities** in the USA.		Results – findings							Research – findings The YALE University – the USA
Similarly, the TED talk **corpus** was compiled by the **most viewed talks** and was **limited** to the three main **tags**.	Corpora – corpus Limitations – limited				The used corpora – The Yale University lectures/the				Corpora – compiled

					TED talk corpus				
Talks outside of these **tags**	Talks Tags								
(which **includes** the […] six **most viewed**)	Most viewed	Cover – includes							
and **talks** not selected due to the practical **considerations** of this **study** were not **included** in the **corpus**.	Talks Considering – considerations Study Includes – included Corpus								
Raw	92	23	0	0	6	6	0	0	23
Normalised	151	16	3	6	3	10	7	9	20

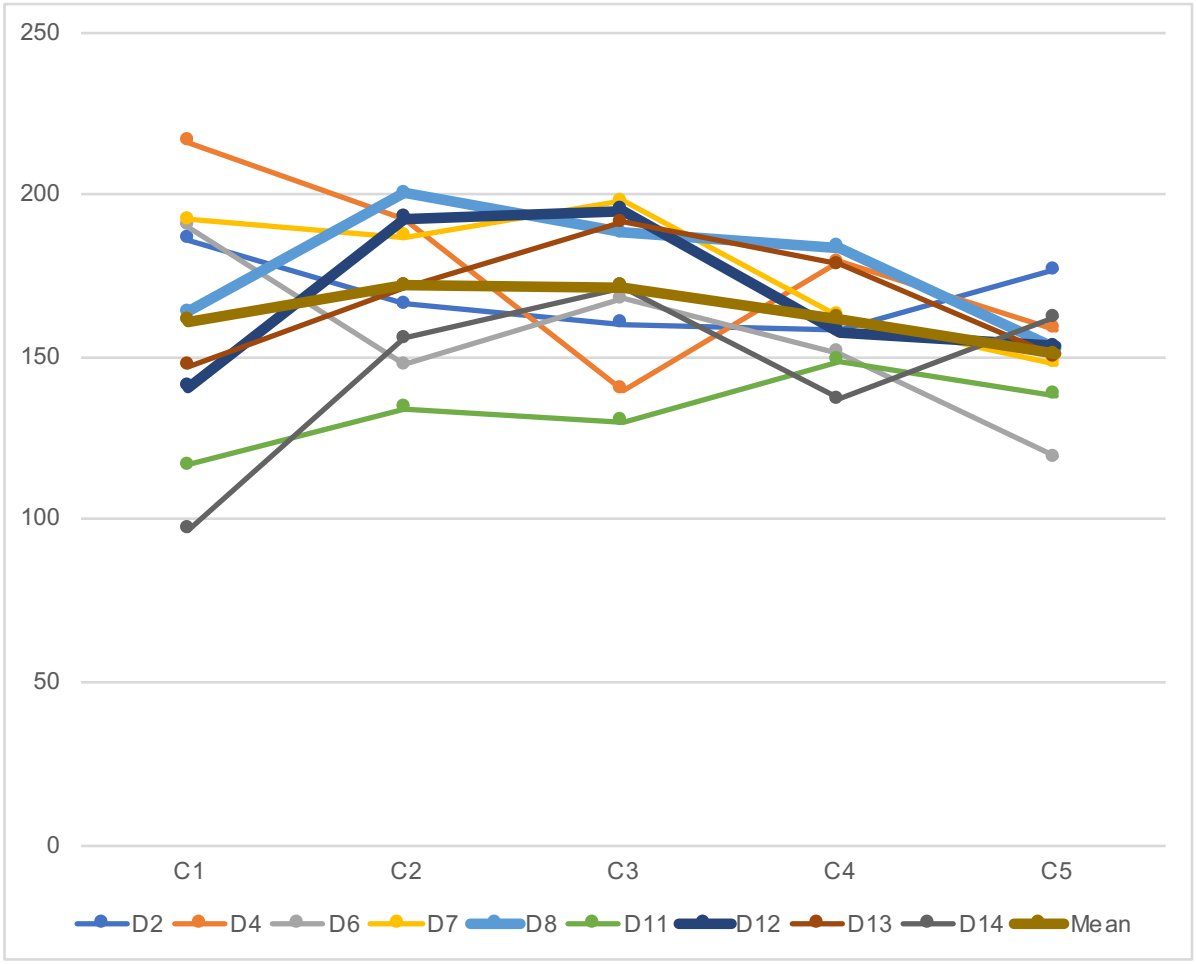
Appendix L

Comparison of repetitive pairs between each dissertation sample and the mean distribution

Table L1 Distribution of repetitive pairs in five sections of each sample and the mean distribution

	C1	C2	C3	C4	C5
D2	186	166	160	159	177
D4	216	193	140	179	158
D6	190	148	168	151	119
D7	192	187	198	163	148
D8	164	200	188	184	152
D11	117	134	130	149	138
D12	141	192	195	158	153
D13	147	171	191	178	150
D14	97	156	172	137	162
Mean	161	172	171	162	151

Figure L2 Distribution of repetitive pairs in five sections of each sample and the mean distribution



Reference

Aisenstadt, E. (1979). Collocability restrictions in dictionaries. *ITL-International Journal of Applied Linguistics*, *45*(1), 71–74.

Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon* (4th ed.). Oxford: Blackwell Publishing Ltd.

Alarcon, J. B., & Morales, K. N. S. (2011). Grammatical cohesion in students' argumentative essay. *Journal of English and Literature*, *2*(5), 114–127.

Allwood, J., Lars-Gunnar, A., & Östen, D. (1977). *Logic in linguistics*. Cambridge & New York: Cambridge University Press.

Badger, R., & White, G. (2000). A process genre approach to teaching writing. *ELT Journal*, *54*(2), 153–160.

Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal*, *47*(1), 56–63.

Bartsch, S. (2004). *Structural and functional properties of collocations in English: A corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Tübingen: Gunter Narr Verlag.

Bednarek, M. (2009). Corpora and discourse: A three-pronged approach to analyzing linguistic data. *Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus*, 19–24. Somerville, MA, USA

Bensemmane, F. (1985). *Lexical cohesion and the determination of text intelligibility by non-native readers of English* (PhD thesis, University College of North Wales, Bangor). Retrieved from: http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.237991

Bouchoux, D. E. (2017). *Aspen Handbook for Legal Writers: A Practical Reference*. New York: Wolters Kluwer.

British Council. (2017). *Country Brief China April 2007*. Retrieved from siem.britishcouncil.org

Bryman, A. (2001). *Social research methods*. Oxford & New York: Oxford University Press.

Burns, A. (1999). *Collaborative action research for English language teachers*. Cambridge: Cambridge University Press.

Butcher, L. S. (1995). *Mother tongue literacy maintenance among the children of recent Chinese immigrants in Brisbane*. Adelaide, South Australia: Helios Art and Book Company.

Cambridge Dictionary. (2019). Retrieved from https://dictionary.cambridge.org/

Cerban, M. (2010). Lexical Cohesion : Aspects of collocation using Halliday and Hasan 's systemic model of cohesion. *Translation Studies: Retrospective and Prospective Views*. Galaţi: Galati University Press, 2010. Available at SSRN: https://ssrn.com/abstract=2799260

Chang, S.-C. (2011). A Contrastive study of Grammar Translation Method and Communicative Approach in teaching English grammar. *English Language Teaching*, *4*(2), 13.

Chanyoo, N. (2018). Cohesive devices and academic writing quality of Thai undergraduate students. *Journal of Language Teaching and Research*, *9*(5), 994–1001.

Charles, M. (2003). "This mystery. . .": A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes*, *2*(4), 313–326.

Chen, X. (2007). Lexical cohesion in Chinese college EFL writing. *CELEA*

Journal, *30*(5), 46–57.
Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.
Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in Education 6th ed. New Delhi, PVT: Vikas publishing House.* (Vol. 55). Abingdon Oxon: Routledge.
Connor, U. (1984). A Study of cohesion and coherence in English as a second language students' writing. *Papers in Linguistics*, 301–316.
Cook, G. (1989). *Discourse*. Oxford: Oxford University Press.
Cowie, A. P. (1981). Measuring nominal scale agreement among many raters. *Applied Linguistics*, *2*(3), 223–235.
Cowie, A. P. (1992). Multiword lexical units and communicative language teaching. In Arnaud P.J.L., Béjoint H. (eds), *Vocabulary and Applied Linguistics* (pp. 1–12). Houndsmills: Macmillan.
Cowie, A. P. (1994). Phraseology. In R. E. Asher (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 3168–3171). Oxford: Pergamon.
Cowie, A. P. (1998). *Phraseology: Theory, analysis and applications*. Oxford: Oxford University Press.
Crystal, D. (2008). *A dictionary of Linguistics and Phonetics*. Oxford: Blackwell Publishing Ltd.
de Beaugrande, R.A. (1980). *Text, discourse, and process :Toward a multidisciplinary science of texts*. London: Longman.
Dörnyei, Z. (2007). *Research methods in Applied Linguistics*. Oxford: Oxford University Press.
Durham University. (2019). Dissertation handbook 2018-2019: MA TESOL & MA Applied Linguistics for TESOL. Durham, the United Kingdom: Durham University.
Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1–4.
Everson, M. E., Chang, K., & Ross, C. (2016). Developing Initial literacy in Chinese. In S. C. Wang & J. K. Peyton (Eds.), *CELIN Briefs Series* (pp. 1–7). New York: Asia Society.
Fillmore, C. J. (1985). Frames and the semantics of understanding 1985. *Quaderni Di Semantica*, *6*(2), 222–254.
Fillmore, C. J., & Baker, C. F. (2001). Frame semantics for text understanding. *Proceedings of WordNet and Other Lexical Resources Workshop*. Pittsburgh: NAACL.
Firth, J. R. (1957). *Papers in Linguistics 1934-1951*. Oxford: Oxford University Press.
Firth, J. R. (1968). *Selected Papers of JR Firth, 1952-59*. London & Harlow: FR Palmer.
Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *75*(5), 378.
Flowerdew, J, & Mahlberg, M. (2009). *Lexical and corpus linguistics*. Amsterdam: John Benjamins Publishing Company.
Flowerdew, John. (2006). Use of signalling nouns in a learner corpus. *International Journal of Corpus Linguistics*, *11*(3), 345–362.
Flowerdew, John, & Forest, R. W. (2015). *Signalling nouns in English: A corpus-based discourse approach*. Cambridge: Cambridge University

Press.
Francis, G. (1986). Anaphoric nouns. *Discourse Analysis Monographs*, *2*.
Francis, G. (1994). Labelling discourse: An aspect of nominal-group lexical cohesion. In M. Coulthard (Ed.), *Advances in Written Text Analysis* (pp. 83–101). London: Routledge.
Gao, M. (2000). *Mandarin Chinese*. Oxford: Oxford University Press.
Ghadessy, M., & Gao, Y. (2001). Small corpora and translation: Comparing thematic organization in two languages. In M. Ghadessy, A. Henry, & R. L. Roseberry (Eds.), *Small Corpus Studies and ELT: Theory and Practice* (pp. 335–359). Amsterdam & Philadelphia: John Benjamins.
Gholami, H., & Alizadeh, F. (2017). A Contrastive study of lexical cohesion in introduction in research articles : Native English and Iranian Applied Linguistics. *Journal of Applied Linguistics and Language Research*, *4*(8), 307–316.
Green, C. F., Christopher, E. R., & Mei, J. L. K. (2000). The incidence and effects on coherence of marked themes in interlanguage texts: A corpus-based enquiry. *English for Specific Purposes*, *19*(2), 99–113.
Greenbaum, S. (1970). *Verb-intensifier collocations in English: An experimental approach*. The Hague: Mouton.
Gries, S. T. (2013). 50-something years of work on collocations: What is or should be next …*. *International Journal of Corpus Linguistics*, 18(*1)*, 137–165.
Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34.
Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In memory of J. R. Firth* (pp. 148–162). London: Longmans Green & Co.
Halliday, M. A. K. (1985). *An introduction to functional grammar*. London: Edward Arnold.
Halliday, M. A. K. (1994). *An introduction to functional grammar (2nd ed.).* London: Edward Arnold.
Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *Introduction to functional grammar (3rd ed.).* London: Hodder Arnold.
Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday' s introduction to Functional Grammar*. London & New York: Routledge.
Haris Fatimah, S. N., & Yunus, M. M. (2014). The use of lexical cohesion among TESL post graduate students, J*ournal of Education and Human Development, 3*(2), 847–869.
Hasan, R. (1984). Coherence and cohesive harmony. In J. Flood (Ed.), *Understanding reading comprehension: Cognition, language, and the structure of prose* (pp. 181–219). Newark, Delaware: International Reading Association.
Hawkins, J. A. (1978). *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Croom Helm.
Hawkins, J. A. (2015). *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. Abingdon Oxon & New York: Routledge.
Helal, F. (2013). Discourse and intercultural academic rhetoric. *Open Journal of Modern Linguistics*, *3*(2), 149–156.

Hinton, P. R., Brownlow, C., McMurray, I., & Cozens, B. (2004). *SPSS Explained*. Hove, East Sussex: Routledge.
Hoek, J., & Scholman, M. C. J. (2017). Evaluating discourse annotation : Some recent insights and new approaches. *Proceedings of the 13th Joint ISO-ACL Workshop on Interoperable Semantic Annotation*. Montpellier, France.
Hoey, M. (1991). *Patterns of lexis in text*. Oxford: Oxford University Press.
Hoey, M. (1994). Patterns of lexis in narrative: A preliminary study. Topics and comments. In S.K. Tanskanen & B. Wårvik (Eds.), *Topics and Comments: Papers from the Discourse Project* (pp. 1–40). Turku: University of Turku.
Hoey, M. (2000). A world beyond collocation: New perspectives on vocabulary teaching. In M. Lewis (Ed.), *Teaching Collocation: Further Developments in the Lexical Approach* (pp. 224–245). Hove: Language Teaching Publications.
Hoffmann, C. R. (2012). *Cohesive profiling: Meaning and interaction in personal weblogs*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Hofstede, G. (n.d.). What about China? Retrieved from https://www.hofstede-insights.com/country/china/
Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind. Intercultural cooperation and its importance for survival*. *Cultures and Organizations* (3rd ed.). Maidenhead, Berkshire: McGraw-Hill Education.
Howarth, P. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making*. Tübingen: Niemeyer.
Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 161–186). Oxford: Oxford University Press.
Hu, G. (2002). English language teaching in the People' s Republic of China. In R. Silver, G. Hu, & M. Lino (Eds.), *English language education in China, Japan, and Singapore* (pp. 1–77). Singapore: National Institute Education, Nanyang Technology University.
Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
Hunston, S., & Francis, G. (1999). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
Hyland, K. (2018). *Metadiscourse: Exploring interaction in writing* (2nd ed.). London: Bloomsbury Publishing.
Ivanic, R. (1991). Nouns in search of a context: A study of nouns with both open- and closed-System characteristics. *International Review of Applied Linguistics in Language Teaching*, *29*(2), 93–114.
Jiang, F., & Hyland, K. (2016). Nouns and academic interactions: A neglected feature of metadiscourse. *Applied Linguistics*, 1-25.
Jiang, F., & Hyland, K. (2017). Metadiscursive nouns: Interaction and cohesion in abstract moves. *English for Specific Purposes*, *46*, 1–14.
Jie, Y., & Keong, Y. C. (2014). College English teaching methodology and language planning : A pilot study in Hefei, China, *Procedia - Social and Behavioral Sciences*, *118*, 495–502.
Jin, L., & Cortazzi, M. (2002). English language teaching in China : A bridge to the future. *Asia Pacific Journal of Education*, *22*(2), 53–64.
Jin, L., & Cortazzi, M. (2006). Changing practices in Chinese cultures of

learning. *Language, Culture and Curriculum*, *19*(1), 5–20.
Jin, W. (2000, September 15-16). *A quantitative study of cohesion in Chinese graduate students' writing : Variations across genres and proficiency levels*. Paper presented at the Symposium on Second Language Writing at Purdue University, West Lafayette, Indiana. Retrieved from https://eric.ed.gov/?id=ED452726
Johnson, M. (2017). Improving cohesion in L2 writing: A three-strand approach to building lexical cohesion. *English Teaching Forum*, *55*(4), 2–13.
Jordan, M.P. (1984). *Fundamentals of technical description*. Malabar Fla.: Krieger Publishing Company.
Jordan, M.P. (1982). The Thread of Continuity in Functional Writing. *Journal of Business Communication*, *19*(4), 5–22.
Kaplan, R. B. (1966). Cultural thought patterns in intercultural education. *Language Learning*, *16*, 1–20.
Keevallik, L. (2009). Collaborating towards coherence: Lexical cohesion in English discourse. Sanna-Kaisa Tanskanen, John Benjamins, Amsterdam/Philadelphia, 2006, 192 pp., $158. *Journal of Pragmatics*, *41*(5), 1071–1073.
Kennedy, G. (2014). *An introduction to Corpus Linguistics*. Abingdon Oxon: Routledge.
Khalil, A. (1989). A study of cohesion and coherence in Arab EFL college students' writing. *System*, *17*(3), 359–371.
Khider, A., & Othman, A. (2019). Investigation of grammatical cohesive devices errors made by Saudi EFL students in written paragraphs : A case study of the University of Tabuk, Saudi Arabia. *European Scientific Journal*, *15*(2), 1–10.
Kjellmer, G. (1994). *A dictionary of English collocations: Based on the Brown corpus*. Oxford: Clarendon Press.
Kramsch, C. (2004). Language, thought and culture. In A. Davies & C. Elder (Eds.), *The handbook of second language acquisition* (pp. 235–265). Oxford & New York: Blackwell Publishing Ltd.
Kurland, D. (2000). How the language really works: The fundamentals of critical reading and effective writing.
Lake, J. (2004). Using "on the contrary": The conceptual problems for EAP students. *ELT Journal*, *58*(2), 137–144.
Langacker, R. W. (1987). *Foundations of cognitive grammar. Vol. I., Theoretical prerequisites*. Standford: Stanford University Press.
Langacker, R. W. (1991). *Foundations of cognitive grammar. Vol Two., Descriptive applications*. Standford: Stanford University Press.
Leech, G. (1981). *Semantics: the study of meaning*. Harmondsworth, Middlesex: Penguin Books Ltd.
Leedham, M. E. (2011). *A corpus-driven study of features of Chinese students' undergraduate writing in UK universities* (PhD thesis, the Open University, the United Kingdom). Retrieved from http://www.open.ac.uk/education-and-languages/main/people/m.e.leedham%5Cnhttp://oro.open.ac.uk/29228/1297/Leedham_Thesis.pdf
Lei, L. (2012). Linking adverbials in academic writing on applied linguistics by Chinese doctoral students. *Journal of English for Academic Purposes*, *11*(3), 267–275.
Léon, J. (2007). Meaning by collocation: The Firthian filiation of corpus

linguistics. In D. A. Kibbee (Ed.), *History of Linguistics 2005: Selected Papers from the Tenth International Conference on History of the Language Sciences, 1-5 September 2005, Urbana-Champaign, Illinois* (pp. 404–415). Amsterdam/Philadelphia: John Benjamins Publishing Company.
Li, C., & Thompson, S. (1981). *Mandarin Chinese: A functional reference grammar*. Berkelley: University of California Press.
Li, H. (2014). The psycholinguistics of Chinese literacy. In *Teaching Chinese literacy in the early years: Psychology, pedagogy and practice*. London: Routledge.
Lindgren, J., & Vogels, J. (2018). Referential cohesion in Swedish preschool children's narratives. *Journal of Pragmatics*, *133*, 45–62.
Littlewood, W. (2007). Communicative and task-based language teaching in East Asian classrooms. *Language Teaching*, *40*(03), 243.
Liu, D. (2000). The use of lexical cohesion among TESL post graduate students. *English Teaching Forum*, *38*(1), 28–33.
Liu, L., & Qi, X. (2010). A contrastive study of textual cohesion and coherence errors in Chinese EFL abstract writing in engineering discourse. *Intercultural Communication Studies*, *XIX*(3), 176–187.
Liu, M., & Braine, G. (2005a). Cohesive features in argumentative writing produced by Chinese undergraduates. *System*, *33*(4), 623–636.
Liu, X., & Jiao, S. (2014). A multidimensional comparison of discourse organization in English and Chinese university students' argumentative writing. *International Journal of Applied Linguistics*, *24(1)*, 74-96.
Lyons, J. (1969). *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.
Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
Mahlberg, M. (2005). *English general nouns: A corpus theoretical approach*. John Benjamins Publishing.
Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge,Massachusetts: MIT Press.
Martin, J. R. (1992). *English Text System and structure*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
Martin, J. R., & Rose, D. (2007). *Working with discourse: Meaning beyond the clause*. London: Continuum International Publishing Group.
McCarthy, M. (1988). Some vocabulary patterns in conversation. In R. Carter & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 181–200). London & New York: Longman.
McCarthy, P. M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.
McGee, I. (2009). Traversing the lexical cohesion minefield. *ELT Journal*, *63*(3), 212–220.
McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, *27*(1), 57–86.
McSweeney, B. (2002). Hofstede's model of national cultural differences and their consequences: A triumph of faith – A failure of analysis. *Human Relations*, *55*(1), 89–118.
Men, H. (2015). *Vocabulary Increase and collocation learning: A corpus-based cross-sectional study of Chinese EFL learners* (PhD thesis, Birmingham

City University, Birmingham). Retrieved from http://www.open-access.bcu.ac.uk/3870/1/2015_Men_646025.pdf
Men, H. (2018). *Vocabulary Increase and Collocation Learning: A corpus-based cross-sectional study of Chinese learners of English*. Singapore: Springer.
Mitchell, T. F. (1975). *Principles of Firthian linguistics*. London: Longman.
Mohammed, A. S. (2015). *Conjunctions as cohesive devices in the writings of English as second language learners*. *Procedia - Social and Behavioral Sciences, 208*, 74-81.
Mohan, B. A., & Lo, W. A. (1985). Academic writing and Chinese students : Transfer and developmental factors. *Tesol Quarterly*, *19*(3), 515–534.
Morris, J., & Hirst, G. (1991). Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, *17*(1), 21–48.
Mu, C., & Carrington, S. (2007). An Investigation of Three Chinese Students ' English Writing Strategies. *Teaching English as a Second or Foreign Language*, *11*(1), 1–23. https://doi.org/10.5829/idosi.wasj.2013.22.11.804
Murphy, M. L. (2003). *Semantic Relations and the Lexicon: Antonymy, Synonymy and other Paradigms*. Cambridge: Cambridge University Press.
Nathan, P. B. (2010). A genre-based study of pedagogical business case reports (Unpublished PhD thesis, Birmingham University, Birmingham).
Nesselhauf, N. (2004). What are collocations? In D. Allerton, N. Nesselhauf, & P. Skandera (Eds.), *Phraseological units: Basic concepts and their application* (pp. 1–21). Basel: Schwabe.
Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Amsterdam: Benjamins.
Neuner, J. (1987). Cohesive ties and chains in good and poor freshman essays. *Research in the Teaching of English*, *21*(1), 92–103.
Nga, C. T. H. (2012). *Analysis of cohesive devices in the ESP textbook on accounting at university of labor and social affairs* (Unpublished MA Thesis, University of Languages and International Studies, Hanoi, Vietnam).
Oxford English Dictionary Online. (2019). Retrieved from https://www.oed.com/
Ong, J. (2011). Investigating the use of cohesive devices by Chinese EFL learners. *The Asian EFL Journal Quarterly*, *13*(3), 42–65.
Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS 6th edition*. Maidenhead, Berkshire.
Palmer, F. R. (1990). *Modality and the English modals, 2nd ed.* London & New York: Longman.
Palmer, H. E. (1933). Second interim report on English collocations. Proceedings of *the Tenth Annual Conference of English Teachers Under the Auspices of the Institute for Research in English Teaching*. Tokyo: Institute for Research in English Teaching.
Palmer, H. E. (1938). *A grammar of English words*. London: Longman.
Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam: Benjamins.
Pincas, A. (1982). *Writing in English: Book 1*. London: Macmillan.
Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Harlow: Longman.
Rapp, R. (2002). The computation of word associations : Comparing syntagmatic and paradigmatic approaches. *Proceedings of COLING 2002: The 19th international conference on Computational linguistics* (pp. 1–7).

Taipei: Association for Computational Linguistics.
Riazi, A. (1997). Acquiring disciplinary literacy: A social-cognitive analysis of text production and learning among Iranian graduate students of education. *Journal of Second Language Writing*, *6*(2), 105–137.
Scarcella, R. C. (1984). *Cohesion in the writing development of native and non-native English speakers*. (PhD Thesis, University of Southern California, Los Angeles, CA). Retrieved from http://digitallibrary.usc.edu/cdm/ref/collection/p15799coll36/id/211599
Schmid, H.J. (1997). Constant and ephemeral hypostatization: Thing, problem and other shell nouns. *Proceedings of the 16th International Congress of Linguists*. Paris.
Schmid, H.J. (2000). *English abstract nouns as conceptual shells: From corpus to cognition*. Berlin&New York: de Gruyter.
Schmid, H.J. (2018). Shell nouns in English – a personal roundup. *Caplletra. Revista Internacional de Filologia*, *64*(64), 109-128.
Schmitz, L., & Weber, W. (2014). Are Hofstede's dimensions valid? A test for measurement invariance of Uncertainty Avoidance. *Interculture Journal*, *13*(22), 11–26.
Scott, M. (2010). *WordSmith tools help*. Liverpool: Lexical Analysis Software Liverpool. Retrieved from http://www.lexically.net/downloads/version6/HTML/index.html
SIL International. (2003). Figurative sense. Retrieved from https://glossary.sil.org/term/figurative-sense
Sinclair, J. (1966). Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In memory of J. R. Firth* (pp. 410–430). London: Longman.
Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
Sinclair, J. (1993). Written discourse structure. In John M. Sinclair, M. Hoey, & G. Fox (Eds.), *Techniques of description: Spoken and written discourse*. New York: Routledge.
Sinclair, J., Jones, S., & Daley, R. (2004). *English collocation studies: The OSTI report*. London & New York: Continuum.
Sinclair, J. (1988). Sense and structure in lexis. In Benson, J. D., Cummings, M. J., & Greaves, W. S. (Eds.), *Linguistics in a systemic perspective* (pp. 73–97). Amsterdam: Benjamins.
Sinicrope, C. (2007). Revisiting cohesive devices in academic L2 English writing : What do successful writers use ? *Journal of Language Teaching and. Research*, *3*(6), 2–25.
Stern, H. H. (1983). *Fundamental concepts of language teaching*. Oxford: Oxford University Press.
Stotsky, S. (1983). Types of lexical cohesion in expository writing: Implications for developing the vocabulary of academic discourse. *College Composition and Communication*, *34*(4), 430–446.
Susan, J., & Sinclair, J. (1974). English lexical collocations: A study in computational linguistics. *Cahiers de Lexicologie*, *24*, 15–61.
Tanskanen, S. (2006). *Collaborating towards coherence*. Amsterdam: John Benjamins Publishing Company.
Tarawhiti, N. W. (2016). *The effect of an explicit strategy instructional approach for the treatment of L2 writing difficulties* (PhD Thesis, Auckland University

of Technology, Auckland, Newzeeland). Retrieved from https://openrepository.aut.ac.nz/handle/10292/10576
The British National Corpus. (2007). Distributed by Bodleian Libraries, University of Oxford. Retrieved from http://www.natcorp.ox.ac.uk/
Tickoo, A. (2002). On the use of "then'/'after that" in the marking of chronological order: Insights from Vietnamese and Chinese learners of ESL. *System*, *30*(1), 107–124.
Todd, R. W., Khongput, S., & Darasawang, P. (2007). Coherence, cohesion and comments on students' academic essays. *Assessing Writing*, *12*, 10–25.
Tuggy, D. (1996). The thing is is that people talk that way. The question is is Why?. In E. H. Casad (Ed.), *Cognitive linguistics in the redwoods. The expansion of a new paradigm in linguistics* (pp. 713–752). Berlin & New York: Mouton de Gruyter.
Vendler, Z. (1967). *Linguistics in philosophy*. Ithaca, NY: Cornell University Press.
Vendler, Z. (1968). *Adjectives and nominalizations*. The Hague: Mouton.
Vyncke, M. (2012). *The Concept and practice of critical thinking in academic writing: An investigation of international students' perceptions and writing experiences* (MA Thesis, King's College London, London). Retrieved from https://englishagenda.britishcouncil.org/sites/default/files/filefield_paths/m_vyncke_0_1.pdf
Wang, J., & Zhang, Y. (2019). Lexical cohesion in research articles. *Linguistics and Literature Studies*, *7*(1), 1–12.
Wang, W., & Wen, Q. (2002). L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing*, *11*(3), 225–246.
Wenden, A. L. (1991). Metacognitive strategies in L2 Writing: A case for task knowledge. In J. E. Alatis (Ed.), *Georgetown University Round Table on Languages and Linguistics 1991* (pp. 302–322). Washington, DC: Georgetown University Press.
William H. Hannon Library. (2019). Cohesion: Does my paragraph flow? Retrieved from http://libguides.lmu.edu/c.php?g=324079&p=2174120
Wingate, U. (2012). "Argument!" helping students understand what essay writing is about. *Journal of English for Academic Purposes*, *11*(2), 145–154.
Winter, E. (1992). The notion of unspecific as one way of analysing the information of a fund-raising letter. In *Discourse Description: Diverse linguistic analyses of a fund-raising text* (pp. 131–170). Amsterdam: John Benjamins.
Wu, S. (2010). Lexical cohesion in oral English. *Journal of Language Teaching and Research*, *1*(1), 97–101.
Xi, Y. (2010). Cohesion studies in the past 30 years: Development, application and chaos. *Language Society and Culture*, (31), 139–147.
Xiao, R. (2015). Collocation. In *The Cambridge Handbook of English Corpus Linguistics* (pp. 106–124). Cambridge: Cambridge University Press.
Yang, A. W. (1989). Cohesive chains and writing quality. *Word*, *40*, 1–2.
Yang, M., Pan, R., & Chen, Z. (2018). Cohesive features in English as a second language students' argumentative writing – Taking Kunming University as an Example. *International Seminar on Education Research and Social Science, 195,* 380–384.

Yang, W., & Sun, Y. (2012). The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education*, *23*(1), 31–48.
Yu, N. (1993). Chinese as a paratactic language. *El Two Talk*, *1*(1), 1–15.
Zaiontz, C. (2019). Fleiss' Kappa. Retrieved from http://www.real-statistics.com/reliability/fleiss-kappa/
Zamel, V. (1983). Teaching those missing links in writing. *ELT Journal*, *37*(1), 22–29.
Zhang, A. (2004). Use of cohesive ties in relation to the quality of compositions by Chinese college students. *Journal of Cambridge Studies*, *5*(2–3), 78–86.
Zhang, M. (2000). Cohesive features in the expository writing of undergraduates in two Chinese universities. *RELC Journal*, *31*(1), 61–95.
Zhang, Z. (2018). Academic writing difficulty of Chinese students : The cultural issue behind Chinese and British academic writing styles. *Studies in Literature and Language*, *17*(2), 118–124.
Zhou, X. (2007). Application of English cohesion theory in the teaching of writing to Chinese graduate students. *US-China Education Review*, *4*(7), 31–37.