

## **Bundle-driven metadiscourse analysis: Sentence initial bundles in Chinese and New Zealand postgraduates' thesis writing**

**Abstract:** Metadiscourse and lexical bundles are two closely related concepts and both operate as overlapping functional units in texts. Metadiscourse analysis always takes a top-down approach, in which discourse analysts begin from pre-determined metadiscourse items down to the analysed texts. Lexical bundle analysis usually uses a bottom-up approach, in which the analysis begins with bundles, extracted automatically from texts, up to generate metadiscourse items to reach an understanding of the discourse. The bundle-driven bottom-up approach is likely to lead to the discovery of longer metadiscourse units and create new categories, while at the same time allowing for the verification of existing researcher-generated metadiscourse lists. While many researchers have focused on examining metadiscourse in academic writing, few studies have used a bottom up approach beginning with lexical bundles in this way to explore the use of metadiscourse. Moreover, research on sentence initial bundles is rare.

The present study explores the metadiscourse functions of generated four-word sentence initial bundles from the corpora of Chinese L2 and New Zealand L1 masters and PhD theses, and compares bundle distributions between L1 and L2 thesis writing. Except for a few propositional bundles, all the other bundles were identified as metadiscourse bundles and two new categories (introduction bundles and condition bundles) were created in to supplement those in Hyland's (2005a, 2005c) metadiscourse model. In contrast to New Zealand thesis writing, both the Chinese masters and PhD corpora were characterised by the heavy use of code gloss bundles (e.g. *In other words, the*), condition bundles (e.g. *In the case of*) and booster bundles (e.g. *It is obvious that*), and a relatively low use of endophoric bundles (e.g. *The use of the*), introduction bundles (e.g. *There are a number*), attitude bundles (e.g. *It is interesting to*), hedge bundles (e.g. *It is possible that*) and self-mention bundles (e.g. *In this chapter, I*). These findings indicate how productive bundle-driven metadiscourse analysis is in expanding the scope of current metadiscourse studies. It also suggests that L2 students could benefit from having attention drawn to lexical bundles as metadiscourse devices to support their academic writing.

**Keywords:** metadiscourse, lexical bundles, L2 academic writing

Li, L., Franken, M., & Wu, S. (2017). Bundle-driven metadiscourse analysis: Sentence initial bundles in Chinese and New Zealand postgraduates' thesis writing. In C. Hatipoglu, E. Akbas, & Y. Bayyurt (Eds.), *Metadiscourse in written genres: Uncovering textual and interactional aspects of texts* (pp. 251-283). Peter Lang.

## **1. Introduction**

Metadiscourse has become a major focus of academic writing research after Vande Kopple's (1985) first comprehensive classification. Studies either analyse the use of metadiscourse devices as a unified whole (e.g. Ädel, 2006; Cao & Wang, 2009; Crismore, Markkanen, & Steffensen, 1993; Heng & Tan, 2010; Hong & Cao, 2014; Marandi, 2003; Mauranen, 1993) or target a particular aspect of metadiscourse such as certainty markers (including hedges and boosters) (e.g. Burrough-Boenisch, 2005; Hinkel, 2005; Hyland & Milton, 1997; Vassileva, 2001; Yang, 2013), self-mentions (e.g. Fløttum, Kinn, & Dahl, 2006; Harwood, 2005a, 2005b; Hyland, 2001; Kuo, 1999) or directives (e.g. Hyland, 2002, 2005b). Many metadiscourse studies largely rely on pre-determined language items in the existing metadiscourse lists (e.g. Hyland's list) to compare the use of metadiscourse. Few studies have adopted a bundle-driven approach to explore the use of metadiscourse, in spite of the close relationship between metadiscourse and lexical bundles.

The present study intends to fill this gap by exploring the metadiscourse functions of generated four-word sentence initial bundles (i.e. lexical bundles at the beginning of sentences). We used a Chinese L2 and New Zealand L1 masters and PhD thesis corpora to investigate the distributions of bundles across metadiscourse categories. We chose to focus on sentence initial bundles because it is more challenging to start a sentence given that a writer needs to have regard to the sequence of the information that follows, and the reader's expectations (Hinkel, 2004; Williams, 2003).

In this paper, we first discuss the relationship between metadiscourse and lexical bundles and select an analytical framework by comparing five popular metadiscourse models proposed by Vande Kopple (1985), Crismore et al. (1993), Hyland (2005a, 2005c), Mauranen (1993) and Ädel (2006). We then describe the procedures of corpus construction, bundle identification and categorisation. Finally we report the findings in regard to the extension of the existing metadiscourse models and bundle distributions between Chinese and New Zealand corpora, together with the implications for writing pedagogy and suggestions for future research.

## **2. Background of the study**

In this section, we discuss the relationship between metadiscourse and lexical bundles, the two overlapping functional units, and explore the possibility of applying a bundle-driven approach

for metadiscourse analysis. We will also compare five popular metadiscourse models to choose one for our study.

## **2.1 The relationship between metadiscourse and lexical bundles**

Metadiscourse has been used as an umbrella term in discourse studies since the 1980s, and includes a great number of discourse devices that writers use to structure their texts, to express their stance, and to engage their readers (Hyland, 2005a). Lexical bundles are products of corpus linguistics, defined as recurrent multiword combinations identified empirically on the frequency of co-occurrence and distribution across texts (Biber, Johansson, Leech, Conrad, & Finegan, 1999). Metadiscourse and lexical bundles are closely related concepts. Both are functional units that exist within texts, and they overlap. Metadiscourse devices are non-propositional and most of lexical bundles excluding content-based ones (which are unlikely to appear in the extracted bundle lists if the corpus contains texts with a wide range of topics) are also non-propositional expressions. The analysis of metadiscourse, as Ädel and Mauranen (2010) argue, often extends beyond pre-determined small search terms and covers larger chunks. Therefore, it is possible to apply a bundle-driven approach to metadiscourse analysis.

The use of lexical bundles and the application of a corpus-driven approach has been recommended by Granger (2014) as “an efficient way of accessing the longer stretches of discourse which are often used to express metadiscourse” (p.59). Metadiscourse analysis takes a top-down approach, in which discourse analysts begin from pre-determined metadiscourse items down to the analysed texts. Lexical bundle analysis usually uses a bottom-up approach, in which the analysis begins with bundles, extracted automatically from texts, up to generate metadiscourse items to reach an understanding of the discourse. This bundle-driven bottom-up approach verifies existing researcher-generated metadiscourse lists, and is likely to lead to the discovery of multiword metadiscourse devices and create new categories.

At the same time, the application of metadiscourse models in bundle research allows writers to access lexical bundles as devices for interpersonal communication. Lexical bundles, as high-frequency and widely distributed linguistic resources, deserve special attention in pedagogy. However, the two extensively-used taxonomies of lexical bundles, Biber and his colleagues’ taxonomy (Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2003, 2004) (i.e. referential, discourse and stance bundles) and Hyland’s (2008a) framework (i.e. research-oriented, text-oriented and participant-oriented bundles), have been initially developed for data analysis

rather than writing pedagogy. Therefore, Biber and his colleagues' taxonomy, generalised from both spoken and written data, contains functions that seem to have little relevance to academic writing (e.g. desire bundles). The research-oriented bundles in Hyland's framework were originally developed from Halliday's (1994) ideational function, but these bundles (except for the content-based topic bundles) could possibly be seen to perform metadiscourse functions. For example, the location bundle *at the same time* indicates a transition within a text. The quantification bundle *one of the most* hedges a statement. Procedure and description bundles like *the use of the* and *the structure of the* can be regarded as endophoric bundles, referring to other parts of the text by means of shell nouns such as *use* and *structure*. The use of a metadiscourse model as analytical framework allows writers to approach lexical bundles as metadiscourse devices, which according to Hyland (2005a) facilitates writing in three different ways:

First, it (metadiscourse) helps them (students) to better understand the cognitive demands that texts make on readers and the ways writers can assist them to process information. Second, it provides them with the resources to express a stance towards their statements. Third, it allows them to negotiate this stance and engage in a community-appropriate dialogue with readers. (Hyland, 2005a, p. 178)

## **2.2 Analytic framework**

Over the past few decades, different metadiscourse models have been developed in various studies. The models of Vande Kopple (1985), Crismore et al. (1993), Hyland (2005a, 2005c), Mauranen (1993) and Ädel (2006), are those among the most popular and widely-cited models. These five models were compared in Appendix A in order to choose one for this study. As presented in Appendix A, the five models can be put on a continuum in terms of their coverage from a broad inclusion of both interactive and interactional functions (e.g. Crismore et al., 1993; Hyland, 2005a, 2005c; Vande Kopple, 1985), to interactive plus interactional functions but excluding stance markers (e.g. Ädel, 2006), and to a narrow perspective of merely interactive functions (e.g. Mauranen, 1993). Appendix A also reflects the development of metadiscourse models from the early versions of Vande Kopple (1985) and Crismore et al. (1993) to the recent model of Hyland (2005a, 2005c). This can be seen from the more detailed categorisation of engagement markers. This categorisation, together with Hyland's (2005a) re-examination of Halliday's metafunctions and proposed lists of metadiscourse items, qualifies Hyland's (2005a, 2005c) model as the most comprehensive metadiscourse model.

Hyland later introduced his framework of lexical bundles, which contains three classifications: research-oriented, text-oriented and participant-oriented bundles (2008a). Both Hyland's (2005a, 2005c) metadiscourse model and his (2008a) framework of lexical bundles are based on Halliday's (1994) systematic functional theory, and regard textual, interpersonal and ideational (propositional) functions as three discrete and separate elements of a text<sup>1</sup>. This results in a close correlation between the two models and implies the possibility of application of Hyland's (2005a, 2005c) metadiscourse model in bundle analysis.

Hyland's (2005a, 2005c) metadiscourse model, as one of the most comprehensive models so far and a model closely related to Hyland's (2008a) framework of lexical bundles, was selected for our bundle-driven metadiscourse analysis. Hyland's (2005a, 2005c) model includes both interactive and interactional resources. According to Hyland (2004), interactive resources "allow the writer to manage the information flow to explicitly establish his or her preferred interpretations", and interactional resources "focus on the participants of the interaction and seek to display the writer's persona and a tenor consistent with the norms of the disciplinary community" (p.129). Interactive resources are currently composed of transitions, frame markers, endophoric markers, evidentials and code glosses; and interactional resources consist of stance and engagement markers. Stance covers four elements: hedges, boosters, attitude markers and self-mentions. Engagement comprises five elements: directives, shared knowledge, questions, reader pronouns and personal asides. Details are presented in Table 1.

---

<sup>1</sup> Halliday originally suggests these three functions are realised simultaneously during writing. Hyland, K. (2005a) has highlighted this misunderstanding in his recent work and borrowed two terms *interactive* and *interactional* from Thompson, G., & Thetela, P. (1995) to replace the original terms of *textual* and *interpersonal* in Halliday's theory for his metadiscourse model.

**Table 1. Hyland's (2005a, 2005c) interpersonal model of metadiscourse**

Category	Function	Examples
<b>Interactive</b>	<b>Help to guide the reader through the text</b>	<b>Resources</b>
Transitions	express relations between main clauses	<i>in addition; but; thus; and finally; to conclude; my purpose is</i>
Frame markers	refer to discourse acts, sequences or stages	<i>noted above; see Fig; in section 2</i>
Endophoric markers	refer to information in other parts of the text	<i>according to X; Z states namely; e.g.; such as; in other words</i>
Evidentials	refers to information from other texts	
Code glosses	elaborate propositional meanings	
<b>Interactional</b>	<b>Involve the reader in the text</b>	<b>Resources</b>
Hedges	withhold commitment and open dialogue	<i>might; perhaps; possible; about</i>
Boosters	emphasize certainty or close dialogue	<i>in fact; definitely; it is clear that</i>
Attitude markers	express writer's attitude to proposition	<i>unfortunately; I agree; surprisingly</i>
Self-mentions	explicit reference to author(s)	<i>I; we; my; me; our</i>
Directives	instruct the reader to perform an action or to see things in a way determined by the writer	<i>note; should; important</i>
Shared knowledge	position readers within apparently naturalized boundaries of disciplinary understandings	<i>we know</i>
Reader Pronouns	pronouns and possessive adjectives referring to the readers	<i>you; your</i>
Questions	the main strategy of dialogic involvement	
Personal asides	allow writers to address readers directly by briefly interrupting the argument to offer a comment on what has been said	

### 3. Methodology

In this section, we describe the procedure for corpus building, setting criteria of bundle identification, and the classifying the extracted bundles. FLAX (<http://flax.nzdl.org>), a self-access language learning and analysis system, documented in Wu (2010), Wu, Franken and Witten (2009, 2010), and Wu, Witten, and Franken (2010) was used to extract sentence initial bundles.

#### 3.1 Corpus construction

Four corpora were built for this study: a Chinese masters thesis corpus, a New Zealand masters thesis corpus, a Chinese PhD thesis corpus and a New Zealand PhD thesis corpus. These four corpora contain theses submitted from 2000 to 2013 in the discipline of general and applied

linguistics. For New Zealand theses, those written by non-native English authors were removed on the basis of the author names and thesis titles. This was not altogether a satisfactory approach but one that was practical and convenient. Only the body of the theses were included; the title page, abstract, acknowledgements, table of contents, lists of tables and figures, references and appendices were excluded from the corpora.

Table 2 provides information on each of these corpora. As can be seen from the table, the average length of the theses was different between each corpus and the New Zealand theses contained comparatively more words than the Chinese ones. The differences in length are likely to have affected the number of certain types of bundles to some extent. For example, the shorter length may raise the number of frame bundles of the same running words, as they are used to signal the boundaries of the arguments (e.g. *The thesis consists of, In this chapter, I, In this section, we*), as they label the stages of texts (e.g. *To sum up, the, In a word, the*) and as they describe text-internal sequences (e.g. *The first of these, This is followed by, First of all, the, Last but not least,*).

**Table 2. Corpus collection**

	CH MA	CH PhD	NZ MA	NZ PhD
<b>Universities</b>	74	12	5	5
<b>Theses</b>	200	67	60	46
<b>Words</b>	c.3,300,000	c.3,800,000	c.2,000,000	c.3,800,000
<b>Length</b>	16,504 words	57,232 words	34,000 words	82,609 words

### 3.2 Bundle identification

The key criteria for generating lexical bundles are the length of word combinations, the frequency threshold and the breadth of distribution (Chen & Baker, 2010). As in most previous studies, four-word bundles were investigated as target bundles because four-word bundles incorporate shorter bundles (e.g. three-word bundles) (Cortes, 2004; Hyland, 2008b) and at the same time occur more frequently with less variation than longer ones. Four-word bundles are sufficient to present productive grammatical structures and tend to be more focused on single instead of multiple functions than longer bundles. In the literature, the frequency threshold usually ranges between 10-40 times per million words and the distribution threshold is at least 3-5 texts (e.g. Ädel & Erman, 2012; Chen & Baker, 2010; Cortes, 2002, 2004, 2013; Hyland, 2008a, 2008b; Wei & Lei, 2011). In this study, as a result of the distinction between sentence initial and non-initial bundles, the less conservative threshold was used against the size of the

corpora and the occurrence of the sentence initial bundles: the cut-off frequency is 5 times per million words and the distribution is at least 5 texts.

As with other studies, altogether 15 content-based bundles (including topic-specific bundles and bundles containing chapter titles, method names and proper names) were removed from the retrieved bundle lists, which resulted in 207 sentence initial bundles. Considering the four corpora were of different sizes, the final frequencies were normalised to 1,000,000 words to conduct a reliable comparison.

### 3.3 Bundle classification

Hyland's (2005a, 2005c) interactive and interactional model of metadiscourse was adapted in this study. It is important to note that a small proportion of sentence initial bundles (i.e. 9%) were multi-functional, acting as both interactive and interactional devices. For example, *Therefore, it is necessary* and *However, it is important* functioned as transition markers and attitude markers. *In this chapter, we* and *In this section, I* performed the functions of frame markers and self-mention devices. *It can be seen* and *As can be seen* served as endophoric markers and directives. *The fact that the* acted as endophoric marker and booster. These bundles were allocated to both categories and each category will be calculated respectively. This categorisation will inflate the total frequencies of both interactive and interactional bundles in terms of type and token; however, it will not affect the comparisons between the four thesis corpora as the categorisation is consistent across the four corpora.

Hyland (2005a) distinguishes text-internal from text-external references and in his view metadiscourse refers only to internal relations of the discourse. Four *study* bundles in this research were identified with both internal and external functions (*In the present/current study, In this study, the* and *The present study is*). For example, *In the present study*, referred to the overall thesis as an internal reference (1); at the same time, this bundles referred to the real research experience as an external reference (2). This type of bundle was classified as other in this study because of the ambiguous functions.

(1) *In the present study, we will study Chinese learners' verb/noun collocating patterns and draw the similarities and difference between the native speakers and Chinese learners with respect to collocation and find out to what extent they have acquired the target language English. (CH MA)*



(2) *In the present study, the combined taxonomy proposed by James (1998) is employed to describe and categorize cc4 errors and some modifications are made in order to deal with cc4 errors properly. (CH MA)*

The statistical software Minitab 17 was used to describe the distributions of the interactive and interactional bundles, and the Chi-square goodness-of-fit test was conducted to measure the differences between the bundle distributions across the four corpora.

## 4. Findings

In this section, we report two major findings. First, this study extends the range of metadiscourse devices to four-word units and adds two new categories to Hyland's (2005a, 2005c) model. Second, this study also presents the results of quantitative analysis of the interactive and interactional bundle distributions between the Chinese and New Zealand corpora.

### 4.1 Extensions of Hyland's (2005a, 2005c) metadiscourse model

This study extends metadiscourse analysis to the bundle-based four-word units, as presented in Table 3. These four-word bundles represent a number of salient linguistic features which are highlighted in the literature on academic writing but are not included in Hyland's (2005a) metadiscourse list, such as the use of demonstratives (e.g. *The first of **these***), shell nouns (e.g. *The **results** of the*) and anticipatory-*it* clauses (e.g. ***It is important/interesting to***). The introduction of these metadiscourse bundles also provides writers, particularly student writers, with a certain amount of automatized prefabricated language, so that writers can construct their sentences and expand their writing from points of fixation, known as "islands of reliability" (Dechert, 1984, p. 223). According to Dechert (1984), the size of a writer's *island* repertoire partially determines his or her writing competence.

**Table 3. Metadiscourse bundles**

Category	Function	Examples
<b>Interactive</b>	<b>guide the reader through the text</b>	<b>Resources</b>
Transition bundles	highlight internal relations between units of texts	<i>On the other hand</i>
Frame bundles	signal coverage, stages or sequences of texts	<i>The first of these</i>
Endophoric bundles	refer to other parts of text	<i>The results of the</i>
Code gloss bundles	elaborate propositional meanings	<i>In other words, the</i>
Condition bundles	specify the pre-conditions of statements	<i>In the case of</i>
Introduction bundles	introduce new information	<i>There are a number</i>
<b>Interactional</b>	<b>Involve the reader in the text</b>	<b>Resources</b>
Attitude bundles	express writer's subjective evaluation or personal feeling	<i>It is important</i> <i>/interesting to</i>
Hedge bundles	address writer's uncertainty	<i>It is possible that</i>
Booster bundles	imply writer's certainty	<i>It is clear that</i>
Self-mention bundles	explicitly refer to writer	<i>In this section, I</i>
Directive bundles	guide readers throughout arguments	<i>It should be noted</i>
Shared knowledge bundles	indicate mutual understanding	<i>As we all know</i>

As shown in Table 3, due to the nature of lexical bundles and the genre of the corpus data, no interactive bundle was identified as evidential and no interactional bundle was used to indicate personal asides, to pose questions or to embed reader pronouns. However, two new interactive subcategories — condition bundles and introduction bundles — were created. Condition bundles present the pre-conditions for the succeeding arguments, signalling the specific contexts, cases, perspectives, etc. Examples are: *In the case of*, *In terms of the*, *In spite of the*, *With regard to the* and *On the basis of*, as indicated in the following sentences.

- (3) ***In the case of*** the family domain, this means that the bilingual children accommodate their language to the speakers of their family. (NZ MA)
- (4) ***In terms of the*** first process, they suggest that the feeling of belonging was an essential condition for maintaining the continuity of identity between the old and the new meanings and for achieving the sense of connectedness with the local community. (NZ PhD)
- (5) ***In spite of the*** findings reported above, there are needs to design and conduct experiments to detect the effect of each variable and the relationships among them through a strict manipulation of different variables in different tests. (CH PhD)

- (6) ***With regard to the*** use of the test, about one-third of the conference participants (34.3%) have no explicit opinion on the question whether the CET is an effective measurement of the implementation of the CES. (CH PhD)
- (7) ***On the basis of*** the logic semantic relations, the connectors are classified into three types: elaboration, extension and enhancement. (NZ MA)

Introduction bundles refer to the initial parts of existential *there* clauses, either used to introduce the subject matters of the upcoming texts (e.g. *There are a number*, *There appears to be*) or to report the writer's inferences of the research results (e.g. *There was no/a significant*), usually followed by *difference*, *correlation*, *association* and *effect*, as in:

- (8) ***There are a number of*** reasons for the choice of these sites for this research. First, as stated in the section above, I am well known to each of the schools and they feel safe with me gathering research data from them. This aligns with Kaupapa Maori Research principles (discussed below) . . . (NZ PhD)
- (9) ***There appears to be*** no research investigating the relationship between the level of teacher qualification and language outcomes for children. (NZ MA)
- (10) ***There was no significant*** difference between the mean retention scores for the two conditions however the Child-Led teaching condition produced a slightly better level of retention for six of the seven children. (NZ MA)
- (11) ***There was a significant*** correlation between gain scores on the written production immediate posttest scored for pronoun form and performance on the working memory test designed to test processing of information ( $r = .489^*$ ). (NZ PhD)

The bundle-driven metadiscourse categorisation confirms most of the categories of Hyland's (2005a, 2005c) metadiscourse model and extends the model by introducing multiword units as metadiscourse devices and by adding another two categories, namely condition bundles and introduction bundles. As discussed above, Hyland's (2005a, 2005c) model is probably the most inclusive and comprehensive model so far, so the results of this study can be considered as the contribution to the development of current understanding of metadiscourse functions. Appendix B provides a summary of the subcategories of interactive dimension found in the data with the sentence initial bundles taken from the corpora, which are composed of transition bundles, frame bundles, endophoric bundles, code glosses bundles, condition bundles and introduction bundles. Appendix C provides a summary of the subcategories of interactional

dimension found in the data with the sentence initial bundles taken from the corpora, which is comprised of attitude bundles, hedge bundles, booster bundles, self-mention bundles, directive bundles and shared knowledge bundles.

## 4.2 Bundle distributions between Chinese and New Zealand corpora

Table 4 presents the number of interactive and interactional bundles in the four corpora. All four groups of students used more interactive than interactional bundles in terms of both type and token. This is in line with Thompson's (2001) argument "interactional signals are typically less frequent and less overt in academic text" (p. 73).

**Table 4. Number of interactive and interactional bundles**

Category		CH MA	CH PhD	NZ MA	NZ PhD
Interactive	Type	62	45	47	35
	Token	712	573	400	305
Interactional	Type	23	20	19	14
	Token	176	167	161	121

Table 5 illustrates the use of interactive bundles with respect to different groups of students. The two groups of Chinese students used more types of bundles (62 compared to 47, 45 compared to 35) with higher mean tokens (11.47 compared to 8.48, 12.72 compared to 8.71) in contrast to their New Zealand counterparts.

**Table 5. Descriptive statistics: Interactive bundles**

Corpus	Types	Mean tokens	StDev
CH MA	62	11.47	10.73
CH PhD	45	12.72	11.13
NZ MA	47	8.48	4.99
NZ PhD	35	8.71	5.06

The result of the Chi-square goodness-of-fit test showed that the functional distributions of interactive bundles were significantly different between each corpus (P-Value < 0.05). Table 6 shows the percentage in each interactive category. There was considerable variation between the writers. Code gloss bundles and condition bundles were found to be more frequent in the Chinese students' writing, while endophoric bundles and introduction bundles were more

common in the New Zealand students' writing. The use of transition bundles was almost the same between the Chinese and New Zealand students.

**Table 6. Distribution of interactive bundles in each corpus (tokens)**

Category	CH MA	CH PhD	NZ MA	NZ PhD
transition bundles	23%	25%	25%	26%
frame bundles	22%	7%	15%	19%
endophoric bundles	19%	18%	36%	23%
code gloss bundles	13%	21%	11%	7%
condition bundles	22%	29%	9%	16%
introduction bundles	1%	0%	6%	9%
Total	100%	100%	100%	100%

*Note.* The highlighted percentages are the percentages consistently different between the two Chinese and the two New Zealand corpora.

Table 7 describes the distribution of interactional bundles in each postgraduate corpus. Both Chinese corpora consisted of more types of interactional bundles (23 compared to 18, 19 compared to 14); however, the mean tokens of both groups of New Zealand bundles were relatively high (8.61 compared to 7.64; 8.65 compared to 8.46).

**Table 7. Descriptive statistics: Interactional bundles**

Corpus	Types	Mean tokens	StDev
CH MA	23	7.64	2.32
CH PhD	19	8.46	2.77
NZ MA	18	8.61	5.23
NZ PhD	14	8.65	4.38

The result of the Chi-square goodness-of-fit test indicated that the functional distributions of interactional bundles differed significantly between each corpus ( $P\text{-Value} < 0.05$ ). Table 8 presents the percentage of bundles in each interactional category. As can be seen from Table 8, a large proportion of data fell into the stance subset, which included attitude bundles, hedge bundles, booster bundles and self-mention bundles, while a few bundles acted as engagement devices, mainly directive bundles. At the same time, the distributions suggested considerable variation between the writers. The two Chinese corpora were characterised by a heavy use of booster bundles and a relatively low use of attitude bundles, hedge bundles and self-mention bundles.

**Table 8. Distribution of interactional bundles in each corpus (tokens)**

Category		CH MA	CH PhD	NZ MA	NZ PhD
<b>Stance</b>	attitude bundles	18%	9%	<b>44%</b>	<b>25%</b>
	hedge bundles	16%	13%	<b>25%</b>	<b>32%</b>
	Booster bundles	<b>43%</b>	<b>28%</b>	5%	11%
	self-mention bundles	4%	14%	<b>4%</b>	<b>16%</b>
<b>Engagement</b>	directive bundles	13%	36%	21%	16%
	shared-knowledge bundles	6%	0%	0%	0%
<b>Total</b>		100%	100%	100%	100%

*Note.* The highlighted percentages are the percentages consistently different between the two Chinese and the two New Zealand corpora.

## 5. Discussion and implications

In this study, we adopted a bundle-driven approach to analyse metadiscourse, and focused on sentence initial bundles and their metadiscourse functions in a Chinese and a New Zealand theses corpora. This bottom-up approach not only verifies the existing metadiscourse models, but also extends metadiscourse devices to multiword units and enriches our understanding of metadiscourse functions. The application of a metadiscourse model sees lexical bundles as interpersonal communication devices, which supports writers to access and utilise lexical bundles as text organisers, stance indicators or engagement markers.

Our investigation of the use of sentence initial bundles in Chinese student writing further confirmed the importance of teaching lexical bundles in relation to their metadiscourse functions. Chinese L2 and New Zealand L1 postgraduates showed different preferences in their use of interactive and interactional bundles. For interactive bundles, Chinese students tended to place code gloss (e.g. *In other words, the*) and condition bundles (e.g. *In the case of*) at the beginning of their sentences, whereas New Zealand students were more likely to start their sentences with endophoric bundles (e.g. *The results of the*). The lack of endophoric bundles in Chinese student writing perhaps indicates insufficient knowledge of shell nouns because most nouns in these bundles are shell nouns (Cortes, 2013). These nouns are pervasive in academic discourse, and carry little or no meaning, but operate to encapsulate the meaning from the anaphoric or cataphoric contexts, that is, the preceding and succeeding clauses or noun phrases. Aktas and Cortes (2008) found the shell nouns in their study of research articles either served a characterisation function (e.g. *the **problem** of this technique*), a temporary concept-formation

function (e.g. *the same **result***) or a linking function (e.g. *this **fact***). The shell nouns identified in the endophoric bundles were found to perform the same functions in facilitating the writers to semantically characterise and conceptualise their research process and outcomes, and at the same time, connecting ideas as cohesive devices. This is illustrated in the following excerpt (12) from a masters student's thesis:

- (12) *Clarke (1988) conducted a comparative study over five months that compared the written progress of children in writing, in four Grade One classrooms. In two classrooms, the children were encouraged to use invented spelling during process writing, while the children in the other two Grade One classrooms were encouraged to write using conventional spelling. **The results of the** study showed that children participating in each teaching approach wrote more words at the end of the five months than at the beginning.*  
(NZ MA)

For interactional bundles, Chinese students preferred to use booster bundles (e.g. *It is clear that*) to express their certainty, while New Zealand students appeared to be more cautious about making statements deploying more hedge bundles. New Zealand students also used more attitude bundles (e.g. *It is interesting to*) and self-mention bundles (e.g. *In this section, I*) in their writing. The findings of booster and hedge bundles were consistent with previous findings of metadiscourse and lexical bundle studies. For example, non-native writers such as the Cantonese writers in Hyland and Milton (1997), Bulgarian writers in Vassileva (2001), Dutch-writers in Burrough-Boenisch (2005) and Chinese writers in Yang (2013) were found to use comparatively more boosters and fewer hedges than native English writers. Non-native writers were also found to use a narrower range of hedge bundles in the bundle studies of Chen and Baker (2010), Ädel and Erman (2012), and Hyland (2008a).

On the basis of this study, we argue that metadiscourse bundles, bundles categorised in terms of metadiscourse functions, which convey the importance of metadiscourse in academic writing and represent the high frequency of lexical bundles, deserve special attention in writing pedagogy. Teachers and textbook writers could possibly provide students with a set of carefully selected target bundles (e.g. sentence initial bundles) together with their metadiscourse functions, or teachers could get students to collect useful bundles before they start writing. As mentioned above, these function-based multiword units can work as “islands of reliability” to facilitate students to expand their writing (Dechert, 1984). If teachers and students can have access to corpora and corpus tools, the corpus-based bundle learning approach can be applied

during bundle learning, which would support students to negotiate the use of unfamiliar bundles such as *It is interesting to*, through to searching its content word (*interesting*), its structure (*It + is + predictive adjective + to*), as well as its multiple contexts.

This study investigated metadiscourse by means of lexical bundles. It should be noted that the corpora for this study were built from postgraduate theses in the discipline of general and applied linguistics. We urge some caution in interpreting and generalising the findings of this study to a broader context, however. The sentence initial bundles analysed in this study only consist of a small proportion of lexical bundles or metadiscourse devices, the ones around four-word length, occurring at the beginning of sentences, and with high frequency. The four-word length bundle identification criterion provides learners with useful four-word metadiscourse bundles. However, this approach ignores other salient metadiscourse items such as individual words (e.g. *also*, *surprisingly*) or shorter word combinations (e.g. *defined as*, *tend to*) as in Hyland's (2005a) list of metadiscourse items, or longer word sequences (e.g. *the purpose of this study is to*, *to determine the effects of*) as in Cortes's (2013) lexical bundle study. The focus on sentence initial bundles leaves out all the bundles occurring at the other parts of sentences. Non-initial bundles perform functions as important as initial ones. For example, *in the context of*, *as well as the* and *more likely to be* act respectively as an endophoric bundle, a transition bundle and a hedge bundle in the following extracts (13-15). The study of these bundles complements the findings of this study and is equally important to extend learners' metadiscourse bundle knowledge.

- (13) *The first of these added elements was to analyse miscommunication and problematic talk **in the context of** a discursive community of practice framework in order to strengthen the sensitivity of the analysis to contextual and situational factors. (NZ PhD)*
- (14) *Next the benefits of bilingualism are discussed as they have been evidenced in the research internationally, **as well as the** implications of that research for Maori medium students and programmes. (NZ MA)*
- (15) *Words that are unknown to learners and are encountered repeatedly in context are **more likely to be** learned (Rodgers & Webb, 2011; Webb & Rodgers, 2009a; Webb, 2008). (NZ PhD)*



## 6. Conclusion

This study is a pilot study in bundle-driven metadiscourse analysis. Future research is greatly needed to explore metadiscourse bundles in different genres, with different lengths, and at different positions. The comparisons between learner bundles and native (proficient) writers' bundles of various types are also needed to inform ESL/EFL writing pedagogy.

## References

- Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Amsterdam, Netherlands: John Benjamins Publishing Company.
- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92. doi:10.1016/j.esp.2011.08.004
- Ädel, A., & Mauranen, A. (2010). Metadiscourse: Diverse and divided perspectives. *Nordic Journal of English Studies*, 9(2), 1-11.
- Aktas, R. N., & Cortes, V. (2008). Shell nouns as cohesive devices in published and ESL student writing. *Journal of English for Academic Purposes*, 7(1), 3-14. doi:10.1016/j.jeap.2008.02.002
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286. doi:10.1016/j.esp.2006.08.003
- Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: an initial taxonomy. In G. N. Leech, T. McEnery, A. Wilson & P. Rayson (Eds.), *Corpus linguistics by the lune*. New York, NY: Peter Lang.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405. doi:10.1093/applin/25.3.371
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London, United Kingdom: Longman.
- Burrough-Boenisch, J. (2005). NS and NNS scientists' amendments of Dutch scientific English and their impact on hedging. *English for Specific Purposes*, 24(1), 25-39. doi:10.1016/j.esp.2003.09.004
- Cao, F., & Wang, X. (2009). 中美大学生英语议论文中的元话语比较研究 [Comparative analysis of the metadiscourse devices in Sino-American college students' English argumentative compositions]. *外语学刊 [Foreign Language Research]*, 5, 97-100.
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.
- Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds.), *Using corpora to explore linguistic variation* (pp. 131-145). Amsterdam, Netherlands: John Benjamins Publishing Company.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423. doi:10.1016/j.esp.2003.12.001
- Cortes, V. (2013). *The purpose of this study is to*: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, 12(1), 33-43. doi:10.1016/j.jeap.2012.11.002

- Crismore, A., Markkanen, R., & Steffensen, M. S. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communication*, 10(1), 39-71.
- Dechert, H. W. (1984). Second language production: Six hypotheses. In H. W. Dechert, D. Mohle & M. Raupach (Eds.), *Second language productions* (pp. 211-230). Tübingen, Germany: Gunter Narr Verlag.
- Fløttum, K., Kinn, T., & Dahl, T. (2006). "We now report on . . ." versus "Let us now see how . . .": Author roles and interaction with readers in research articles. In K. Hyland & M. Bondi (Eds.), *Academic discourse across disciplines* (pp. 203-224). Bern, Switzerland: Peter Lang.
- Granger, S. (2014). A lexical bundle approach to comparing languages: Stems in English and French. *Languages in Contrast*, 14(1), 58-72. doi:10.1075/lic.14.1.04gra
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London, United Kingdom: Edward Arnold.
- Harwood, N. (2005a). 'Nowhere has anyone attempted . . . In this article I aim to do just that': A corpus-based study of self-promotional *I* and *we* in academic writing across four disciplines. *Journal of Pragmatics*, 37(8), 1207-1231. doi:10.1016/j.pragma.2005.01.012
- Harwood, N. (2005b). 'We do not seem to have a theory . . . The theory I present here attempts to fill this gap': Inclusive and exclusive pronouns in academic writing. *Applied Linguistics*, 26(3), 343-375. doi:10.1093/applin/ami012
- Heng, C. S., & Tan, H. (2010). Extracting and comparing the intricacies of metadiscourse of two written persuasive corpora. *International Journal of Education and Development using Information and Communication Technology*, 6(3), 124-146.
- Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hinkel, E. (2005). Hesging, inflating, and persuading in L2 academic writing. *Applied Language Learning*, 15(1 & 2), 29-53.
- Hong, H., & Cao, F. (2014). Interactional metadiscourse in young EFL learner writing: A corpus-based study. *International Journal of Corpus Linguistics*, 19(2), 201-224. doi:10.1075/ijcl.19.2.03hon
- Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, 20(3), 207-226. doi:10.1016/S0889-4906(00)00012-0
- Hyland, K. (2002). Directives: Argument and engagement in academic writing. *Applied Linguistics*, 23(2), 215-239. doi:10.1093/applin/23.2.215
- Hyland, K. (2004). *Disciplinary discourse: Social interactions in academic writing*. Ann Arbor, MI: University of Michigan Press.
- Hyland, K. (2005a). *Metadiscourse: Exploring interaction in writing*. London, United Kingdom: Continuum.
- Hyland, K. (2005b). Representing readers in writing: Student and expert practices. *Linguistics and Education*, 16(4), 363-377. doi:10.1016/j.linged.2006.05.002
- Hyland, K. (2005c). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192.
- Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62. doi:10.1111/j.1473-4192.2008.00178.x
- Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21. doi:10.1016/j.esp.2007.06.001

- Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, 6(2), 183-205. doi:10.1016/S1060-3743(97)90033-3
- Kuo, C.-H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes*, 18(2), 121-138. doi:10.1016/S0889-4906(97)00058-6
- Marandi, S. (2003). Metadiscourse in Persian/English master's thesis: A contrastive study. *IJAL*, 6(2), 23-42.
- Mauranen, A. (1993). Contrastive ESP rhetoric: Metatext in Finnish-English economics texts. *English for Specific Purposes*, 12(1), 3-22. doi:10.1016/0889-4906(93)90024-I
- Thompson, G. (2001). Interaction in academic writing: Learning to argue with the reader. *Applied Linguistics*, 22(1), 58-78.
- Thompson, G., & Thetela, P. (1995). The sound of one hand clapping: The management of interaction in written discourse. *TEXT*, 15(1), 103-127.
- Vande Kopple, W. J. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication*, 36(1), 82-93. doi:10.2307/357609
- Vassileva, I. (2001). Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes*, 20(1), 83-102. doi:10.1016/S0889-4906(99)00029-0
- Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELJ Journal*, 42(2), 155-166. doi:10.1177/0033688211407295
- Williams, J. M. (2003). *Style: Ten lessons in clarity and grace* (7th ed.). New York, NY: Addison-Wesley.
- Wu, S. (2010). *Supporting collocation learning* (Doctoral dissertation). University of Waikato, Hamilton, New Zealand. Retrieved from <http://hdl.handle.net/10289/4885>
- Wu, S., Franken, M., & Witten, I. H. (2009). Refining the use of the web (and web search) as a language teaching and learning resource. *Computer Assisted Language Learning*, 22(3), 249-268.
- Wu, S., Franken, M., & Witten, I. H. (2010). Supporting collocation learning with a digital library. *Computer Assisted Language Learning*, 23(1), 87-110.
- Wu, S., Witten, I. H., & Franken, M. (2010). Utilizing lexical data from a Web-based corpus to expand productive collocation knowledge. *ReCALL*, 22(1), 83-102.
- Yang, Y. (2013). Exploring linguistic and cultural variations in the use of hedges in English and Chinese scientific discourse. *Journal of Pragmatics*, 50(1), 23-36. doi:10.1016/j.pragma.2013.01.008

## Appendix A. Summary of the metadiscourse models

Vande Kopple (1985)		Crismore et al. (1993)		Mauranen (1993)	Hyland (2005a, 2005c)		Ädel's (2006)		
Category	Subcategory	Category	Subcategory	Category	Category	Subcategory	Category (personal)	Category (impersonal)	
Textual metadiscourse		Textual metadiscourse		Metatext	Interactive metadiscourse		Metatext	Metatext	
Text connectives	Sequencers	Textual markers	Sequencers		Frame markers			Phoric markers	
	Logical connectives		Logical connectives	Connectors	Transitions				
	Reminders		Remainders	Reviews	Endophoric markers		Reminding		
	Topicalizers		Topicalizers		Code glosses		Focusing, Exemplifying		
	Announcements	Interpretive markers	Announcements	Previews	Endophoric markers				
Code glosses			Code glosses		Defining, Clarifying (writer-reader interaction)		Code glosses		
Illocution markers			Illocution markers	Action markers	Frame markers		Saying, Introducing topic, Concluding, Adding, Arguing, Contextualising	Discourse labels	
Narrators					Evidentials				
					Frame markers				References to the text/code

Li, L., Franken, M., & Wu, S. (2017). Bundle-driven metadiscourse analysis: Sentence initial bundles in Chinese and New Zealand postgraduates' thesis writing. In C. Hatipoglu, E. Akbas, & Y. Bayyurt (Eds.), *Metadiscourse in written genres: Uncovering textual and interactional aspects of texts* (pp. 251-283). Peter Lang.

Vande Kopple (1985)		Crismore et al. (1993)		Mauranen (1993)	Hyland (2005a, 2005c)		Ädel's (2006)	
Category	Subcategory	Category	Subcategory	Category	Category	Subcategory	Category (personal)	Category (impersonal)
<b>Interpersonal metadiscourse</b>		Interpersonal metadiscourse			Interactional metadiscourse			
<b>Validity markers</b>	Hedges	Hedges			Hedges			
	Emphatics	Emphatics			Boosters			
	Attributors	Attributors						
<b>Attitude markers</b>		Attitude markers			Attitude markers			
							Writer-reader interaction	
<b>Commentaries</b>		Commentaries			Engagement markers	Directives	Aligning perspectives, Imagining scenarios, Appealing to the reader	
							Shared knowledge	
							Questions	
							Reader Pronouns	
							Personal asides	Anticipating the reader's reaction, Hypothesising about the reader
					Self-mentions			

Li, L., Franken, M., & Wu, S. (2017). Bundle-driven metadiscourse analysis: Sentence initial bundles in Chinese and New Zealand postgraduates' thesis writing. In C. Hatipoglu, E. Akbas, & Y. Bayyurt (Eds.), *Metadiscourse in written genres: Uncovering textual and interactional aspects of texts* (pp. 251-283). Peter Lang.

## Appendix B: Interactive categories and sentence initial bundles

	CH MA		CH PhD		NZ MA		NZ PhD	
<b>transition bundles</b>	On the other hand,	62	On the other hand,	65	On the other hand,	34	On the other hand,	28
	On the one hand,	17	On the one hand,	27	As a result of	12	In addition to the	14
	As a result, the	15	In addition to the	15	In addition to the	10	As a result of	7
	In addition to the	8	As a result, the	11	However, it is important	7	On the one hand,	6
	Therefore, it is necessary	7	As a result of	5	In addition to this,	7	In contrast to the	6
	However, it is not	7	However, it should be	5	In contrast to the	6	At the same time,	18
	As a result, it	5	At the same time,	18	As a result, the	6		
	So it is necessary	5			At the same time,	12		
	At the same time,	38			At the same time	6		
<b>frame bundles</b>	In the process of	28	To sum up, the	9	At the time of	10	In this chapter I	8
	In this chapter, the	19	In this chapter, we	9	The chapter concludes with	9	The first of these	6
	At the end of	17	In this section, we	7	By the end of	9	In this section I	6
	At the beginning of	13	In this section, I	7	At the end of	9	In this section, I	5
	During the process of	12	In this section, the	7	In this chapter I	7	At the end of	15
	To sum up, the	8			The next chapter will	6	At the time of	7
	Last but not least,	7			In this section the	5	At the beginning of	6
	In this chapter, we	7			This chapter describes the	5	This is followed by	5
	In this section, the	7						

	The thesis consists of	6						
	In a word, the	6						
	First of all, the	6						
	In this part, the	6						
	The first one is	6						
	This thesis consists of	5						
	In the course of	5						
<b>code gloss bundles</b>	That is to say,	51	In other words, the	39	That is to say,	10	In other words, the	11
	In other words, the	20	That is to say,	37	In other words, the	9	This is not to	6
	For example, in the	7	To be more specific,	9	This is because the	7	This is not a	5
	In other words, they	7	To put it another	6	In other words the	6		
	That is to say	6	For example, in the	6	This is not to	6		
	It means that the	5	This means that the	6	This suggests that the	6		
			In other words, it	6				
			In other words, they	5				
			This suggests that the	6				
<b>endophoric bundles</b>	It can be seen	13	Look at the following	11	The results of the	23	As discussed in Chapter	17
	As is shown in	10	The following is a	9	The results of this	13	As can be seen	9
	The following table shows	8	As shown in Table	7	As can be seen	13	The results of the	12
	The results of the	29	As is shown in	7	The purpose of this	12	The results of this	6

	The purpose of this	12	It can be seen	15	The majority of the	11	The analysis of the	6
	The purpose of the	8	As can be seen	14	The aim of the	8	The purpose of this	7
	The result of the	8	The results of the	24	The purpose of the	6	The aim of this	6
	The main purpose of	7	The analysis of the	7	The findings of this	6	The fact that the	7
	The following are some	6	The following is an	6	The results from the	6		
	As shown in Table	5	The following are some	5	As discussed in Chapter	6		
	From the above table,	5			The fact that the	8		
	The following is a	5			The aim of this	6		
	We can see from	5			The limitations of the	5		
	The following is the	5			The findings of the	5		
	As can be seen	5			The use of the	5		
					It can be seen	5		
					The analysis of the	6		
<b>condition bundles</b>	On the basis of	22	In the case of	20	In the case of	14	In the case of	18
	With the development of	21	On the basis of	18	In terms of the	5	In the context of	7
	Based on the above	10	In terms of the	17	For the purpose of	9	In terms of the	6
	With the help of	10	As far as the	14	In spite of the	7	With regard to the	6
	From the perspective of	8	From the perspective of	12			On the basis of	5
	In the light of	8	With regard to the	11			For the purposes of	8



	As far as the	7	With respect to the	8			
	In this way, the	13	In this case, the	8			
	When it comes to	6	In this way, the	16			
	As one of the	6	In the field of	6			
	With regard to the	6	When it comes to	6			
	In view of the	5	For the sake of	8			
	In order to make	16	In this sense, the	17			
	In order to get	8	In spite of the	7			
	In order to find	6					
introduction bundles	There is no doubt	6		There appears to be	8	There are a number	8
				There are a number	8	There was a significant	7
				There was no significant	7	There was no significant	7
						There were no significant	6

## Appendix C: Interactional categories and sentence initial bundles

	CH MA		CH PhD		NZ MA		NZ PhD	
<b>attitude bundles</b>	Therefore, it is necessary	7	It is important to	7	It is important to	26	It is important to	17
	It is important to	7	It is necessary to	7	It is interesting to	13	It is interesting to	8
	It is necessary to	7			It is important that	7	It is difficult to	6
	It is evident that	5			It is difficult to	6		
	So it is necessary	5			However, it is important	7		
					It is interesting that	6		
					It was important to	5		
<b>Hedge bundles</b>	One of the most	8	It seems that the	7	It is possible that	14	It is possible that	20
	It is suggested that	9	It is argued that	6	It may be that	8	It is not clear	7
	The results indicate that	6	This suggests that the	6	There appears to be	8	It may be that	6
	It is hoped that	5	It is hoped that	8	It would appear that	5	It is also possible	6
					It is possible to	5		
					This suggests that the	6		
<b>Booster bundles</b>	It is obvious that	12	It is clear that	10	The fact that the	8	The fact that the	7
	It is clear that	7	It is true that	8			It is clear that	6
	There is no doubt	6	It is obvious that	7				
	It is believed that	7	The results showed that	6				
	It is found that	7	It should be pointed (out)	6				
	The results show that	9	As a matter of	7				
	The results showed that	7						

	The following table shows	8						
	As a matter of	11						
<b>self-mention bundles</b>	In this chapter, we	7	In this chapter, we	9	In this chapter I	7	In this chapter I	8
			In this section, we	7			In this section I	6
			In this section, I	7			In this section, I	5
<b>Directive bundles</b>	It can be seen	13	It should be noted	14	It should be noted	10	It should be noted	11
	We can see from	5	It can be seen	15	As can be seen	13	As can be seen	9
	As can be seen	5	As can be seen	14	It can be seen	5		
			Look at the following	11	It must be noted	5		
			We can see that	5				
<b>shared knowledge bundles</b>	As we all know,	11						