



THE UNIVERSITY OF  
**WAIKATO**  
*Te Whare Wānanga o Waikato*

Research Commons

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

## Research Commons at the University of Waikato

### Copyright Statement:

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

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

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

# **Sentence initial bundles in L2 thesis writing:**

A comparative study of Chinese L2 and New Zealand L1 postgraduates' writing

**A thesis submitted in fulfilment  
of the requirements for the degree**

**of**

**Doctor of Philosophy (PhD)**

**at**

**The University of Waikato**

**by**

**Liang Li**



THE UNIVERSITY OF  
**WAIKATO**  
*Te Whare Wānanga o Waikato*

2016



## Abstract

Multiword combinations perform a crucial role in signifying fluency, accuracy and idiomaticity in academic writing. Lexical bundles are recurrent, but not salient, multi-word combinations, for example, *on the other hand*, *the fact that the*, and *it should be noted*. They are important as they act as discourse frames to relate to new information or as interactional devices to mark the involvement of the writer and the reader. These functions can also be regarded as metadiscoursal functions, represented by metadiscoursal models.

The use of lexical bundles in L2 academic writing has been the focus of a number of recent studies, but few studies distinguish bundles in different sentence positions, investigate bundles from the perspective of metadiscoursal functions, and explore the reasons underlying the bundle choices of L2 writers.

The present study sought to fill these gaps by comparing the use of sentence initial bundles (i.e. bundles at the beginning of sentences) in Chinese L2 and New Zealand L1 thesis writing in the discipline of general and applied linguistics. Four collections were built: a Chinese masters thesis corpus, a New Zealand masters thesis corpus, a Chinese PhD thesis corpus and a New Zealand PhD thesis corpus. In comparing these four corpora, this study provided a detailed picture of the use of sentence initial bundles in Chinese postgraduate writing and an overall picture of variation in bundle use across different postgraduate levels of students in terms of frequency, structure and function. Semi-structured interviews with six Chinese postgraduates were conducted after the text analysis to understand the reasons for Chinese students' bundle choices. The interviews were based on the expressions in participants' original drafts, which were completely or partially overlapped with the sentence initial bundles generated from the corpus data.

Chinese masters and PhD students were found to rely more heavily on sentence initial bundles, particularly interactive bundles. They preferred to start sentences with PP-based bundles, VP-based bundles, and *conjunction + clause fragment* bundles; but were less aware of the importance of NP-based bundles and anticipatory-*it* bundles. With regard to function, both the Chinese PhD and masters

---

corpora were characterised by a heavy use of condition bundles and booster bundles; and a relatively low use of endophoric bundles, attitude bundles, hedge bundles, self-mention bundles and directive bundles of cognitive acts.

In regard to bundle development, both groups of masters students were found to use more bundles than their PhD counterparts. However, the two PhD groups shared more bundles. More research-related NP-based bundles occurred in masters corpora, and more PP-based bundles and anticipatory-*it* bundles appeared in PhD students' writing. A functional analysis showed that both groups of PhD students used more transition bundles, condition bundles, section-level frame bundles and self-mention bundles, but fewer attitude bundles.

Interviews with six Chinese postgraduates revealed possible reasons for Chinese students' bundle selection and use, which included but were not limited to interlingual transfer, classroom learning, noticing in reading, a lack of rhetorical confidence, and misunderstanding of rhetorical conventions. The findings suggest the need to go beyond the teaching of lexical bundles as a list of fixed multiword expressions. Teachers and learners are advised to address the pedagogical implications of bundle studies, and to use corpus-based tools (e.g. FLAX) to approach bundles as lexico-grammatical frames in which slots can be filled with a variety of words.

## Acknowledgements

A PhD journey is often regarded as a kind of compensation. I always asked myself what my PhD compensated for: a strong desire for knowledge, a deep sense of inequality, or a relative lack of self-confidence. Maybe all of these. As an ordinary girl from a normal family, a tight budget for education taught me to cherish all the valuable opportunities for learning. As an adult Chinese woman with multiple family and social roles, the conflicting values within and outside the family during the transition period of China challenged my sense of self-identity. As a young teacher with a high workload and multiple responsibilities, the fear of failure to satisfy students' learning needs always lingered in my mind. All of these triggered the start of my PhD journey.

After a four-year journey, looking back, I feel so fortunate to be one of the few who have been able to enjoy the luxury of the PhD learning and exploring experience. In addition, the invaluable opportunity to be involved in the exciting FLAX project has greatly increased the beauty of my journey. I deeply appreciate all the great support that I have received during this journey. Without all these important opportunities and persons associated with them in my life, my dream would never have come true.

Associate Professor Margaret Franken, my chief supervisor, has been guiding me for more than ten years. She is the person who introduced me to the area of applied linguistics over ten years ago, and who led me to another fascinating area, corpus linguistics during my PhD. It is she who has opened the doors to treasures of knowledge so that I could enjoy the pleasure of collecting.

Dr. Shaoqun Wu is my second supervisor, a computer science expert and lead researcher of the FLAX project. Her wide knowledge in digital libraries and computer assisted language learning expanded my understanding of information technology, and allowed me to trial different corpus-driven approaches and to enjoy the beauty of technology in corpus analysis and language learning.

---

Professor Ian H. Witten, the leader of the FLAX project, always impressed me with his rich experience, generous help and optimistic attitude. I would like to express my deep gratitude to him for his trust, support and guidance, and for providing a space and a position for me. It was such a pleasant and rewarding experience to work in the digital library lab and for the FLAX project, to spend time with my multicultural friends Anupama Krishnan, Katherine Don and John Thompson, and to meet visiting scholars Jennifer Thøegersen and Daniil Mirylenka. Jennifer kindly helped me to identify New Zealand theses written by L1 writers for my corpus building.

I am also deeply impressed by the support I have received from the university. I wish to express special thanks to my subject librarian Alistair Lamb for the enormous amount of time he devoted to helping me to search for thesis data, narrow down research literature, use the reference tool Endnote and format my thesis. My thanks also go to interlibrary service librarian Maria McGuire for her significant work on interloaning over 100 books and articles for me from across the world. Thanks also to the staff in Student Learning, especially Andrea Haines and Dawn Marsh. Andrea was my first individual tutor and her dedicated work on editing my writing advanced my knowledge of academic writing and built my confidence as a L2 writer. Dawn's feedback on writing often echoed the findings of my own PhD work, and her L1 writer's intuition and rich first-hand editing experience promoted my understanding of the data of Chinese postgraduate writing. Besides the individual support, I greatly benefited from a wide range of learning and networking opportunities, which were provided at different levels in the forms of workshops and seminars, such as Doctoral Writing Conversations, Postgraduate Development Workshops, Teaching Development Workshops, FEDU Doctoral Support Sessions and Workshops, WMIER Seminars, Research Bites, Writing Breakfasts and Research Group Meetings. All the best memories thereof are also shared with my peer students Jinah Lee, Yi Wang, Susan Pudín-Baduk, Nhue Nguyen, Lula Mengesha and Ignasia Mligo.

Going beyond my own university, my work was highly promoted by a series of conversations with international scholars in the field during international conferences or through emails. My sincere thanks to Professor Ken Hyland,

---

Professor Fang Xu, Dr. Lynne Flowerdew and Professor Paul Baker. Also I am grateful to Richard Lawrence, Jono Ryan, Jenny Field and Farrah Jin, the staff at the Centre for Languages at WINTEC; Shuping Wang, Yunzhi Shi, Weiping Ren, Xikui Zhu, Ying Guo and Likun Cai, the lecturers in China, who generously offered me suggestions from their years of English teaching experience. I am also grateful for the development of distance education, especially MOOC (Massive Open Online Course) developments. The distance learning opportunities granted me the freedom to build up my own expertise from courses such as *Corpus linguistics: Method, analysis, interpretation* offered by Lancaster University, *Writing in the sciences* provided by Stanford University and *Research commercialisation* delivered by e-Grad School (Australia).

I am greatly indebted to my beloved family. My grandparents cared and supported me for so many years. My parents downloaded Chinese postgraduate theses for my corpus building from China. My sister Xin Li is my best and most constructively critical friend, and her thought-provoking questions often deepened my understanding. My husband Kun Cao supported my study both emotionally and financially since our engagement in 2003. My beloved daughter Ziqi Cao gave me the courage to be a mother. Also thanks to the youngest one in the family, our little unborn baby Ziming Cao, who stayed healthy and behaved well until the submission of my PhD.

Last, but certainly not least, I would like to acknowledge my anonymous participants and the interesting stories they shared with me, the enormous help from my personal writing and analysing assistant FLAX (a free online language learning system), and the financial support granted by the University of Waikato, which included a Doctoral Merit Award and Doctoral Scholarship.

It is nearly the end of my PhD journey, but not the end of my dream. Thanks to all professional guidance, generous support, enjoyable company and even brief encounters during my PhD journey, which allowed me to dive into the knowledge of my subject area, to establish a new identity, and to build confidence in my own abilities. *Ko Te Tangata*, for the people, is the motto of the University of Waikato, which will be mine for the rest of my life.





---

## Contents

<b>Abstract</b> .....	<b>i</b>
<b>Acknowledgements</b> .....	<b>iii</b>
<b>Contents</b> .....	<b>vii</b>
<b>Tables</b> .....	<b>xi</b>
<b>Figures</b> .....	<b>xv</b>
<b>Chapter 1 Introduction</b> .....	<b>1</b>
1.1 Motivation for the study.....	1
1.1.1 Nature of lexical bundles .....	2
1.1.2 Limited resources of lexical bundles .....	3
1.1.3 Lack of connection between research and pedagogy .....	5
1.2 Objectives.....	6
1.3 Contributions.....	7
1.3.1 Potential contributions to theory .....	7
1.3.2 Potential contributions to methodology .....	8
1.3.3 Potential contributions to pedagogy.....	8
1.4 Thesis outline .....	9
<b>Chapter 2 Corpus linguistics and academic discourse analysis</b> .....	<b>13</b>
2.1 Corpus linguistics and corpora.....	13
2.2 Corpus linguistics and word lists .....	15
2.3 Corpus linguistics and academic discourse analysis .....	15
2.3.1 Corpus linguistic research on different languages .....	17
2.3.2 Corpus linguistic research on registers .....	18
2.3.3 Corpus linguistic research on written genres .....	19
2.3.4 Corpus linguistic research on disciplines.....	20
2.4 Corpus linguistics and contrastive interlanguage analysis.....	22
<b>Chapter 3 Lexical bundles</b> .....	<b>25</b>
3.1 The Concept of lexical bundles.....	25
3.2 Lexical bundles and academic writing.....	28
3.3 Studies on lexical bundles .....	30
3.3.1 Frequency-based analysis .....	31
3.3.2 Structural analysis .....	33
3.3.3 Functional analysis.....	36
3.3.4 Possible explanations of L2 student bundle choices.....	43
3.4 Limitations of the existing research .....	44
<b>Chapter 4 Metadiscourse</b> .....	<b>47</b>
4.1 The concept of metadiscourse .....	47
4.2 The relationship between metadiscourse and lexical bundles.....	49

---

4.3 Metadiscourse models .....	50
4.3.1 Vande Kopple's metadiscourse classification .....	51
4.3.2 Crismore, Markkanen and Steffensen's metadiscourse system .....	52
4.3.3 Mauranen's metatext model .....	52
4.3.4 Hyland's metadiscourse model.....	53
4.3.5 Ädel's taxonomy of metadiscourse .....	55
4.3.6 Comparisons between the metadiscourse models .....	58
4.4 Studies on metadiscourse.....	62
4.4.1 Studies on metadiscourse as a whole.....	62
4.4.2 Studies on specific aspects of metadiscourse .....	64
4.4.3 Studies on writer interpretations of metadiscourse use .....	66
4.5 Limitations of the existing research.....	67
<b>Chapter 5 Methodology .....</b>	<b>69</b>
5.1 Corpus-based analysis .....	70
5.1.1 Corpus building .....	70
5.1.2 Bundle identification .....	73
5.1.3 Structural categories .....	75
5.1.4 Functional categories.....	76
5.2 Semi-structured interviews .....	79
5.2.1 Background of participants.....	79
5.2.2 Interview data analysis .....	81
<b>Chapter 6 Frequency-based and structural analysis .....</b>	<b>83</b>
6.1 Frequency-based analysis .....	83
6.2 Structural analysis.....	86
6.2.1 NP-based bundles .....	90
6.2.2 PP-based bundles.....	93
6.2.3 VP-based bundles .....	95
6.2.4 Clause-based bundles .....	97
6.2.5 Other bundles.....	100
6.3 Summary.....	100
6.3.1 Differences between the Chinese and New Zealand writing.....	100
6.3.2 Differences between the masters and PhD writing.....	101
<b>Chapter 7 Interactive functions of the bundles.....</b>	<b>103</b>
7.1 Transition bundles .....	104
7.1.1 Shared transition bundles .....	105
7.1.2 Transition bundles in the Chinese students' writing .....	109
7.2 Frame bundles.....	111
7.2.1 Boundary bundles .....	113
7.2.2 Discourse-label bundles.....	114
7.2.3 Sequence bundles .....	115
7.3 Endophoric bundles .....	118

---

7.3.1 Shared endophoric bundles .....	119
7.3.2 Endophoric bundles in the New Zealand students' writing .....	120
7.3.3 Endophoric bundles in the Chinese students' writing.....	121
7.4 Code gloss bundles.....	122
7.4.1 Shared code gloss bundles .....	123
7.4.2 Code gloss bundles in the New Zealand students' writing .....	125
7.4.3 Code gloss bundles in the Chinese students' writing.....	126
7.5 Condition bundles .....	127
7.5.1 Shared condition bundles .....	128
7.5.2 Condition bundles in the New Zealand students' writing.....	130
7.5.3 Condition bundles in the Chinese students' writing .....	130
7.6 Introduction bundles.....	134
7.6.1 Introduction bundles in the New Zealand students' writing .....	134
7.6.2 Introduction bundles in the Chinese students' writing.....	135
7.7 Summary .....	137
7.7.1 Differences between the Chinese and New Zealand writing .....	137
7.7.2 Differences between the masters and PhD writing .....	138
<b>Chapter 8 Interactional functions of the bundles .....</b>	<b>139</b>
8.1 Attitude bundles .....	140
8.1.1 Shared attitude bundle.....	142
8.1.2 Attitude bundles in the New Zealand students' writing.....	144
8.1.3 Attitude bundles in the Chinese students' writing .....	145
8.2 Hedge bundles .....	148
8.2.1 Hedge bundles in the New Zealand students' writing .....	149
8.2.2 Hedge bundles in the Chinese students' writing.....	151
8.3 Booster bundles .....	155
8.3.1 Shared booster bundle .....	156
8.3.2 Booster bundle in the New Zealand students' writing.....	157
8.3.3 Booster bundles in the Chinese students' writing.....	157
8.4 Self-mention bundles .....	163
8.5 Directive bundles .....	166
8.5.1 Shared directive bundles .....	167
8.5.2 Directive bundles in the New Zealand students' writing.....	168
8.5.3 Directive bundles in the Chinese students' writing .....	169
8.6 Shared knowledge bundles.....	170
8.7 Summary .....	171
8.7.1 Differences between the Chinese and New Zealand writing .....	172
8.7.2 Differences between the masters and PhD writing .....	172
<b>Chapter 9 Discussion and conclusion .....</b>	<b>175</b>
9.1 Discrepancies and reasons of discrepancies.....	176
9.1.1 Discrepancies in frequency .....	176

---

9.1.2 Discrepancies in structure.....	177
9.1.3 Discrepancies in function .....	180
9.1.4 Reasons for discrepancies.....	182
9.2 Limitations and suggestions for future research.....	186
9.2.1 Limitations.....	187
9.2.2 Suggestions.....	189
9.3 Implications .....	190
9.3.1 Theoretical implications .....	191
9.3.2 Methodological implications .....	192
9.3.3 Pedagogical implications.....	193
9.4 Concluding remarks.....	206
<b>References.....</b>	<b>207</b>
<b>Appendix A: Adaptations of Biber and his colleagues' taxonomy.....</b>	<b>227</b>
<b>Appendix B: Ädel's (2006) taxonomy of personal metadiscourse.....</b>	<b>231</b>
<b>Appendix C: Bundles identified in the four postgraduate corpora .....</b>	<b>233</b>
<b>Appendix D: Interactive categories and sentence initial bundles .....</b>	<b>237</b>
<b>Appendix E: Interactional categories and sentence initial bundles.....</b>	<b>241</b>
<b>Appendix F: Ethical approval .....</b>	<b>243</b>
<b>Appendix G: Interview questions.....</b>	<b>245</b>
<b>Appendix H: The 50 most frequent sentence initial bundles in each corpus.....</b>	<b>247</b>

---

## Tables

Table 1. Overlap between Ädel (2006) and Hyland (2005a).....	49
Table 2. Vande Kopple's (1985) classification of metadiscourse.....	51
Table 3. Crismore, Markkanen & Steffensen's (1993) system of metadiscourse.	52
Table 4. Mauranen's (1993) model of metatext.....	53
Table 5. Hyland's (2005a) interpersonal model of metadiscourse .....	54
Table 6. Hyland's (2005c) model of engagement in academic writing .....	55
Table 7. Ädel's (2006) taxonomy of personal metadiscourse.....	56
Table 8. Ädel's (2006) taxonomy of impersonal metadiscourse .....	57
Table 9. Summary of the metadiscourse models .....	59
Table 10. Comparison of metadiscourse categorical labels .....	61
Table 11. Corpus collection .....	72
Table 12. Bundle exclusion.....	75
Table 13. Major categories and structural patterns of sentence initial bundles ....	76
Table 14. Overview of six Chinese participants .....	80
Table 15. Descriptive statistics: sentence initial bundles.....	83
Table 16. Number of interactive and interactional bundles .....	85
Table 17. Proportion of interactive and interactional bundles (tokens).....	85
Table 18. Top 10 frequent sentence initial bundles in each corpus in rank order.	86
Table 19. Distribution of sentence initial bundles in thesis writing.....	88
Table 20. Distribution of sentence initial bundles in each corpus (types) .....	89
Table 21. Distribution of sentence initial bundles in each corpus (tokens) .....	90
Table 22. NP-based bundles in each corpus in rank order.....	91
Table 23. Z's interview on his use of <i>noun phrase</i> .....	93
Table 24. V's interview on her use of <i>noun phrase</i> .....	93
Table 25. Distribution of the PP-based bundles in each corpus.....	94
Table 26. J's interview on her use of <i>multiple preposition phrase</i> .....	94
Table 27. V's interview on her use of <i>multiple preposition phrase</i> .....	95
Table 28. A's interview on his use of <i>to-phrase fragment</i> .....	97
Table 29. V's interview on her use of <i>to-phrase fragment</i> .....	97
Table 30. W's interview on her use of <i>to-phrase fragment</i> .....	97
Table 31. Descriptive statistics: Interactive bundles.....	103
Table 32. Distribution of interactive bundles in each corpus (tokens) .....	104
Table 33. Transition bundles in the New Zealand and Chinese corpora .....	105
Table 34. Locations of the three shared transition markers .....	106
Table 35. J's interview on her use of <i>on the one hand</i> and <i>on the other hand</i> ....	108
Table 36. V's interview on her use of <i>on the one hand</i> and <i>on the other hand</i> ..	108
Table 37. S's interview on his use of <i>however</i> and <i>therefore</i> .....	110
Table 38. V's interview on her use of <i>however</i> .....	111

---

Table 39. Frame bundles in the New Zealand and Chinese corpora .....	112
Table 40. Scope distribution of boundary bundles (token) .....	114
Table 41. Z's interview on his use of sequence markers.....	117
Table 42. W's interview on her use of sequence markers .....	117
Table 43. V's interview on her use of sequence markers.....	118
Table 44. Endophoric bundles in the New Zealand and Chinese corpora.....	119
Table 45. Code gloss bundles in the New Zealand and Chinese corpora.....	123
Table 46. Frequency of <i>In other words</i> and <i>That is to say</i> (pmw) .....	124
Table 47. Z's interview on his use of <i>for example</i> .....	126
Table 48. V's interview on her use of <i>for example</i> .....	126
Table 49. V's interview on her use of <i>to be specific</i> .....	127
Table 50. Condition bundles in the New Zealand and Chinese corpora .....	128
Table 51. Positions of the four shared condition bundles .....	129
Table 52. A's interview on his use of <i>from the perspective of</i> .....	131
Table 53. V's interview on her use of <i>when elder is mentioned</i> .....	131
Table 54. A's interview on his use of <i>way</i> .....	132
Table 55. W's interview on her use of <i>way</i> .....	133
Table 56. J's interview on her use of <i>in this way</i> .....	133
Table 57. Z's interview on his use of <i>with the development of</i> .....	133
Table 58. V's interview on her use of <i>with the development of</i> .....	134
Table 59. W's interview on her use of <i>with the development of</i> .....	134
Table 60. Introduction bundles in the New Zealand and Chinese corpora.....	134
Table 61. J's interview on her use of <i>there be</i> .....	136
Table 62. V's interview on her use of <i>there be</i> .....	136
Table 63. W's interview on her use of <i>there be</i> .....	136
Table 64. A's interview on his use of <i>there be</i> .....	136
Table 65. Descriptive statistics: Interactional bundles .....	139
Table 66. Distribution of interactional bundles in each corpus (tokens).....	140
Table 67. Attitude bundles in the New Zealand and Chinese corpora .....	141
Table 68. Distribution of <i>It is important to</i> .....	143
Table 69. V's interview on her use of <i>necessary</i> .....	147
Table 70. V's interview on her use of <i>interesting</i> .....	148
Table 71. J's interview on her use of <i>it is difficult to</i> .....	148
Table 72. Hedge bundles in the New Zealand and Chinese corpora.....	149
Table 73. Functions of <i>It is possible that</i> .....	150
Table 74. Z's interview on his use of <i>hope</i> and <i>suggest</i> .....	153
Table 75. V's interview on her use of <i>indicate</i> and <i>show</i> .....	154
Table 76. V's interview on her use of <i>one of the most</i> .....	155
Table 77. W's interview on her use of <i>one of the most</i> .....	155
Table 78. Booster bundles in the New Zealand and Chinese corpora.....	156
Table 79. J's interview on her use of <i>clear</i> and <i>obvious</i> .....	159

---

Table 80. Z's interview on his use of <i>clear</i> and <i>obvious</i> .....	159
Table 81. V's interview on her use of <i>clear</i> and <i>obvious</i> .....	159
Table 82. J's interview on her use of <i>believe</i> .....	161
Table 83. V's interview on her use of <i>undoubted</i> .....	162
Table 84. Self-mention bundles in the New Zealand and Chinese corpora .....	164
Table 85. W's interview on her use of <i>I</i> .....	165
Table 86. Directive bundles in the New Zealand and Chinese corpora .....	166
Table 87. V's interview on her use of <i>note</i> and <i>notice</i> .....	168
Table 88. J's interview on her use of <i>note</i> .....	169
Table 89. A's interview on his use of <i>see</i> .....	170
Table 90. V's interview on her use of <i>As we all know</i> .....	171
Table 91. Metadiscourse bundles .....	191





---

## Figures

Figure 1. The relationship between lexical bundles, collocations and formulaic sequences .....	28
Figure 2. Biber and Barbieri's (2007) functional taxonomy of lexical bundles ...	37
Figure 3. Hyland's (2008a) functional framework of lexical bundles .....	41
Figure 4. Collocations of <i>important</i> and <i>necessary</i> in Wikipedia as displayed in FLAX .....	146
Figure 5. Search for <i>perspective</i> bundle in FLAX .....	198
Figure 6. Search of <i>knowledge</i> in <i>Learning Collocations</i> collection in FLAX ..	200
Figure 7. Search of <i>knowledge</i> in <i>Web Phrases</i> collection in FLAX.....	201
Figure 8. Sentence initial bundles in the New Zealand PhD thesis corpus.....	202
Figure 9. Context sentences of the bundles <i>It is important to</i> .....	202
Figure 10. Sentence initial bundles in BAWE .....	204
Figure 11. Function-based sentence initial bundle list in BAWE.....	204
Figure 12. Sentences containing <i>important</i> at the beginning, grouped by pattern .....	205
Figure 13. Sentences with the same pattern <i>It is important to + verb</i> .....	206



## Chapter 1 Introduction

Language is formulaic. As early as the 1970s, Bolinger (1976) suggested that “our language does not expect us to build everything starting with lumber, nails, and blueprint, but provides us with an incredibly large number of prefabs” (p. 1). Biber, Johansson, Leech, Conrad, and Finegan (1999), and Erman and Warren (2000) found that prefabricated language constituted 21-52.3% of written text. Therefore, formulaic language, especially that which occurs with high frequency, deserves attention (Nation, 2013). As an important component of formulaic language, lexical bundles, which combine three or more words and occur repeatedly in a given register, should be a focus of language pedagogy and should be taught earlier. However, what are the target bundles for learning? Where and how are these bundles used in text? What are the reasons for L2 (second language) learners’ bundle choices? The answers to these questions need to be addressed by researchers before lexical bundles can be integrated into pedagogy effectively by teachers and learners.

The present study explores answers to these questions with regard to Chinese postgraduate L2 thesis writers. The study uses sentence initial bundles in Chinese L2 and New Zealand L1 (first language) thesis writing in the discipline of general and applied linguistics as a point of comparison, and explores Chinese postgraduate students’ reasons for their bundle choices. This chapter introduces the motivation, objectives and possible contributions that the study may make, as well as a description of the organisation of the thesis.

### 1.1 Motivation for the study

Lexical bundles (e.g. *on the other hand, the fact that the, it should be noted*), as recurrent multiword combinations, are extremely common discourse building blocks (Biber et al., 1999) and usually carry specific metadiscourse functions. The use of lexical bundles facilitates writers’ language production, improves idiomaticity, accuracy and fluency of academic writing, and indicates writers’ membership in a particular academic community. Therefore, these bundles deserve special attention in discourse analysis and pedagogy, and bundle research should

---

seek to inform language teaching and learning. However, teaching and learning lexical bundles remains relatively peripheral (Byrd & Coxhead, 2010; Cortes, 2006; Eriksson, 2012; Jones & Haywood, 2004). This is possibly due to the following three reasons: the nature of lexical bundles, limited learning resources, and the lack of a connection between research and pedagogy, which are discussed below.

### **1.1.1 Nature of lexical bundles**

Lexical bundles are features of text, but have become identifiable through corpus linguistics, and thus are the *product* of corpus linguistics. A common method of generating them involves a computer programme automatically processing a collection of texts and identifying word chunks with three or more words that occur repetitively with relatively high frequency and wide distribution across texts. As a result of this type of investigation in corpus linguistics, a vast number of lexical bundles have been generated from many different sources. Biber and his colleagues (1999) report that three-word bundles occur more than 60,000 times per million words and four-word bundles occur over 5,000 times per million words in academic prose. In addition, the length of bundles usually varies from three to six words, though four-word bundles are the most popular bundles under investigation. The number and length of bundles pose difficulties for learners wanting to choose their target bundles for learning.

Another direct result of corpus-based analysis is that most bundles (i.e. 85% in conversation and 95% in academic prose) are incomplete structural units (Biber et al., 1999). Biber and his colleagues (1999) reveal that a large number of bundles in conversation are composed of a pronominal subject followed by a verb phrase plus the start of a complement clause (e.g. *I don't know what*). Bundles in academic prose usually contain parts of noun phrases and prepositional phrases (e.g. *the nature of the, as a result of*). This appears to contradict traditional grammar-based pedagogy, which usually focuses on complete structural units.

Lexical bundles are not perceptually salient or easily noticed within text, and bundle identification is largely confined to the availability of corpora and corpus-based tools. Learners with little access to corpora and corpus-based tools may find it difficult to decide on and extract target bundles from a particular corpus. Although

---

there are some researcher-generated bundle lists, such as the conversation and academic bundle lists produced by Biber and his colleagues (1999), Hyland's (2008b) discipline-based bundle lists and Simpson-Vlach and Ellis's (2010) Academic Formulas List, learners with little background information and limited access to context may feel confused, and find it difficult to decide on the most relevant and valuable bundles to study (Byrd & Coxhead, 2010).

In addition, a large majority of bundles are transparent in meaning and consist of well-known words (e.g. *in the case of, it is interesting to note*). This feature makes them very unlikely to capture learners' attention: learners may regard many bundles as their acquired vocabulary knowledge and may not attend to them. As Byrd and Coxhead (2010) claim, lexical bundles lack face validity for learners.

### **1.1.2 Limited resources of lexical bundles**

Language resources, no matter whether they are traditional resources such as dictionaries, or newly-developed ones such as corpus-based tools, often fail to offer sufficient help for bundle learning. There are dictionaries of collocations and idioms, but few including examples of lexical bundles. Many corpus-based tools present lexical bundles as frozen chunks which does not address any variation. The nature of available resources is discussed below.

#### *1.1.2.1 Dictionaries for learning prefabricated language*

A number of dictionaries are compiled for collocation or idiom learning. Most dictionaries of collocations target intermediate to advanced learners who already have a repertoire of individual words but lack the knowledge of co-occurring words. For example, *LTP Dictionary of Selected Collocations* (Hill & Lewis, 1997) provides intermediate and advanced learners with five kinds of word combinations: adjective + noun, verb + noun, noun + verb, adverb + adjective and verb + adverb. These word combinations are grouped into two sections: the noun section and the adverb section. The first section contains 50,000 collocations for 2,000 essential nouns and the second section lists the combinations of 5,000 adverbs with over 1,200 verbs and adjectives. *The BBI Dictionary of English Word Combinations* (Third Edition) (Benson, Benson, & Ilson, 2010) contains 20,000 entries and

---

110,000 collocations, including both grammatical and lexical collocations. Grammatical collocations incorporate a dominant word (e.g. noun, adjective or verb) and a preposition or grammatical construction (e.g. infinitive or clause). Lexical collocations are combinations of nouns, adjectives, verbs and adverbs, which do not contain a dominant word. *The Macmillan Collocations Dictionary for Learners of English* (Rundell, 2010) is a corpus-based dictionary with collocations generated from a two billion word corpus of modern English, and grouped on the basis of grammatical structures and semantic meanings. It is particularly designed for upper intermediate to advanced learners with a focus on academic or professional English.

Like collocation dictionaries, many idiom dictionaries remain popular, although different positions have been taken with regard to the teaching of idioms (O'Keefe, McCarthy, & Carter, 2007). For example, *Oxford Idioms Dictionary for Learners of English* (Second Edition) (Parkinson & Francis, 2007) covers 10,000 British and American idioms. *Oxford Dictionary of Idioms* (Second Edition) (Siefring, 2004) includes more than 5,000 idioms from English-speaking countries. *Cambridge Idioms Dictionary* (Second Edition) (Walter, 2006) presents around 7,000 idioms with examples from the Cambridge International Corpus.

Possibly due to the nature of lexical bundles discussed above, few dictionaries have been compiled for bundle learning. In other words, language learners are not able to consult dictionaries for direct bundle reference.

#### *1.1.2.2 Corpus-based tools for learning prefabricated language*

The last twenty years have seen an increase in research in the area of data-driven learning (DDL). This is a term coined by Johns (1991) to refer to the idea of learners as language researchers. The development of electronic corpora and the application of corpus-based tools have created the potential for language learners to explore various patterns in a somewhat independent way. Many researchers have investigated the possibility of applying the DDL approach to various multiword combinations (e.g. Boulton, 2009, 2010, 2012; Chambers & O'Sullivan, 2004; Chan & Liou, 2005; Chang, 2014; Chen, 2011; Daskalovska, 2015; Geluso & Yamaguchi, 2014; O'Sullivan & Chambers, 2006; Yeh, Li, & Liou, 2007; Yoon, 2008; Yoon & Hirvela, 2004). Positive responses have strongly suggested that corpus use not only

---

facilitates learning and writing, but also raises learners' awareness of multiword combinations and increases their writing confidence.

At the same time, many corpus-based tools are limited to presenting bundles as frozen chunks. For example, the generated bundle *It is important to note* only represents one variation of the pattern *It is important to + verb* and the verb slot can be filled with alternative verbs such as *consider, remember, examine, analyse* and *recognise*. Flowerdew (2014) argues that “a drawback of a lexical bundle approach is that the automatic analysis does not capture variation” (p. 37). Some writing teachers have expressed their worries about the overly repetitive use of bundles in student writing. As a result, they may hesitate to introduce these fixed chunks to their students (L. Flowerdew, personal communication, June 12, 2015).

### **1.1.3 Lack of connection between research and pedagogy**

The poor connection between bundle research and pedagogy further limits the application of research findings. Factors that contribute to this disconnection include the taxonomies used in research, the lack of context-based qualitative study, and the scant attention to the reasons for learner bundle choices. For example, the two most popular functional taxonomies of lexical bundles (i.e. Biber and his colleagues' taxonomy and Hyland's framework) have initially been developed for data analysis with specialised linguistic terminology (e.g. epistemic stance bundles) or categories that are somewhat difficult to apply to writing (e.g. research-oriented bundles). Most bundle studies have placed more weight on overall quantitative bundle comparison, and insufficient information has been provided to learners with a few bundles introduced in context examples. The reasons for typical learner production have mostly been explored on the basis of researchers' perceptions rather than empirical research, which undermines the implications of learner bundle research.

Lexical bundle research is a new area with a short history of about twenty years, and only a limited number of studies have focused on learner bundles in academic prose. These studies are clearly insufficient to support language pedagogy considering the wide diversity of learners: their different first languages, proficiency levels, genres of writing and contexts of study. Moreover, many



---

teachers and researchers, though they have learner data available, are unable to explore the features of learner bundles due to resistance to new technologies, little knowledge of corpus linguistics, and the inaccessibility of corpora and corpus-based tools (Boulton, 2012; Kilgarriff, 2009; Kilgarriff & Grefenstette, 2003).

## 1.2 Objectives

In this study, I compare the use of sentence initial bundles (i.e. bundles at the beginning of sentences) in Chinese and New Zealand thesis writing and investigate the reasons for typical bundles used by Chinese postgraduates. Chinese postgraduate theses are selected because Chinese students comprise the largest proportion of FL (foreign language) (if they study in a non-English-speaking country including mainland China) or L2 students (if they study in any English-speaking countries). I, as a native speaker of Chinese and a previous university lecturer in China, have received most of the education in mainland China. Therefore, I am particularly familiar with the education system and able to interpret the data as an insider. The two primary objectives of this study are as follows:

1. To identify the differences in the use of sentence initial bundles between Chinese and New Zealand postgraduates (including both masters and PhDs), and between masters and PhD levels of study in terms of frequency of occurrence, grammatical structure and discourse function; and
2. To explore the reasons for the use of those typical bundles in the Chinese postgraduate theses.

To achieve these two objectives, the following research questions were developed:

1. What are the frequencies of four-word sentence initial bundles in the Chinese and New Zealand masters and PhD corpora? Are there any differences between Chinese and New Zealand postgraduates, or between masters and PhD levels of study in the use of sentence initial bundles?
2. What are the salient structures of these bundles in the Chinese and New Zealand corpora? Are there any differences between Chinese and New Zealand postgraduates, or between masters and PhD levels of study in the distribution of structures?

- 
3. What are the metadiscourse functions of these bundles in the Chinese and New Zealand corpora? Are there any differences between Chinese and New Zealand postgraduates, or between masters and PhD levels of study in the distribution of functions?
  4. What reasons do Chinese postgraduates give for their sentence initial bundle choices in their thesis writing?

In order to answer these questions, four thesis corpora were built: a Chinese masters thesis corpus, a New Zealand masters thesis corpus, a Chinese PhD thesis corpus and a New Zealand PhD thesis corpus. 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 automatically generate lexical bundles from the corpora. The structural categories and patterns of this study were developed from the studies of Biber and his colleagues (Biber, Conrad, & Cortes, 2004; Biber et al., 1999) and Chen and Baker (2010). The functional analysis was adapted from Hyland's (2005a, 2005c) interactive and interactional model of metadiscourse. Semi-structured interviews with Chinese postgraduates were conducted after the text analysis to understand the reasons for Chinese students' bundle choices. The interviews were based on the expressions in participants' original drafts, which completely or partially overlapped with the sentence initial bundles generated from the corpus data.

### **1.3 Contributions**

This study seeks to contribute to the existing theory, methodology and pedagogy of the work of lexical bundles. The following sections will discuss each aspect in detail so as to provide a rationale for the study.

#### **1.3.1 Potential contributions to theory**

Metadiscourse has been a focus of writing research and different models have been developed to inform writing pedagogy. For example, the models of Vande Kopple (1985), Crismore, Markkanen, and Steffensen (1993), Hyland (2005a, 2005c), Mauranen (1993) and Ädel (2006) are among the most popular and widely-cited models. However, most existing models take a top-down approach and their

---

categorisations are largely determined by the pre-determined items, mostly individual words.

This study adopts a bottom-up approach and conducts metadiscourse analysis from the perspective of lexical bundles. It highlights multiword units as metadiscourse devices, and verifies and extends the existing metadiscourse functions. More detailed discussion can be found in Section 4.2 The relationship between metadiscourse and lexical bundles.

### **1.3.2 Potential contributions to methodology**

Lexical bundle research has been conducted for about twenty years. Despite development in scope, the methodologies of all the studies and the taxonomies used for functional analysis are largely the same. Most researchers have investigated lexical bundles regardless of their position in sentences. Functional analysis tends to either follow the taxonomy of Biber and his colleagues (Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2003; Biber et al., 2004), or to use the framework of Hyland (2008a). Little interest has been devoted to writers' interpretations of corpus data.

This study distinguishes sentence initial and non-initial bundles, based on the recognition that these two types of bundles pose different challenges for learners and perform different functions in sentences. It employs Hyland's (2005a, 2005c) metadiscourse model to investigate sentence initial bundles, taking into consideration the shared features between lexical bundles and metadiscourse devices. The application of metadiscourse model allows researchers to view and introduce lexical bundles as writing devices, alongside lexical chunks under investigation, and for this reason I have used the model in the present study. This study combines corpus-based analysis and semi-structured interviews. The involvement of participant writers provides insights into corpus data. See Chapter 5 Methodology for further discussion.

### **1.3.3 Potential contributions to pedagogy**

The development of electronic corpora and the application of corpus-based tools have created the potential for language learners to explore various multiword

---

combinations. Many researchers have investigated the possibility of applying corpus-based approaches to different levels of learners (e.g. Boulton, 2009, 2010, 2012; Chambers & O'Sullivan, 2004; Chan & Liou, 2005; Chang, 2014; Chen, 2011; Daskalovska, 2015; Geluso & Yamaguchi, 2014; O'Sullivan & Chambers, 2006; Yeh et al., 2007; Yoon, 2008; Yoon & Hirvela, 2004). However, despite the fact that corpus-based approaches have been proved effective, “language learners rarely have hands-on experience with corpora in mainstream education” (Leńko-Szymańska & Boulton, 2015, p. 3).

This study seeks to support the arguments for corpus-based multiword learning approaches. It is anticipated that a range of practical suggestions will be provided for L2 student writers and their teachers on the basis of its findings. Details can be found in Section 9.3.3 Pedagogical implications.

#### **1.4 Thesis outline**

This thesis is structured in nine chapters. Following this introductory chapter, chapter 2 introduces the concept of corpus linguistics and corpus, overviews the application of corpus linguistic approaches in the development of word lists and the studies of academic discourse — two major areas of corpus research. Since corpus-based academic discourse research is the approach taken in this study, the research in this area has been extensively explored under the headings in which research has been undertaken: languages, registers, genres, and disciplines. At the end of the chapter, the roles of learner corpora and the contributions of contrastive interlanguage analysis as well as its limitations have been discussed in relation to the present study.

Chapter 3 presents a key concept in the current research, *the lexical bundle*, defines its characteristics, and distinguishes it with the other two closely-related concepts, collocations and formulaic sequences. The significance of lexical bundles in academic writing has also been highlighted, followed by a comprehensive review of lexical bundle studies in the area of L2 academic writing in terms of frequency of occurrence, grammatical structure and discourse function. Alongside these studies, this chapter also evaluates the most popular structural categories — Biber and his colleagues' (1999) structural patterns, and two widely-used functional

---

taxonomies — Biber and his colleagues' taxonomy (Biber & Barbieri, 2007; Biber et al., 2003, 2004) and Hyland's (2008a) framework. Although some researchers have included possible interpretations of student bundle choices, these studies have largely ignored student voices as they have been the interpretations of researchers, not student writers. The limitations of the existing research and taxonomies are explored at the end of the chapter.

Chapter 4 introduces metadiscourse, another important concept in academic writing and in this study, and justifies the use of metadiscourse in sentence initial bundle study. Then, it compares the commonly-used metadiscourse models. Among them, Hyland's (2005a, 2005c) model of metadiscourse, as the most inclusive, comprehensive and relevant model so far, has been chosen as the model to guide the present study. Studies on metadiscourse are reviewed, which include investigations of metadiscourse as a unified whole, and the examination of a specific aspect of metadiscourse (e.g. hedges). Unlike bundle research, a few metadiscourse studies have included writer interpretations. Limitations of metadiscourse research have been addressed at the end of this chapter.

Chapter 5 is the methodology chapter. It first states the objectives of this research with four primary research questions. Then, it mainly introduces the procedures of corpus-based analysis, which involve corpus building, bundle identification and the development of frameworks for structural and functional analysis. Four corpora were built from online databases: a Chinese masters thesis corpus, a New Zealand masters thesis corpus, a Chinese PhD thesis corpus and a New Zealand PhD thesis corpus. The same criteria were applied across the corpora for generating bundles. A small number of non-applicable categories of the selected frameworks were excluded and several new ones were added on the basis of the empirical data. The last section of this chapter focuses on semi-structured interviews, including the recruitment of participants, the backgrounds of participants and the process of thematic analysis.

Chapters 6, 7 and 8 are finding chapters. Frequency, structure and function are the three foci of lexical bundle research: chapter 6 covers the findings of frequency-based and structural analysis, and chapters 7 and 8 report the findings of functional

---

analysis with respect to interactive functions and interactional functions. During the discussion, similarities and differences in the use of sentence initial bundles have been explored, together with possible reasons drawn from the literature and/or interview data.

Chapter 6 describes the overall distribution of sentence initial bundles in the four corpora and presents the number of shared bundles between the corpora. This chapter also illustrates the bundle distribution in relation to five identified structural categories, NP-based, PP-based, VP-based, clause-based and other bundles; and explores the different structural distributions between the four corpora. Possible reasons have also been discussed.

Chapter 7 analyses sentence initial bundles with interactive functions, which consist of transition bundles, frame bundles, endophoric bundles, code gloss bundles, condition bundles and introduction bundles. This chapter describes the distribution of interactive bundles in each corpus, examines the shared and different bundles between Chinese and New Zealand students, and considers the major discrepancies in the use of bundles between masters and PhDs. It also suggests possible interpretations of those identified typical bundles in Chinese student writing.

Chapter 8 is devoted to interactional bundles, which include attitude bundles, hedge bundles, booster bundles, self-mention bundles, directive bundles and shared knowledge bundles. As in the previous chapter, bundles are compared within each category and possible sources of the identified bundles are presented.

Chapter 9 relates the findings of sentence initial bundles to the literature on corpus linguistics, lexical bundles and metadiscourse to verify previous research and to highlight new findings, particularly the findings of interviews. Limitations of the present study are outlined and suggestions are provided for future research. Implications for theory, methodology and pedagogy are discussed. This study is a unique study using metadiscourse models to explore lexical bundles, and employing interview data to interpret student bundle choices. Drawing on the findings, this study makes a strong case for the combination of corpus-based discourse analysis with data-driven learning as an effective approach to language teaching and learning.

---

The following three chapters, chapters 2, 3 and 4, are literature review chapters. I will review the work in corpus linguistics, lexical bundle studies and metadiscourse analysis in each chapter. Chapter 2 is an overview of corpus linguistics approaches.

---

## Chapter 2 Corpus linguistics and academic discourse analysis

Corpus linguists have taken a variety of approaches to understanding and addressing the phenomena of academic discourse. This chapter sets the scene by defining corpus linguistics and corpus, categorising different types of corpora, providing an overview of corpus linguistic approaches in relation to word lists and academic discourse (two important areas of corpus research), and exploring the use of learner corpora in contrastive interlanguage analysis.

### 2.1 Corpus linguistics and corpora

Corpus linguistics, either defined as a methodology (Gray & Biber, 2013; McEnery, Xiao, & Tono, 2006) or a theory (Baker, 2010; Tognini-Bonelli, 2001), is the study of linguistic variation on the basis of large collections of real language data. The term *corpus* is the Latin word for *body* and here refers to a collection of texts. McEnery and his colleagues (2006) highlight three qualities of a modern corpus that contribute to the quality of analysis: machine-readable, authentic and representative, which contribute to assuring the efficiency, reliability and generalisability of corpus analysis.

Baker (2010) distinguishes between general corpora (also known as reference corpora) and specialised corpora: a general corpus is “normally very large”, “with texts collected from a wide range of sources”, “representing many language contexts” (p. 12); whereas a specialised corpus is designed to address specific research questions with “clear restrictions placed on the texts” (p. 14). General corpora usually provide the language norms for specialised corpora. Butterfield (2009) divides general corpora into three generations on the basis of corpus size and computer technology. The first generation includes the one-million-word Brown Corpus (BROWN) of the 1960s and the Lancaster-Oslo-Bergen Corpus (LOB) of the 1970s. Examples of the larger second-generation corpora are the 450-million-word Corpus of Contemporary American English (COCA) of the 2010s and the 100-million-word British National Corpus (BNC) of the 1990s. The third generation corpora comprise over one billion words, examples being the Cambridge



---

International Corpus, the Oxford English Corpus and the World Wide Web, the last considered as a type of general corpus (Kilgarriff & Grefenstette, 2003).

Examples of specialised corpora are the 1.8-million-word Michigan Corpus of American Spoken English (MICASE), the 2.6-million-word Michigan Corpus of Upper-Level Student Papers (MICUSP) and the 6.5-million-word British Academic Written English Corpus (BAWE). Learner corpora, containing texts produced by L2 or FL learners, are an important type of specialised corpora. The Cambridge Learner Corpus, developed by Cambridge University Press, is “the world’s largest learner corpus”, containing “over 200,000 exam scripts from students speaking 148 different languages living in 217 different countries or territories” (Cambridge University Press, 2015). Granger, Dagneaux, Meunier, and Paquot (2009) built the International Corpus of Learner English (ICLE), which contains 3.7 million words of EFL (English as a foreign language) writing from intermediate to advanced learners representing 16 mother tongue backgrounds (i.e. Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Turkish and Tswana). Wen, Liang, and Yan (2008) developed the Spoken and Written English Corpus of Chinese Learners (SWECCL) with a spoken and a written sub-corpus. The spoken one comprises over-one-million-word spoken texts from the Test for English Majors, Bands 4 and 8, and the written corpus is a collection of argument or expository essays produced by Chinese undergraduates from more than 20 universities.

As Baker (2010) argues, the distinction between general corpora and specialised corpora is blurry and “all corpora are specialised in some way” (p.14). For example, BNC, a large general corpus, can also be regarded as a specialised corpus as it is a collection of British English of the late 20<sup>th</sup> century.

McEnery and Hardie (2012) adopt a different approach to Baker (2010) in dividing corpora according to data collection processes. They have two categories: monitor corpora and sample corpora (or balanced corpora). Monitor corpora grow over time and items are selected on the basis of pre-determined criteria; in contrast, sample corpora represent the language at a particular point of time. Well-known examples of monitor corpora are the Bank of English (BoE) and the Web. Examples of sample

---

corpora are the Brown corpus, the LOB corpus, the BAWE corpus and the SWECCL corpus.

## **2.2 Corpus linguistics and word lists**

Corpus linguistic research started (as first generation corpora) with a focus on vocabulary and various corpora have been built or adapted to generate different word lists. Since then, the size of corpora has increased dramatically over recent years enabled by computer technology.

Many corpora of vocabulary have been built *on the back of* previous corpora. For example, the early version of the General Service List (West, 1953) was derived from a 5-million-word corpus, and the Academic Word List (Coxhead, 2000) was compiled from a 3.5-million-word corpus of written academic texts. In contrast, the recently-developed New General Service List (Brezina & Gablasova, 2015) was developed from four corpora (i.e. LOB, BNC, BE06 and EnTenTen12) with a total of over 12 billion words, the Academic Keyword List (Paquot, 2012) is based on two professional writing corpora (i.e. Micro-Concord Corpus Collection B and the Baby BNC Academic Corpus) and two student writing corpora (i.e. the Louvain Corpus of Native Speaker Essays and the BAWE corpus), and the Academic Vocabulary List (Gardner & Davies, 2014) was generated from a 120-million-word academic sub-corpus of COCA.

Recently, corpus linguists start to develop lists of multiword combinations. Simpson-Vlach and Ellis (2010) developed an Academic Formulas List (AFL) with 3-, 4-, and 5-grams, using MICASE, BNC and Hyland's (2004a) research article corpus. Ackermann and Chen (2013) compiled an Academic Collocation List (ACL) with 2,468 most frequent and pedagogically useful entries from 25.6-million-word written curricular component of the Pearson International Corpus of Academic English (PICAE).

## **2.3 Corpus linguistics and academic discourse analysis**

Another important area of corpus linguistics began to emerge in the late 20<sup>th</sup> century, that of uncovering recurrent lexical-grammatical patterns in language use (e.g. Biber et al., 1999; Nattinger & DeCarrico, 1992; Sinclair, 1991). Many studies in

---

this area have been conducted in relation to discourse analysis, particularly academic discourse analysis.

The term *discourse* has been conceptualised in many different ways. Following Schiffrin, Tannen, and Hamilton (2001), Gray and Biber (2013) group the definitions into three major categories, and the first two of these frame the concept of discourse as it is used here:

1. discourse as language in use, which investigates variation in the use of linguistic forms and traditional linguistic constructs;
2. discourse as language structure above the sentence level, which focuses on the broader text structure, that is, on the systematic ways that texts are constructed; and
3. discourse as social practices and ideologies associated with language and/or communication, focusing on the general characteristics and participants of a particular discourse community. (p.138)

As a sub-category of discourse, *academic discourse* refers to the ways of thinking and using language in the context of the academy (Hyland, 2009). Corpus linguistic research on academic discourse has mainly investigated the languages (i.e. first, second or foreign language) used in different academic settings such as journal articles, textbooks, essays, theses, classroom teaching, study groups and office hours (Suomela-Salmi & Dervin, 2009). Individual corpus studies can be situated on a continuum from “bottom-up (more corpus-based)” to “top-down (more discourse-analytic)” (Charles, Pecorari, & Hunston, 2009, p. 5).

The findings from corpus linguistic studies on academic discourse are reviewed in the following sections under the headings of languages, registers, genres, and disciplines as a way of narrowing down the complexity of the many comparative studies conducted. There is much confusion between the terms *register* and *genre* in the literature (Lee, 2001). Here I take Biber and Conrad’s (2009) distinction: “[r]egister variation focuses on the pervasive patterns of linguistic variation across such situations, in association with the functions served by linguistic features; genre variation focuses on the conventional ways in which complete texts of different types are structured” (p.23). Studies have looked at comparisons and contrasts

---

between the texts: produced by different language groups, in both written and spoken registers in English, in different academic written genres, and in different disciplines within the same genre. The studies comparing L1 and L2 texts are excluded from this review at this point but will be reviewed under the title lexical bundles and metadiscourse in Chapters 3 and 4 because they are the subject of the present study and as such deserve detailed attention. The following sections review frequency, structure and function of language items because these are the three foci of this study.

### **2.3.1 Corpus linguistic research on different languages**

Many studies have focused on the features of English and only three studies have investigated the structural variations of lexical bundles in other languages: Spanish in Tracy-Ventura, Cortes, and Biber (2007), Korean in Kim (2009) and Japanese in Kaneyasu (2012). Noun-phrase-based and preposition-phrase-based bundles in Spanish (Tracy-Ventura et al., 2007) and noun-phrase-based bundles in Korean (Kim, 2009) are generally more common than verb-phrase-based bundles. This distribution differs from English, in which verb-phrase-based bundles are more frequent (Biber et al., 1999). The majority of the Japanese bundles in three spoken registers (i.e. conversation, interview and speech) are verb-phrase-based, the same as those in English conversation.

Texts in other languages have also been examined in comparison to English texts with regard to metadiscourse functions, namely the textual (or interactive) and interpersonal (or interactional) functions. For examples, Jiang (2009), Kim and Lim (2013) compared Chinese introductions of research articles with those of English ones; Dahl (2004) investigated textual metadiscourse of English, French and Norwegian articles in the disciplines of economics, linguistics and medicine; Molino (2010) analysed personal and impersonal authorial reference in linguistics research articles in English and Italian; and Marandi (2003) contrasted the introduction and discussion sections of English and Persian masters theses. Metadiscourse devices were generally found more frequent in English texts, but some subcategories of metadiscourse such as connectives (e.g. *however*), attributors (e.g. *according to John*) and persona markers (e.g. *strangely*) occur more often in

---

Persian masters theses. The higher density of metadiscourse devices in English texts may be the result of the research design of these studies. The employment of the English metadiscourse taxonomies (e.g. Crismore et al., 1993; Hyland, 2005a; Vande Kopple, 1985), may result in the identification of more metadiscourse items in English texts.

### 2.3.2 Corpus linguistic research on registers

Biber and his colleagues have examined the differences between various written and spoken registers, particularly registers at U.S. universities. Biber (2009), Biber and Gray (2010), and Biber, Gray, and Poonpon (2011) compared the grammatical complexity of academic writing with that of conversation. Academic writing consists of formulaic frames with an internal variable slot predominated by content words, mostly nouns, to form noun or prepositional noun phrase fragments (e.g. *the end of the, in the case of*), whereas conversation is distinguished by continuous fixed sequences with a preceding or following variable slot usually filled by function words to indicate clause fragments (e.g. *but I don't know, I don't know if*) (Biber, 2009).

Unlike conversation, academic writing tends to employ noun phrases instead of dependent clauses for structural elaboration, including adjectives or nouns as pre-modifiers (e.g. *theoretical orientation, system perspective*) and prepositional phrases as post-modifiers, among which, many are *of*-phrases (e.g. *the participant perspective of members of a lifeworld*) (Biber & Gray, 2010). Noun phrases, primarily prepositional post-modified phrases, reflect the complexity of academic writing; while finite dependent clauses (e.g. *if, because, that* and *WH* clauses) significantly contribute to the structural complexity of conversation (Biber et al., 2011).

In regard to function, Biber, Conrad, Reppen, Byrd, and Helt (2002) adopted Biber's (1988) five major dimensions of variation (i.e. involved versus informational production, narrative versus nonnarrative discourse, situation-dependent versus elaborated reference, overt expression of persuasion, and nonimpersonal versus impersonal style) and undertook a multidimensional (MD) analysis of spoken and written registers at U.S. universities. They indicate "a strong

---

polarization between (university) spoken and written registers” (p.41). The written registers (e.g. textbooks, course packs, course management and other campus writing) are informationally dense with the extensive use of nouns, long words, prepositions, and attributive adjectives; whereas the spoken registers (e.g. class sessions, office hours, study groups and on-campus service encounters) are characterised largely as involvement and persuasion, with frequent modal and semimodal verbs (e.g. *will, should, have to*), suasive verbs (e.g. *command, propose, insist*) and conditional subordination (e.g. *if you want*). Biber et al. (2004), and Biber and Barbieri (2007) identified the use of more stance devices in the spoken registers in terms of modal verbs (e.g. *will, can, must*), stance adverbs (e.g. *actually, possibly, generally*) and stance complement clauses (e.g. *we recognize, you need to, it is also clear*), these being associated with epistemic evaluation or personal attitudes. Written registers, particularly textbooks and academic prose, are dominated by referential devices such as place references (e.g. *in the college of*), time references (e.g. *at the time of*) and what they termed intangible framing attributes (e.g. *the nature of the*) because writers need to constantly remind readers of the attributes of an entity, being presented in different contexts. Classroom teaching, unlike other university spoken registers, relies heavily on discourse organizing bundles (e.g. *want to talk about*) and referential bundles (e.g. *those of you who*), in addition to stance bundles (e.g. *I don't know what*) (Biber & Barbieri, 2007; Biber et al., 2004). A detailed explanation of Biber and his colleagues' functional taxonomy of lexical bundles can be found in Section 3.3.3.1 Biber and his colleagues' taxonomy.

### **2.3.3 Corpus linguistic research on written genres**

Researchers have examined rhetorical strategies in various written academic genres (e.g. research theses, textbooks, popular science articles and opinion articles) and compared them with the strategies used in research articles (Fu & Hyland, 2014; Hyland, 2004a, 2010; Koutsantoni, 2006). Koutsantoni (2006) found that thesis writers hedged more and employed more strategic hedges than authors of research articles. Strategic hedges here are used to indicate limitations of method, limitations of the scope of the paper, limited knowledge, agreement with other research and limitations of the study (Koutsantoni, 2006). Research article authors, however, use

---

considerably more personally attributed hedges than thesis writers (i.e. 16% compared to 0.6% of all the hedges), that is, the hedges with personal pronouns (e.g. *we, our*).

Hyland (2004a) adopted Crismore and her colleagues' (1993) metadiscourse taxonomy (see Section 4.3.4 Hyland's metadiscourse model for more details) and identified more textual than interpersonal devices in textbooks. Compared with research articles, textbooks are characterised by the greater use of transitions and the low occurrence of hedges, self-mentions and citations (Hyland, 2012). Hyland (2010), and Fu and Hyland's (2014) examinations of opinion articles, popular science articles and research articles reveal the highest level of engagement and degree of certainty in opinion articles with the greatest use of interactional devices such as engagement features, boosters and self-mentions. Research articles, on the other hand, are marked with the lowest frequency of interactional devices (i.e. attitude markers, reader pronouns and questions) and the highest use of hedges. This is not surprising given the fact that opinion articles need to "establish a more intimate relationship with readers and claim an individual credit for arguments" (Fu & Hyland, 2014, p. 141), while research articles aim to minimise subjective elements in the texts and present arguments with caution.

Corpus linguistic research has also been carried out on learner writing of different genres. Hong and Cao (2014) compared two different genres of learner writing: argumentative and descriptive essays written by three groups of EFL learners from mainland China, Poland and Spain. The argumentative essays show a significant use of hedges and self-mentions, but there is little difference in the use of boosters, attitude markers and engagement markers between these two genres.

### **2.3.4 Corpus linguistic research on disciplines**

Disciplinary variation is another concern of corpus-based research. Comparisons between disciplinary practices, particularly the practices of *soft* and *hard* sciences<sup>1</sup>,

---

<sup>1</sup>Hyland (2004a) discusses hard-soft distinction between knowledge fields, which categorises disciplines of social sciences and humanities such as business studies and applied linguistics as *soft* sciences and those of applied and pure sciences such as electronic engineering and microbiology as *hard* sciences.

---

are popular among researchers, for example, Cortes (2004), Biber (2006), Hyland (2004a, 2007b, 2008b) and Durrant (2015).

Cortes (2004) compared lexical bundles between history and biology articles. Hyland (2008b) examined the use of lexical bundles in research articles, PhD dissertations and masters theses across four disciplines: electronic engineering, microbiology, business studies and applied linguistics. Both Cortes (2004) and Hyland (2008b) found that noun phrase (e.g. *the majority of the, the power of the*) and prepositional phrase bundles mostly with an embedded *of* (e.g. *on the basis of, in the case of*) were more prevalent in the disciplines of social sciences, such as history, business studies and applied linguistics; in contrast, passive verb phrases (e.g. *is shown in Fig., are summarised in Table*) and anticipatory-*it* patterns (e.g. *it is possible that, it is found that*) were important features in the science and engineering writing.

As to discourse functions, Hyland (2004a) found that interactional features were underused in science and engineering texts. Hyland (2007b) further proposes that except for directives the other markers such as hedges (e.g. *may*), boosters (e.g. *definitely*), self-mentions (e.g. *our*), reader pronouns (e.g. inclusive *we*) and questions, are less common in engineering and microbiology papers than in the *soft* fields of marketing, philosophy, sociology and applied linguistics. For directives, cognitive ones (which “instruct readers how to interpret an argument”, e.g. *note, consider*) are predominant in the *hard* sciences to direct knowledge construction and textual ones (which “direct readers to another part of the text or to another text”, e.g. *refer to table 1*) are dominant in the *soft* sciences, leading readers to a reference (Hyland, 2007b, p. 96).

Biber (2006) and Hyland (2008b) investigated the functions of lexical bundles across disciplines. Biber (2006) found there was no difference in the use of discourse organising bundles in his textbook corpus, but stance bundles occurred most frequently in business, and referential bundles were prevalent in the natural and social sciences. Hyland (2008a) classified lexical bundles into three broad categories: research-oriented bundles serve an ideational function in describing real-world research experiences (e.g. *the use of the*); text-oriented bundles fulfil a



---

textual function, concerned with the organisation of the text (e.g. *on the other hand*); and participant-oriented bundles perform an interpersonal function in representing the existence of the writer and the reader of the text (e.g. *as can be seen*). On the basis of this functional framework, Hyland (2008b) identified almost half of the bundles in science and engineering texts as research-oriented bundles and two-thirds of the applied linguistics and business studies bundles as text-oriented bundles. For participant-oriented bundles, the social science articles are concerned with indicating the writer's stance; whereas the *hard* science articles place emphasis on engaging readers. Durrant (2015) adopted Hyland's (2008a) framework to analyse lexical bundles in the corpus of British Academic Written English (BAWE). He confirms Hyland's (2008b) findings and further argues that Science and Technology bundles and Humanities and Social Sciences bundles perform different functions even when they belong to the same category as text-oriented bundles, and the same bundles (e.g. *the centre of the*,) are sometimes used differently by Science and Technology writers or Humanities and Social Sciences writers.

## **2.4 Corpus linguistics and contrastive interlanguage analysis**

The potential of learner corpora has been recognised by many corpus linguists. Flowerdew (2001) explored the role of learner corpora in uncovering learner difficulties in the areas of collocational patterning, pragmatic appropriacy and discourse features and suggests that "insights gleaned from learner corpora need to be employed to complement those from expert corpora for syllabus and materials design" (p. 364). Gilquin, Granger, and Paquot (2007) highlight the advantages of learner corpora over other types of learner data:

The corpora are usually quite large and therefore give researchers a much wider empirical basis than has ever been available before; they can be submitted to a wide range of automated methods and tools which make it possible to quantify learner data, to enrich them with a wide range of linguistic annotations and to manipulate them in various ways in order to uncover their distinctive lexico-grammatical and stylistic signatures. (Gilquin et al., 2007, p. 322)

---

One approach of learner corpus analysis is contrastive interlanguage analysis (CIA) (Granger, 1996), inspired by contrastive analysis (CA) theory. According to Granger (1996), this approach, targeting learner language, involves two types of comparison:

1. NL (native language) vs IL (interlanguage), i.e. the comparison of native and non-native varieties of one and the same language.
2. IL vs IL, i.e. the comparison of different interlanguages of the same language: the English of French learners (E2F), German learners (E2G), Swedish learners (E2S), Japanese learners (E2J), etc. (Granger, 1996, p. 44)

Hunston (2002) proposes that CIA brings two advantages in comparison to other learner language analysis approaches:

1. [I]t makes the basis of the assessment entirely explicit: learner language is compared with, and if necessary measured against, a standard that is clearly identified by the corpus chosen.
2. [T]he basis of assessment is realistic, in that what the learners do is compared with what native/expert speakers actually do rather than what reference books say they do (Hunston, 2002, p. 212).

Therefore, CIA has been very popular among learner corpus researchers (Gilquin et al., 2007). For example, Granger (1998) analysed the use of amplifiers in French students' writing; Shih (2000) examined Taiwanese learners' use of synonyms *big*, *large* and *great*; and Nesselhauf (2003), Marco (2011), and Wang and Zhou (2009) all investigated verb-noun collocations for different learner groups as German, Spanish and Chinese students respectively.

Recently, Granger (2015) has proposed a new version of CIA, CIA<sup>2</sup>. It replaces NL with reference language varieties to cover dialectal variables (e.g. World Englishes, Lingua Franca Englishes) and diatypic variables (e.g. journal articles, undergraduate dissertations), and expands IL to interlanguage varieties so that learner variables and task variables can be highlighted. The new model addresses the criticism that CIA has privileged native norms.

---

The present study will adopt the CIA<sup>2</sup> approach in comparing the use of English sentence initial bundles between New Zealand L1 writers' theses and Chinese L2 writers' theses and comparing the theses written by masters and doctoral students.

It would seem that to trace the IL development, undertaking a longitudinal study of a same group of learners would be the most ideal. As suggested by Huat (2012),

Having collections of data by the same (group of) learner(s), gathered at several points in time, has the extra advantage of illuminating how each of the several states of learner language looks like and relates to others, and how and the extent to which various sub-systems of learner language interact and change over time. (Huat, 2012, p. 196)

Otherwise, for practical reasons, an alternative is to collect samples of language from learners at different developmental stages. It would also be possible to observe salient features of IL at different stages in this case because only high-frequency items, in other words, learner-shared features, are generated as data for corpus analysis.

In the next chapter, I will introduce the concept of lexical bundles, examine their important role in academic writing, and review the work on lexical bundles.

---

## Chapter 3 Lexical bundles

Multiword combinations, with distinctive structures and discourse functions, have been recognised as an essential aspect of vocabulary knowledge and an important focus to support language production (Firth, 1957; Lewis, 2008; Nation, 2001; Nattinger & DeCarrico, 1992; Sinclair, 1991). Lexical bundles, as an important component of multiword combinations, have attracted considerable attention in recent years (e.g. Biber et al., 1999; Chen & Baker, 2010; Cortes, 2004; Hyland, 2008a; Xu, 2012). This study expands the findings of previous research in the area of lexical bundles by focusing on sentence initial bundles, employing Hyland's metadiscourse model, and including L2 writers' interpretations on their own bundle productions.

In this chapter, first, I will compare lexical bundles with the other two popular types of lexico-grammatical associations — collocations and formulaic sequences — to clarify these closely related terms. Second, I will review and evaluate the studies on lexical bundles especially in the area of L2 academic writing in terms of how they account for frequency, structure and function. Third, I will examine the features of two widely-used functional taxonomies: Biber and his colleagues' taxonomy (i.e. referential bundles, discourse bundles and stance bundles) and Hyland's framework (i.e. research-oriented, text-oriented and participant-oriented bundles).

### 3.1 The Concept of lexical bundles

Lexical bundles such as *On the other hand* and *It is important to* are the focus of the present study. The pioneering work of Altenberg (1993, 1998) developed a methodology for generating lexical bundles. Biber and his colleagues (1999), first coined the term *lexical bundles* and defined lexical bundles as recurrent multiword combinations of three or more words, identified empirically on the frequency of co-occurrence and distribution across texts. Biber and Barbieri (2007) propose three major characteristics of lexical bundles: (1) "lexical bundles are by definition extremely common"; (2) "most lexical bundles are not idiomatic in meaning and not perceptually salient"; and (3) "lexical bundles usually do not represent a complete structural unit" (pp. 269-270).

---

Lexical bundles are extremely common discourse building blocks in a given register, and they act as discourse frames to connect to new information (Biber & Barbieri, 2007) or as interactional devices for the involvement of the writer and engagement of target readers (Hyland, 2005c, 2008c). Examples of discourse frames are *the fact that the*, *the results of the* and *in the case of*. Examples of interactional bundles are *as can be seen*, *it should be noted* and *it is interesting to*. As Nation (2013) argues, the criteria of frequency and range of a language item determine its value in learning. Therefore, these bundles, as highly frequent and widely distributed items, deserve sufficient attention in research and pedagogy.

Lexical bundles differ from idioms in frequency and transparency. Lexical bundles are frequency-based linguistic products, mostly occurring 10-40 times per million words. Idioms, such as *kick the bucket* or *raining cats and dogs*, rarely occur in texts. Another difference between lexical bundles and idioms is transparency. Unlike idioms, most bundles are transparent in meaning, but not perceptually salient. A factor is that a large majority of lexical bundles (i.e. 85% of the lexical bundles in conversation and more than 95% in academic prose) are not complete structural units and often do not begin or end at phrase or clause boundaries (Biber et al., 1999). As a result, there is a need to look at the text beyond lexical bundles, that is, the preceding or succeeding words, to provide learners with a broad context. Thus it is important to identify them in relation to registers, disciplines, or genres to inform learners of these ready-made chunks that can potentially serve as language resources.

A number of terms in the literature are closely associated with the term *lexical bundles*, such as *collocations* (Firth, 1957), *formulaic sequences* (Schmitt, 2004), *lexical phrases* (Nattinger & DeCarrico, 1992) and *prefabs* (Erman & Warren, 2000). Among them, *collocations* and *formulaic sequences* are two most popular concepts and there is a need to examine the boundaries between *collocations*, *formulaic sequences* and *lexical bundles* here.

The term *collocation* has been defined diversely by different scholars. There are three major approaches taken to the notion and identification of *collocation*: a frequency-oriented approach, a syntax-oriented approach and a collocability-

---

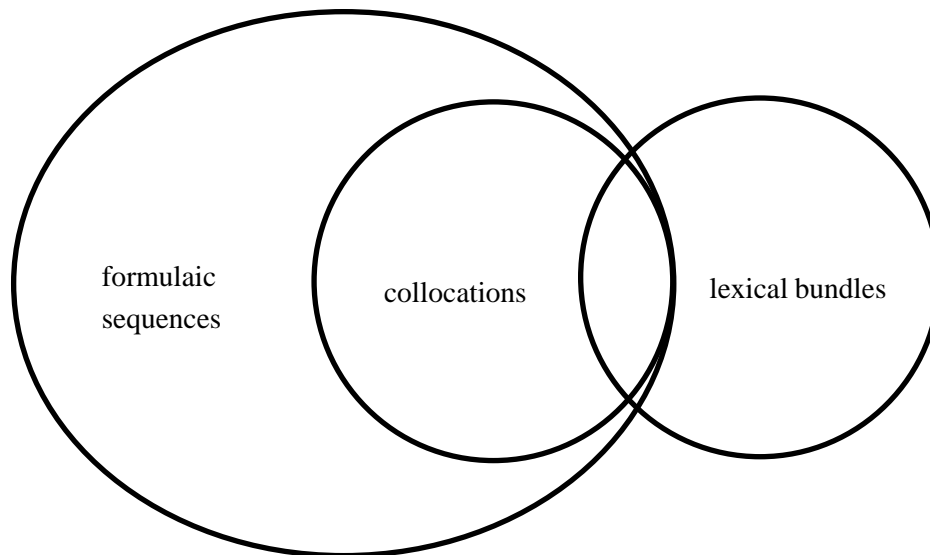
oriented approach. Most definitions incorporate two or more approaches and cover shared important criteria. A frequency-oriented approach regards *collocation* as the statistically significant co-occurrence of words within a short distance (Firth, 1957; Lewis, 2008; Nation, 2013; Nattinger & DeCarrico, 1992; Nesselhauf, 2004; Sinclair, 1991). A syntax-oriented approach emphasises the grammatical structure of *collocation* (Firth, 1957; Nation, 2013; Nattinger & DeCarrico, 1992; Nesselhauf, 2004; Sinclair, 1991) and identifies collocations by syntactic structures. According to Benson (1990), who uses a syntax-oriented approach, collocations can be further divided into lexical collocations and grammatical collocations according to their grammatical structures. Lexical collocations normally consist of nouns, verbs, adjectives and adverbs and none of these words are dominant words (e.g. *tackle the problem, heavy smoker, widely available*). Grammatical collocations consist of a dominant word (e.g. noun, verb or adjective) and a preposition or grammatical structure such as an infinitive or clause (e.g. *prepare for, necessary to work, agreement that*). A collocability-oriented approach highlights the mutual expectancy between words, i.e. the likelihood that items will co-occur (Lewis, 2008; Nation, 2013; Nattinger & DeCarrico, 1992; Nesselhauf, 2004). Collocations are positioned on a continuum, with completely invariant combinations (e.g. *by the way*) at one end and freely combining phrases (e.g. *drink tea*) at the other.

Formulaic sequences cover a wide range of formulaic language that occurs in a sequence and as a whole (Schmitt & Carter, 2004; Wray, 2002). Formulaic sequences can be continuous (e.g. *by and large*) or discontinuous (e.g. *the greater X, the better Y*). They can be as long as a whole sentence (e.g. *You can choose your friends, but you can't choose your family.*) or as short as a couple of words (e.g. *blonde hair*) (Schmitt & Carter, 2004). Along with collocations, formulaic sequences consist of idioms (e.g. *kick the bucket*), polywords (e.g. *by the way*), institutionalized expressions (e.g. *How are you?*), phrasal constraints (e.g. *a month/year ago*) and sentence builders (e.g. *not only X, but also Y*). Details of categories and definitions can be found in Nattinger and DeCarrico (1992).

From the above definitions of lexical bundles, collocations and formulaic sequences, it is clear that the most significant similarity between these three types of sequences is fixedness, which means they are somewhat *frozen* semantically and

---

grammatically. This distinguishes them from free combinations. However, these three terms vary in degree of fixedness: lexical bundles are completely fixed, collocations are semi-fixed and formulaic sequences can be both fixed and semi-fixed. In addition, lexical bundles differ from the other two terms in their arbitrary frequency-based identification criteria. All the bundles are generated on the basis of their cut-off criteria (i.e. the minimal number of occurrences and distribution of texts), arbitrarily set by the researchers. Lexical bundles also differ from the other two terms in their incomplete structural units. Most bundles, such as *the end of the* and *the extent to which*, only represent part of phrases or clauses. The relationships between these three terms are shown in Figure 1 below.



**Figure 1. The relationship between lexical bundles, collocations and formulaic sequences**

### **3.2 Lexical bundles and academic writing**

As discussed above, lexical bundles play an important role in signifying fluency, accuracy and idiomaticity in academic writing:

- Lexical bundles are a major component of academic writing; therefore, they deserve special attention.

- 
- Lexical bundles serve as “points of fixation” to construct academic writing.
  - Appropriate academic writing requires bundle knowledge to achieve idiomaticity, accuracy and fluency.
  - Lexical bundles are important indicators of one’s membership of a specific discourse community because they comply with conventional expressions.

Many researchers suggest that a large proportion of language is constituted by pre-assembled lexical chunks (Bolinger & Sears, 1981; Erman & Warren, 2000; Nattinger & DeCarrico, 1992; Schmitt & Carter, 2004). As a type of lexical chunk, lexical bundles form a major component of academic writing, which can be observed from their frequency and coverage. Biber and his colleagues (1999) report that three-word bundles occur over 60,000 times per million words and four-word bundles, over 5,000 times in academic prose. The most frequent three-word bundles (e.g. *in order to*, *one of the*, *the fact that*) feature over 200 times per million words and four-word bundles (e.g. *in the case of*, *on the other hand*, *as well as the*), over 100 times. These recurrent multiword combinations make up of about 21% of the 5.3 million academic prose collection of the Longman Spoken and Written English Corpus. Hyland (2008b) extracted 240 bundles from his 3.5 million academic corpus and among them, the most frequent one, *on the other hand*, occurs about 200 times per million words. As Nation (2013) highlights, more frequent items (e.g. lexical bundles) deserve special attention by teachers and learners.

Lexical bundles function as starting points of texts, where writing can be expanded. This idea is supported by the concept of “islands of reliability” introduced by Dechert (1984). He argues that possessing a certain amount of automatized prefabricated language is necessary so that writers can have “points of fixation, anchoring ground to start from and return to” (Dechert, 1984, p. 223). A writer’s competence is influenced by the size of his or her *island* repertoire. Coxhead and Byrd (2007) further explain that student writers, equipped with the word sets of advanced writers, write more efficiently because they do not need to create sentences word by word.



---

Lexical bundles have great potential in fostering idiomaticity, accuracy and fluency to meet the requirements of appropriate academic writing. Pawley and Syder (1983) argue that the fixedness of prefabricated language (including lexical bundles) explains two linguistic capacities — nativelike selection and nativelike fluency. The former refers to the ability to pick up a natural expression from a range of grammatically correct paraphrases and the latter is the ability to encode one clause at a time to produce fluent utterances. Sinclair (1991) proposes two models of language processing: the open-choice principle and the idiom principle, and suggests “the first mode to be applied is the idiom principle” (p.114). The open-choice principle is also known as a “slot-and-filler” model, that is, any word can be filled in each slot of a text and grammaticalness is the only constraint. Contrary to the open-choice principle, the idiom principle suggests “a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments” (p. 110). The use of lexical bundles is in accordance with Sinclair’s idiom principle and bundle knowledge will equip writers with the ability to apply the idiom principle.

The use of lexical bundles as conventional expressions reflects one’s membership of his or her discourse community. Stubbs (2001) argues that “[r]epeated patterns show that evaluative meanings are not merely personal and idiosyncratic, but widely shared in a discourse community” (p. 215). Wray (2002) regards familiarity with formulaic sequences as the signal of gaining membership of the target community. Jones and Haywood (2004) further claim that the use of these expressions allows the writer to express technical ideas economically, display the necessary level of formality and mark stages in a text; while their absence may indicate inadequate writing and peripheral participation of the community.

### **3.3 Studies on lexical bundles**

There has been a sharp rise in the study of lexical bundles after the pioneering study of Altenberg (1993, 1998). Bundles have been investigated in relation to languages (Kaneyasu, 2012; Kim, 2009; Tracy-Ventura et al., 2007), registers (Biber, 2006; Biber & Barbieri, 2007; Biber et al., 2004; Biber et al., 1999; Herbel-Eisenmann & Wagner, 2010; Jablonkai, 2010; Neely & Cortes, 2009; Nesi & Basturkmen, 2006;

---

Schnur, 2014), genres (Chen, 2010; Hyland, 2008a; Qin, 2014; Xu, 2012), disciplines (Biber, 2006; Cortes, 2004; Hyland, 2008b; Pecorari, 2009), academic competence (Chen & Baker, 2010, 2014; Cortes, 2002, 2004; Karabacak & Qin, 2013; Qin, 2014), varieties of English (Liu, 2012), moves (Cortes, 2013) and language proficiency (Ädel & Erman, 2012; Allen, 2009; Chen & Baker, 2010; Hyland, 2008a; Karabacak & Qin, 2013; Pang, 2009; Pérez-Llantada, 2014; Salazar, 2014; Staples, Egbert, Biber, & McClair, 2013; Wei & Lei, 2011; Xu, 2012).

The methodology of all these studies and the frameworks or taxonomies that are used or developed for data analysis are largely the same. Frequency, structure and function are three typical research foci. Most researchers have investigated four-word lexical bundles with an occurrence of more than 20 times per million words across 3 to 20 texts. The categories in Biber et al. (1999) have been most commonly adopted to analyse structural patterns. Biber and his colleagues' taxonomy (Biber & Barbieri, 2007; Biber et al., 2003, 2004) (i.e. referential, discourse and stance bundles) or Hyland's (2008a) framework (i.e. research-oriented, text-oriented and participant-oriented bundles) has been extensively used to examine functions. The reason to undertake function-based analysis, alongside structure-based analysis, is that these bundles are not merely high-frequent multi-unit combinations, but "contribute to text meaning, and give each genre its characteristic identity by serving particular functions" (Qin, 2014, p. 224). As the focus of the current research is on Chinese masters and PhD thesis writing, the following review mainly covers studies in the area of L2 academic writing.

### **3.3.1 Frequency-based analysis**

Frequency-based<sup>2</sup> comparisons have been included in a number of studies to investigate the differences in the use of bundles in academic writing between non-native writers and native (or advanced) writers (e.g. Ädel & Erman, 2012; Chen & Baker, 2010; Hyland, 2008a; Pang, 2009; Pérez-Llantada, 2014; Staples et al., 2013; Wei & Lei, 2011; Xu, 2012). Three studies, Chen and Baker (2010) Ädel and Erman

---

<sup>2</sup>Chen and Baker (2010) distinguish the type and token of lexical bundles in their study with the former referring to unique bundles and the latter describing total occurrences of bundles. However, only the number of types are presented and compared in most studies, so "frequency" in this section refers to the number of bundle types.

---

(2012) and Staples et al. (2013), calculated the number of bundles both before and after the removal of the content-dependent bundles, which refer to the bundles like *financial and non financial*, *between men and women* and *the Second World War*, but other studies do not include this data refinement stage. In addition, this refinement stage with a few bundles excluded will not change the general trend in most cases (see Table 4 in Y.-H. Chen & Baker 2010 for more details). Therefore, the comparisons in this section are based on raw data, that is, the number of bundles before the removal of content bundles.

The correlation between frequency of bundles and writers' English proficiency has been examined in many studies, in correlation with L2 writers' academic levels. It has been found that from undergraduates to academics, the frequency of lexical bundles in L2 writers' texts forms a bell shape with masters level students using the most bundles. From undergraduate to taught masters level, students generally lack knowledge and awareness of lexical bundles, and bundles are consistently underused in the student writing. Chen and Baker (2010) and Ädel and Erman (2012) identified fewer types of bundles in Chinese and Swedish student writing, compared to the number in the work of their native peers, L1 English students in British universities (i.e. 90 compared to 120, 60 compared to 130). In line with their findings, Xu (2012) also found fewer bundles in her Chinese undergraduate texts than those in published research articles. The only exception is P. Pang's (2009) work. She reports the significantly greater use of bundles in the essays written by university undergraduates in China than their native counterparts in the British and American universities (i.e. 861 compared to 263). However, as she acknowledges, the range of essay topics in her two student corpora, LOCNESS (Louvain Corpus of Native English Essays) and WECCL (Written English Corpus of Chinese Learners), might greatly affect the number of generated bundles. LOCNESS covers about 100 topics; in contrast, WECCL, more than three times as large as LOCNESS, only includes essays of 17 topics. As a result, probably more topic-related bundles have been extracted from WECCL.

At the higher levels, the number of bundles decreases as the level of study increases from masters, to PhD, to academics and the changes can be observed in Hyland (2008a), Wei and Lei (2011), Xu (2012) and Pérez-Llantada (2014). Hyland (2008a)

---

examined masters and PhD theses at five Hong Kong universities in comparison to published research articles. His analysis indicates that a considerable higher reliance on bundles among the less proficient writers, particularly masters students (i.e. 149, 95 compared to 71). Wei and Lei (2011) support Hyland's (2008a) finding and retrieved 154 bundles in their doctoral dissertation corpus and 87 bundles in the corresponding journal article corpus in the discipline of applied linguistics. Xu (2012) compared both masters and PhD theses with journal articles in the same discipline (i.e. linguistics or applied linguistics) and also found that theses produced by Chinese learners generally contained more bundles (i.e. 367, 168 compared to 169), with the most bundles occurring in the masters theses. Pérez-Llantada (2014) investigated English research articles written by Spanish scholars and native scholars. She found that the Spanish scholars incorporated more bundles in their academic writing (i.e. 77 compared to 56).

At the same time, Hyland (2008a) and Xu (2012) argue that the greatest use at masters level will not result in more overlapping bundles between these writers' texts and published writing. Qin's (2014) study also affirms that the number of overlapping bundles steadily increases along with levels of study from non-native masters and PhDs to native academics.

### **3.3.2 Structural analysis**

Many studies have investigated structural patterns using Biber and his colleagues' categories. Significant differences were identified between the bundles used in L2 students' writing (mainly Chinese learner writing) and those used in native writing or professional works. What has been found to be significant are the patterns that feature in academic texts: noun phrases, prepositional phrases, passive verb phrases and anticipatory-*it* patterns.

#### *3.3.2.1 Structural categories*

On the basis of the Longman Spoken and Written English Corpus, Biber and his colleagues (1999) identified twelve widely-used structural patterns in academic prose, which are:

1. noun phrase with *of*-phrase fragment

- 
2. noun phrase with other post-modifier fragment
  3. prepositional phrase with embedded *of*-phrase fragment
  4. other prepositional phrase fragment
  5. anticipatory *it* + verb phrase/adjective phrase
  6. passive verb + prepositional phrase fragment
  7. copula *be* + noun phrase/adjective phrase
  8. (verb phrase +) *that*-clause fragment
  5. (verb/adjective +) *to*-clause fragment
  6. adverbial clause fragment
  7. pronoun/noun phrase + *be* (+ ...)
  8. other expressions

Biber et al. (2004) later developed three broad structural categories to group their structural patterns featuring in conversation, university teaching, textbooks and academic prose. These categories were: bundles incorporating verb phrase fragments, dependent clause fragments, and noun or prepositional phrase fragments. Along with Biber et al. (2004) but only focusing on academic writing, Chen and Baker (2010) distinguished another three major categories: noun phrase based (NP-based), preposition phrase based (PP-based), and verb phrase based (VP-based) bundles.

### 3.3.2.2 *Studies on structural analysis*

In spoken discourse, about 90% lexical bundles incorporate elements of verb phrases and among these, 50% begin with a personal pronoun followed by a verb phrase (e.g. *I want you to*), 19% are extended verb phrase fragments (e.g. *take a look at*) and 17% are question fragments (e.g. *do you want to*) (Biber et al., 2004; Biber et al., 1999). In academic prose, however, noun phrases (e.g. *the use of the*) and prepositional phrases (e.g. *in the present study*) comprise over 60% lexical bundles (Biber et al., 2003, 2004; Biber et al., 1999). This supports the generally held view that academic writing contains many noun and preposition phrases. Together with passive verb phrases (e.g. *can be found in*) and anticipatory-*it* patterns (e.g. *it is important to, it was found that*), these four structures are the most common patterns of lexical bundles in academic writing (Hyland, 2008a).

---

Therefore, in this section, I will mainly report the identified differences between L2 writing (particularly Chinese learner writing) and native or professional writing in regard to these four patterns.

Noun phrases mostly with an embedded *of* were found to occur more frequently in essays written by native writers or in journal articles (Chen & Baker, 2010; Hyland, 2008a; Pang, 2009; Xu, 2012). Chinese undergraduates and masters students do not appear to recognise the importance of this structure (Chen & Baker, 2010; Hyland, 2008a; Pang, 2009; Xu, 2012), but the distribution of noun phrase bundles in Chinese PhD writing tends to be fairly close to that of published writing (Wei & Lei, 2011).

The use of prepositional phrase bundles in Chinese student writing has been found to increase along with their levels of study. At the undergraduate level, Chinese students use considerably fewer bundles than native writers (Pang, 2009). From undergraduate to taught masters level, they employ a similar proportion of PP-based bundles to native and expert writers, slightly higher than their native peers but lower than expert writers (Chen & Baker, 2010). At the PhD level, they turn to rely more heavily on PP-based bundles in comparison to research masters and expert writers (Hyland, 2008a).

Passive verb patterns were rarely found in low-level L2 students' writing (i.e. Chinese and Swedish university writing) (Ädel & Erman, 2012; Chen & Baker, 2010), but frequent in high-level Chinese students' writing (i.e. masters and PhD theses) (Hyland, 2008a; Wei & Lei, 2011).

The use of anticipatory-*it* structures differs across studies. Hyland (2008a), and Ädel and Erman (2012) found that anticipatory-*it* patterns were more common in Hong Kong and Swedish students' writing. In contrast, Xu (2012), and Wei and Lei (2011) report that Chinese learners employ fewer anticipatory-*it* bundles than expert writers.

Differences between the Chinese students' writing and native or professional writing are also evident in the use of another two patterns: *to*-clause fragments and existential *there* constructions. Chinese students show a strong preference for *to*-

---

clause fragments, especially the structure *(in order) to + verb* (Chen & Baker, 2010; Pang, 2009). Native writers, compared with Swedish students in Ädel and Erman's (2012) study, use *there be* bundles to a much greater extent.

The review of these studies has presented a general picture of L2 learners', particularly Chinese students' bundle distribution in terms of grammatical structures. Functional analysis, a supplement to structural analysis, will be discussed in the next section.

### 3.3.3 Functional analysis

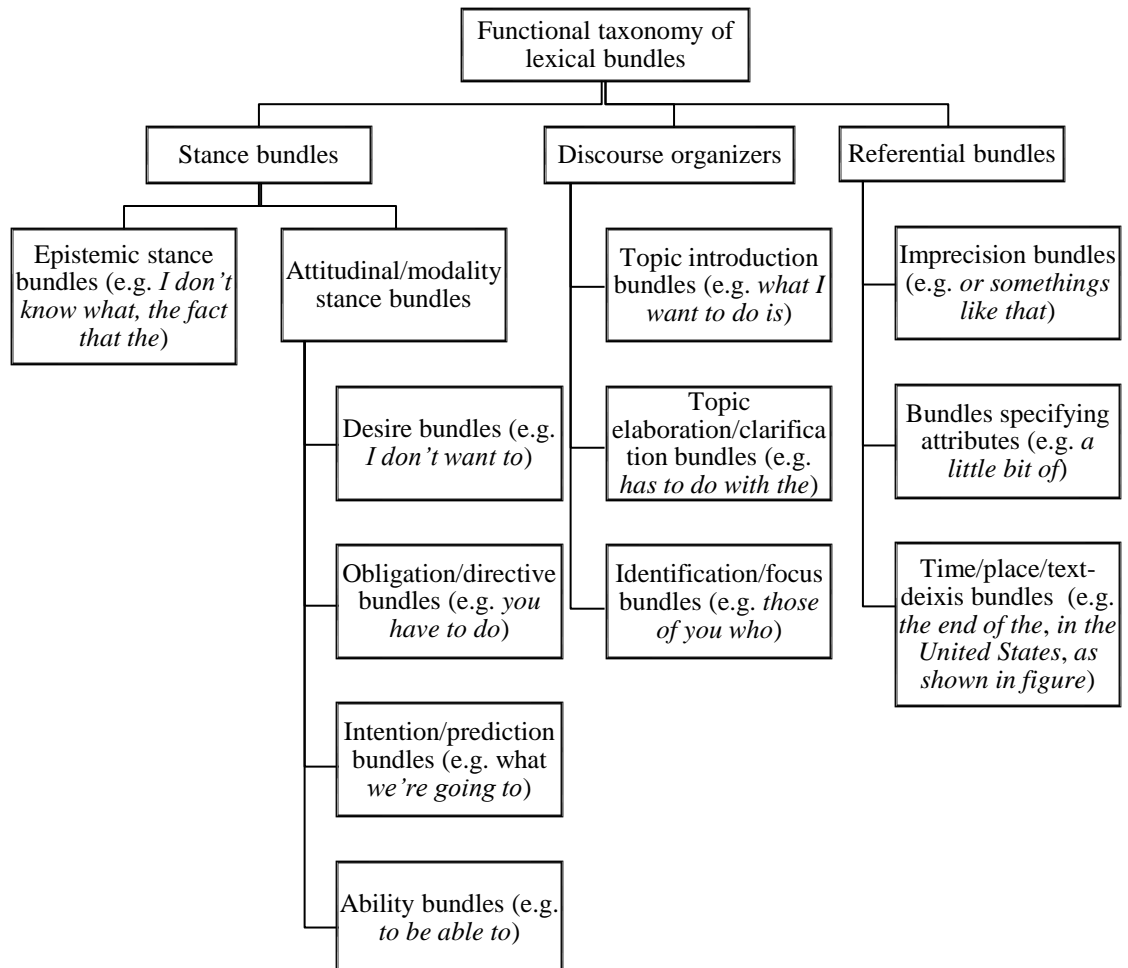
Functional analysis, focusing on the intrinsic functions of lexical bundles, is another major perspective in the area of lexical bundle research. Two existing taxonomies — Biber and his colleagues' taxonomy (i.e. referential, discourse and stance bundles) and Hyland's taxonomy (i.e. research-oriented, text-oriented and participant-oriented bundles) — have been widely adopted or adapted in almost all studies and the frequency of lexical bundles have been compared in each category.

#### 3.3.3.1 Biber and his colleagues' taxonomy

A series of studies (e.g. Ädel & Erman, 2012; Biber & Barbieri, 2007; Biber et al., 2003, 2004; Chen & Baker, 2010; Cortes, 2004, 2013) have used and developed Biber and his colleagues' taxonomy (Biber & Barbieri, 2007; Biber et al., 2003, 2004). The taxonomy distinguishes three primary functions: stance expressions, discourse organisers and referential expressions.

Stance bundles express attitude or assessment of certainty. The former are known as attitudinal or modality stance bundles (e.g. *I don't want to*) and the latter are epistemic stance bundles (e.g. *the fact that the*). Attitudinal bundles are further categorised as desire bundles, obligation/directive bundles, intention/prediction bundles and ability bundles. Discourse organisers, the second function, indicate text-internal relationships, which include topic introduction bundles, topic elaboration/clarification bundles, inferential bundles, contrast/comparison bundles, framing bundles, etc. Referential bundles, the third group, perform four main functions in indicating imprecision, introducing quantities, specifying attributes and referring to particular times, places or units of texts. Figure 2 shows the categories

and sub-categories of Biber and Barbieri's (2007) taxonomy along with sample bundles from their work. Biber and Barbieri's (2007) taxonomy is a development of the taxonomies in Biber et al. (2003) and Biber et al. (2004), which maintains the three primary categories but differs in some sub-categories.



**Figure 2. Biber and Barbieri's (2007) functional taxonomy of lexical bundles**

Biber and his colleagues' taxonomy, as pioneering work on functional analysis, provides a comprehensive view of major discourse functions of lexical bundles. However, it should be noted that the development of this taxonomy is largely based on the lexical bundles in spoken rather than written registers because a greater range of lexical bundles was generated from the spoken registers used in the corpus. This



---

is demonstrated in Biber et al.'s (2004) study, which aims to develop the functional taxonomy. The number of lexical bundles in each corpus is presented as follows: 43 in conversation, 84 in classroom teaching, 27 in textbooks and 19 in academic prose. The strong bias of this taxonomy towards spoken registers can also be seen from the overwhelming proportion of personal bundles in the category of stance expressions.

According to Ädel (2010), spoken discourse is distinct from written discourse in at least two ways: simultaneous output and the presence of an audience. Therefore, it may not be appropriate to adopt Biber and his colleagues' taxonomy to analyse written academic texts.

Besides the differences in the nature of the corpora on which the taxonomies are based, it is important to note that some criteria and categories of these bundles are not consistent between researchers and the inconsistencies can be found in three aspects:

1. Different sub-categories are created to refer to same functions.
2. The same sub-categories are grouped into different categories.
3. The same bundles are placed into different sub-categories.

Appendix A summarises how Biber and Barbieri's (2007) functional taxonomy has been expanded or altered by researchers.

First, different sub-categories are created to refer to the same functions. Inferential bundles (in Ädel & Erman, 2012; Biber et al., 2003; Chen & Baker, 2010; Cortes, 2004) and contrast/comparison bundles (in Biber et al., 2003; Cortes, 2004) indicate the logical relationships between units of texts, which are, in Biber and Barbieri's (2007) work, the components of topic elaboration/clarification bundles. Frame bundles (in Ädel & Erman, 2012; Biber et al., 2003; Chen & Baker, 2010; Cortes, 2004, 2013) identify textual conditions. Quantifying bundles (in Ädel & Erman, 2012; Chen & Baker, 2010; Cortes, 2004) mainly introduce quantities. In Biber and Barbieri's (2007) study, both frame bundles and quantifying bundles are referred to as bundles specifying attributes.

Second, the same sub-categories are grouped into different categories. In earlier studies, frame bundles were grouped into discourse organisers (e.g. Biber et al., 2003; Cortes, 2004), but in recent studies, frame bundles are regarded as referential expressions (e.g. Ädel & Erman, 2012; Chen & Baker, 2010; Cortes, 2013). Focus bundles (e.g. *one of the most*, *one of the things*) are referred to as both discourse organisers and referential expressions. If researchers are highlighting the function beyond the bundle, that is, introducing or summarising the main points of the texts, they classify focus bundles as discourse organisers (e.g. Ädel & Erman, 2012; Biber & Barbieri, 2007; Chen & Baker, 2010). If researchers are focusing on the function within the bundle, that is, the bundle is used to emphasise its succeeding noun phrase, they regard focus bundles as referential expressions (e.g. Biber et al., 2004; Cortes, 2013). One special case is Cortes (2004). The focus bundles in her study are to pinpoint the importance or difficulties posed by the succeeding statements, and so are categorised as discourse organisers.

Third, the same bundles are placed into different sub-categories. Some examples can be found in Appendix A. *On the basis of* belongs to both inferential bundles and frame bundles, *the extent to which* appears in frame bundles and quantifying bundles, and *one of the most* are labelled as quantifying bundles and focus bundles. One of the reasons for the vague divisions is that lexical bundles are always polypragmatic in that one bundle often serves more than one function simultaneously (K. Hyland, personal communication, March 15, 2014).

### 3.3.3.2 Hyland's framework

Hyland (2008a), more recently, introduced another framework, which is based on Halliday's (1994) theory of systemic functional linguistics and includes three broad metafunctions of language. These are:

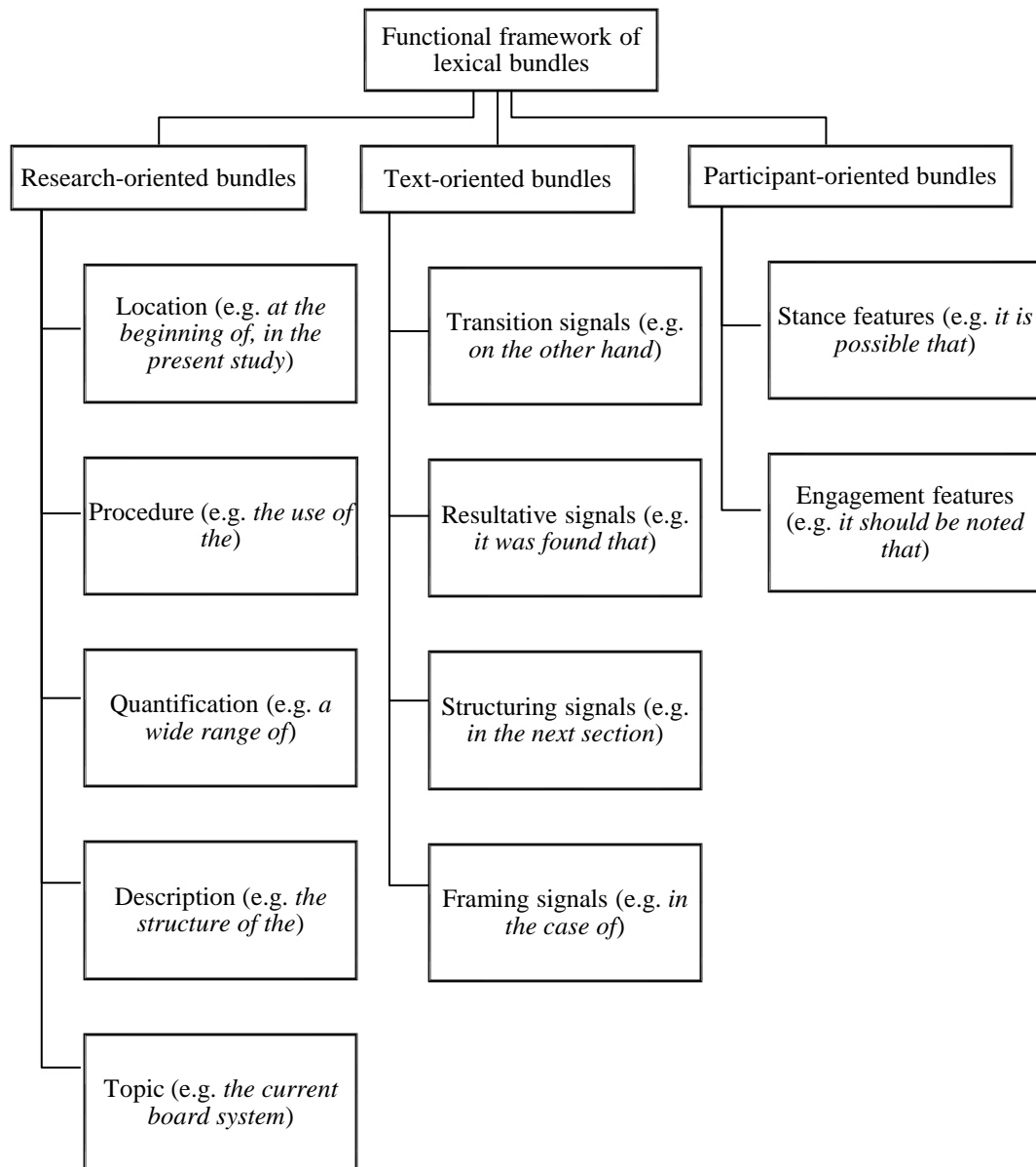
Ideational metafunction: the use of language to construe real-world experience or ideas.

Interpersonal metafunction: the use of language to encode interaction, indicating personal feelings and evaluations or engaging with audiences.

Textual metafunction: the use of language to organise texts to create cohesion and continuity.

---

Hyland's (2008a) framework has drawn on Biber and his colleague's classification, but differs from Biber's taxonomy in that it specifically reflects the characteristics of written research-focused genres on the basis of his three electronic corpora. These corpora are composed of 120 research articles, 80 doctoral dissertations and 80 masters theses from four disciplines: electrical engineering, business studies, applied linguistics and microbiology. Hyland's (2008a) framework also contains three classifications: research-oriented, text-oriented and participant-oriented bundles. As shown in Figure 3, research-oriented bundles serve an ideational function in describing real-world research experiences such as location, procedure, quantification, description and topic. Text-oriented bundles fulfil a textual function, concerned with the organisation of the text, which include transition signals, resultative signals, structuring signals and framing signals. Participant-oriented bundles perform an interpersonal function in representing the existence of the writer and the reader of the text by means of stance features and engagement features. The criteria of each sub-category can be found in Hyland (2008a).



**Figure 3. Hyland's (2008a) functional framework of lexical bundles**

Despite the popularity of both Biber and Hyland's taxonomies, Byrd and Coxhead (2010) criticise their complexity as merely research-oriented systems, which are not applicable to classroom teaching and learning. At the same time, they point out the similarity between Biber and Hyland's categories, suggesting that "whatever terms are used, these systems generally include three basic categories: 'presentation of

---

content’ and ‘organization of the discourse/text’ and ‘expression of attitudes by the writer/speaker’” (p. 42).

### 3.3.3.3 *Studies on functional analysis*

Functional analysis has been included in almost all lexical bundle studies. Many studies have used Biber and his colleagues’ taxonomy due to the fact that their taxonomy is the pioneering work in this area. Two recent studies on academic writing have adopted Hyland’s framework.

Biber et al. (2004), and Biber and Barbieri (2007), using their own taxonomy, investigated the use of lexical bundles in a variety of university spoken and written registers (e.g. classroom teaching, textbooks and academic prose), and revealed that stance bundles and discourse organisers were common in conversation, while referential bundles were normally used in academic texts. Many other researchers have been interested in using this taxonomy to compare lexical bundles in student academic writing and published works. For example, Chen and Baker (2010) compared the Chinese student essays produced in British universities with the native speaker students’ writing and published academic texts. Ädel and Erman (2012) examined the English writing of British and Swedish students. Xu (2012) focused on Chinese undergraduate, masters and PhD theses in mainland China and compared them with published journal articles. Pérez-Llantada (2014) investigated the use of bundles in English articles written by English L1 and Spanish L1 scholars. Pang (2009) compared the English essays written by American and Chinese university students. A similar bundle distribution was identified across the first four studies: stance and referential bundles were used extensively in the native student and published texts, and discourse organisers were prevalent in the L2 student or published writing. Pang’s (2009) study, however, found both discourse organisers and referential bundles were heavily used in the American student essays in comparison to the Chinese student corpus.

The application of Hyland’s framework was mostly found in the analysis of academic writing. Hyland (2008a) studied the masters theses and PhD dissertations written by Cantonese L1 writers at Hong Kong universities and compared their lexical bundles with that of research articles. Wei and Lei (2011) analysed the PhD

---

dissertations written by Chinese L1 writers in mainland China and compared them with published journal articles. The findings of these two studies show that L2 masters theses contain the most research-oriented, and the least text-oriented bundles and participant-oriented bundles. In contrast, research articles contain the least research-oriented, and the most text-oriented bundles and participant-oriented bundles. The distribution of bundles in L2 PhD dissertations is closer to that of research articles than masters theses. This suggests attempts of PhD students to consider audience rather than merely report research.

In addition to overall comparisons, Chen and Baker (2010), Ädel and Erman (2012), and Hyland (2008a) highlight the difference between native and non-native writers in the use of hedge bundles: native writing manifests a wider range of hedging expressions. Hyland (2008a) also found the relative absence of stance bundles in his advanced L2 student texts. This is not surprising given the fact that these students may not feel comfortable and confident to explicitly express their personal evaluations.

### **3.3.4 Possible explanations of L2 student bundle choices**

As seen above, the research on lexical bundles has been very much text focused. However, in line with the “social turn” (Block, 2003) in applied linguistics, learners’ choices and use should complement textual analysis. A few scholars have attempted to explore the reasons for the discrepancy of L2 student bundle choices (e.g. Allen, 2009; Cortes, 2004; Hyland, 2008a; Paquot, 2013; Qin, 2014; Wei & Lei, 2011). Except for Paquot’s (2013) statistical measure (i.e. ANOVA test and Dunnett’s tests) of L1 (French) transfer effects on English texts, other researchers mainly subjectively suggest the reasons for discrepancy.

Factors that possibly contribute to student bundle production include familiarity with linguistic items (Cortes, 2004), content issues (Cortes, 2004), noticing in reading (Cortes, 2004; Wei & Lei, 2011), learning experience (Hyland, 2008a; Wei & Lei, 2011), cultural preference (Hyland, 2008a; Wei & Lei, 2011), rhetorical confidence (Hyland, 2008a), text length (Allen, 2009; Qin, 2014), interlingual transfer (Allen, 2009) and reader awareness (Qin, 2014). Cortes (2004) suggests that students tend to use more familiar bundles and avoid unfamiliar ones such as

---

some referential bundles (e.g. *in the course of*). Bundles related to specific issues (e.g. *on the evolution of*) are also absent from student writing. Cortes (2004), and Wei and Lei (2011) believe that students lack sufficient exposure to readings and conscious learning of target bundles. Hyland (2008a), and Wei and Lei (2011) attribute impersonality in Chinese student writing to teaching materials and practices, and to cultural preferences. Hyland (2008a) considers confidence as another factor for the absence of stance bundles in Hong Kong PhD student dissertations. Allen (2009) and Qin (2014) propose the length of student writing possible affects the number of text organisers (e.g. *in the next section*). Allen (2009) also acknowledges the role of linguistic transfer and he believes the bundle *it can be said* in his Japanese student corpus is the result of interlingual transfer from the Japanese expression *to iwareru*. Qin (2014) adds audience awareness as another reason by arguing that raising students' awareness will possibly help them to produce clearer and more consistent writing.

An investigation of the reasons for L2 student bundle choices could provide further evidence towards learner language production and better inform L2 academic writing pedagogy. There is undoubtedly a need to further explore learner interpretations of their own bundle choices in different contexts.

### **3.4 Limitations of the existing research**

Lexical bundles are an important component of academic writing and useful for language production. A range of studies has investigated their use in various genres of academic writing in terms of frequency, structure and function. These studies have provided a justification for investigating and teaching lexical bundles. However, they have ignored the differences between genres, overlooked the influences of academic competence, conflated sentence initial and non-initial bundles, and have failed to consider the writers of texts and their reasons for selecting particular bundles.

Many studies, such as Cortes (2004), Hyland (2008a), and Wei and Lei (2011), have focused on comparing various genres regardless of the differences between genres and writers' academic competence (e.g. published research articles compared to doctoral dissertations). These studies have overlooked the factors such as

communicative goals, text lengths and writer identities; therefore, it is difficult to attribute the different distribution of bundles to any variable (e.g. language proficiency).

Another limitation is caused by the shortcomings of computer software used in these studies, such as AntConc, a free corpus analysis toolkit for concordancing and text analysis. Bundles at the beginning or second part of sentences (i.e. sentence initial and non-initial bundles) have been conflated, although they perform distinctly different functions. See for example Cortes's (2013) discussion of triggers or complements.

The studies are further limited by the focus on overall comparisons rather than comparisons within each subcategory (e.g. comparisons of epistemic stance bundles), on global quantitative approach rather than in-depth context-based qualitative analysis. The latter would provide learners with a good knowledge of when and how to use these bundles. The knowledge of lexical bundles in professional writing can effectively facilitate learner writing. Understanding of learner bundles can support learners' self-reflection of their own language production and at the same time can help learners to consciously avoid inappropriate expressions.

Possible explanations of student bundle choices have been covered in a few studies; however, only Paquot's (2013) research directly tested one of the hypotheses, the L1 effects. Little research has explored the reasons for discrepancy in student bundle choices. The present study intends to fill this gap by interviewing a group of Chinese postgraduates to find out these L2 learners' interpretations. It is important to find out the sources of learner bundles as this would complement the existing bundle knowledge and provide useful first-hand information for EFL or ESL (English as a second language) pedagogy.

In the following chapter, I will explore the concept of metadiscourse, discuss the relationship between metadiscourse and lexical bundles, and review the work on metadiscourse.





---

## Chapter 4 Metadiscourse

Hyland's metadiscourse model, in comparison to Biber and Hyland's taxonomies on lexical bundles, appears to be more useful to analyse the functions of sentence initial bundles and more applicable to classroom teaching and learning. In this chapter, I will introduce the concept of metadiscourse, discuss the relationship between metadiscourse and lexical bundles, outline the development of metadiscourse models and review the studies on metadiscourse.

### 4.1 The concept of metadiscourse

Metadiscourse has been used as an umbrella term in discourse studies since the 1980s. Williams (1981) defines it as "writing about writing, whatever does not refer to the subject matter being addressed" (pp. 211-212). Vande Kopple (1985), from a reader's perspective, suggests that metadiscourse serves as cues to "help our readers organize, classify, interpret, evaluate, and react to such material (propositional content)" (p. 83). Crismore et al. (1993), from a writer's perspective, propose that "metadiscourse allows writers to show readers how different parts of the text are related and how they should be interpreted. Metadiscourse also permits writers to express their attitudes toward the propositional content of the text and toward their readers" (p. 40). Hyland (2005a) regards metadiscourse as a facilitator of interpersonal communication, "assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community" (p. 37). Ädel (2006) defines metadiscourse as reflexive linguistic expressions, "[displaying] an awareness of the current text or its language use *per se*" (p. 20).

Metadiscourse is a fuzzy category with blurred boundaries, which "can only be defined by taking into account a fairly large number of different criteria ranging from necessary conditions to characteristic properties" (Ädel, 2006, p. 22). Therefore, Hyland (2005a) highlights three key principles to set the criteria for metadiscourse "1. that metadiscourse is distinct from propositional aspects of discourse; 2. that metadiscourse refers to aspects of the text that embody writer-reader interactions; [and] 3. that metadiscourse refers only to relations which are internal to the discourse" (p. 38). The first principle is a shared feature between metadiscourse devices and lexical bundles (except for content-based ones) and one

---

of the main reasons to adapt a metadiscourse model to analyse sentence initial bundles in the current study. The second principle distinguishes the concept of metadiscourse from a narrow view of metatext, which according to Mauranen (1993) refers only to items of textual organisation and excludes devices facilitating writer-reader interactions. In other words, metatext covers Hyland's (2005a) interactive rather than interactional category of metadiscourse, which will be further discussed in the next section. The third principle differentiates between text-internal and text-external references. The following two examples are from Hyland (2005a). The first *therefore* is a metadiscourse resource in signalling the consequence of the preceding text (1), while the second *therefore* expresses the relation between activities in the real world, which are external to the text (2).

- (1) *The poll was taken just after this month's messy reshuffle and puts the Tories on 33 points, Labour on 32 and the Liberal Democrats on 25. **Therefore**, on today's results the Tories would gain an extra 41 seats and the Lib Dems 20 in the next election, leaving Blair with an uncomfortably narrow majority.*  
(newspaper article)
- (2) *We understand that the idea of moving your account to us may be daunting, **therefore** we will do most of it for you.* (bank advertisement)

Ädel (2006) specifies five features of metadiscourse: explicitness, world of discourse, current discourse, writer *qua* writer, and reader *qua* reader. Explicitness refers to the words used in text, not typographical marking such as italics and boldface. World of discourse refers to discourse-internal rather than discourse-external phenomena and metadiscourse excludes the references to the real world. Current discourse is distinct current text from other texts and metadiscourse solely refers to the current text. Writer *qua* writer and reader *qua* reader apply specifically to personal metadiscourse, which distinguish the roles of writer/reader from experiencers in the real world. These features, except for explicitness, overlap with Hyland's (2005a) three principles, as indicated in Table 1.

**Table 1. Overlap between Ädel (2006) and Hyland (2005a)**

Ädel (2006)	Hyland (2005a)
Explicitness	
world of discourse	metadiscourse refers only to relations which are internal to the discourse
current discourse	metadiscourse is distinct from propositional aspects of discourse
writer <i>qua</i> writer reader <i>qua</i> reader	metadiscourse refers to aspects of the text that embody writer-reader interactions

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

Metadiscourse and lexical bundles are closely related. Both are overlapping functional unit that exist within texts. 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. In addition, both Biber and his colleagues' functional taxonomy and Hyland's framework of lexical bundles are strongly related to metadiscourse models (see Section 4.3 for detailed discussion), as discourse organisers and text-oriented bundles can be allocated to textual (interactive) metadiscourse devices; and stance bundles and participant-oriented bundles can be regarded as interpersonal (interactional) metadiscourse devices. Therefore, it is possible to apply metadiscourse models to the study of larger chunks, in this case lexical bundles.

The application of metadiscourse models in bundle research is useful not only for researchers but also for students. It allows students to perceive lexical bundles as metadiscourse devices, devices of interpersonal communication (Hyland, 2005a). Bundles can become valuable metadiscourse resources, which will facilitate writing in three ways according to Hyland (2005a):

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)

The use of lexical bundles and the application of corpus-driven approach is also “an efficient way of accessing the longer stretches of discourse which are often used to express metadiscourse” (Granger, 2014, 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-based bottom-up approach can verify existing researcher-generated metadiscourse lists and is likely to lead to the discovery of other metadiscourse devices and create new categories, which will add to previous metadiscourse studies.

### **4.3 Metadiscourse models**

Different metadiscourse models have been developed in various studies. The models of Vande Kopple (1985), Crismore et al. (1993), Hyland (2005a, 2005c), Mauranten (1993) and Ädel (2006), as the most popular and widely-cited models, will be introduced and compared in this section. Except for Ädel’s (2006) model, the other four metadiscourse models all draw on Halliday’s systematic functional theory, which has been mentioned earlier in relation to Hyland’s framework of lexical bundles in Chapter 3. These metadiscourse models regard textual, interpersonal and propositional (ideational) functions as three discrete and separate elements of a text, although Halliday himself suggests that these functions are realised simultaneously during writing. Hyland (2005a) has highlighted this misunderstanding in his recent work and borrowed two terms *interactive* and *interactional* to replace the original terms of *textual* and *interpersonal* in Halliday’s theory. Unlike the other researchers, Ädel (2006) has based her model of metadiscourse on Jakobson’s (1980) functional model of language, as a reflective triangle with three foci: the text/code, the writer and the reader.

### 4.3.1 Vande Kopple's metadiscourse classification

Vande Kopple's (1985) classification is the first comprehensive classification of metadiscourse. By comparing and expanding Williams' (1981) and Lautamatti's (1978) work, Vande Kopple (1985) identifies seven kinds of metadiscourse and categorises them into textual and interpersonal metadiscourse. They are text connectives, code glosses, illocution markers, narrators, validity markers, attitude markers and commentaries. The first four belong to textual metadiscourse and the remainders are interpersonal metadiscourse (Table 2).

**Table 2. Vande Kopple's (1985) classification of metadiscourse**

Category	Function	Subcategory	Example
<b>Textual metadiscourse</b>			
Text connectives	help readers recognize how texts are organized and how different parts of a text are connected	Sequencers	<i>first</i>
		Logical connectives	<i>however</i>
		Reminders	<i>as I noted in Chapter One</i>
		Announcements	<i>what I wish to do now is develop the idea that</i>
		Topicalizers	<i>there are</i>
Code glosses	help readers grasp the appropriate meanings of items in texts		<i>i.e.</i>
Illocution markers	inform readers of the speech or discourse acts performed at certain points of texts		<i>to sum up</i>
Narrators	emphasize who said or wrote something		<i>according to X</i>
<b>Interpersonal metadiscourse</b>			
Validity markers	express the truth-value of the propositional content and the writer's degrees of commitment	Hedges Emphatics Attributors	<i>perhaps</i> <i>clearly</i> <i>according to</i>
Attitude markers	reveal the writer's attitudes toward the propositional content		<i>surprisingly</i>
Commentaries	directly address readers		<i>most of you will oppose the idea that</i>

### 4.3.2 Crismore, Markkanen and Steffensen's metadiscourse system

Crismore and her colleagues' (1993) metadiscourse system was generated from actual student writing data (i.e. American and Finnish university student persuasive texts). They retained the two major categories of Vande Kopple's (1985) classification (i.e. textual and interpersonal metadiscourse), but revised the subcategories. Their classification ends up with textual markers (including logical connectives, sequencers, reminders and topicalizers) and interpretive markers (including code glosses, illocution markers and announcements) as textual metadiscourse; and hedges, certainty markers, attributors, attitude markers and commentary as interpersonal metadiscourse (Table 3).

**Table 3. Crismore, Markkanen & Steffensen's (1993) system of metadiscourse**

Category	Subcategory	Example
<b>Textual metadiscourse</b>		
Textual markers	Logical connectives	<i>however</i>
	Sequencers	<i>first</i>
	Reminders	
	Topicalizers	
Interpretive markers	Code glosses	<i>X means Y</i>
	Illocution markers	<i>to sum up</i>
	Announcements	
<b>Interpersonal metadiscourse</b>		
Hedges		<i>perhaps</i>
Certainty markers		<i>clearly</i>
Attributors		<i>according to X</i>
Attitude markers		<i>surprisingly</i>
Commentaries		<i>you may not agree that</i>

### 4.3.3 Mauranen's metatext model

Mauranen (1993) also used actual writing data (i.e. Finnish and Anglo-American academic papers) but unlike Crismore and her colleagues (1993), her analysis merely focuses on textual metadiscourse (i.e. metatext). Her narrow perspective excludes another equally important component, interpersonal metadiscourse, and

can only partially reflect the use of metadiscourse. As indicated in Table 4, she divides four major types: connectors, reviews, previews and action markers.

**Table 4. Mauranen's (1993) model of metatext**

Category	Function	Example
Connectors	show relationships between propositions	<i>however</i>
Reviews	refer back to previous stated texts	<i>So far we have assumed that X</i>
Previews	refer forward to coming texts	<i>We show below that X</i>
Action markers	indicate the discourse acts of texts	<i>the explanation is</i>

#### 4.3.4 Hyland's metadiscourse model

Hyland (2005a) offers a more comprehensive model of metadiscourse with proposed metadiscourse items in each category. Hyland's (2005a) model follows Crismore and her colleagues' (1993) system and uses the terms of *interactive* and *interactional resources* from Thompson and Thetela (1995) and Thompson (2001). According to Hyland (2004a), metadiscourse is a matter of interpersonal communication, which covers:

interactive resources which allow the writer to manage the information flow to explicitly establish his or her preferred interpretations, and  
interactional resources which 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. (Hyland, 2004a, p. 129)

Transitions, frame markers, endophoric markers, evidentials and code glosses are subcategories of interactive resources; and hedges, boosters, attitude markers, self-mentions and engagement markers are subcategories of interactional resources (Table 5).



**Table 5. Hyland's (2005a) interpersonal model of metadiscourse**

<b>Category</b>	<b>Function</b>	<b>Examples</b>
<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</i>
Frame markers	refer to discourse acts, sequences or stages	<i>finally; to conclude; my purpose is</i>
Endophoric markers	refer to information in other parts of the text	<i>noted above; see Fig; in section 2</i>
Evidentials	refers to information from other texts	<i>according to X; Z states</i>
Code glosses	elaborate propositional meanings	<i>namely; e.g.; such as; in other words</i>
<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>
Engagement markers	explicitly build relationship with the reader	<i>consider; note; you can see that</i>

*Note.* Adapted from *Metadiscourse: Exploring interaction in writing* (p.49), by K. Hyland, 2005, London, United Kingdom: Continuum. Reprinted with permission granted by Bloomsbury Continuum, an imprint of Bloomsbury Publishing Plc.

As part of his metadiscourse model, Hyland (2005c) further developed the categories of interactional metadiscourse, which includes stance (i.e. writer-oriented interaction) and engagement (i.e. reader-oriented interaction). The use of stance serves three purposes: to emphasise or withdraw the writer's commitment to the reliability of his or her propositions, to express a broad range of personal and professional attitudes towards his or her propositions, and to establish the presence of him or herself in the text. These purposes are realised through four elements: hedges, boosters, attitude markers and self-mentions, which have been covered in Hyland's (2005a) model. The use of engagement fulfils two purposes: to position and guide readers throughout the text and to acknowledge the need to meet readers' expectations. Engagement comprises five elements: directives, shared knowledge,

questions, reader pronouns and personal asides. These new subcategories have been presented in Table 6.

**Table 6. Hyland's (2005c) model of engagement in academic writing**

Category	Function	Examples
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	

As stated in Section 4.3.6, Hyland's metadiscourse model is the most inclusive and comprehensive model so far; however, it is still impossible to claim exhaustive coverage of metadiscourse functions. A top-down approach of discourse analysis with pre-determined metadiscourse items is likely to miss salient metadiscourse functions and devices, and a corpus-driven bottom-up approach is needed as a complement.

#### 4.3.5 Ädel's taxonomy of metadiscourse

On the basis of Jakobson's (1980) functional model of language, Ädel (2006) created her model of metadiscourse as a reflective triangle with three foci: the text/code, the writer and the reader, and their corresponding functions: the metalinguistic, the expressive and the directive. Ädel (2006) considers metadiscourse as explicit linguistic reference to the current text or to the writer/reader, and she divides metadiscourse devices into two categories: personal and impersonal. The use of personal pronouns indicates personal metadiscourse. Among them the most typical ones are the first and second personal pronouns (e.g. *I, we* and *you*) because these pronouns address the writer and reader of the current

text. Tables 7 and 8 summarise Ädel's (2006) taxonomy of personal and impersonal metadiscourse. The texts used in her study are university student essays written by Swedish and native writers.

As shown in Table 7, the taxonomy of personal metadiscourse falls into two major categories: metatext and writer-reader interaction. Metatext consists of ten discourse functions: defining, saying, introducing topic, focussing, concluding, exemplifying, reminding, adding, arguing and contextualising. Writer-reader interaction includes six discourse functions: anticipating the reader's reaction, clarifying, aligning perspectives, imagining scenarios, hypothesising about the reader and appealing to the reader. For a full description of the functions in Ädel's (2006) taxonomy, please refer to Appendix B.

**Table 7. Ädel's (2006) taxonomy of personal metadiscourse**

Category	Function	Examples
<b>Metatext</b>		
Code	Defining	<i>What do we mean by . . . then?</i> <i>We have to consider our definition of . . .</i>
	Saying	<i>What I am saying is . . .</i> <i>A question I ask myself is . . .</i>
Text: focus on structure of essay	Introducing the topic	<i>In the course of this essay, we shall attempt to analyse whether . . .</i> <i>I will discuss . . .</i>
	Focussing	<i>Now I come to the next idea which I presented in the beginning . . .</i> <i>I will only discuss the opponents of . . .</i>
	Concluding	<i>In conclusion, I would say that . . .</i>
	Exemplifying	<i>As an example of . . . , we can look at . . .</i> <i>If we take . . . as an example</i>
	Reminding	<i>As I mentioned earlier, . . .</i> <i>As we have seen, . . .</i>
	Adding	<i>I would like to add that . . .</i>
	Arguing	<i>The . . . which I argue for is . . .</i>
	Contextualising	<i>I have chosen this subject because . . .</i> <i>I could go on much longer, but . . .</i>

<b>Writer-reader interaction</b>		
Participant: focus on writer and/or reader of current text	Anticipating the Reader's Reaction	<i>I do realise that all this may sound . . .</i> <i>You probably never heard of . . . before either</i>
	Clarifying	<i>I am not saying . . . , I am merely pointing out that . . .</i> <i>By this I do not mean that . . .</i>
	Aligning perspectives	<i>If we [consider/compare] . . . , we [can/will] [understand/see] . . .</i>
	Imagining Scenarios	<i>If you consider . . . , you can perhaps imagine . . .</i> <i>Think back to when you were . . .</i>
	Hypothesising about the Reader	<i>You have probably heard people say that . . .</i>
	Appealing to the Reader	<i>I hope that now the reader has understood . . .</i> <i>In order for . . . , you and I must keep our minds open</i>

*Note.* Adapted from *Metadiscourse in L1 and L2 English* (pp.60-61), by A. Ädel, 2006, Amsterdam, Netherlands: John Benjamins Publishing Co. Reprinted with permission.

Table 8 provides the taxonomy of impersonal metadiscourse, developed from the pre-selected 61 terms based on the literature and student essays. These items fall into four categories: phoric markers, references to the text/code, code glosses and discourse labels.

**Table 8. Ädel's (2006) taxonomy of impersonal metadiscourse**

<b>Category</b>	<b>Function</b>	<b>Examples</b>
Phoric markers	point to various portions in the current text	<i>first, second, third, here, now</i>
References to the text/code	refer to the whole or part of text and the metalinguistics units below the paragraph level	<i>text, essay, paragraph, sentence, phrase, word</i>
Code glosses	give cues the proper interpretation of elements, comment on ways of responding to elements in text or call attention to or identify a style	<i>brief, i.e./i e/ie, mean, namely</i>
Discourse labels	refer to the names of discourse acts in text	<i>aim, state, question, answer</i>

Like Hyland's model, Ädel's (2006) model is based on pre-selected items, which are largely determined by the researcher herself. Ädel (2010) extends the model of personal metadiscourse to cover both written and spoken data, but the new taxonomy has the bias in favour of spoken data with a large majority of items found in the spoken corpus (5,000 compared to 800). Therefore, this study only introduces Ädel's (2006) model.

#### **4.3.6 Comparisons between the metadiscourse models**

Table 9 summarises the above discussed metadiscourse models and these 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). This study intends to take a broad approach in investigating both interactive and interactional functions of sentence initial bundles.

**Table 9. Summary of the metadiscourse models**

<b>Vande Kopple (1985)</b>		<b>Crismore et al. (1993)</b>		<b>Mauranen (1993)</b>	<b>Hyland (2005a, 2005c)</b>		<b>Ädel's (2006)</b>	
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		Announcements	Previews	Endophoric markers			
Code glosses		Interpretive 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

**Table 9. continued**

<b>Vande Kopple (1985)</b>		<b>Crismore et al. (1993)</b>		<b>Mauranen (1993)</b>	<b>Hyland (2005a, 2005c)</b>		<b>Ädel's (2006)</b>	
Category	Subcategory	Category	Subcategory	Category	Category	Subcategory	Category (personal)	Category (impersonal)
Interpersonal metadiscourse		Interpersonal metadiscourse			Interactional metadiscourse			
Validity markers	Hedges	Hedges			Hedges			
	Emphatics	Emphatics			Boosters			
	Attributors	Attributors						
Attitude markers	Attitude markers			Attitude markers				
Commentaries		Commentaries			Engagement markers	Directives	Writer-reader interaction	
							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			

Table 9 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.

Table 10 provides information for the comparison of categorical terms to minimise the confusion in different labels. There is consistency between the studies of Vande Kopple (1985) and Crismore et al. (1993) in labelling metadiscourse subcategories, and Crismore and her colleagues (1993) only regrouped the subcategories. However, a number of divergent categorical names are chosen by the subsequent researchers.

**Table 10. Comparison of metadiscourse categorical labels**

<b>Subcategories in Crismore et al., 1993 &amp; Vande Kopple, 1985</b>	<b>Alternative labels in other studies</b>
logical connectives	connectors (Mauranen, 1993) transitions (Hyland, 2005a)
Reminders	reviews (Mauranen, 1993) reminding (Ädel, 2006)
announcements	previews (Mauranen, 1993)
illocution markers	action markers (Mauranen, 1993) discourse labels (Ädel, 2006)
Emphatics	boosters (Hyland, 2005c)
narrators (Vande Kopple, 1985)	evidentials (Hyland, 2005a)
commentaries	engagement markers (Hyland, 2005a)

In addition to the relabelling, the boundaries of some categories, especially the ones in Hyland (2005a), differ from others. Frame markers include sequencers (e.g. *first*), part of illocution markers (e.g. *to summarize*) and references to the text (e.g. *In this section*). Endophoric markers refer not only to reminders (e.g. *as noted above*) and announcements (e.g. *refer to the next section*) but also to other parts of the text (e.g. *see Figure 2*).

On the basis of the above comparisons, I have chosen Hyland's model of metadiscourse to investigate sentence initial bundles in the current study because



---

as proposed before, it is a more inclusive model than Mauranen's (1993) and Ädel's (2006), and a more comprehensive model compared with Vande Kopple (1985) and Crismore and her colleagues' (1993). Another reason for discarding Ädel's (2006) taxonomy is the extensive use of personal pronouns, which rarely occur in the sentence initial bundles of academic writing. In addition, both Hyland's framework of lexical bundles and metadiscourse are developed from Halliday's three metafunctions (i.e. ideational, interpersonal and textual functions) and they are closely correlated. Therefore, Hyland's (2005a, 2005c) metadiscourse model is the most appropriate and rational model in analysing the sentence initial bundles for this study.

#### **4.4 Studies on metadiscourse**

Metadiscourse has become a major focus of academic writing research after Vande Kopple's (1985) classification. Studies either analyse the use of metadiscourse devices as a unified whole or target a particular aspect of metadiscourse such as hedges, boosters, self-mentions or directives. Unlike lexical bundle studies, which rely heavily on quantitative comparisons, studies on metadiscourse usually combine both quantitative and qualitative approaches. Quantitative approaches are used to statistically describe and compare the distributions of metadiscourse devices in different texts and qualitative approaches serve to illustrate the categories and functions of metadiscourse using examples. There is, however, little research exploring the underlying reasons for metadiscourse variation between writers.

##### **4.4.1 Studies on metadiscourse as a whole**

Comparisons of metadiscourse devices have been made in relation to various factors including gender (Crismore et al., 1993), time period of abstract writing (Gillaerts & Velde, 2010), genre (Hong & Cao, 2014), discipline (Abdi, 2002; Cao & Hu, 2014; Dahl, 2004; Khedri, Heng, & Ebrahimi, 2013), the language of writing (Dahl, 2004; Jiang, 2009; Kim & Lim, 2013; Marandi, 2003; Molino, 2010), the context of writing (Li & Wharton, 2012), the quality of writing (Intaraprawat & Steffensen, 1995; Liu & Braine, 2005), the level of students (Xu & Gong, 2006; Yang & Sun, 2012), and language and cultural background of writers (Ädel, 2006;

---

Cao & Wang, 2009; Crismore et al., 1993; Heng & Tan, 2010; Hong & Cao, 2014; Marandi, 2003; Mauranen, 1993).

This section focuses on comparing the use of metadiscourse between L2 learners and native writers, in other words, the studies of contrastive interlanguage analysis (Granger, 1996, 2015) because this is the focus of the current investigation. L2 writing collected from different countries have been compared with American or British writing by native speakers. For comparison, the texts composed by Finnish (Crismore et al., 1993; Mauranen, 1993), Iranian (Marandi, 2003), Swedish (Ädel, 2006), Chinese (Cao & Wang, 2009) and Malaysian learners (Heng & Tan, 2010) have been used.

The most popular genre under investigation is argumentative essays, but researchers have also examined other genres such as academic research reports (Mauranen, 1993) and masters theses (Marandi, 2003). The number of the selected research texts varies from four up to around seven hundred. Manual and computer-assisted approach have been used for data analysis. Most researchers have compared the total number of metadiscourse devices employed by L2 and native writers, the use of textual (interactive) and interpersonal (interactional) metadiscourse, or the number of metadiscourse devices in a specific subcategory.

L2 writers generally deploy more metadiscourse devices than native writers. Crismore et al. (1993) and Ädel (2006) identified big differences in terms of density per line or normalised frequency. According to Crismore et al. (1993), Finnish student texts contain nearly 30% more metadiscourse devices than the U.S. student texts (1.358 compared to 1.08 per line). Ädel (2006) calculated both personal and impersonal metadiscourse. In regard to personal metadiscourse, she extracted more than twice the number of expressions in the Swedish learner writing compared with those in the American student writing, which were, in turn, at least twice as many as those in the British student texts. In the case of impersonal metadiscourse, the Swedish learners employed about 50% more expressions than both native groups. Cao and Wang (2009), and Heng and Tan (2010) found small differences between non-native and native student writing: Chinese and Malaysian student essays contained slightly more metadiscourse devices than their American or British

---

counterparts (65.13 compared to 63.17 per 100 words, and 673.5 compared to 621 per 10,000 words). The only exception is Mauranen's (1993) study and she found a lower proportion of metatext in Finnish writers' texts than that of native-English writers' texts (22.6% compared to 54.2%). However, the writers of Mauranen's (1993) texts are expert instead of student writers and her research focus is on metatext, not all metadiscourse devices.

The use of textual (interactive) and interpersonal (interactional) metadiscourse, however, has been found to differ in L1 and L2 writing due to the diverse research foci and approaches. Crismore and her colleagues (1993) concluded that both Finnish and American groups used more interpersonal than textual metadiscourse. Cao and Wang (2009) identified more textual metadiscourse in the Chinese student writing and more interpersonal metadiscourse in the American student writing. In contrast, Heng and Tan (2010) found more interactional metadiscourse in the Malaysian corpus and more interactive metadiscourse in the BAWE corpus.

Most studies have also compared more salient subcategories of metadiscourse such as hedges and textual markers. Unlike native writers, hedges are underused by most L2 learners; in contrast, textual markers dominate L2 writing (Cao & Wang, 2009; Heng & Tan, 2010; Marandi, 2003). The distributions of hedges and textual markers echo the results of bundle studies in which the underuse of hedge bundles and the overuse of discourse organiser bundles were generally identified in L2 writing. One different finding has been reported in Crismore et al. (1993), in which the Finnish students hedge more than the American students and the American students deploy more text markers than their Finnish counterparts. However, the students in Crismore and her colleagues' research wrote in their native language.

#### **4.4.2 Studies on specific aspects of metadiscourse**

Recent research has seen an increasing focus on the interactional aspect of metadiscourse, including hedges, boosters, self-mentions and directives. The investigated genres include research articles, student writing (especially L2 student writing), and other written texts, such as textbooks. Certainty markers including hedges and boosters have been mostly examined. Many studies compare L2 writers' interlanguage development with native writers' English production (e.g. Burrough-

---

Boenisch, 2005; Hinkel, 2005; Hyland & Milton, 1997; Vassileva, 2001; Yang, 2013) and a few studies focus on variations between languages, particularly between Chinese and English (e.g. Hu & Cao, 2011; Vassileva, 2001; Yang, 2013). It has been generally found that while writing in English non-native writers use comparatively fewer hedges and more boosters than native English writers, such as Cantonese writers in Hyland and Milton (1997), Bulgarian writers in Vassileva (2001), Dutch writers in Burrough-Boenisch (2005) and Chinese writers in Yang (2013). It has also been found that abstracts or research articles written in Chinese contain fewer hedges than the corresponding English texts (Hu & Cao, 2011; Yang, 2013). However, this might be due to the bias in favour of English texts. Since there is no ready-made list of Chinese hedges, the translation of pre-defined English items into Chinese may exclude some Chinese hedges.

Most work on self-mentions targets published research articles (e.g. Fløttum, Kinn, & Dahl, 2006; Harwood, 2005a, 2005b; Hyland, 2001b; Kuo, 1999) and only two studies have been found that investigate L2 academic writing: Hyland (2002a) and Xu (2011). Both studies compared the use of first person pronouns between L2 theses and published research articles. Hyland (2002a) observed four times fewer first person pronouns in his corpus of 64 Hong Kong undergraduate reports than in the published writing. In addition, he found that undergraduates preferred pronouns performing *low-risk* functions (i.e. stating a purpose, explaining a procedure) and avoided those with *high-risk* functions (i.e. stating results/claims, elaborating an argument). Interestingly, plural forms were also common in these single-authored theses. Xu (2011) not only investigated the difference between Chinese student L2 writing and published writing, but also examined the development across undergraduate, masters and PhD theses in their use of first person pronouns. She proposes that first-person pronoun sequences are generally underused in Chinese student writing, especially in PhD theses. This is in line with Hyland's (2001b) finding.

Hyland built three parallel corpora to investigate directives in Hong Kong undergraduate reports: an undergraduate corpus, a research article corpus and a textbook corpus (Hyland, 2002b, 2005b). He found the number of directives in the L2 student writing was only half of that in the journal articles and one third of that

---

in the textbooks. He further divided directives into three categories according to the leading rhetorical activities. From least to most imposing functions, they are textual acts, physical acts and cognitive acts: textual acts steer readers to another part of the text or to another text, physical acts instruct readers to perform some real world action, and cognitive acts direct readers to understand a point in a certain way. On the basis of this categorisation, he discovered that cognitive acts were least used by the students, whereas they tended to use directives to guide their readers through research procedures or to steer their attention to non-linear information in their texts (e.g. tables, examples and appendices).

#### **4.4.3 Studies on writer interpretations of metadiscourse use**

A few studies have incorporated writer interpretations in order to gain insights into metadiscourse data. Hyland (2002b) reported both students and staff members' perceptions of the use of directives in research articles, textbooks and undergraduate project reports. Hyland (2004b) sought explanations from 24 masters and PhD students for variations in metadiscourse use across degrees and disciplines. Hyland (2005b) conducted focus group discussions of 23 final year Hong Kong undergraduates to explore students' interpretations of the use of reader engagement features (e.g. reader pronouns, directives and questions). Kim and Lim (2013) invited Chinese writers to provide insights into the results of metadiscourse comparison of English and Chinese research article introductions. Lewin (2005) required her respondents to identify the hedges in their own published articles and to provide the motivation for each hedge. Except for Lewin's research, the interviewees of the other studies are not the writers of the text data.

Among these studies, only Hyland's work includes student voices. Student writers are generally reported to be more tentative and more reluctant in signalling their presence, addressing readers and directing readers to act or think (Hyland, 2002b, 2004b, 2005b). More advanced students, such as PhD students show more awareness of readers and feel more comfortable to use self-mentions (Hyland, 2004b). The informants of Hyland's studies are not the same writers of their corpus data and there is no evidence that these informants' writing embeds the same features with the texts in the analysed corpora.

## **4.5 Limitations of the existing research**

Metadiscourse, as an umbrella term, covers all devices of interpersonal communication. In recent years, more attention has been devoted to the use of metadiscourse in academic writing and a series of models have been developed to investigate different textual (interactive) and interpersonal (interactional) functions. Among them, the most popular models are Vande Kopple's (1985) classification, Crismore and her colleagues' (1993) system, Mauranen's (1993) work, Hyland's (2005a, 2005c) model and Ädel's (2006) taxonomy. The research on metadiscourse uncovers the ways that writers formulate arguments, present themselves and engage their readers.

However, most studies have drawn on pre-determined items, mainly individual words. The word-based analysis fails to provide an accurate picture of metadiscourse devices because it is sometimes insufficient to determine the functions of texts by means of single words. The same words may function differently within different contexts. This top-down approach with pre-determined items is also likely to miss some salient features of academic writing. In addition to the limitations of metadiscourse approach, most researchers have solely focused on the comparisons between texts and few studies have investigated the reasons for variations.

In the next chapter, Methodology, I will provide the details and explain the procedures of this research, which include corpus-based analysis and semi-structured interviews.



---

## Chapter 5 Methodology

Drawing on previous research, this study compares the use of sentence initial bundles in Chinese and New Zealand thesis writing, and in addition to exploring possible influences on bundle choices by Chinese postgraduates. Four collections 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. This study compared both masters and PhD theses because comparing the same genre at the same level is likely to eliminate some other factors. In comparing these four corpora, this study aims to provide a detailed picture of the use of sentence initial bundles in advanced Chinese student writing and an overall picture of variation in bundle use across different levels of students. This study focuses 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). Sentence initial bundles also function differently from non-initial bundles as they serve the function of triggers and complements: the former act as sentence starters to trigger the statements (e.g. *It should be noted*), and the latter, to complete clauses or provide additional information (e.g. *the extent to which*) (Cortes, 2013; Williams, 2003).

The research questions are as follows:

1. What are the frequencies of four-word sentence initial bundles in the Chinese and New Zealand masters and PhD corpora? Are there any differences between Chinese and New Zealand postgraduates, or between masters and PhD levels of study in the use of sentence initial bundles?
2. What are the salient structures of these bundles in the Chinese and New Zealand corpora? Are there any differences between Chinese and New Zealand postgraduates, or between masters and PhD levels of study in the distribution of structures?
3. What are the metadiscourse functions of these bundles in the Chinese and New Zealand corpora? Are there any differences between Chinese and New Zealand postgraduates, or between masters and PhD levels of study in the distribution of functions?



- 
4. What reasons do Chinese postgraduates give for their sentence initial bundle choices in their thesis writing?

This study involved two forms of data collection: corpus collection and semi-structured interviews. The triangulation between these two types of data collection methods will contribute to the trustworthiness of the data (Brown, 2014). This chapter introduces the procedures of corpus building, the criteria of bundle identification, the frameworks for structural analysis and functional analysis, the backgrounds of six Chinese interviewees and the stages of interview data analysis.

## **5.1 Corpus-based analysis**

This section introduces corpus building, bundle identification, and frameworks for structural analysis and functional analysis. Both quantitative and qualitative approaches were taken in corpus-based analysis. The use of a quantitative approach yields the number of bundles in type and token, so that it is possible to obtain an overview through comparing total occurrences and bundle distributions across corpus, structure and function. A qualitative approach enables the analysis of structures and functions in relation to contexts, so that the extended units, locations and discourse functions of some typical bundles are explored and compared between each corpus. Both structural analysis and functional analysis were used because lexical bundles are not merely lexico-grammatical patterns, but high-frequency overlapping language chunks serving particular metadiscourse functions.

### **5.1.1 Corpus building**

The present study initially involved the building of four postgraduate thesis corpora. These four corpora contain theses submitted from 2000 to 2013 in the discipline of general and applied linguistics. The corpora are a Chinese masters corpus, a Chinese PhD corpus, a New Zealand masters corpus and a New Zealand PhD corpus. The main and practical reason for the discipline selection is that only the Chinese students at faculties, schools or departments of foreign languages are expected to complete their theses in English.

The Chinese masters and PhD theses were downloaded from one of the most prominent and accessible academic databases in China: Wanfang Data Knowledge

---

Service Platform (<http://www.wanfangdata.com.cn/>). The China Knowledge Resource Integrated Database (CNKI) (<http://www.cnki.net/>), the largest academic database in China, was considered as the first choice. However, it has been proved to be impossible to transform the CAJ format that the texts are published in on CNKI into computer readable documents. The number of available theses on Wanfang is sufficient to comprise corpora.

The Chinese masters corpus comprises the theses randomly chosen from four popular topic areas (i.e. task-based language learning, learning strategies, teaching mode and corpus-based lexical analysis) in order to avoid an overwhelming number of theses within one particular topic area and at the same time to guarantee sufficient data for corpus building. Only a small number of PhD theses are written in English in China, so there was no need to narrow down the selection. The New Zealand masters and PhD theses were randomly selected and directly downloaded from the university library websites and only open-access New Zealand theses were collected for this study. Among them, theses written by non-native authors were excluded on the basis of the author names and thesis titles. This was not altogether a satisfactory approach but one that was practical and convenient.

As all the Chinese and New Zealand theses were in PDF format, FineReader 11, which is an optical character recognition (OCR) software for PDF conversion, was used to transform the Chinese theses and New Zealand theses into Word documents, ready for processing. Only the body of the texts were transformed; the title page, abstract, acknowledgements, table of contents, lists of tables and figures, references and appendices were not included in the corpora.

Table 11 provides information on each of these corpora. According to Gray and Biber (2013), corpus size and representativeness are the two concerns of corpus building. The Chinese masters corpus, totalling 3.3 million words, was composed of 200 theses from 74 universities and the average length of each thesis was 16,504 words. The Chinese PhD corpus contained 67 theses from 12 universities with a totalling of 3.8 million words and 57,232 words each thesis. The New Zealand masters corpus consisted of 60 theses collected from 5 universities, altogether 2 million words and 34,000 words each. The New Zealand PhD corpus included 46

theses collected from 5 universities, amounting to 3.8 million words and the average text length was 82,609 words. The sampling procedure resulted in corpora of different sizes, but this is not likely to affect the cross-corpora comparison because the cut-off frequency of sentence initial bundles is same and the final frequency of these bundles was normalised to 1,000,000 words, as will be discussed in the next section. 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*). Therefore, I chose to compare the percentage differences between the four corpora. The percentage reflects the bundle distribution within any corpus, and the comparison between percentage differences will not be affected by the different lengths of texts in different corpora, the different numbers of texts between corpora and any different bundle generation criteria between corpora.

**Table 11. 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

Similarities such as genre, discipline and level of writing ensure broad comparability of these two pairs of corpora. It should also be noted that the purpose of this comparison is not to present the Chinese students' linguistic deviations from the native norm, but to highlight the different writing practices between these L2 and L1 postgraduates, to reveal the socio-cultural norms in these two particular writing contexts.

---

### 5.1.2 Bundle identification

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 in this study. FLAX can automatically generate lexical bundles from corpora and display them with their frequencies and their context sentences. Besides, the inbuilt corpora (e.g. Wikipedia) of FLAX and the British National Corpus (BNC) in the BNCweb (<http://bncweb.lancs.ac.uk/>) were chosen as the reference corpora to validate the findings from the comparison of the four thesis corpora, for example, to search for collocations in Wikipedia, the contemporary English corpus, or to check the frequency of a particular word in BNC, the general English corpus.

FLAX reads from the first word of each text in the corpus and advances one word at a time. Along with the reading process, FLAX stores every four-word sequence and checks against its previously identified sequences. For this study, FLAX generates the sequences with at least 3 occurrences across more than 3 texts as the raw data.

There are two differences between FLAX and the programmes used in Biber et al. (2003). First, instead of calculating all the same bundles as one group, FLAX categorises the retrieved lexical bundles into sentence initial and non-initial bundles according to their position. Second, FLAX treats both uninterrupted word sequences and sequences containing a punctuation mark as lexical bundles. The reason to include punctuation-embedded bundles is to cover the sequences incorporating linking words and shorter fixed or semi-fixed phrases (e.g. *However, it is not, In other words, the, In this section, I*), which are part of sentence starting strategies.

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 four-word bundles occur more frequently with less variation than longer ones. Biber et al. (1999) report that four-

---

word bundles occur about 10 times as frequently as five-word bundles. 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. For example, both three-word bundles *on the other* and *the other hand* are a part of the four-word complex preposition bundle *on the other hand*. This four-word bundle acts as a transition marker; however, its corresponding five-word bundle *on the other hand, it* serves two functions as a transition marker and an endophoric reference, as shown in example 1.

(1) ***On the other hand**, it is difficult for the LI English speaker to acquire this new distinction when learning Spanish. (CH PhD)*

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 FLAX, the frequency and distribution threshold is pre-set as 3 occurrences across 3 texts to avoid individual author idiosyncrasies. 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. The FLAX-generated complete bundle lists (including both sentence initial and non-initial bundles) and all the texts of the corpora (available to view at the sentence, paragraph and thesis levels) are available for search and analysis. This allows for side-by-side comparison between bundles at different positions and for further exploration into the contexts of bundles.

As with other studies, content-based bundles (including topic-specific bundles and bundles containing chapter titles, method names and proper names) were removed from the retrieved bundle lists. This is because (a) these bundles do not show much pedagogical value, being confined to a specific subject; (b) it is almost impossible to compare these bundles between corpora due to their uniqueness. Table 12 presents these exclusion criteria along with all the excluded bundles from the four corpora. Altogether 15 different bundles were removed from the initial bundle lists. Among them were 13 bundles from the Chinese masters corpus, one from the New

Zealand masters corpus and another one from the New Zealand PhD corpus. The four domain-specific sub-corpora of the Chinese masters corpus may result in more overlapping themes between the texts, which attributes to the comparatively larger number of discarded bundles. This is particularly interesting as it indicates that the narrower the scope of text selection is, the more content-based bundles appear in the text collection. Appendix C includes four comprehensive lists of bundles identified in these four corpora. Considering the four corpora were of different sizes, the final frequencies were normalised to 1,000,000 words to conduct a reliable comparison.

**Table 12. Bundle exclusion**

<b>Exclusion criteria</b>	<b>Excluded bundles</b>
Topic-specific bundles	<i>Language learning strategies are (CH MA)</i> <i>Most of the students (CH MA)</i> <i>In this way, students (CH MA)</i> <i>The students in the (CH MA)</i> <i>For example, the teacher (CH MA)</i> <i>All of the participants (NZ PhD)</i>
Bundles containing chapter titles	<i>Chapter Two Literature Review (CH MA)</i> <i>Chapter Four Results and (CH MA)</i> <i>Chapter Three Research Design (CH MA)</i>
Bundles containing method names	<i>t-test for Equality of (CH MA)</i> <i>Equal variances not assumed (CH MA)</i> <i>The mean score of (CH MA)</i> <i>*Correlation is significant (CH MA)</i> <i>Levene's Test for Equality (CH MA)</i>
Bundles containing proper names	<i>The Ministry of Education (NZ MA)</i>

### 5.1.3 Structural categories

The structural categories and patterns of this study were developed from the studies of Biber and his colleagues (Biber et al., 2004; Biber et al., 1999) and Chen and Baker (2010). Five major categories were identified: NP-based, PP-based, VP-based, clause-based and other bundles. As a result of the specific nature of the bundles in the current study (i.e. sentence initial bundles), three new patterns were created: *there be-clause fragment*, *noun phrase + verb phrase fragment* and *conjunction + clause fragment* and four initial patterns were discarded: *copula be + noun phrase/adjective phrase*, *(verb phrase +) that-clause fragment*, *adverbial clause fragment* and *pronoun/noun phrase + be (+ ...)*. In addition, another two

original patterns, *passive verb + prepositional phrase fragment* and *(verb/adjective +) to-clause fragment*, were amended into *active or passive verb + noun/preposition phrase fragment* and *(in order) to-clause fragment* to fit the data. Table 13 presents examples of each pattern.

**Table 13. Major categories and structural patterns of sentence initial bundles**

Category	Pattern		Example
NP-based	noun phrase with post-modifier fragment	of	<i>The results of the</i>
		other	<i>The fact that the</i>
PP-based	preposition + noun phrase fragment	of	<i>In the case of</i>
		other	<i>On the other hand,</i>
VP-based	VP with	active verb	<i>Look at the following</i>
		passive verb	<i>Based on the above</i>
Clause-based	anticipatory it +	(in order) to-clause fragment	<i>To sum up, the</i>
		adjectiveP	<i>It is important to</i>
		VP	<i>It should be noted</i>
		there be-clause fragment	<i>There are a number</i>
		noun phrase + VP	<i>This is not to</i>
Other	other expressions	conjunction + clause fragment	<i>As can be seen</i>
			<i>That is to say,</i>

#### 5.1.4 Functional categories

The functional analysis is based on Hyland's (2005a, 2005c) interactive and interactional model of metadiscourse rather than the two extensively-used taxonomies of lexical bundles, Biber and his colleagues' taxonomy (i.e. referential, discourse and stance bundles) and Hyland's (2008a) framework (i.e. research-oriented, text-oriented and participant-oriented bundles). On the one hand, it is because the two lexical bundle taxonomies 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*. On the other hand, as discussed in Section 4.3.6, Hyland's (2005a, 2005c) model is the most comprehensive metadiscourse model so far. This metadiscourse model is also closely related to Hyland's (2008a) framework of lexical bundles.

In this study no interactive bundle was identified as evidential, but two new 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: *On the basis of*, *In the case of*, *In terms of the* and *With regard to the*. Introduction bundles refer to the initial parts of existential *there* clauses, for example, *There are a number*, *There was no significant* and *There appears to be*, which draw the reader's attention to new information, research results or writers' conclusions. Appendix D 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.

The majority of interactional bundles fell into the stance category and the bundles classified as engagement devices were mainly directive bundles. Only one bundle, *As we all know*, was used to label shared knowledge and there was no bundle indicating personal asides, questions or embedding reader pronouns. Appendix E 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.

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.

As previously discussed, Hyland (2005a) distinguishes text-internal from text-external references and in his view metadiscourse refers only to internal relations of the discourse. Several *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 in this study to describe the distributions of all the bundles, 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.

---

## **5.2 Semi-structured interviews**

Interviews were used to mainly explore the possible reasons for Chinese postgraduates' bundle choices or avoidance in comparison with New Zealand L1 writers. Little effort was put into interpreting the discrepancies between the masters' and doctoral theses as it is beyond the scope of this study.

One-on-one interviews were conducted after the text analysis. This is because "corpus data does not interpret itself" (Baker, 2006, p. 18). It is needed to interrogate text users to "understand how and why language users make the choices they do when they speak/write" (Hyland, 2011, p. 106). Ethical approval was obtained from the Faculty of Education Research Ethics Committee at the University of Waikato (Appendix F).

Six Chinese postgraduates studying at the University of Waikato were recruited as participants. Their original drafts, drafts with no editing from the supervisors or other language tutors, were collected. The expressions in these participants' writing that completely or partially overlapped with the sentence initial bundles of the corpus data were manually identified. Semi-structured interviews, as "a balance between structure and openness" (Gillham, 2005, p. 79), were conducted on the use of particular bundles in the participants' writing to evoke these participants' perspectives on and learning experiences of these bundles. Appendix G is an example of the interview questions asked on the basis of the identified expressions in one participant's writing.

### **5.2.1 Background of participants**

Table 14 provides an overview of the participants' information and experiences, which were considered closely relevant to the current research.

**Table 14. Overview of six Chinese participants**

<b>Participant</b>	<b>Z</b>	<b>A</b>	<b>S</b>	<b>J</b>	<b>W</b>	<b>V</b>
<b>Age</b>	40+	30+	25+	30+	25+	25+
<b>Gender</b>	Male	Male	Male	Female	Female	Female
<b>Level</b>	PhD	PhD	PhD	PhD	PhD	Master
<b>Discipline</b>	Applied Linguistics	Knowledge Management	Psychology	Tourism Management	Management Communication	Applied Linguistics
<b>Months of English-medium study</b>	3	32	48	16	6	16
<b>Any help received with the language problems in academic writing</b>	help from the supervisors	little help from the supervisors who does not regard this as a major focus	search in FLAX while writing	help from the chief supervisor with grammar	little help from the supervisors and have never been to student learning centre	help from both the supervisor and student learning centre
<b>Relevant language teaching &amp; learning experience</b>	14 years of lecturing at foreign language department of a Chinese university				half-a-year experience of writing English correspondence s	4 years of English language teaching experience, teaching Cambridge English, New Concept English; having attended TOEFL and IELTS for several times

The current learning context of these six participants is different from that of the Chinese writers of my thesis corpora. These participants were studying in an English-medium New Zealand university; in contrast, the Chinese writers composed their theses in mainland China. Li and Wharton (2012) argue that “academic literacy needs to be seen as a locally situated practice” (p. 353) and the expectations of the institutions and supervisors largely influence the writing practices. This should be taken into consideration when interpreting the interview data. However, the six participants all received their primary, secondary and undergraduate education in mainland China. Except for S and V, the other participants have completed their masters degrees in China. The years of formal education in China have schooled them in the expectations of Chinese context and

---

the writing practices in the Chinese community. This is evident from the fact that many overlaps were identified between the typical bundles used solely in the Chinese student theses and the expressions in the participants' writing.

Six participants were recruited not only from the discipline of general and applied linguistics, which is the same as the discipline of the four thesis corpora, but also from the disciplines of management and psychology. One practical reason is that it was impossible to recruit sufficient participants in the discipline of general and applied linguistics due to the limited number of Chinese postgraduates in that particular discipline area. Another reason to recruit participants from other social science disciplines is the academic writing in social science bears many similarities (Hyland, 2008b).

Besides disciplines, these participants also differed in age, gender, the level of study, the length of English-medium education, the received language support during thesis writing and the work or test experience related to English writing. Details can be found in Table 14. It is interesting to note that the participants in the disciplines of management and psychology had received fairly limited language support from their supervisors or other language tutors. This reflects the subordinate position of English instruction in mainstream education and it seemed that sentence-level accuracy was not a strong focus for the supervisors. In this case, these Chinese students relied on their learned English expressions from the Chinese context. Typical Chinese bundles featured in their writing, although they were studying in a New Zealand context. Z and V, from the discipline of applied linguistics, received comparatively more feedback on their language problems; however, typical Chinese bundles were still prevalent in their writing.

### **5.2.2 Interview data analysis**

All interviews were conducted in Mandarin, both the participants' and the researcher's L1. The interview data were transcribed before analysis. All the unclear points were clarified with the participants through emails. Only key interviews in Chinese were translated and if some words were untranslatable, transliteration — the original Chinese words along with their closest English meanings given in brackets — were adopted (Halai, 2007). Around 10% of all the

translation was double-checked by a Mandarin-speaking peer. All interview data will be reported along with their the corpus data to interpret the use of particular sentence initial bundles in the Chinese postgraduates' corpora.

Thematic analysis, "a method for identifying, analyzing and reporting patterns (themes) within data" (Braun & Clarke, 2006, p. 79), was used to summarise and categorise the interview data in Section 9.1.4 Reasons for discrepancies. Braun and Clarke's (2006) six phases of thematic analysis (i.e. familiarising yourself with your data, generating initial code, searching for themes, reviewing themes, defining and naming themes, and producing the report), and Fereday and Muir-Cochrane's (2006) six stages of data coding (i.e. developing the code manual, testing the reliability of codes, summarising data and identifying initial themes, applying template of codes and additional coding, connecting the codes and identifying the themes, and corroborating and legitimating coded themes) were referred to while coding, generating and refining the themes.

I will present the key findings in the following three chapters, Chapters 6, 7 and 8. Chapter 6 will cover the findings of frequency-based analysis and structural analysis. Chapters 7 and 8 will report the findings of functional analysis within two major categories: interactive and interactional bundles. Interview data will be embedded in the corpus data to provide possible reasons for Chinese students' bundle choices.

---

## Chapter 6 Frequency-based and structural analysis

Frequency, structure and function are three foci of lexical bundle research. In this chapter, I will report the findings of the frequency-based and structural analysis, and the functional analysis will be covered in Chapters 7 and 8. I will first present the frequencies and salient structures of four-word sentence initial bundles in the Chinese and New Zealand masters and PhD corpora. Then I will highlight the differences between Chinese and New Zealand thesis writing, or between masters and PhD levels of study in terms of bundle distribution. I will also discuss the possible reasons drawn from the literature and/or interview data. Student interview data will be embedded when possible and appropriate.

### 6.1 Frequency-based analysis

Table 15 describes the distribution of sentence initial bundles in the four corpora: Chinese masters and PhD corpus and New Zealand masters and PhD corpus. Consistent with many previous studies (Hyland, 2008a; Pang, 2009; Pérez-Llantada, 2014; Staples et al., 2013; Wei & Lei, 2011; Xu, 2012), the students with lower levels of English proficiency and less experience in English writing appeared to rely more on lexical bundles. The Chinese writers used more types of bundles than their New Zealand counterparts (80 compared to 63, 60 compared to 44), the masters students used more types of bundles than their PhD counterparts (80 compared to 60, 63 compared to 44). The tokens of the Chinese student bundles were significantly higher than those of the New Zealand student bundles (P-Value < 0.05). However, there was no significant difference between the tokens of the masters and PhD bundles (P-Value > 0.05).

**Table 15. Descriptive statistics: sentence initial bundles**

Corpus	Types	Mean tokens	StDev
CH MA	80	10.86	9.6
CH PhD	60	11.67	10.01
NZ MA	63	8.683	5.067
NZ PhD	44	8.955	5.044

---

This might be explained by Dechert's (1984) concept of *islands of reliability*: less competent writers are more dependent on "points of fixation", in this case the prefabricated chunks of words, to build up their writing. The less variety and greater frequency of bundles used by the Chinese students and masters students could possibly suggest their limited vocabulary repertoire. It may have been that these L2 learners and lower-level students had to stick to a limited number of familiar clusters to start their sentences. The Chinese writers' deficiency in vocabulary knowledge might also be interpreted from their overreliance on one salient discourse marker *On the other hand* the most frequent bundle in the four corpora: its occurrences in the two Chinese corpora were twice those in the New Zealand corpora.

Appendix H lists the 50 most frequent sentence initial bundles in each corpus. Only 44 bundles were retrieved from the New Zealand PhD corpus; therefore, the subsequent 6 bundles with lower cut-off frequencies (i.e. over 4 times per million words) but with the same text distribution threshold (i.e. 5 texts) were included to complete the New Zealand PhD top 50 bundle list. Two bundles, *On the other hand* and *In other words, the*, were shared across all four corpora. The use of *On the other hand*, indicates the particular need to demonstrate alternative views in argument writing. The use of *In other words, the*, reflects another strategy in academic writing, rephrasing or elaborating.

More shared bundles were found in the two PhD corpora than in the two masters corpora (19 compared to 11), that is, more convergence was identified in the higher-level writing than the less advanced master-level writing. This suggests a greater degree of familiarity with the conventional expressions of the target academic community and this may also be interpreted as an indicator of greater English writing competence of these Chinese doctoral students. For further details on the shared bundles between the Chinese and New Zealand PhD and masters corpora will be reported in Chapter 7 Interactive functions of the bundles and Chapter 8 Interactional functions of the bundles.

The identified bundles were divided into two functional categories on the basis of Hyland's (2005a, 2005c) interactive and interactional model of metadiscourse, All

students used more interactive than interactional bundles in terms of both type and token (Table 16). This is in line with Thompson’s (2001) argument “interactional signals are typically less frequent and less overt in academic text” (p. 73).

**Table 16. 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 17 displays the proportions of interactive and interactional bundles and the different proportions between interactive and interactional bundles in each corpus in terms of tokens. The distributions were similar in all four corpora: around three-quarters of the bundles were interactive bundles, which largely exceeded those of interactional bundles (i.e. from 20% to 29%). However, greater differences were found in the two Chinese corpora, whereas the two New Zealand corpora showed a more balanced distribution of interactive and interactional bundles (i.e. 60% compared to 42%, 54% compared to 44%).

**Table 17. Proportion of interactive and interactional bundles (tokens)**

Category	CH MA	CH PhD	NZ MA	NZ PhD
Interactive	80%	77%	71%	72%
Interactional	20%	23%	29%	28%

The distribution of interactive and interactional bundles can also be seen from the 10 most frequent sentence initial bundles in each corpus. As shown in Table 18, the majority of the bundles were interactive bundles; and the interactional bundles, such as *It is possible that*, *It is important to* and *It is interesting to*, were popular in the two New Zealand corpora.



**Table 18. Top 10 frequent sentence initial bundles in each corpus in rank order**

CH MA	CH PhD	NZ MA	NZ PhD
On the other hand,	On the other hand,	On the other hand,	On the other hand,
That is to say,	In other words, the	It is important to	It is possible that
At the same time,	That is to say,	The results of the	In the case of
The results of the	On the one hand,	It is possible that	At the same time,
In the process of	The results of the	In the case of	It is important to
On the basis of	In the case of	The results of this	As discussed in Chapter
With the development of	In the present study,	As can be seen	At the end of
In other words, the	At the same time,	It is interesting to	In addition to the
In the present study,	On the basis of	As a result of	The results of the
In this chapter, the	In this sense, the	The purpose of this	In other words, the

## 6.2 Structural analysis

Developed from the studies of Biber and his colleagues (Biber et al., 2004; Biber et al., 1999) and Chen and Baker (2010), the five major structural categories used in the current study consisted of NP-based bundles, PP-based bundles, VP-based bundles, clause-based bundles and other bundles. Three new patterns (*there be-clause fragment*, *noun phrase + verb phrase fragment* and *conjunction + clause fragment*) were identified and added in regard to the specific features of sentence initial bundles. Details are as follows:

NP-based bundles. This category refers to any noun phrases with post-modifier fragments. In this study, 90% NP-based bundles in the Chinese and New Zealand corpora comprised *of*-phrase fragments (e.g. *The results of the*, *The purpose of this*, *The analysis of the*) and the remaining 10% were NP-based bundles with post-nominal clause fragments (e.g. *The fact that the*) or any other preposition phrase fragments (e.g. *The results from the*).

PP-based bundles. This category refers to preposition phrases or preposition phrases plus noun phrase fragments. More than one-third of PP-based bundles in the Chinese and New Zealand corpora consisted of *of*-phrase fragments functioning as post-modifiers of nouns (e.g. *In the case of*). The other two-thirds of PP-based bundles were mostly fixed or semi-fixed phrases (e.g. *On the other*

---

*hand, In the current study*) or phrases plus articles or personal pronouns (e.g. *In other words, the, In this section, I*).

VP-based bundles. This category is composed of verb phrase fragments, including the two amended patterns *active or passive verb + noun/preposition phrase fragment* and *(in order) to-clause fragment*. VP-based bundles only appeared in the Chinese students' writing and included two verb bundles (*Look at the following, Based on the above*) and a small number of *to-clause* fragment bundles (e.g. *To sum up, the, In order to make*).

Clause-based bundles. This category begins with independent or dependent clauses featuring four major patterns: *anticipatory it-clause fragment, there be-clause fragment, noun phrase + verb phrase fragment* and *conjunction + clause fragment*. Unlike Chen and Baker (2010), the pattern *anticipatory it-clause fragment* was grouped into clause-based rather than VP-based category because the fragment was the starter of a main clause rather than a verb phrase. The three newly-created patterns all fall into clause-based bundles and they are *there be-clause fragment, noun phrase + verb phrase fragment* and *conjunction + clause fragment*. The most prevalent clause-based pattern was *anticipatory it-clause fragment*, which accounted for 40% of the clause-based bundles.

Other bundles. This category refers to idiomatic phrases, such as *That is to say, Last but not least and First of all, the*. The bundles included in this category comprised those which did not fit into the other patterns that had been identified.

Table 19 illustrates the bundle distribution of each structural pattern, showing the percentage in terms of both type and token in these four thesis corpora. It is not surprising to find that PP-based constructions (e.g. *On the basis of, In addition to the, In this chapter, we*), anticipatory-*it* patterns (e.g. *It is important to, It should be noted*) and post-modified noun phrase fragments (e.g. *The results of the, The fact that the*) are the most frequent forms of the bundles, as these forms were also found to be dominant in academic texts in the previous studies (Biber et al., 1999; Hyland, 2008a). Noun phrase + verb phrase fragments are another frequent form in this study. This is because of the nature of sentence initial bundles and many recurrent

subject + verb combinations were generated (e.g. *The chapter concludes with, This is not to*).

**Table 19. Distribution of sentence initial bundles in thesis writing**

Category	Pattern		% of all types (frequency)	% of all tokens (frequency)
NP-based	noun phrase with post-modifier fragment	of other	12% (29)	11% (268)
PP-based	preposition + noun phrase fragment	of other	42% (103)	50% (1256)
VP-based	VP with active/passive verb (in order) to-clause fragment		1% (2) 3% (7)	1% (21) 2% (62)
Clause-based	anticipatory it +	adjectiveP VP	17% (42)	15% (366)
	there be-clause fragment		3% (8)	2% (56)
	noun phrase +VP		13% (32)	8% (204)
	conjunction + clause fragment		7% (18)	6% (148)
Other	other expressions		2% (6)	5% (117)
Total			100% (247)	100% (2498)

The following sections in this chapter will focus on investigating the similarities and differences between the Chinese and New Zealand English writers, as well as between the masters and PhD theses with regard to the identified five structural categories: NP-based, PP-based, VP-based, clause-based and other bundles.

It is important to note a major difference between the four groups of bundles in the use of the demonstratives *this* and *these*. The New Zealand students employed nearly 10% more demonstrative bundles than the Chinese students (i.e. 19% compared to 10%, 23% compared to 15%) and the doctoral students used approximately 5% more than the masters students (i.e. 15% compared to 10%, 23% compared to 19%). These demonstratives have an immediate referential function, which enhances the text cohesion of academic writing (Biber et al., 1999; Halliday & Hasan, 1976; Hinkel, 2004). The greater use of *this* and *these* suggest a stronger sense of coherence in the New Zealand and doctoral students' texts, as presented in the examples 1 and 2.

- (3) *The results of this study showed a pedagogic mismatch was evident, highlighting the difference between the teacher and learner perceptions of the short-or long-term instructional objectives of language learning tasks. (NZ MA)*
- (4) *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)*

Tables 20 and 21 compare the proportions of sentence initial bundles of each pattern between four corpora in terms of type and token.

**Table 20. Distribution of sentence initial bundles in each corpus (types)**

Category	Pattern		CH	CH	NZ	NZ
			MA	PhD	MA	PhD
NP-based	noun phrase with post-modifier fragment	of	8%	3%	19%	14%
		other	0%	0%	3%	2%
PP-based	preposition + noun phrase fragment	of	15%	15%	13%	20%
		other	26%	33%	22%	23%
VP-based	VP with active/passive verb (in order) to-clause fragment		1%	2%	0%	0%
			5%	5%	0%	0%
Clause-based	anticipatory it + adjectiveP VP		6%	8%	13%	16%
			6%	8%	8%	5%
		there be-clause fragment	1%	0%	5%	9%
		noun phrase +VP	16%	15%	11%	7%
	conjunction + clause fragment	10%	8%	5%	5%	
Other	other expressions		5%	2%	2%	0%
Total			100%	100%	100%	100%

**Table 21. Distribution of sentence initial bundles in each corpus (tokens)**

Category	Pattern		CH	CH	NZ	NZ
			MA	PhD	MA	PhD
NP-based	noun phrase with post-modifier fragment	of	8%	4%	20%	11%
		other	0%	0%	2%	2%
PP-based	preposition + noun phrase fragment	of	19%	14%	14%	20%
		other	34%	45%	24%	28%
VP-based	VP with active/passive verb (in order) to-clause fragment		1%	2%	0%	0%
			4%	3%	0%	0%
Clause-based	anticipatory it +	adjectiveP	4%	5%	15%	17%
		VP	5%	7%	6%	4%
		there be-clause fragment	1%	0%	4%	7%
		noun phrase +VP	10%	8%	8%	4%
		conjunction + clause fragment	7%	5%	5%	7%
Other	other expressions		8%	5%	2%	0%
Total			100%	100%	100%	100%

The inclusion of both type and token data can provide a complete picture of bundle distribution in the four corpora with type data indicating the number of different types of bundles and token data showing the total number of bundles.

### 6.2.1 NP-based bundles

Academic writing is considered to be “nouny” (Halliday, 1985), in which there is a prevalence of nouns and noun phrases. This can be explained in part by the conceptual rather than action-oriented nature of academic text. Biber and his colleagues found that the intensive use of noun phrases, primarily prepositional post modified phrases (e.g. the dominant use of *of*-phrases), was correlated with the grammatical complexity of academic writing (Biber, 2009; Biber & Gray, 2010; Biber et al., 2011). Unlike conversation, academic writing employs noun phrases instead of dependent clauses for structural elaboration.

In line with Biber and his colleagues’ finding, 90% NP-based bundles in this study were comprised *of*-phrase fragments, with the rest ending with other post-modifier fragments. Table 22 below lists the NP-based bundles in each corpus. It is clear to

see the pattern *The + N + of* in the *of*-phrase group and within this pattern three nouns, *results*, *purpose* and *analysis*, were shared between the Chinese and New Zealand students. However, the New Zealand students employed a considerably wider range of nouns (*results*, *findings*, *aim*, *purpose*, *analysis*, *limitations*, *use*). These were used to characterise and anticipate the results or findings, aim or purpose, analysis and limitations of their research or to describe the use of particular methods. Like the two New Zealand student corpora, the two masters corpora also manifested an extensive use of research-related nouns (*results*, *findings*, *aim*, *purpose*, *analysis*, *limitations*, *use*) compared to the PhD texts (*results*, *aim*, *purpose*, *analysis*), but this difference was not as marked.

**Table 22. NP-based bundles in each corpus in rank order**

CH MA	CH PhD	NZ MA	NZ PhD
<b>The results of the</b>	<b>The results of the</b>	<b>The results of the</b>	<b>The results of the</b>
<b>The purpose of</b>	<b>The analysis of</b>	The results of this	<b>The purpose of</b>
<b>this</b>	<b>the</b>	<b>The purpose of this</b>	<b>this</b>
<b>The purpose of the</b>		The majority of the	The first of these
One of the most		The aim of the	The results of this
The result of the		<b>The purpose of the</b>	<b>The analysis of the</b>
The main purpose		The findings of this	The aim of this
of		<b>The analysis of the</b>	The fact that the
		The aim of this	
		The limitations of	
		the	
		The findings of the	
		The use of the	
		The fact that the	
		The results from the	

*Note.* The bolded bundles represent the overlap between the Chinese and New Zealand corpora.

No use of the *noun phrase + other post-modifier* bundle was found in the Chinese corpora, whereas two occurred in the New Zealand texts, *The fact that the* and *The results from the*. *The fact that the* was always followed by a complementing noun clause and was popular in both New Zealand masters and PhD writing (i.e. 8 and 7 times per million words). However, the Chinese students did not use this bundle. This supports Aktas and Cortes's (2008) argument that non-native writers at masters and PhD level use fewer *the fact + noun clause* structures than the writers of published research articles.

According to Cortes (2013), most nouns in these bundles are shell nouns. Shell nouns are also known by various names: general nouns (Halliday & Hasan, 1976), anaphoric nouns (Francis, 1986), carrier nouns (Ivanič, 1991), enumerative nouns (Hinkel, 2001, 2002, 2004) signalling nouns (Flowerdew, 2003) and stance nouns (Jiang & Hyland, 2015). Examples of shell nouns are *fact*, *result*, *problem*, *approach* and *purpose*. 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 research-related shell nouns identified in the sentence initial 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 1 from a masters student's thesis<sup>3</sup>:

(1) *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)*

However, during interviews, it was found that the Chinese informants were unaware of the power of these nouns and noun phrases, although they employed a few shell nouns in their texts. Z considered his use of the noun phrase *The complaints from my colleagues and the results of the meetings* as an inferior choice and a temporary choice because it resulted in a long subject (Table 23). V was more conscious of the need to avoid word repetition, as a result of her learning and testing experiences,

---

<sup>3</sup> All examples are the original texts of the students with spelling, grammatical, lexical and punctuation mistakes unedited.

rather than the characterisation and linking functions performed by her selected shell nouns *definition*, *measurement* and *identity* (Table 24). The use of synonyms instead of the same word in her short text was likely to increase the cognitive load of her readers and undermine the cohesion and coherence of her text.

**Table 23. Z's interview on his use of *noun phrase***

Text	Interpretation
<i>The complaints from my colleagues and the results of the meetings often linger in my mind.</i> (Z)	I could not find a better sentence structure at the time of writing, so I used this phrase. The subject of this sentence is too long. There should be some other better expressions. (Z)

**Table 24. V's interview on her use of *noun phrase***

Text	Interpretation
<i>However, the definition of old varies from one society to another. The common measurement which is used to define old or ageing or elder is chronological age, but this is incorrect and misleading. The identity of old age is not only culturally different, but also distinct by class and gender.</i> (V)	I am changing the nouns in this paragraph to avoid repetition. These words are the same meaning. (V) My teachers suggested that I should not repeat words. They would change the word for me if I used one word repetitively. (V) The use of a wide range of vocabulary is also necessary to obtain higher marks in English tests, such as TOFEL and IELTS. (V)

### 6.2.2 PP-based bundles

The largest proportion of the sentence initial bundles were PP-based bundles. As shown in Tables 20 and 21 above, the proportions of the PP-based bundles in both the Chinese student corpora were generally higher than those in the New Zealand student corpora, and the two PhD corpora also contained higher proportions of PP-based bundles compared to the corresponding masters writing.

A preliminary analysis of PP-based bundles revealed that some bundles allowed the writer to mark logical relations between the elements. These bundles functioned as *complex prepositions* (Hinkel, 2004), consisting of multiword preposition sequences (e.g. *In the case of*, *On the basis of*, *As a result of*, *On the other hand*, *At*



*the same time*) or extended complex prepositions (e.g. *In other words, the, In addition to the, As a result, the*) to make texts cohere. Other bundles were used to identify time periods (e.g. *At the beginning of, At the end of, In the process of*) or discourse or research contexts (e.g. *In the present study, In this chapter, I, In this section, I*). Table 25 shows the percentage of these two groups of PP-based bundles in each corpus. Both the Chinese students and the PhD students relied more on complex prepositions to elaborate logical connections between their texts. The differences were not consistent between Chinese and New Zealand writers when time and context bundles were compared.

**Table 25. Distribution of the PP-based bundles in each corpus**

	CH MA	CH PhD	NZ MA	NZ PhD
Logical relation bundles (type)	26%	40%	22%	27%
Logical relation bundles (token)	35%	52%	26%	34%
Time & context bundles (type)	15%	8%	13%	16%
Time & context bundles (token)	17%	7%	12%	14%

The Chinese informant J provided two reasons for her use of sentence initial preposition phrases (Table 26). One was the writing habit developed from writing in Chinese and the other was her personal preference to achieve balance in her sentences. This can be seen in her interpretation below.

**Table 26. J's interview on her use of *multiple preposition phrase***

Text	Interpretation
<i>By means of 16 depth interviews with senior managers and staff in the local DMO as well as other stakeholders with diverse roles, five categories of critical specialities are identified: culture awareness, stakeholder partnerships, networking coordination, leadership and interest reciprocity. (J)</i>	I habitually place adverbial modifiers at the beginning of sentences. This is possibly the influence from my mother tongue. (J) As to this sentence, the adverbial modifier is too long, which is too heavy to put at the end of the sentence. (J)

Another interesting finding is the Chinese masters students were less likely to use complex prepositions with embedded *of*-phrases, although they demonstrated their ability to use prepositional units without the embedding *of*-phrase fragments (e.g.

*On the other hand, In other words, the, At the same time, In addition to the*). For example, the highly used bundles in the other three corpora, *In the case of, In terms of the* and *As a result of*, were largely underused in the Chinese masters writing.

According to informant V, these underused preposition phrases have been highly marginalised in English teaching and ignored during academic reading (Table 27). Therefore, Chinese students, like V, were not competent and confident enough to include the phrases such as *in the case of* and *in terms of* in their writing.

**Table 27. V's interview on her use of *multiple preposition phrase***

Text	Interpretation
<p><i>Situational barriers means participants' personal situations during daily life do not provide the condition for learning or have the contradiction with learning activities. For example, younger adult learners may "lack of time due to their job or home responsibilities". (V)</i></p>	<p>I know the phrases <i>with regard to, in the case of</i> and <i>in terms of</i>, but I do not know how to use them. (V)</p> <p>With your (the researcher's) suggestion, I know I can use these phrases here, but I will not choose them myself because I am not familiar with them and I may make a mistake. (V)</p> <p>During my learning, the teachers have rarely explained these phrases and they always suggest us to use <i>for example</i>. (V)</p> <p>I have never noticed these phrases while reading journal articles. I have to put great efforts to understand the meaning of reading, so I have paid little attention to these phrases. (V)</p>

### 6.2.3 VP-based bundles

VP-based bundles were only found in the Chinese students' writing. A number of Chinese students chose to start their sentences with verb phrases (*Based on the above* or *Look at the following*) or *in order to* or *to*-phrase fragments (e.g. *In order to make* or *To sum up, the*). *To sum up, the*, was the only shared bundle between the Chinese masters and Chinese PhD corpus. Other VP-based bundles performed apparently different functions in the Chinese masters and Chinese PhD writing: the bundles of the masters corpus (*Based on the above, In order to make/get/find*) indicated the pre-conditions of their main clauses; the bundles *To be more specific* and *To put it another (way)* of the PhD corpus were parts of fixed clusters to express

---

additional information. It is possible that more specialised and comprehensive PhD research requires more explanation and elaboration.

There were no VP-based sentence initial bundles in the New Zealand corpora. A comparative examination of sentence initial and non-initial bundles revealed that the New Zealand students usually employed the VP-based bundles in the second part of their sentences to add complementary information to their main clauses. The difference can be seen from examples 2 and 3 below:

- (2) ***In order to make** the participants get a main idea of task-based teaching method and make them have an understanding of what they should do in the class, the author briefly introduced task-based teaching method to the experimental class before the experiment briefly. (CH MA)*
- (3) *In chapter five, implications from the existing student data and responder and student interviews are drawn together **in order to make** some recommendations about the impact of socio-cultural contexts for mediating the learning of second language learners within the context of responsive written feedback. (NZ MA)*

As Williams (2003) points out, long introductory phrases hinder understanding and readers “have to hold in mind that the subject and verb of the main clause are still to come” (p. 138). Therefore, it is more appropriate to start a sentence with its topic rather than the wordy (*in order*) *to* phrase as in example (2). Vande Kopple (1989) recommends the end of a sentence as the place to express the most important information.

In contrast to this advice, the Chinese informants A, V and W regarded sentence initial (*in order*) *to*-infinitive phrases as an effective strategy to write concisely (Table 28), to highlight purposes (Table 28) and to reduce the information in the main clauses (Table 30). Both A and V attributed their use of sentence initial (*in order*) *to*-infinitive phrases to the transfer of Chinese (Tables 28 & 29). Not surprisingly, none of them have ever noticed the position of these phrases in their reading, and as V stated, nobody had picked up the sentence initial position as a mistake (Tables 28-30).

**Table 28. A's interview on his use of *to*-phrase fragment**

Text	Interpretation
<i>To interpret numbers, graphs and charts are used to show the meaning from the great amount of numbers which have more details but lower cognition load.</i> (A)	I believe it is concise to use these <i>to</i> -infinitive phrases, which indicate the purposes of these sentences. I put them at the beginning of sentences to highlight the purposes. (A)
<i>To be aware of, and to identify the skills that need to be learned from others, the participants used the knowledge stock in their mind which came from the manuals or the observations/experience and facilitated the awareness and recognition process.</i> (A)	This is my writing habit, maybe learned from my Chinese writing. Chinese sentences usually start with the indications of purposes. (A) I have never noticed the position of <i>to</i> -infinitive phrases in my reading. (A)

**Table 29. V's interview on her use of *to*-phrase fragment**

Text	Interpretation
<i>In order to make the interviews operating smoothly, some questions were prepared beforehand as prompts for interviews (see Appendix A).</i> (V)	I habitually used these verb phrases at the beginning of the sentences, maybe because of the transfer of the Chinese expression 为了 (in order to) (V)
<i>To help them involve more in learning activities, there are some important conceptions we should know:</i> (V)	I have never noticed the position of <i>in order to</i> in my reading. Nobody has ever picked up my sentence initial ( <i>in order to</i> - infinitive phrases as a mistake. (V)

**Table 30. W's interview on her use of *to*-phrase fragment**

Text	Interpretation
<i>In order to understand whether consumers' understanding and perceptions of purchasing a real estate go align or clash with the ideologies inferred in the advertising representations. Interviews will work for deeper probing into the complexity of consumers' behaviors and better address the issue.</i> (W)	I started this sentence with <i>in order to</i> because this sentence is very long. If I take out the modification, in this case, the <i>in order to</i> phrase, the main clause will become shorter. I always change the positions of sentence components according to the length of sentences, so that my readers can better understand my writing. (W) I have never noticed the position of <i>in order to</i> in my reading. (W)

#### 6.2.4 Clause-based bundles

There were four patterns of clause-based bundles: *anticipatory it-clause fragment*, *there be-clause fragment*, *noun phrase + verb phrase fragment* and *conjunction +*

clause fragment. As summarised in Table 20 and 21 above, the two New Zealand corpora contained a high proportion of clause-based bundles with the structures of anticipatory it + adjective phrase fragment and there be-clause fragment; whereas these two patterns occurred much less frequently in the two Chinese corpora, which had more bundles falling into the structures of noun phrase + verb phrase fragment and conjunction + clause fragment. Moreover, the two groups of PhD students used slightly more anticipatory it + adjective phrase fragment bundles and fewer noun phrase + verb phrase fragment bundles than their corresponding masters writers did.

The pattern anticipatory it-clause fragment can be further divided into anticipatory it + adjective phrase fragment and anticipatory it + verb phrase fragment. As proposed by Hyland and Tse (2005), the anticipatory it-clause fragment bundles highlight the writer's stance towards the argument but at the same time conceal the writer's identity and reduce the writer's responsibility for the argumentation. This is indicated in examples 4 and 5.

(4) ***It is important to recognise that voluntary migration can still result in communicative practices that can disempower citizens within New Zealand society and raises the challenge of how to integrate newcomers into the school environment. (NZ MA)***

(5) ***It is suggested that collocation be included into English exams and syllabus, thus learner can combine grammatical rules and lexical knowledge in a more scientific way and the improvement of their productive skills can be facilitated. (CH MA)***

There be-clause fragment pattern in the New Zealand student writing introduced the results of the research (*There was/were no/a significant, There appears to be*) (6) or acted as a topic sentence to inform the reader of the upcoming text (*There are a number*) (7). Only one *There be*-bundle, *There is no doubt*, appeared in the Chinese student corpora, and unlike those *There be*-bundles in the New Zealand student writing, this bundle expressed the writer's certainty towards his or her statement (8).

- 
- (6) **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)
- (7) Why would one assume that there is some kind of pairing across sets? **There are a number** of possible reasons, including (1) regular phonological alternations between specific cross-set pairs in some language or English. (2) One set being a structural mirror image of the other one, once we normalise for the factor that distinguishes the two sets (i.e. the structural relations within one set are exactly the same as those in the other one). (3) Articulatory/acoustic similarity of pairs of vowels across sets. (4) Spelling, which uses (in some instances at least) the same symbols for pairs of vowels. (NZ PhD)
- (8) **There is no doubt** that collocations can pose daunting problems to foreign language users and learners. (CH MA)

Short subjects (e.g. *This, The results, The present study*) attributed to the occurrences of the noun phrase + verb phrase fragment bundles. The higher frequency of these bundles explained the reason for the underuse of NP-based bundles in the Chinese students' writing. As can be seen in examples 9 and 10, noun phrases without modification (e.g. *the results*) were less clear and specific than modified ones (e.g. *the results of the writing behaviours discussed below*).

- (9) **The results show that** most Chinese English learners have the awareness of using strategies, but with different frequency. (CH MA)
- (10) **The results of the** writing behaviours discussed below did not show clearly or conclusively change during the treatment phases relative to the baseline phases across the seven children in the study. (NZ MA)

The pattern conjunction + clause fragment reflected the extensive use of single-word conjunctions. The conjunction *as* was the only shared conjunction across all four corpora (e.g. *As can be seen, As discussed in Chapter, As shown in table, As is shown in*). Besides *as*, the Chinese students used a wide range of other conjunctions such as *Therefore* (*Therefore, it is necessary*), *However* (*However, it is not, However, it should be*), *When* (*When it comes to*) and *So* (*So it is necessary*) to start

their sentences. Among them, only one conjunction *However* (However, it is important) appeared in the New Zealand masters texts.

### 6.2.5 Other bundles

In the “other” category, the bundle *That is to say*, was a top bundle in both Chinese corpora, which ranked as the second frequent bundle in the masters corpus with an occurrence of 57 times per million words (i.e. the total occurrences of *That is to say*, plus *That is to say*) and the third frequent bundle in the PhD corpus with an occurrence of 37 times per million words. In contrast, it occurred with a comparatively low frequency in the New Zealand masters and PhD corpus, 10 times and 3 times per million words respectively. The other two bundles — *Last but not least* and *First of all, the* — used for enumeration of units of texts, appeared only in the Chinese masters corpus.

## 6.3 Summary

In this chapter, the sentence initial bundles have been examined in terms of frequency and grammatical structures. Biber and his colleagues’ (1999) taxonomy was adjusted to investigate the structural patterns of sentence initial bundles in the Chinese and New Zealand student corpora. On the basis of the generated data, five major categories were created, which were NP-based, PP-based, VP-based, clause-based and other bundles. Three new structural patterns were added to the taxonomy and they were *there be-clause fragment*, *noun phrase + verb phrase fragment* and *conjunction + clause fragment*.

### 6.3.1 Differences between the Chinese and New Zealand writing

It was found that the New Zealand students used a considerably wide range of research-related nouns as shell nouns and a high proportion of anticipatory *it*, existential *there* clauses and demonstrative *this*. The Chinese students’ writing, on the other hand, was characterised by a relatively frequent use of sentence initial complex prepositions, verb phrases, conjunctions and enumerating linking adverbials. This violates Vande Kopple’s (1989) general principle of sentences writing: “topics often appear early in sentences” (p.52) and the end of sentences is usually used to express the most important information. Therefore, many of the

---

sentence initial elements in the Chinese student writing are likely to cause confusion and fail to convey key points. Moreover, the Chinese masters students rarely used complex prepositions with embedded *of*-phrases, which possibly reflected the more limited productive language repertoire of these L2 writers.

### **6.3.2 Differences between the masters and PhD writing**

The masters students employed more research-focused nouns in their texts. One possible reason is that as emerging researchers, they were likely to put more emphasis on the research-related activities “to showcase their ability to handle research methods appropriately and to demonstrate their familiarity with the subject content of the discipline” (Hyland, 2008a, p. 55). The PhD students used more anticipatory-*it* clauses and PP-based bundles. As more experienced researchers, they might be more confident and competent to incorporate interpretations and evaluations in their theses. The length of a PhD thesis also requires more work on cohesion and coherence.

In the next chapter, I will discuss the findings of interactive bundles and uncover the differences between Chinese and New Zealand thesis writing, or between masters and PhD levels of study in terms of sentence initial bundles. Interview data from Chinese postgraduates will also be presented to provide the interpretations for corpus data.





---

## Chapter 7 Interactive functions of the bundles

The analysis of functions in the present study has been based on Hyland's (2005a, 2005c) metadiscourse model. On the basis of this model, sentence initial bundles have been classified into two groups: interactive or interactional bundles. I will report on interactive bundles in this chapter and interactional bundles in Chapter 8. Identified interactive bundles consist of transition bundles, frame bundles, endophoric bundles, code glosses bundles, condition bundles and introduction bundles. I will present these bundles within each category and compare the use of bundles between Chinese and New Zealand postgraduates, or between masters and PhD levels of study. Possible interpretations of the identified typical Chinese bundles will also be provided.

Table 31 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, and the two masters corpora also contained a larger range of bundles (62 compared to 45, 47 compared to 35) in comparison to the PhD collections. There was a wider dispersion of the tokens in the two Chinese corpora. This can also be seen from the tokens of *On the other hand*, the most frequent interactive bundle in each corpus, which occurred almost twice in the Chinese texts than the New Zealand ones (62 vs 34, 65 compared to 28).

**Table 31. 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 32 shows the percentage in each interactive category. I

calculated the percentage in terms of tokens because tokens (here referring to the total occurrences of bundles) are a better way than type to reflect bundle distribution.

**Table 32. Distribution of interactive bundles in each corpus (tokens)**

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

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

There was considerable variation between the writers and the genres. Code gloss bundles and condition bundles were found to be more frequent in the Chinese students' writing, while transition bundles, endophoric bundles and introduction bundles were more common in the New Zealand students' writing. Transition bundles and condition bundles were the two categories consistently different amongst the two groups of PhD and masters writing, occurring more in the PhD corpora. The following is a close examination and comparison of the use of typical bundles in the four corpora.

### 7.1 Transition bundles

Transition bundles are expressions highlighting internal relations — addition, comparison or consequence — between units of texts. Table 33 summarises transition bundles in the New Zealand and Chinese corpora. The number 0 means no bundle has been generated from the corpus according to the pre-set criteria, but the string may still exist in the corpus. As shown in Table 33, three transition bundles were shared between all four corpora: *On the other hand*, *At the same time* and *In addition to the*. Another three bundles were shared between the Chinese and New Zealand texts but not in all corpora and they were *On the one hand*, *As a result of* and *As a result, the*. All transition bundles were divided into multiword transition bundles and single-word transition bundles according to the number of conjunction

words in the bundle. The most obvious difference between the Chinese and New Zealand writing was in the use of single-word transition bundles. Bundles, such as *Therefore, it is necessary*, *However, it is not* and *So it is necessary*, were heavily used in the Chinese students' writing.

**Table 33. Transition bundles in the New Zealand and Chinese corpora**

	NZ	Chinese
multiword transition bundles	<b>On the other hand</b> , (34, 28); <b>In addition to the</b> (10, 14); <b>At the same time</b> , (12, 18); <b>As a result of</b> (12, 7); <b>On the one hand</b> (0, 6); <b>As a result, the</b> (6, 0); In contrast to the (6, 6); In addition to this, (7, 0); At the same time (6, 0)	<b>On the other hand</b> , (62, 65); <b>In addition to the</b> (8, 15); <b>At the same time</b> , (38, 18); <b>As a result of</b> (0, 5); <b>On the one hand</b> (17, 27); <b>As a result, the</b> (15, 11); As a result, it (5, 0)
single-word transition bundles	However, it is important (7, 0)	Therefore, it is necessary (7, 0); However, it is not (7, 0); However, it should be (0, 5); So it is necessary (5, 0)

*Note.* The numbers in the brackets give the tokens in each corpus: the first number the masters corpus and the PhD corpus. Shared bundles between the New Zealand and Chinese corpora are in bold.

### 7.1.1 Shared transition bundles

Three transition bundles, *On the other hand*, *At the same time* and *In addition to the*, were shared across all four corpora. Therefore, it is possible to compare their locations in these texts and the comparison of locations may explain the comparatively high frequency of these three bundles in the Chinese student writing. Bundles can be located at the beginning, at the end or in the middle of the sentences. The location “at the beginning” means the sentence starts with the bundle, “at the end” indicates the sentence ends with the bundle and “in the middle” includes all the other cases. It is regarded as a middle bundle if the transition bundle immediately follows the subject of the sentence (1).

(1) *English politeness, **on the other hand**, relies on different space-giving devices such as register and indirectness, he claims. (NZ PhD)*

While calculating, the tokens of similar strings with different punctuation marks (e.g. *at the same time* and *at the same time.*) were added up. Table 34 displays the percentage of these three shared bundles at different parts of the sentences. The New Zealand texts were analysed first to identify the locations of these bundles in native speaker writing; therefore, the use of these bundles in Chinese writing could be addressed through comparison.

For the bundles *On/on the other hand* and *In/in addition to the*, the most common location was initial location and the second common location was medial location. It was rare for the former and impossible for the latter to occur in final location. The bundle *At/at the same time*, different from the other two bundles, was more often placed in medial location than initial location and also occurred in final location.

**Table 34. Locations of the three shared transition markers**

	on the other hand			at the same time			in addition to the	
Position	initial	medial	final	initial	medial	final	initial	medial
NZ MA	68%	32%	0%	32%	54%	13%	71%	29%
NZ PhD	51%	47%	2%	35%	54%	11%	55%	45%
CH MA	74%	26%	0%	45%	40%	15%	78%	22%
CH PhD	69%	28%	3%	32%	54%	14%	67%	33%

Both Chinese groups showed a preference for using these bundles as sentence initial bundles and the only exception was the use of *at the same time* in the Chinese PhD theses, which was similar to the distribution in the New Zealand PhD corpus. Another significant difference was the use of *on the other hand*. Both New Zealand and Chinese students used it in medial location. However, in the New Zealand students' writing, this bundle often immediately followed the subject, while in the Chinese student texts especially the masters level writing this bundle frequently occurred before the second part of the coordinate clause. The following examples (2, 3) illustrate this respectively.

(2) *All interviewees had been brought up on dairy farms in South Taranaki, had attended primary schools in the region followed by attendance at Opunake High School. All had spent the early years of their adulthood away from the area and had returned to take up dairy farming in their midtwenties. On leaving*

---

high school the men attended polytechnic in New Plymouth or in one instance Massey University in Palmerston North, to obtain agricultural and/or trade certificates/diplomas and they worked in New Plymouth as tradesmen for a few years. The women **on the other hand** moved to New Plymouth and worked in offices, banks or hair dressing salons. (NZ PhD)

(3) On one hand, in order to make students more aware of how to learn more efficiently and effectively, the learning strategies must be instructed, **on the other hand**, the teacher has to complete his or her teaching work according to the curriculum that involves no specific training of learning strategies. (CH MA)

Vande Kopple (1989) suggests putting transition signals early in sentences but not as the first element, unless more emphasis is placed on the contrasting point. Williams (2003) interprets the relationship between topics and coherence as:

Readers judge a passage coherent to the degree that they quickly and easily see two things:

- the topics of individual sentences and clauses.
  - how the topics in a whole passage constitute a related set of concepts.
- (Williams, 2003, p. 85)

In this case, the New Zealand PhD student began his or her sentence with the short simple noun phrase *The women* (underlined in the text), as both the subject and topic of this sentence. At the same time, this noun phrase, together with the previous sentence topics *All interviewees*, *All* and *the men* (underlined in the text), formed a set of related concepts. In other words, the New Zealand student relied more on the noun phrase *The women* rather than the transition marker *on the other hand* to create a sense of cohesive and coherent flow. The Chinese student, on the other hand, solely depended on the transition marker to connect his or her poorly structured clauses, with *the learning strategies* (underlined in the text) as the subject and topic of the first clause; *the teacher* (underlined in the text), a weakly related concept as the subject of the second clause.

Another three bundles were partially shared between the Chinese and New Zealand texts and they were *On the one hand*, *As a result of* and *As a result, the*. Unlike *On the other hand*, *On the one hand* was a much less popular bundle among the New Zealand students, but it occurred frequently in the Chinese student writing. Many Chinese students, like the informants J and V (Tables 35 & 36), who learned English from their L2 teachers and course materials, and lacked access to first-hand knowledge of the language. Therefore, the rules were likely to be partially learned and incorrectly generalised.

**Table 35. J's interview on her use of *on the one hand* and *on the other hand***

Text	Interpretation
<i>On the one hand</i> , networking relationships, or informal social networks, are believed to be used predominantly by the DMO in China society within the traditional collectivism system . . . . <i>On the other hand</i> , there is increasingly demand of rule-based governance as the development of rural tourism being into the way of modern governance . . . . (J)	I always use them ( <i>on the one hand</i> and <i>on the other hand</i> ) in pairs and I do not know <i>on the other hand</i> can be used alone. I have never learned this. I learned from my teacher how to use them. I have not even noticed this usage while reading newspapers. By the way, if I only use <i>on the other hand</i> , can my readers understand there are two contrasting points? (J)

**Table 36. V's interview on her use of *on the one hand* and *on the other hand***

Text	Interpretation
<i>On the one hand</i> , policy makers encourage lifelong learning initially because of economic reasons (Field, 2012); <i>on the other hand</i> , lifelong learning would bring more benefits to individuals, especially for elder people to maintain a meaningful active later life. (V)	These two phrases should occur in pairs. My teacher said so and they always appeared together in my English exercises. I do not know <i>on the other hand</i> can be used alone so I used them as a pair. (V)

The New Zealand and Chinese students showed their different preferences towards the use of *As a result* bundles: *As a result of* (12 per million words in New Zealand masters and 7 per million words in New Zealand PhD writing compared to 0 per million words in Chinese masters and 5 per million words in Chinese PhD writing) and *As a result, the* (6 per million words in New Zealand masters and 0 per million

words in New Zealand PhD writing compared to 15 per million words in Chinese masters and 11 per million words in Chinese PhD writing). The New Zealand students chose *As a result of* together with shell nouns (e.g. *demand* and *processes*) to specify the causes and to closely link to the preceding propositions (4, 5). The Chinese students merely addressed the consequences with *As a result, the* (6, 7).

- (4) *In Aotearoa New Zealand, immersion Māori education initiatives, like the one observed in Petone Central School, have grown from Māori effort. The success of these programmes can be measured in their increasing number: as of 1991, 1 percent of Māori primary school students were enrolled in kura kaupapa Māori; as of 1993, 49.2 percent of Māori children enrolled in pre-school were at a kōhanga reo (Ministry of Education 2004). **As a result of** the demand, the number of immersion schools has more than doubled between 1999 and 2003 (Ministry of Education 2004). (NZ MA)*
- (5) *Different approaches and strategies were employed by the participants in the study for re-constructing their previous identity and gaining new qualities that allowed them to claim agency and co-ownership of socio-cultural resources in the society. **As a result of** these processes, new meanings were created by some of them which constructed novel frameworks for articulating immigrant identity. (NZ PhD)*
- (6) *In this class, every one had the chance to express and show in the face of the other classmates. **As a result, the** students leaved the classes with a great sense of achievement because they discovered abilities they did not know they had. (CH MA)*
- (7) *Remember that in section 3.2 we mentioned Passive is in effect more marked than It Extra, and Existential there constructions in that its derivation involves one movement while the derivation of the latter two does not. **As a result, the** production of Passive is supposed to be less than that of It Extra, and Existential there constructions. (CH PhD)*

### 7.1.2 Transition bundles in the Chinese students' writing

A group of Chinese student bundles started with one-word conjunctions such as *Therefore*, *However* and *So*. One possible reason was that the Chinese students



lacked the knowledge of a wide variety of cohesive devices and had to depend on single-word conjunctions to connect their ideas. The New Zealand students, on the other hand, were more competent at using alternative linking devices, for example, concept-related nouns as discussed above. Another possible reason was that the Chinese students preferred to begin their sentences with conjunctions to immediately illustrate the particular relations to the preceding sentences (8), whereas the New Zealand students were more likely to place conjunctions in the first part rather than exactly at the beginning of their sentences when the conjunctions showed connections between sentences (9).

- (8) *Thirdly, limited pixels of computer screens may cause problems of recognition when it comes to images. **Therefore it is advisable** to include an image enlargement function. (CH PhD)*
- (9) *These prosodic markings characteristic of disagreements have the potential to offend an interlocutor and threaten the addressee's face. **It is therefore important** to look at the second key feature of this study, namely how to counteract potential offence through the use of politeness strategies and modification devices. (NZ PhD)*

The informant S's words provided two additional reasons: his habit of marking logical relations between sentences and his familiarity with these conjunctions (Table 37).

**Table 37. S's interview on his use of *however* and *therefore***

Text	Interpretation
<p><i>After running an EFA, it suggested there are two factors in this scale. <b>However</b>, a CFA didn't prove that was a good model, as the fit indices were outside of the acceptable range. (S)</i></p> <p><i>Hair et al. (2006) recommended values of .60 to .70 are deemed the lower limit of acceptability. <b>Therefore</b>, in this research, it seems reasonable that the 3-item positive factor with sufficient sample size generated a relatively lower Cronbach alpha of .65. (S)</i></p>	<p>While writing, after expressing one idea, I would like to use some connector to connect it with the following idea; otherwise, I will feel the logical relations within paragraphs are not clear. With these conjunctions, it is much easier for readers to understand my writing. (S)</p> <p>I also use multiword connectors, but these single-word conjunctions always occur to my mind first because <i>however</i> and <i>therefore</i> often appear in my reading and the revisions from my teachers. (S)</p>

Like the use of *however* in the Chinese student corpora, initial position was also the informant V's favourite position for *however* (Table 38). She believed her English proficiency and linguistic confidence were two reasons for this. These two reasons confirmed Paquot's (2012) argument that "[p]ositional variation of connectors is usually not taught, and learners use the sentence-initial position as a safe bet" (p. 203). The informant V intuitively favoured *subject + however* over sentence initial *however* and labelled the former position as "authentic", but she has not given the underlying reason.

**Table 38. V's interview on her use of *however***

Text	Interpretation
<i>However, the definition of old varies from one society to another.</i> (V)	I am used to putting <i>however</i> at the beginning of my sentences. It has been years. I am confident with it. (V)
<i>However, ageing or old age does not mean dull or stagnant.</i> (V)	I'm afraid I will make a mistake if I put
<i>However, after about 20 years of development, elder people's participation in learning activity experiences a significant rise and both the researchers and curriculum designers started to concern about this special group.</i>	I put <i>however</i> after the subject of the sentence, so I try not to use it in this way. (V)
<i>However, it still has a long way to go.</i> (V)	It sounds authentic when you put <i>however</i> after the subject. This is better, but I could not think out this pattern myself. (V)

## 7.2 Frame bundles

Frame bundles function as signposts, signalling the boundaries of arguments (e.g. *The thesis consists of, In this chapter, I, In this section, we*), labelling the stages of texts (e.g. *To sum up, the, In a word, the*) and describing text-internal (e.g. *The first of these, This is followed by, First of all, the, Last but not least,*) or text-external sequences (e.g. *At the beginning of, At the end of, At the time of, In the process of*). Therefore, frame bundles were further classified into boundary bundles, discourse-label bundles and sequence bundles.

Hyland (2005a) defines frame markers as text-internal references. However, in this study it was found that the bundles used to sequence research processes (e.g. *At the beginning of, At the end of, In the process of*) could also function as frame markers ordering units of texts. As in example 10 below, the bundle *At the end of* echoed the time marker *At the start of* in the first sentence, sequencing the stages of peer

review and connecting the two pieces of texts into a cohesive paragraph. The only difference identified between the text-internal and external reference bundles was the genre of the texts: internal reference bundles ordered claims, evidence, explanations and argumentations, and external reference bundles sequenced narratives.

(10) *At the start of the peer review, students exchanged essays with a partner of their own choosing and then answered about their partner's essay a series of questions on a handout. The handout contained 13 different questions and asked the student to do such things as identify the topic and purpose of the essay from the introduction, name the supporting arguments in the body and count the number of citations used in the essay. At the end of the activity, the students gave the essays back to their partners and discussed with their partners their revision ideas for writing the final draft. (NZ PhD)*

Table 39 presents the shared and unique bundles in the New Zealand and Chinese corpora. One boundary bundle *In this section, I* and two sequence bundles *At the end of* and *At the beginning of* were employed by both L1 and L2 writers. Apart from these, most bundles were used differently. No discourse-label bundle, the second type of frame bundles, was found in the New Zealand students' writing. The following sections will examine the boundary bundles, discourse-label bundles and sequence bundles in the Chinese and New Zealand students' writing.

**Table 39. Frame bundles in the New Zealand and Chinese corpora**

	NZ	Chinese
boundary bundles	<b>In this section, I (0, 5)</b> In this chapter I (7, 8); The chapter concludes with (9, 0); The next chapter will (6, 0); This chapter describes the (5, 0); In this section the (5, 0); In this section I (0, 6)	<b>In this section, I (0, 7)</b> In this chapter, we (7, 9); In this section, the (7, 7); In this chapter, the (19, 0); In this part, the (6, 0); In this section, we (0, 7); The thesis consists of (6, 0); This thesis consists of (5, 0)
discourse-label bundles		To sum up, the (8, 9); In a word, the (6, 0)
sequence bundles	<b>At the end of (9, 15);</b> <b>At the beginning of (0, 6)</b> At the time of (10, 7); By the end of (9, 0); The first of these (0, 6); This is followed by (0, 5)	<b>At the end of (17, 0);</b> <b>At the beginning of (13, 0)</b> In the process of (28, 0); During the process of (12, 0); Last but not least (7, 0); First of all, the (6, 0); The first one is (6, 0); In the course of (5, 0)

### 7.2.1 Boundary bundles

Boundary bundles signal the scope of the text. The bundle *In this section, I* was the only shared boundary bundle between the Chinese and New Zealand students' writing to introduce or summarise the main ideas of the section (11, 12).

- (11) *In this section, I will survey briefly the thought of some of the major figures in the pragmatic and cognitive study of conditionals and mention some of the contemporary findings in them. (CH PhD)*
- (12) *In this section, I have established that there are complex issues involved in describing the dimension of rhoticity in English phonological systems. (NZ PhD)*

Many other boundary bundles in the Chinese and New Zealand texts, though shared the scope-indicating words (e.g. section, chapter), exhibited slight variations.

Bunton's (1999) levels of scope were adapted to investigate the boundary bundles in the Chinese and New Zealand theses, that is, how much text is referred to (e.g. sentence, paragraph, section, chapter and thesis). The bundles were divided into three levels in terms of scope: section (e.g. *In this section, I, In this part, the*), chapter (e.g. *In this chapter I, The chapter concludes with*) and thesis (e.g. *The thesis consists of*). Table 40 shows the distribution of the boundary bundles across these three levels. The difference between the Chinese and New Zealand students' writing was not significant, but the PhD students included a considerably larger proportion of section-level bundles (75% compared to 33%, 67% compared to 20%) and smaller proportion of chapter-level bundles (25% compared to 33%, 33% compared to 80%) compared to their masters counterparts. The length of PhD theses might require more introductions and summaries at section level. At thesis level only two bundles occurred in the Chinese masters corpus (*The/This thesis consists of*), but none in the other corpora.

**Table 40. Scope distribution of boundary bundles (token)**

	CH MA	CH PhD	NZ MA	NZ PhD
Section level	25%	70%	16%	58%
Chapter level	51%	30%	84%	42%
Thesis level	24%	0%	0%	0%
Total	100%	100%	100%	100%

### 7.2.2 Discourse-label bundles

Discourse-label bundles are used to mark the stages of text development. No discourse-label bundle was found in the New Zealand students' writing and two summarisation bundles were identified in the Chinese student texts, among which, *To sum up, the* was a shared bundle between the Chinese masters and PhD writing (13, 14) and *In a word, the* was commonly employed in the Chinese masters writing (15), but did not occur in the New Zealand thesis corpora.

(13) *To sum up, the most important requirement for the teachers is that the teachers should observe students' intellectual characteristics, capabilities, interests, etc, as carefully as possible, and take up more alternative assessment techniques to suit different kinds of students. (CH MA)*

(14) *To sum up, the case study revealed quantitative and qualitative differences in motivational regulation between high and low achievers. (CH PhD)*

(15) *Without the carrier, the language users might find it difficult to recall word by word. In a word, the mnemonic is more memorable than the target material, and so is more likely to be recalled successfully. (CH MA)*

The informant V described her learning experience of *in a word*, which was introduced as a language point along with and equally important as *in other words* and reinforced by error correction exercises.

I have learned *in a word* and *in other words* as two parallel patterns. *In a word* introduces one word, while *in other words* starts a sentence. I have done a kind of exercises named error correction and one of the questions is about the use of

---

this phrase. I have never used *in a word* because I cannot generalise my ideas into one word. (V)

However, the frequency of *in other words* in the BNC is 32.43 per million words, whereas that of *in a word* is merely 1.59 per million words. It is not worthwhile to introduce such a low-frequency item to L2 students. As can be seen from example 15, the parallel introduction confused the student, who inappropriately used *in a word* to start a sentence.

### 7.2.3 Sequence bundles

Sequence bundles are used to order texts. The two shared bundles between the Chinese and New Zealand students were *At the end of* and *At the beginning of*, and both of them were used as text-external sequence bundles (16-19).

- (16) *At the end of the writing lesson, when the children had completed their writing and had read to another child, the children were able to illustrate their stories. (NZ MA)*
- (17) *At the end of this semester, the students of the two classes took the final exam that was used as the source of the posttest. (CH MA)*
- (18) *At the beginning of the third process interview, which followed the completion of the final draft, the participants were instructed to recreate on paper their writing process from start to finish. (NZ PhD)*
- (19) *At the beginning of the experiment, the two classes of students with the same English level were chosen to participate in the experiment. (CH MA)*

The number of sequence bundles in the Chinese masters corpus, including both text-internal (e.g. *Last but not least, First of all, the, The first one is*) and text-external sequence bundles (e.g. *In the process of, At the end of, At the beginning of*), far exceeded that of the rest three groups. There was no sequence bundle in the Chinese PhD corpus, only text-external sequence bundles in the New Zealand masters corpus and far fewer text-internal and external sequence bundles in the New Zealand PhD corpus.

The text-internal bundles of the Chinese masters writing were general sequence signposts without any specific reference (20-22), while the text-internal bundles of the New Zealand PhD writing, *The first of these* (23), *This is followed by* (24), with the use of demonstratives, *these* or *this*, immediately referred back to the preceding text and introduced the succeeding information at the same time. In the following examples, *The first of these added elements* linked back to *three further elements*, and the demonstrative pronoun *this* in *This is followed by the concluding chapter* referred to *Chapter Eight* in the previous sentence. The use of demonstratives in the sequence bundles, as noted before, improved textual cohesion.

- (20) ***Last but not least**, reviewing the practiced learning strategies is necessary to ensure the training success. (CH MA)*
- (21) ***First of all, the** author will draw a basic distinction between the real-world or target tasks and pedagogical task. (CH MA)*
- (22) *In this paper, there are three questions to be studied. **The first one is** why the author has chosen the adult learners as the participants in the training institutions? (CH MA)*
- (23) *This added three further elements to the analytic model, made possible by the intensive case study research design. **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)*
- (24) *Chapter Eight is the discussion chapter where the key results from all three case studies will be discussed. **This is followed by** the concluding chapter (Chapter Nine), which discusses the educational implications of the findings of this research. (NZ PhD)*

The Chinese informants of this study expressed divergent attitudes towards sequence markers. The informant Z recalled his experience of learning sequence markers (Table 41). The training and assessment of IELTS writing in China were obviously a crucial factor, in which not only had he first encountered the idea of writing framework, but the effectiveness of writing framework had been proved by his final IELTS mark. As a result, he believed it was necessary to number the arguments in writing.

**Table 41. Z's interview on his use of sequence markers**

Text	Interpretation
<p><i>Firstly, research shows that most scholars' tri-multilingual studies to date have been conducted largely in culturally western settings . . . . Secondly, Chinese tri-multilingual education emerges from bilingualism sharing the characteristics of all the nationalities' education . . . . Thirdly, this study investigates trilingual education in China . . . . Fourthly, this study will help to solve the confusions of my colleagues . . . . Lastly, this special study will be a contribution toward trilingualism . . . . (Z)</i></p>	<p>I think it is logical and clear to number the arguments; otherwise, the relations between my arguments become obscure. I learned to sequence my arguments from IELTS writing. It is popular in China to adopt some ready-made frameworks to IELTS writing. I have gained 5.5 points in writing when I first attended IELTS. I was very surprised. It was impossible. The second time, I followed its framework and then received 7 points, much higher. (Z)</p>

Unlike Z, the informant W tried to avoid using sequence markers in her writing, although she had also encountered these markers in her learning and reading (Table 42). She did not regard sequence markers as effective frame markers and only reserved them as the last selection when she could not find any alternative.

**Table 42. W's interview on her use of sequence markers**

Text	Interpretation
<p><i>Firstly, quantitative content analysis will be adopted . . . . Secondly, selecting the representative samples and deconstruct the visual persuasion device . . . . Thirdly, qualitative interviewing to probe into . . . . Lastly, compare and examine whether . . . . (W)</i></p>	<p>I do not like to use these sequence markers in my writing because my writing is not an instruction. I prefer to choose a more natural and cohesive way. Here I am using them because I do not know any alternative way. (W) I learned these sequence markers from my teachers and my reading. (W)</p>

The informant V agreed with Z and W that Chinese teachers put effort into introducing sequence markers as a general strategy to achieve cohesion and coherence of English writing (Table 43). She also expressed her personal opinion about these sequence markers — *too rigid* and indicated her opposite finding from her reading — *journal articles rarely used sequence markers and hardly put the markers at the beginning of sentences.*



**Table 43. V's interview on her use of sequence markers**

Text	Interpretation
<i>The last but not least, I will analyse some possible suggestions for future research.</i> (V)	I learned this phrase in China. We do not use so many conjunctions (连词) in Chinese writing, but English speakers like to connect ideas. My teacher told us cohesion and coherence were important in English and these conjunctions were fairly important. Chinese teachers put efforts on teaching conjunctions in English writing classes. If I do not use conjunctions in my writing, I will lose marks. (V) My writing is not cohesive because I do not like to use sequence markers such as <i>firstly</i> and <i>secondly</i> . I think it is too rigid. (V) While reading, I found journal articles rarely used sequence markers and hardly put the markers at the beginning of sentences. Chinese students favour sequence markers and like to put them at sentence initial position. I have already noticed the difference and I am trying to avoid the Chinese way of writing. (V)

### 7.3 Endophoric bundles

Endophoric bundles refer the reader to other parts of the text, which include the previews, reviews or overviews of the unfolding texts (e.g. *As discussed in Chapter*) or the additional materials such as tables, figures, examples, extracts, etc. (e.g. *As shown in Table*). Shell noun bundles, as discussed in Chapter 6, are an important component of endophoric markers, referring to preceding or succeeding clauses or noun phrases.

Table 44 shows the different distribution of endophoric bundles between the New Zealand and Chinese corpora. The New Zealand students used more shell noun bundles, whereas the Chinese students relied heavily on other types of endophoric bundles. As discussed in the Section 6.2.1 NP-based bundles, the use of shell noun bundles in the Chinese writing was limited to *result(s)*, *purpose* and *analysis* bundles, while the shell noun bundles deployed by the New Zealand students

contained a wide variety of research-related nouns. Chinese students were found to lack knowledge of shell nouns.

This section will not focus on shell noun bundles but on other endophoric bundles, including the shared bundles (*As can be seen* and *It can be seen*), the unique New Zealand bundle (*As discussed in Chapter*) and the two prevailing Chinese patterns (*As shown in Table/As is shown in* and *The following is/are a/an/the/some*). Shell noun bundles have been covered in Section 6.2.1 NP-based bundles.

**Table 44. Endophoric bundles in the New Zealand and Chinese corpora**

	NZ	Chinese
shell noun bundles	<b>The results of the (23, 12);</b> <b>The purpose of this (12, 7);</b> <b>The purpose of the (6, 0);</b> <b>The analysis of the (6, 6)</b> The results of this (13, 6); The aim of this (6, 6); The fact that the (8, 7); The majority of the (11, 0); The aim of the (8, 0); The findings of this (6, 0); The results from the (6, 0); The limitations of the (5, 0); The findings of the (5, 0); The use of the (5, 0)	<b>The results of the (29, 24);</b> <b>The purpose of this (12, 0);</b> <b>The purpose of the (8, 0);</b> <b>The analysis of the (0, 7)</b> The result of the (8, 0); The main purpose of (7, 0)
other bundles	<b>As can be seen (13, 9);</b> <b>It can be seen (5, 0)</b> As discussed in Chapter (6, 17)	<b>As can be seen (5, 14);</b> <b>It can be seen (13, 15)</b> As is shown in (10, 7); The following are some (6, 5); The following is a (5, 9); As shown in Table (5, 7); The following table shows (8, 0); From the above table, (5, 0); We can see from (5, 0); The following is the (5, 0); Look at the following (0, 11); The following is an (0, 6)

### 7.3.1 Shared endophoric bundles

*As can be seen* was an expression shared by all four student corpora. As an endophoric bundle of textual acts, *As can be seen* was usually followed by the

preposition *from* or *in*, steering the reader to tables, figures, examples, data or other additional sources (25, 26).

- (25) *As can be seen from Table 8 children made more verbal initiations to peers than to adults. (NZ MA)*
- (26) *As can be seen in Figure 6.2, over two thirds ( $43.04 + 24.48 = 67.52\%$ ) of the teacher educators did not support an 'English-only' policy in the classroom. (NZ PhD)*

*It can be seen* was not popular among the New Zealand PhDs but shared by the other three groups of students, especially the Chinese postgraduates. *It can be seen*, like the bundle *As can be seen*, instructed readers to different textual sources, for example, tables in example 27. Another important function of this bundle was to draw the reader's attention to the writer's conclusion, contained in the subsequent *that*-clause, as in examples 28 and 29. In other words, *It can be seen* also evoked cognitive acts of the reader.

- (27) *It can be seen from these two tables that the participant teachers and their students differ substantially in all sections. (CH PhD)*
- (28) *It can be seen that after a period of strategy training, the students from the experimental classes have improved some. (CH MA)*
- (29) *It can be seen that word classes are not represented evenly throughout a text. (NZ MA)*

### 7.3.2 Endophoric bundles in the New Zealand students' writing

*As discussed in Chapter* was a typical endophoric bundle in the New Zealand students' writing. As in example 30, it reminded the reader of the relevant information in the previous Chapter (i.e. the multilingual practices at almost all Luxembourgish and German banks in Chapter four) and at the same time provided the pre-condition for the present argument. That is to say, the case discussed in the current Chapter (i.e. the de facto policy at Bank George and Bank Ivan) could be generalised to a larger context (i.e. almost all Luxembourgish and German banks). A sense of the whole, therefore, was effectively created through the use of *As*

*discussed in Chapter*. However, this bundle rarely occurred in the Chinese student texts.

(30) *The context of Luxembourg influenced top down de facto policy considerably and a value for multilingualism was not limited to Bank George and Bank Ivan. As **discussed in chapter** four, managers at almost all Luxembourgish and German banks of various sizes recruited multilingual staff and made flexible use of multilingual mechanisms of recruitment and language courses within their banks. (NZ PhD)*

### 7.3.3 Endophoric bundles in the Chinese students' writing

There were two popular patterns in the Chinese students' writing: *As shown in Table/As is shown in* and *The following is/are a/an/the/some*. *As is shown in* was a rare expression in native speaker writing, occurring 0.22 times per million words in the BNC. It was similar to *As shown in Table* (31) but used as a multi-reference expression with a much broader reference scope to tables (32), figures (33) or even chapters (34). As in example 34, however, the reference function could be better achieved through the use of another more appropriate cluster such as the New Zealand students' bundle *As discussed in Chapter*.

(31) ***As shown in Table** 5.16, with respect to both academic vocabulary and vocabulary at other levels, the subjects with higher proficiency also shows a higher P/R ratio than the subjects with lower proficiency. (CH PhD)*

(32) ***As is shown in** Table 7, the most powerful predictor of the dependent variable is L2 writing self-efficacy, which has the highest absolute B value of.233. (CH MA)*

(33) ***As is shown in** Figure 2.5, three main sets of affective strategies exist: lowering your anxiety, encouraging yourself, and taking your emotional temperature. (CH MA)*

(34) ***As is shown in** chapter one, the listening and speaking instruction of the postgraduates in China is not that satisfactory. (CH MA)*

*The following is/are a/an/the/some* was a cataphoric reference and the same to *As is shown in*, this bundle was also deployed as a general multi-functional reference, informing the reader of a fairly wide range of issues including reviews (35), analyses (36), problems (37), examples (38), tables (39), etc. The use of these multi-function bundles reflects that these Chinese students may only have a very limited range of reference strategies under control.

(35) ***The following is a*** *brief review of major studies both at home and abroad examining the nature and characteristics of language learning strategies. (CH MA)*

(36) ***The following is an*** *analysis as to why the adoption of TBLT in the Integrated English teaching leads to the changes in these four aspects. (CH MA)*

(37) ***The following are the*** *main problems among others that affect the appropriateness of style. (CH PhD)*

(38) ***The following are some*** *examples from BNC: (CH PhD)*

(39) ***The following is the*** *table presenting pretest results of EG and CG. (CH MA)*

#### 7.4 Code gloss bundles

According to Hyland (2007a), code gloss bundles elaborate on meanings through reformulation or exemplification:

Reformulation is a discourse function whereby the second unit is a restatement or elaboration of the first in different words, to present it from a different point of view and to reinforce the message.

Exemplification is a communication process through which meaning is clarified or supported by a second unit which illustrates the first by citing an example. (Hyland, 2007a, pp. 269-270)

As distinguished in Hyland (2007a), reformulation either expands the reader's understanding or narrows down the scope of interpretation. Exemplification mostly offers a more accessible item or a case from real life.

*In other words, the, That is to say, and This suggests that the* were three shared bundles between the Chinese and New Zealand corpora (Table 45). Apart from these, the code gloss bundles in the New Zealand writing all started with demonstrative *this*, reformulating the anaphoric texts (*This is not to, This is because the* and *This is not a*). None of them was used in the Chinese corpora and the Chinese students relied on other strategies to restate their meanings, strategies that included idiomatic phrases (*For example, in the, To be more specific* and *To put it another*) and the *mean* bundles (*It/This means that the*).

**Table 45. Code gloss bundles in the New Zealand and Chinese corpora**

	NZ	Chinese
code gloss bundles	<b>In other words, the (9, 11);</b> <b>That is to say, (10, 0)</b> <b>This suggests that the (6, 0)</b> This is not to (6, 6); This is not a (0, 5); This is because the (7, 0); In other words the (6, 0)	<b>In other words, the (20, 39);</b> <b>That is to say, (51, 37)</b> <b>This suggests that the (0, 6)</b> For example, in the (7, 6); In other words, they (7, 5); That is to say (6, 0); It means that the (5, 0); To be more specific, (0, 9); To put it another (0, 6); This means that the (0, 6); In other words, it (0, 6)

#### 7.4.1 Shared code gloss bundles

*In other words, the, That is to say, and This suggests that the* were shared between the Chinese and New Zealand thesis corpora. Both *In other words* and *That is to say* served similar functions in elaborating their preceding statements. The sense of equivalence between the preceding and succeeding texts was conveyed through *In other words* and *That is to say*. The equivalent, probably simpler or more exact information was provided to enhance the reader's knowledge construction with a further explanation (40), illustration (41) or conclusion (42, 43).

(40) *According to Richards (1976), Nation (1990, 2001) and Laufer (1990b, 2002), to know a word does not imply to know only the basic meaning of it. Knowing a word involves knowing its form (spoken and written form), position (grammatical patterns and collocations), function (frequency and*

*appropriateness) and meaning (concept and associations) (Nation, 1990). In other words, word knowledge is multi-dimensional and learning a word means learning the various types of word knowledge. (CH PhD)*

(41) *The influences on teacher educators were therefore assumed to be related to the wider social context. In other words the model predicts that teacher educators are influenced by the prevailing ideology about bilingualism and language diversity, particularly as it is expressed in the education system and the specific ethnolinguistic vitality of various groups in the community. (NZ PhD)*

(42) *The teacher should carry out his teaching according to the teaching plan. But he should also adjust his teaching according to the concrete situation. That is to say he can deal with what may happen unexpectedly. (CH MA)*

(43) *The effect of this is that if the brain is a purely syntactic engine, it is entirely plausible for it to create a chain of thoughts that do not exhibit content coherence. That is to say, the content of the thoughts would not make sense in regards to one another. (NZ MA)*

However, *in other words* and *that is to say* varied in frequency and the former is a far more frequent item in the BNC (32.43 compared to 7.4 per million words). Table 46 calculated the tokens of *In other words* and *That is to say* bundles regardless of different ending words (e.g. *In other words, the; In other words, they; In other words, it*) and punctuation marks (e.g. *That is to say; That is to say,*). As can be seen, their tokens in the Chinese texts overwhelmingly exceeded those of the New Zealand corpora. The Chinese students, as non-native writers, may feel it necessary to rephrase their words to secure understanding and agreement or may lack variety in their expressions.

**Table 46. Frequency of *In other words* and *That is to say* (pmw)**

	CH MA	CH PhD	NZ MA	NZ PhD
<i>In other words, the/they/it</i>	27	50	15	11
<i>That is to say(,)</i>	57	37	10	3

---

The bundle *This suggests that the* was shared between the Chinese PhD and New Zealand masters corpus, which expanded the previous statement with an explanation (44) or implication (45).

- (44) *The analysis of the repeated measures ANOVA revealed no significant main effects or interactions for the within-subjects effects (all  $p > .05$ ). **This suggests that the** subjects' lexical decision errors are not affected substantially by the variables listed in Table 6.7. (CH PhD)*
- (45) *It is interesting to note Fay's comment about some languages (vernaculars) not having tenses. **This suggests that the** teachers may not have a very good linguistic understanding of the vernaculars that they or their children speak. (NZ MA)*

#### 7.4.2 Code gloss bundles in the New Zealand students' writing

All the other code gloss bundles in the New Zealand theses were reformulation bundles, starting with *this*, *This is not to (say/suggest that)*, *This is not a* and *This is because the*, to immediately shut down the alternative interpretation of the anaphoric argument (46, 47) or to offer the reason for the previous statement (48).

- (46) *Marae-based te reo regeneration focuses primarily on internal change and development. Thus, analysis of the marae environment will produce more insight into influences on whānau / hapū language practice than will analysis of other environments. **This is not to** say the other environments are not important but simply that less time should be spent on gathering relevant information in relation to these environments. (NZ MA)*
- (47) *Many teachers in New Zealand have little background knowledge about the workings of language. **This is not a** criticism of teachers but an acknowledgement that teaching about language has not been consistently available to all. (NZ PhD)*
- (48) *Reading seems to be a different case from writing and L2 proficiency is a more critical factor within successful L2 reading. **This is because the** complexity of the language in a reading text cannot be manipulated by the reader but must be comprehended. (NZ PhD)*



### 7.4.3 Code gloss bundles in the Chinese students' writing

Code glosses in the Chinese students' writing were classified as exemplifiers (*For example, in the*) and re-formulators (*To be more specific, To put it another, It/This means that the*). *For example* was a common signal of exemplification in the Chinese students' writing, but far less frequent in the New Zealand corpora. The overuse was caused either by repeated encounters in the reading in the case of the informant Z (Table 47) or by over-emphasis from the teacher, as stated by the informant V (Table 48).

**Table 47. Z's interview on his use of *for example***

Text	Interpretation
<i>For example, in public Mongolian – Tibetan training schools in Tibetan areas, Mandarin Chinese, Mongolian, and Tibetan were required courses (Su, 1999).</i> (Z)	<i>For example</i> often appears in my reading. Therefore, I like to use it in my writing. (Z)

**Table 48. V's interview on her use of *for example***

Text	Interpretation
<i>For example, the institutions might ask a full-time tuition fee even if learners just participate part-time.</i> (V)	Our teacher has told us the only expression of exemplification we need to know is <i>for example</i> . (V)

All the re-formulators supplied the reader with additional information and none of them reduced the reader's interpretation to specific cases. Examples 49 to 52 are the extracts from the student theses.

- (49) *Actually, this puzzle mainly comes from the vague understanding of the distinction between initial topics, medial topics and final topics. To put it another way, most researches on topics are mainly based on the medial topics of a neutral order text sentence, but not initial or final topics in a starting or ending text sentence in a discourse.* (CH PhD)
- (50) *In this model, teaching activities such as practice in English listening, speaking, reading, writing and translation can be conducted via either*

computers or classroom teaching. **To be more specific**, the listening course is taught mainly in a computer-based environment, writing and translation courses are taught mainly in the classroom and speaking and reading courses are conducted in both computer-based environment and classroom context. (CH PhD)

(51) Thus, some researchers use the frequency of a lexical item as a signification of its conventionality. **It means that the** more frequently a lexical string occurs, the more likely it is to be habitual and conventional in native speakers' language. (CH MA)

(52) In this clause, the Subject is the nominal group a number of boys, whose Head is number, which is in singular form, but the predicate verb were is in plural form. **This means that the** predicate verb, were, does not accord with the Head, number, but with another element, boys, within the nominal group. (CH PhD)

*To be more specific* and *To put it another (way)* may be negative interlingual transfer from Chinese 具体来说 and 换句话说. As in the case of the informant V, the English translation of the online dictionary was believed to influence her language production (Table 49).

**Table 49. V's interview on her use of *to be specific***

Text	Interpretation
<i>The majority of participants learn English in older age because of interests. However, this interest is not for English language itself but for the usage of the language. <b>To be specific</b>, Xu would like to spread Chinese culture with his New Zealander neighbours, Lee wants to talk with young family members. (V)</i>	Here, I wanted to give a specific example, so I used <i>to be specific</i> . While writing, I like to use the Chinese-English online dictionary 金山词霸. The English equivalent expression of 具体来说 is <i>to be specific</i> . (V)

## 7.5 Condition bundles

Condition bundles present the pre-conditions for the preceding or succeeding statements, signalling the specific contexts, cases, purposes, perspectives, etc. As shown in Table 50, five condition bundles were shared across the corpora, *In the case of*, *In terms of the*, *In spite of the*, *With regard to the* and *On the basis of*.

Besides these five shared ones, three bundles appeared in the New Zealand writing: *For the purpose of*, *For the purposes of* and *In the context of*. A range of other condition bundles was identified in the Chinese student corpora and among them, four were shared between the masters and PhD students: *From the perspective of*, *As far as the*, *In this way, the* and *When it comes to*.

**Table 50. Condition bundles in the New Zealand and Chinese corpora**

	NZ	Chinese
condition bundles	<b>In the case of (14, 18);</b> <b>In terms of the (5, 6);</b> <b>In spite of the (7, 0)</b> <b>With regard to the (0, 6);</b> <b>On the basis of (0, 5)</b> For the purpose of (9, 0); For the purposes of (0, 8); In the context of (0, 7)	<b>In the case of (0, 20);</b> <b>In terms of the (0, 17);</b> <b>In spite of the (0, 7)</b> <b>With regard to the (6, 11);</b> <b>On the basis of (22, 18)</b> From the perspective of (8, 12); As far as the (7, 14); In this way, the (13, 16); When it comes to (6, 6); With the development of (21, 0); Based on the above (10, 0); With the help of (10, 0); In the light of (8, 0); As one of the (6, 0); In view of the (5, 0); In this sense, the (0, 17); With respect to the (0, 8); In this case, the (0, 8); In the field of (0, 6) In order to make (16, 0); In order to get (8, 0); In order to find (6, 0)

### 7.5.1 Shared condition bundles

Five bundles, *In the case of*, *In terms of the*, *In spite of the*, *With regard to the* and *On the basis of*, appeared in both New Zealand and Chinese student texts (53-57).

(53) ***In the case of** the family domain, this means that the bilingual children accommodate their language to the speakers of their family. (NZ MA)*

(54) ***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)*

- (55) *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)*
- (56) *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)*
- (57) *On the basis of the logic semantic relations, the connectors are classified into three types: elaboration, extension and enhancement. (NZ MA)*

Table 51 presents the percentage of their distribution with regard to their locations. The Chinese students showed their general preference to place these condition bundles at the beginning of their sentences: *in the case of* (44% compared to 42%, 38% compared to 35%), *in terms of the* (23% compared to 16%), *in spite of the* (61% compared to 30%), *with regard to the* (53% compared to 18%, 55% compared to 28%) and *on the basis of* (80% compared to 10%).

**Table 51. Positions of the four shared condition bundles**

Position	in the case of		in terms of the		in spite of the		with regard to the		on the basis of	
	initial	medial	initial	medial	initial	medial	initial	medial	initial	medial
CH MA	44%	56%	9%	91%	50%	50%	53%	47%	22%	78%
CH PhD	38%	62%	23%	77%	61%	39%	55%	45%	80%	20%
NZ MA	42%	58%	12%	88%	68%	32%	18%	82%	22%	78%
NZ PhD	35%	65%	16%	84%	30%	70%	28%	72%	10%	90%

The Chinese students preferred to locate condition bundles at the beginning of their sentences, which was supported by the informant S's explanation:

I would like to present pre-conditions first and then main ideas. I think it might be transferred from my Chinese mother tongue. We Chinese would like to articulate pre-conditions first and I feel uncomfortable to start a sentence with my main idea. Another reason is the length of pre-conditions. If it is short, I will put it at the beginning of sentences; otherwise, I will put it as the second part of my sentences. (S)

According to S, the pre-condition-first convention of Chinese sentence composition was transferred to his English writing.

### 7.5.2 Condition bundles in the New Zealand students' writing

With regard to the condition bundles in the New Zealand students' writing, an interesting finding was the use of singular or plural form of the word *purpose* in the masters or PhD bundle *For the purpose(s) of* (58, 59). A large majority of cases were followed by *the/this study*; therefore, one possible explanation was that PhD thesis as a more intensive and extensive work filled multiple knowledge gaps with more than one research purpose.

(58) ***For the purpose of*** the study which occurred in Spain, 35 teachers and 459 students answered a questionnaire about the influence of native and non-native teachers in the English language classroom. (NZ MA)

(59) ***For the purposes of*** this study, a contemporary view of identity is used which characterizes identity as flexible, variable, a social accomplishment, about self and other and constructed through discourse. (NZ PhD)

### 7.5.3 Condition bundles in the Chinese students' writing

Four condition bundles were consistently used by the Chinese masters and PhD students, which were *From the perspective of*, *As far as the*, *When it comes to* and *In this way, the*. Examples 60, 61 and 62 are the student texts of the first three bundles:

(60) ***From the perspective of*** cognition, understanding the culture of the target language is to understand the thinking model of the target language nations. (CH MA)

(61) ***As far as the*** scope is concerned, a good theory covers either a large number of situations for a narrow domain or a large number of domains for a narrow situation. (CH PhD)

(62) ***When it comes to*** the language learning strategy research in China, Wen Qiufang is one of the most important researchers who has done a lot of work in the field and construct a framework for English learning strategy. (CH MA)

The bundles *From the perspective of*, *As far as the* (..... is/are concerned) and *When it comes to* have the equivalences such as *With regard to*, *In terms of* and *In the case of* and the highly marked expressions may be consciously used by the students to overcome their language deficiency. Table 52 and 53 listed two sequences in the informants' (A and V) writing and their interpretations of these two expressions: *from the perspective of* and *when elder is mentioned*. The second one *when elder is mentioned* was chosen because it bore some similarity to the bundle *When it comes to*, which rarely occurred in the informants' writing. According to A and V, the direct Chinese translation led to their use of *from the perspective of* and *when elder is mentioned*. The informant A also highlighted his difficulty in language production, that is, *have to use the same expression repetitively and cannot use a range of expressions flexibly*, as a result of his limited L2 language repertoire.

**Table 52. A's interview on his use of *from the perspective of***

Text	Interpretation
<p><i>From this perspective, big pictures can be understood as "mental pictures". (A)</i>  <i>From the perspective of the learners, less cognitive investment means more chance to synthesis and reflection, and more chance to get the knowledge through, which is especially important when the knowledge carriers have to present their knowledge quickly and efficiently. (A)</i></p>	<p>English is my second language, so my language repertoire is limited. I have to use the same expression repetitively and cannot use a range of expressions flexibly. (A)            The first one equals to the Chinese phrase 从这个方面来讲, and the second one is the translation of 站在学习者的角度上来考虑. (A)</p>

**Table 53. V's interview on her use of *when elder is mentioned***

Text	Interpretation
<p><i>For example, when elder is mentioned, it is often associated with the description of weakness or sickness or reduced energy. (V)</i></p>	<p>This is the direct translation from Chinese to English, 当提到老年人的时候. (V)</p>

Unlike the above three bundles, the reason for the extensive use of *In this way, the* in the Chinese student texts appeared to be that the demonstrative determiner *this* could effectively link back to the anaphoric unit of text and the vague noun *way*

could refer to a number of specific concepts such as the use of *Collocate* in example 63.

(63) *To retrieve such recurring units from the corpus, the author has used Collocate (Barlow, 2004) to get n-grams combinations which have strong evaluative potential. In this way, the study demonstrates that some meaningful patterns can be easily extracted from the corpus to show the realization of evaluative meaning by compiling some patterns such as it would have been and it should be noted that. (CH PhD)*

As indicated by the informants A and W, the prevalence of this vague bundle reflected the Chinese students' familiarity with the high-frequency and transparent word *way*, their lack of more appropriate and specialised vocabulary and their avoidance of choosing unfamiliar words (Tables 54 & 55). According to Hasselgren (1994), A and W clutched *way* here as their *lexical teddy bear*, the word they felt safe with (p.237). In the same vein, the use of *in this way* as their *phrasal teddy bear* (Ellis, 2012) was also a result of familiarity, although this expression sometimes failed to convey the writer's original meaning, as in the case of the informant J (Table 56).

**Table 54. A's interview on his use of way**

Text	Interpretation
<p><i>The natural way to start doing a job by architects is sketching. (A)</i></p> <p><i>For data which always come from machine and are too many thus hard for people to digest, so graphs come into use to find the trends, the relationships, etc. in an simultaneous way which then facilitate the perception and synthesis process. (A)</i></p> <p><i>Graphics are used this way to provide short term information, thus help overcome the limitation of our working memory. (A)</i></p>	<p>I have not realised that I have used so many ways until you pointed out. This word has many different meanings, right? (A)</p> <p>I agree that way is an "empty" word, which does not carry much meaning. (A)</p> <p>I think it is my writing habit. I am familiar with this word so I use it frequently. It should be more accurate and concise if I thought it over. (A)</p>

**Table 55. W's interview on her use of *way***

Text	Interpretation
<i>The way of combining several research methods is employed in different study fields, applied in different subjects and also supported by a few academic researchers . . . .</i> (W)	I could not find an appropriate word to express the meaning 方式 ( <i>style</i> ), 方法 ( <i>method</i> ), or 途径 ( <i>approach</i> ), so I used <i>way</i> , the one comes into my mind. (W) I am not sure whether the other words are appropriate, so I chose <i>way</i> , the most common one, although it does not sound academic. (W)

**Table 56. J's interview on her use of *in this way***

Text	Interpretation
<i>Culture is the collective social experience perpetuated by a symbolic system and individual memories (Fei, 1947). In this way, cultural heritage in Zhu Jiayu is the core and competitive tourism resource in markets.</i> (J)	I want to express the meaning <i>from this perspective</i> here. At the time of writing, I could not recall the phrase <i>from this perspective</i> , so I habitually used <i>in this way</i> . (J)

Besides the above four typical Chinese bundles, the bundle *With the development of* was another bundle found to be pervasive in the Chinese masters corpus and the informants' texts. The informants Z, V and W all employed this expression in their writing and they attributed their familiarity to either previous learning or reading experiences (Tables 57–59).

**Table 57. Z's interview on his use of *with the development of***

Text	Interpretation
<i>With the development of education of Yunnan Province, bilingual education developed to some degree.</i> (Z)	I have learned this phrase from my course book 许国璋英语 and <i>with the development of</i> occurred frequently in the course book. I have also learned it from the English newspaper in China such as <i>China Daily</i> : the phrases like <i>with the reform and opening up</i> and <i>with the development of China's economy</i> are prevailing. (Z)



**Table 58. V's interview on her use of *with the development of***

Text	Interpretation
<i>With the development of economics and the change of the family structures, family support for elder people also experiences a decrease trend.</i> (V)	I must have learned this phrase before, many times. <i>With</i> here means <i>along with</i> . (V)

**Table 59. W's interview on her use of *with the development of***

Text	Interpretation
<i>With the development of new media technology, some scholars compare the both advantages and disadvantages of placing real estate advertisements on print media and new media.</i> (W)	<i>New media technology</i> is a kind of social phenomenon and I always use <i>with the development of</i> to modify social phenomena. (W) I learned this phrase while learning English. (W)

## 7.6 Introduction bundles

Introduction bundles are another new category developed from Hyland's (2005a) original model, created on the basis of the current thesis data, which solely include *there be* pattern. As presented in Table 60, introduction bundles were an important feature of the New Zealand students' writing; in contrast, no introduction bundles were found in the Chinese corpora.

**Table 60. Introduction bundles in the New Zealand and Chinese corpora**

	NZ	Chinese
introduction bundles	There are a number (8, 8); There was no significant (7, 7); There was a significant (0, 7); There were no significant (0, 6); There appears to be (8, 0)	

### 7.6.1 Introduction bundles in the New Zealand students' writing

Introduction bundles were an important component of the New Zealand thesis corpora and were used to introduce the subject matters of the upcoming texts (e.g.

*There are a number, There appears to be*) (64, 65) 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* (66, 67).

(64) ***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)

(65) ***There appears to be*** no research investigating the relationship between the level of teacher qualification and language outcomes for children. (NZ MA)

(66) ***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)

(67) ***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)

### 7.6.2 Introduction bundles in the Chinese students' writing

No introduction bundles were found in both Chinese student corpora. Interestingly, the four Chinese informants, J, V, W and A, held different attitudes towards *there be* pattern. J, V and W preferred to use *there be* sentences in their writing (examples in Tables 61-63). According to them, *there be* was a high-frequent sentence pattern in daily conversation, a key language point in English teaching and a better expression than *have*, which had the same Chinese equivalent (i.e. 有) with *there be*. In contrast, the informant A tried to exclude *there be* in his writing and the example in Table 64 was the only *there be* sentence in his 20-page text. He regarded it as *complex, useless and unclear*.

**Table 61. J's interview on her use of *there be***

Text	Interpretation
As <b>there are</b> great amounts of folk customs in rural areas in China, sometimes <b>there is</b> great difference within 1 kilometer distance between two villages. (J)	I often use <i>there is/are</i> to express the idea of existence because it often occurs in my daily conversation and it can convey my meaning here. I have never thought whether I can use a more exact word to replace it. (J)

**Table 62. V's interview on her use of *there be***

Text	Interpretation
<b>There is</b> growing recognition that learning is important to elder people. (V) To help them involve more in learning activities, <b>there are</b> some important conceptions we should know: (V) <b>There are</b> two main models in U3A system: the French model and the British model. (V) <b>There are</b> cultural variations in the definition of old. (V) During conducting the research, <b>there are</b> some limitations that have to take into consideration. (V)	I have used many <i>there be</i> sentences in my writing. I want to express the Chinese concept 有( <i>there be, have/has/had</i> ). The high frequency patterns in my writing are all key language points of English teaching in China. I myself have been an English teacher in a training centre before. <i>There be</i> pattern is a very important language point. For example, <i>there be</i> pattern is covered in several lessons of <i>New Concept English</i> , one of the popular course books in China and students are trained to distinguish <i>there be</i> from <i>have</i> . Chinese students are good at <i>there be</i> pattern because we have learned too much. (V)

**Table 63. W's interview on her use of *there be***

Text	Interpretation
<b>There are</b> research studies focusing on features of placing real estate advertisements on newspaper and use multi-regression analysis to examine the effects. (W)	I think <i>there be</i> is similar to but better than <i>have</i> , so I like to use <i>there be</i> . (W)

**Table 64. A's interview on his use of *there be***

Text	Interpretation
On this whiteboard, <b>there is</b> a sketch in the centre showing the movement and design for an equipment, while on the left side and above it <b>there are</b> other graphs showing the trend and relationship. (A)	I do not like using <i>there be</i> pattern in my writing. It is an inverted sentence pattern, complex, useless and unclear. It is more vivid to write as <i>Four perspectives on knowledge will be presented</i> . (A)

## 7.7 Summary

This chapter focused on the frequency, structure and function of interactive bundles in the Chinese and New Zealand theses. The differences between the Chinese and New Zealand writing and between the masters and PhD texts were analysed on the basis of Hyland's model of discourse organisers.

### 7.7.1 Differences between the Chinese and New Zealand writing

Differences were identified between the interactive bundles used in the Chinese students' writing and those of the New Zealand student corpora in each subcategory. Code gloss bundles and condition bundles were found to be more frequent in the Chinese students' writing, while transition bundles, endophoric bundles and introduction bundles were more common in the New Zealand students' writing. The Chinese students positioned transition bundles (e.g. *on the other hand*, *However*, *it is not*), code gloss bundles (e.g. *For example*, *in the*, *In other words*, *they*) and condition markers (e.g. *With regard to the*, *On the basis of*) as sentence initial to rely on these bundles to connect their ideas. In contrast, the New Zealand students largely depended on related themes to achieve cohesion and coherence, making effective use of shell nouns. The frame markers (e.g. *Last but not least*, *First of all*, *the*, *The first one is*), endophoric markers (e.g. *As is shown in*, *The following is/are a/an/the/some*) and condition markers (e.g. *In this way*, *the*) in the Chinese student texts appeared vague and lacking specific references, which led to loose connections and ambiguous statements. In addition, there were a range of bundles unique to the New Zealand and the Chinese students' writing, for example, *As discussed in Chapter*, *This is not to (say/suggest that)* and *There was no/a significant* in the New Zealand corpora and *In a word*, *the*, *To be more specific*, *To put it another (way)*, *From the perspective of*, *As far as the (... is/are concerned)* and *When it comes to* in the Chinese corpora. The typical New Zealand bundles can be included as the learning objectives of the Chinese students and the deviant Chinese bundles can be highlighted while teaching so that these L2 learners can consciously avoid these expressions when writing.

---

### **7.7.2 Differences between the masters and PhD writing**

There were a few differences between the masters and PhD level of students' writing. Compared to the masters students, the PhDs deployed more transition bundles and condition bundles and included the considerably larger proportion of section-level frame bundles and smaller proportion of chapter-level bundles in their thesis writing. These highly advanced students appeared to put more efforts into linking, modifying and elaborating their ideas at the local level.

In the next chapter, I will follow the same structure of this chapter to discuss the findings of interactional bundles. I will discuss the differences between Chinese and New Zealand thesis writing, or between masters and PhD levels of study in terms of sentence initial bundles. Interview data from Chinese postgraduates will also be presented to provide the interpretations for corpus data.

## Chapter 8 Interactional functions of the bundles

On the basis of Hyland's model (2005a, 2005c) of writer-reader interaction in academic writing, I will examine interactional bundles in this chapter, which include attitude bundles, hedge bundles, booster bundles, self-mention bundles, directive bundles and shared knowledge bundles. Like the structure in the previous chapter, I will first present sentence initial bundles within each category, compare the use of bundles between Chinese and New Zealand postgraduates, or between masters and PhD levels of study, and then offer possible interpretations of the identified typical Chinese bundles.

Table 65 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). Two New Zealand bundles, *It is important to* in the masters corpus and *It is possible that* in the PhD corpus, occurred with particularly high frequencies (26, 20), ranking among the top 5 bundles in each corpus. In contrast, no interactional bundles in both Chinese corpora appeared as top-5 or even top-10 bundles.

**Table 65. 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-Value < 0.05). Table 66 presents the percentage of bundles in each interactional category.

**Table 66. Distribution of interactional bundles in each corpus (tokens)**

		CH MA	CH PhD	NZ MA	NZ PhD
Stance	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>
Engagement	directive bundles	13%	36%	21%	16%
	shared-knowledge bundles	6%	0%	0%	0%
Total		100%	100%	100%	100%

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

As can be seen from Table 66, 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. The bundle distributions in the two masters corpora were also different from those of the PhD texts with more attitude bundles and fewer self-mention bundles. The following discussion will provide a close examination and comparison of these bundles within the contexts, together with possible interpretations from the Chinese student informants.

### 8.1 Attitude bundles

Attitude bundles express the student's subjective evaluations of his or her arguments or personal feelings towards his or her research-related experiences. All the bundles in this category were part of anticipatory-*it* clauses, used to depersonalise the writers' opinions. Adjectives, such as *important*, *necessary*, *interesting* and *difficult*, were used in the structure *It + is + predictive adjective + to/that*. Table 67 grouped these attitude bundles into two sub-categories: subjective evaluation and personal feeling. Subjective evaluation was comprised of *important* bundles (*It is important to/that*, *However, it is important*, *It was important to*), *necessary* bundles (*It is necessary to*, *Therefore, it is necessary*, *So it is necessary*)

and one *evident* bundle (*It is evident that*). Personal feeling consisted of *interesting* bundles (*It is interesting to/that*) and a *difficult* bundle (*It is difficult to*). The bundle *It is important to* was the only shared bundle between all four corpora. Apart from it, the New Zealand and Chinese postgraduates showed their different preferences for *important* and *necessary* bundles and no personal feeling bundles appeared in the Chinese corpora.

**Table 67. Attitude bundles in the New Zealand and Chinese corpora**

	NZ	Chinese
subjective evaluation bundles	<b>It is important to (26, 17)</b> It is important that (7, 0); However, it is important (7, 0); It was important to (5, 0)	<b>It is important to (7, 7)</b> It is necessary to (7, 7); Therefore, it is necessary (7, 0); So it is necessary (5, 0); It is evident that (5, 0)
personal feeling bundles	It is interesting to (13, 8); It is interesting that (6, 0); It is difficult to (6, 6)	

It should be noted that Hyland (2005a) classifies the single adjective words *important* and *interesting* as attitude markers, whereas Hyland (2002b), on the basis of his extended context analysis, categorises the sequences *It is important/necessary to* into his compiled directive list. This inconsistency may contribute to the variations in the unit of analysis: the function of single words or even four-word sequences can be easily identified as attitude markers, while the function of extended texts beyond four-word sequences is highly likely to be different, such as directive in this case. This also supports my selection of four-word bundles instead of longer ones as the target bundles in this study — longer sequences tend to carry more than one function.

One of the aims of this research is to provide a list of useful sentence initial bundles to L2 writers in terms of their metadiscourse functions. It may be easier for the learners if the functional categories directly correspond to the internal functions of these bundles. Therefore, I chose to categorise these bundles according to their internal functions rather than the functions of their extended texts. The following sections will discuss these functions in more detail. I will first present the shared



attitude bundles between New Zealand and Chinese students' writing and then examine their unique bundles respectively.

### 8.1.1 Shared attitude bundle

The only shared attitude bundle, *It is important to*, occurred in two distinctive structures: *It is important to + verb + that-clause* and *It is important to + verb + object*. The first structure highlighted the subsequent activities, mostly cognitive activities, introduced by the infinitive verbs. These verbs were mental verbs (e.g. *note*, *remember* and *recognise*), used to describe the process of receiving the stated information, as in:

- (1) *The first active step of this stage was to compile all the relevant information into a MS Excel spreadsheet (see Appendix F). **It is important to note** at this stage **that** five Likert scale response categories were reverse scored from the questionnaire. (NZ MA)*
- (2) *The Cambridge Certificate of Proficiency in English takes approximately 40 minutes and contains 28 items. **It is important to remember that** these studies involve listening to multiple short passages and the total test time includes time given to answer questions. (NZ PhD)*
- (3) ***It is important to recognise that** immigrant parents face significant challenges in educating their children. As Eberly, Joshi and Konzal point out, "Increasing diversity in the student population intensifies the need for and the difficulties of establishing culturally sensitive and meaningful communication between teachers and parents" (2007, p. 7). (NZ MA)*

The second structure imposed an obligation to take action suggested by the embedded activity verbs (e.g. *create*, *distinguish* and *reunite*), as a result of the preceding or succeeding statements, as in:

- (4) *The new syllabus and the new course advocate that learning a foreign language is a process of moving from the unfamiliar to the familiar, from imperfection to perfection. Therefore, it is wise for the teacher to tolerate minor mistakes which will not effect the verbal communication. **It is important to create** a more*

- 
- relaxing atmosphere for the students in which they can dispel fears and nervousness and use the language with more confidence and courage. (CH MA)*
- (5) *Tasks are defined in terms of what a language learner would do inside the classroom rather than in the outside world. **It is important to distinguish** what might be defined as pedagogic tasks and real-world tasks. (CH MA)*
- (6) ***It is important to reunite** synchronic and diachronic analyses, since the two approaches cannot be rigidly separated from each other (CH PhD)*

Table 68 shows the distribution of these two distinctive structures in each corpus. The Chinese masters deployed the lowest proportion of *important* bundles in the structure *It is important to + verb + that-clause* (36%), while the proportion in the New Zealand PhD corpus was the highest (87%). The proportions of the New Zealand bundles in the structure *It is important to + verb + that-clause* were generally higher than those of the Chinese bundles.

**Table 68. Distribution of *It is important to***

	<b>It is important to + verb + that-clause</b>	<b>It is important to + verb</b>
CH MA	36%	64%
CH PhD	79%	21%
NZ MA	77%	23%
NZ PhD	87%	13%

Anticipatory-*it* pattern *It is important to* removes the human subject who is expected to take the suggested action. The real human subjects of the structure *It is important to + verb + that-clause* are readers, such as supervisors, examiners and imaginary readers in the case of thesis writing. On the contrary, the subjects of the structure *It is important to + verb + object* are student writers themselves, who are expected to prove their competence in undertaking research independently. The different distributions of the Chinese and New Zealand *important* bundles may reflect the divergent writer-reader relationships in the two cultural and academic contexts. In comparison to the New Zealand students, the Chinese students, particularly the students at the masters level, appeared to be less inclined to attempt to influence the evaluation of their readers. Instead, they felt more comfortable to highlight their judgements on research procedures.

### 8.1.2 Attitude bundles in the New Zealand students' writing

*Important* bundles were pervasive in the New Zealand students' writing. Along with *It is important to*, another three *important* bundles appeared in the New Zealand masters corpus: *It is important that* (7), *However, it is important (to)* (8) and *It was important to* (9). These *important* bundles served the same function as *It is important to* in highlighting the succeeding propositions or actions.

- (7) ***It is important that*** schools and teachers are aware of the most effective programmes so they can then make informed decisions about how spelling should be taught in their schools to provide the best outcome for children. (NZ MA)
- (8) ***However, it is important*** to be aware of the fact that the learners continued with their classroom lessons during the time between the immediate and delayed post-tests as well as having many opportunities to hear the target language outside of the classroom. (NZ MA)
- (9) ***It was important to*** ensure the learners felt safe which meant there was a willingness to be open and share their experiences. (NZ MA)

The personal feeling bundles, *It is interesting to/that* and *It is difficult to*, were another feature of the New Zealand students' writing. *It is interesting to note that* was the only extension of *It is interesting to* bundle, serving to affect the attitude of their readers towards the findings in the succeeding *that*-clauses (10).

- (10) *Studies of passives also vary in their approach. However, they are usually interpretive, involving critical argument and induction, that is, the process of observing facts to generate theories. **It is interesting to note that**, whichever research methods have been used, there is considerable debate on the findings.* (NZ MA)

The bundle *It is interesting that* only appeared in the masters writing, expressing the writer's sheer excitement about his or her findings (11).

---

(11) *It is interesting that all three sets of students wrote more than twice as many adventurous words in English than they did in Maori at pre programme assessments. (NZ MA)*

The *difficult* bundles, on the other hand, were usually followed by a range of action verbs (e.g. *assess, distinguish, explain* and *relate*) to indicate different obstacles encountered during selecting, comparing, interpreting and evaluating data in the process of research (12).

(12) *However, an important finding in this thesis is that, almost without exception, the inhabitants of these small New Zealand towns are geographically mobile. It is difficult to distinguish individuals who are more likely than others to have brought these innovations into their community. Almost all speakers have the opportunity to do so. (NZ PhD)*

### 8.1.3 Attitude bundles in the Chinese students' writing

Unlike *important* bundles, *necessary* bundles (i.e. *It is necessary to, Therefore, it is necessary, So it is necessary*) were popular among the Chinese students, especially at the masters level. These bundles employed the same two structures as the *important* bundles, using mental verbs (e.g. *note*) in the structure *it is necessary to + verb + that-clause* (13) or action verbs (e.g. *offer, reduce* and *provide*) in the structure *it is necessary to + verb + object* (14). The functions of these two patterns were also same with the *important* bundles: the first pattern served to affect the judgement of the implied readers and the second one justified the research activities. Compared with *important* bundles, the use of *necessary* bundles had a stronger bias towards research activities. Only 7 out of 87 tokens of the *necessary* bundles (1.6% in the masters corpus and 6% in the PhD corpus) performed the function of reader involvement, whereas a large majority described research activities.

(13) *From our data and statistics it seems the more politically fanning or fermenting, the more focus is put on the means of Judgment or the means of Affect, as Bush, Blair and bin Laden have done. The more negotiable, the more focus is put on the means of Appreciation as the two peace-lovers have done.*

*It is necessary to note that since our data is too limited, this summary is only preliminary, examinations and discussions on more different data are needed for a more precise conclusion. (CH PhD)*

(14) *For students learning to write, the ability to write readable paper requires a similarly broadened view and an ability to shift from the perspective of the writer to that of the reader. Therefore, it is necessary to offer students some authentic reading materials before the writing and let them take reading as a model for writing. (CH MA)*

Figure 4 displays the most frequent *important/necessary + to + verb* collocations in Wikipedia as a result of a search using the tool FLAX, discussed previously in Section 5.1.2 Bundle identification. The numbers stand for the frequencies of collocations. The adjective *important* mostly collocates with mental verbs as *note, understand, remember, know, consider* and *realize*; in contrast, *necessary* links with action verbs, such as *make, ensure, keep, prevent, use, maintain, protect* and *build*. The different collocates partially explained the overuse of *necessary* bundles by the Chinese students. Unlike their New Zealand counterparts, the Chinese students appeared to be keen to justify their own research procedures rather than cognitively engage their readers.

important + to + verb	 important to note	2017	 important to know	171
	 important to understand	370	 important to consider	167
	 important to remember	288	 important to add	133
	 important to be	236	 important to keep	131
	 important to have	181	 important to realize	123
				>>> more
necessary + to + verb	 necessary to make	447	 necessary to use	312
	 necessary to have	352	 necessary to maintain	306
	 necessary to ensure	337	 necessary to be	292
	 necessary to keep	333	 necessary to protect	289
	 necessary to prevent	313	 necessary to build	241
				>>> more

**Figure 4. Collocations of *important* and *necessary* in Wikipedia as displayed in FLAX**

The interview with the informant V (Table 69) echoed this finding of the Wikipedia-based search. V believed her use of the word *necessary* indicated the compulsory nature of the described action *discuss* in her text.

Her words revealed another reason of the Chinese students' overuse of *necessary* — to avoid word repetition. As a L2 learner, she has been informed by the teachers or English tests with the clear expectation on the use of words in writing — to show her richness of vocabulary. However, the distinction between words, particularly between synonyms (e.g. *important* compared to *necessary*), have been largely neglected or at least less emphasised. This may lead to inappropriate word replacements. For example, instead of *necessary*, *important* should be a more appropriate word to use in the above student sentence (13).

**Table 69. V's interview on her use of *necessary***

Text	Interpretation
<p><i>Chinese elder immigrants have a totally different culture background with native speakers in New Zealand, therefore as a factor that has an inseparable relationship with language, it is necessary to discuss how culture affect language learning for immigrants.</i> (V)</p>	<p>The word <i>necessary</i> does not convey the same meaning with <i>important</i>. I want to express <i>something that I need to do or have to do</i> here. (V)</p> <p>I have already used many <i>important</i> in this thesis, so I chose <i>necessary</i> this time. My previous teachers have suggested avoiding using one word repetitively. When they find word repetition in my writing, they will replace it with another word. I have also attended many English tests, like TOEFL and IELTS. How to gain higher marks in English writing? Try not to repeat. Although you want to express the same meaning, you need to choose another word. (V)</p>

Lack of personal feeling bundles is another feature of Chinese student writing. No *interesting* or *difficult* bundles were found in both Chinese corpora. The informants in this study also rarely used the *interesting* or *difficult* bundles in their texts. The informant V's words (Table 70) provided the reason: as a Chinese student researcher, she was fairly conservative in describing feelings, as she considered it inappropriate to convey attitudes in her academic writing. A close examination of V's text disclosed another reason. She employed neither of the two bundles in the New Zealand writing (i.e. *It is interesting to* and *It is interesting that*) but the

sequence *the interesting is*. That is to say, Chinese students, like V, might not have the ready-made bundle repertoire to draw from.

**Table 70. V's interview on her use of *interesting***

Text	Interpretation
<i>However, the interesting is, although some other participants agreed that memory might decline with aging, they believed it was not the main reason which could stop English learning. (V)</i>	I rarely use <i>interesting</i> in my writing. This word expresses my own feeling. Academic writing should be neutral however. I use it only to describe the extremely interesting stuff. (V)

The interview with J supports this statement. The informant J recalled the guidance she received while learning English writing (Table 71). The recommendation in a popular writing book was not on the use of anticipatory-*it* patterns but on using gerunds as subjects. As a result, her writing featured by gerund-subjects and she lacked the knowledge of anticipatory-*it* bundles. In the following example, the sentence can be better expressed with a shorter subject *it* and the theme *difficult* can be highlighted, using the anticipatory-*it* sentence: *it may be more difficult for the DMO in rural area to raise the awareness of local cultures*.

**Table 71. J's interview on her use of *it is difficult to***

Text	Interpretation
<i>Establishing the awareness of local cultures may be more difficult to the DMO in rural areas. (J)</i>	I remember while learning to write in English I read Xiaoyi Shen's book on IELTS writing, a very popular book. She warned us not to use expressions like <i>it is difficult to</i> . She strongly recommended gerunds as subjects and suggested it would increase our marks on writing. (J)

## 8.2 Hedge bundles

Hedge bundles address the writer's uncertainty and express his or her cautiousness about making statements or claims. They "imply that a statement is based on plausible reasoning rather than certain knowledge, indicating the degree of confidence it is prudent to attribute to it. . . . Equally importantly, hedges also allow writers to open a discursive space where readers can dispute their interpretations" (Hyland, 2005c, p. 179).

The Chinese and New Zealand postgraduates showed different preferences in the use of hedge bundles: the New Zealand students used anticipatory-*it* clause embedding possibility adjectives (*possible* and *not clear*) or certainty copula (*may be*, *appear* and *would appear*), whereas the Chinese students mainly relied on reporting verbs (*suggested*, *argued*, *hoped* and *indicate*) to express tentativeness (Table 72). Generally, the New Zealand students employed more hedge bundles than their Chinese counterparts. This supports Y. Yang's (2013) finding on the frequency of hedges between English articles and Chinese-authored English articles. No hedge bundles were shared between the Chinese and New Zealand corpora.

**Table 72. Hedge bundles in the New Zealand and Chinese corpora**

	NZ	Chinese
possibility adjective bundles	It is possible that (14, 20); It is possible to (5, 0); It is also possible (0, 6); It is not clear (0, 7)	
certainty copula bundles	It may be that (8, 6); There appears to be (8, 0); It would appear that (5, 0)	It seems that the (0, 7)
reporting verb bundles		It is suggested that (9, 0); It is argued that (0, 6); It is hoped that (5, 8); The results indicate that (6, 0)
<i>of</i> -phrase bundle		One of the most (8, 0)

### 8.2.1 Hedge bundles in the New Zealand students' writing

A range of hedge bundles was used extensively in the New Zealand masters and PhD writing. Among them, *It is possible that* was shared between two New Zealand corpora and ranked as one of the top 5 bundles in both corpora. As shown in Table 73, *It is possible that* allowed writers to predict research findings or contradictory findings (15, 16), to infer the underlying reasons (17), to suggest alternative approaches (18), to negotiate the conclusions (19) and to propose possible solutions (20). The other hedge bundles that appeared in the New Zealand students' writing were less frequent, but performed similar functions to *It is possible that*.



**Table 73. Functions of *It is possible that***

Function	Example
predict research findings or contradictory findings	<p>(15) <i>I wanted to find out from the students themselves whether they think there is an issue here. And if there is, how do they describe its nature? Since schools are required to implement the revised curriculum from 2010, it seems timely to explore these changes that are already happening. <b>It is possible that</b> the students themselves may have some ideas which will assist teachers to manage the transition to a new learning area. (NZ MA)</i></p> <p>(16) <i>There are, however, five limitations that apply to the thesis as a whole. First, the participants in this research viewed only episodes from a single television program. <b>It is possible that</b> the results of the studies may have been different if another television program had been utilized. (NZ PhD)</i></p>
infer the underlying reasons	<p>(17) <i>Although Andy described reading on the Internet as part of his process for writing the second draft of his essay, none of the information or text in Andy's second draft could be identified with a website. Therefore, it is unclear what role this Internet reading played in this stage of Andy's writing. <b>It is possible that</b> the textual analysis I conducted, as described in the methodology chapter, simply did not detect the uses of Andy's Internet reading within the second draft. However, <b>it is also possible that</b> Andy's Internet reading provided background information to support Andy's comprehension and interpretation of other English texts. Or, <b>it is possible that</b> Andy could not comprehend enough of the Internet texts to integrate them into his developing understanding of the essay task and the argument he was building. (NZ PhD)</i></p>

---

suggest alternative methods	(18) <i>Polysemy is too complex a topic to be dealt with here at length, although the issue of how to preserve constructive ambiguity is of course central in literary translation. <b>It is possible that</b> by addressing its negative counterpart here, some light may be shed on how to deal with polysemy as well, but such discussion could form the basis of an entire thesis in its own right. (NZ PhD)</i>
negotiate the conclusions	(19) <i>Diachronic studies of the get-passive may also clarify its purpose. The array of views outlined in Chapter 2 on its role may prove to be largely historic. Frequency counts of modern written usage of the get-passive, combined with an analysis of that use, may provide a more consistent picture than the one that exists in the literature. <b>It is possible that</b> the construction is undergoing a change of use. (NZ MA)</i>
propose possible solutions	(20) <i>Secondly, language learning strategies are, themselves, able to be learnt, which allows for the possibility that individual students may be able to improve their language learning effectiveness by choosing appropriate strategies. <b>It is possible that</b> teachers might be able to facilitate the development of language learning strategies by raising awareness of strategy possibilities, by making strategy instruction both implicit and explicit and by providing encouragement and practice opportunities. (NZ PhD)</i>

---

### 8.2.2 Hedge bundles in the Chinese students' writing

Three types of hedge bundles were found in the Chinese students' writing: bundles consisting of certainty copula (e.g. *seem*), reporting verbs (e.g. *suggest*, *argue*, *indicate* and *hope*) or *of*-phrase (e.g. *one of the most*). The bundle of the Chinese PhD writing *It seems that the* (21) performed a similar function to the New Zealand masters bundle *It would appear that* (22), in softening the writer's assertions.

(21) ***It seems that the** limited processing capacities of our learners force them to select some aspects of the story rather than others, starting with foreground*

---

*episodes, followed by setting the scene and finally background episodes. (CH PhD)*

(22) ***It would appear that*** *bilinguals often outperform their monolingual peers in special awareness tasks. (NZ MA)*

Why did the Chinese and New Zealand students show different preferences for word selection? The distributions of *seem* and *appear* in the British National Corpus (BNC) of the BNCweb (<http://bncweb.lancs.ac.uk/>) may explain the Chinese students' habitual selection of *seem* and the New Zealand students' intuitive preference of *appear*. Generally, *seem* is a more popular word, compared with *appear* (168 compared to 109 per million words). This might result in the earlier introduction in pedagogy and more encounters during the learning process, leading to Chinese learners being more familiar with the word *seem* than *appear*. However, *seem* occurs slightly more often in written than in spoken texts (170 compared to 145 per million words), whereas *appear* is far more prevalent in written texts (118 compared to 31 per million words). This may be why the New Zealand students preferred to *appear* while composing theses. The Chinese students, on the other hand, may not be conscious of the register constraint.

The Chinese students relied heavily on reporting verbs to withdraw full responsibility for the credibility of the presented research results (e.g. *The results indicate that*) (23), propositions (e.g. *It is argued that*) (24) or pedagogical implications (e.g. *It is suggested that, It is hoped that*) (25, 26); and at the same time to avoid using self-mentions.

(23) ***The results indicate that*** *as far as the overall frequency of make is concerned, there is no significant difference between the second year Chinese learners and the native speakers, but the fourth year Chinese learners use less than native speakers. (CH MA)*

(24) ***It is argued that*** *the use of cohesive devices can distinguish a text from a series of disconnected sentences and such cohesive devices can function to establish relationships across sentence boundaries. (CH PhD)*

(25) ***It is suggested that*** *collocation be included into English exams and syllabus, thus learner can combine grammatical rules and lexical knowledge in a more*

*scientific way and the improvement of their productive skills can be facilitated.*

(CH MA)

- (26) ***It is hoped that*** *this exploration of the relationships among aspects of word knowledge and developmental features of these essential types of word knowledge when the learners progress from the lower to the advanced learning stages in tertiary institutions would contribute to our understanding of L2 vocabulary acquisition and development and thereafter the construction of an explanatory model of L2 vocabulary acquisition in classroom settings in the future.* (CH PhD)

Both the informants Z and V explained their selections of reporting verbs to manipulate tentativeness in writing (Tables 74 & 75). As an experienced language teacher and a more advanced writer, Z articulated two reasons for his careful selection of *hope* and *suggest*: the linguistic transfer from Chinese and the modesty culture of China. Unlike Z, V's use of reporting verbs was more like a random choice among synonyms. She may not have a clear understanding of the degrees of certainty indicated by reporting verbs.

**Table 74. Z's interview on his use of *hope* and *suggest***

Text	Interpretation
<p><b><i>I also hope that</i></b> <i>my research will have some implications for multilingual education theory and research as discussed in the next section.</i> (Z)</p> <p><b><i>It might be sensible to suggest that</i></b> <i>culture has significant influences on tri-multilingual education.</i> (Z)</p>	<p>I think the use of <i>hope</i> here is a direct translation from the corresponding Chinese expression. We often say 我希望 (<i>I hope</i>) rather than 我确信 (<i>I am sure</i>) or 我坚信 (<i>I believe</i>) in Chinese. (Z)</p> <p>The items <i>hope</i>, <i>believe</i>, <i>I am sure</i>, <i>I consider</i> and <i>I think</i> vary in degree of certainty. I think the selection of <i>hope</i> and <i>suggest</i> is also the result of Chinese modesty. We Chinese like to express our ideas in a relatively modest way, opening some space for disputes. So we seldom say 我坚信 (<i>I strongly believe</i>), 我相信 (<i>I believe</i>) or 我认为 (<i>I consider</i>), but prefer to use more modest expressions such as 暗示有 (<i>It suggests</i>) or 我希望 (<i>I hope</i>). (Z)</p>

**Table 75. V's interview on her use of *indicate* and *show***

Text	Interpretation
<i>The results indicate although there are some studies have already overturned the misconceptions towards elder people, some of the stereotypes still exist deeply in people's minds and the influential cannot be underestimated.</i> (V)	I did not choose the verb carefully. To me, <i>indicate</i> , <i>show</i> and <i>express</i> are all synonyms. I think they are almost the same. (V)
<i>The results show the participant who started early in learning English has the same identity for themselves when choosing the level of their English ability.</i> (V)	

Another interesting hedge bundle in the Chinese writing was *One of the most*, only appearing in the masters theses. As presented in example 27 below, this masters student relied on *one of* to degrade the superlative *the most time-consuming activities*, cautiously modifying the complement of the linking verb, *the preparation of appropriate teaching materials*.

(27) *One of the most time-consuming activities for many ESP programs is the preparation of appropriate teaching materials.* (CH MA)

As indicated in the informants V and W's interviews (Tables 76 & 77), they aimed to use *one of* to soften the superlative degree and to avoid having to find the literature. With this modification, the superlative form had been turned into universal truth. V also stated the source of this expression and she learned this strategy from her previous writing courses. This reflects the Chinese students' writing-from-personal-knowledge L2 learning experience: since most English writing tests are based on personal knowledge, teachers neglect the role of reading materials in writing. Writing and reading have been mostly taught as two isolated skills.

**Table 76. V's interview on her use of *one of the most***

Text	Interpretation
<i>In fact, Hall (1959) defined culture as “a complex message system by which the members of the community exchange messages by which co-operation, cohesion, and survival of the community are effected” (Nababan, 1974, p. 19). No matter this information is the attitude, the behaviour or some other emotion that people have, <b>one of the most effective</b> way to transfer it (culture) is to use the language. (V)</i>	The use of <i>one of</i> hedges my subsequent statement. I am not sure whether language is the top-one effective way to transmit culture. I cannot say so because it must be wrong. If I modify this statement with <i>one of</i> , then it becomes acceptable: one of the most effective way to transmit culture is the use of language. Without <i>one of</i> , it is nonsense or I have to find the literature. With <i>one of</i> , I do not need to search for the literature anymore. This is learned from my previous writing courses. (V)

**Table 77. W's interview on her use of *one of the most***

Text	Interpretation
<i>An influx of capital tap into market, cultivated by value of “life and work in peace and contentment”, Chinese regard “housing” as <b>one of the most</b> indispensable things for a life-long time. (W)</i>	I used the superlative degree <i>the most</i> because housing is particularly important in China. I used <i>one of</i> because there must be something else important. It is absolutely right to put both of them together <i>one of the most indispensable</i> . (W) I have no time to search for the reference. (W)

### 8.3 Booster bundles

Booster bundles express the writer's certainty towards his or her proposition, and function to close down alternative voices. By means of manipulating the weight of hedges and boosters, the writer balances “objective information, subjective evaluation and interpersonal evaluation” (Hyland, 2005c, p. 180).

Booster bundles, contrary to hedge bundles, occurred much more frequently in the Chinese students' writing. The Chinese and New Zealand students showed their opposite positions while making claims and involving readers, with the completely different weight put on hedge and booster bundles. This corroborates previous research on non-native academic writing and non-native writers were found to hedge less than their native counterparts (Gilquin & Paquot, 2008; Hyland & Milton, 1997; Yang, 2013). According to Williams (2003), confident writers use more

hedges than boosters because they are more cautious with their arguments; while inexperienced writers misinterpret the aggressive style as persuasive. Gilquin and Paquot (2008) regard overuse of boosters as the influence of the informal style of speech, which tends to be less tentative than academic prose. Hyland and Milton (1997) attribute Chinese students' overuse to their misinterpretation of the conventions of explicitness and directness in English. Yang (2013) suggests another two reasons typical to Chinese writers: unfamiliarity with the hedge devices and different beliefs in Chinese academic discourse: "the researchers should be authoritative and their results should be as rigorous as possible" (p. 30).

As summarised in Table 78, *It is clear that* was the only booster bundle shared between the New Zealand and Chinese students. Another booster bundle in the New Zealand writing was *The fact that the*. Booster bundles in the Chinese writing were composed of certainty adjectives (*clear, obvious* and *true*), reporting verbs (*show, found* and *believe*) and other booster items (*fact, should* and *no doubt*).

**Table 78. Booster bundles in the New Zealand and Chinese corpora**

	NZ	Chinese
certainty adjective bundles	<b>It is clear that (0, 6)</b>	<b>It is clear that (7, 10)</b> It is obvious that (12, 7); It is true that (0, 8)
reporting verb bundles		The results show that (9, 0); The results showed that (7, 6); The following table shows (8, 0); It is found that (7, 0); It is believed that (7, 0)
other bundles	The fact that the (8, 7)	As a matter of (fact) (11, 7); There is no doubt (6, 0) It should be pointed (out) (0, 6);

### 8.3.1 Shared booster bundle

*It is clear that* was the only booster bundle shared between the Chinese masters, Chinese PhD and New Zealand PhD students. This bundle either presented objective research results (28, 29) in which *clear* means "apparent, easy to notice

or understand” or conveyed the writer’s certainty towards his or her argument (30, 31) with the meaning of *clear* as “certain, impossible to doubt”.

- (28) ***It is clear that*** most of investigated students think that it is necessary to learn western culture in English learning. (CH MA)
- (29) ***It is clear that*** one (55.2%) and we (29.1%) are the most commonly used Projector-realizing resources. (CH PhD)
- (30) ***It is clear that*** Japanese was seen as a more acceptable FL to study than French, even if students ultimately wished to learn French. (NZ PhD)
- (31) ***It is clear that*** these contexts, together with the resulting positive relationships they engender, constitute an important element of effective oral language learning for students with limited language achievement. (NZ PhD)

### 8.3.2 Booster bundle in the New Zealand students’ writing

*The fact that the* was another booster bundle used in the New Zealand student texts. The shell noun *fact* was equal to the succeeding appositive *that*-clause. As argued by Jiang and Hyland (2015), this *fact*-clause either highlights the reality or expresses the epistemic judgement of certainty. As illustrated in examples 32 and 33, the use of *fact* allowed the writer to present his or her research result *language impaired participants and younger participants made the smallest gains* or epistemic evaluation *Goal is necessary to the Process* as an unarguable objective evidence.

- (32) ***The fact that the*** language impaired participants and younger participants made the smallest gains indicates that vocabulary proficiency is related to vocabulary gain. (NZ PhD)
- (33) ***The fact that the*** Goal is necessary to the Process can be seen in two ways. (NZ PhD)

### 8.3.3 Booster bundles in the Chinese students’ writing

Booster bundles were used extensively in the Chinese students’ writing. These Chinese students strengthened their statements through certainty adjectives, reporting verbs and other linguistic strategies, which include the idiomatic



expression *As a matter of (fact)*, modal of obligation *should* and negative sentence *There is no doubt*.

The Chinese students, especially at the masters level, showed their preference to the *obvious* bundle *It is obvious that* to present mostly objective facts (34, 35).

(34) *It is obvious that the traditional teaching methods in China mainly dominate English classroom teaching although these methods made great contribution to the English teaching patterns nowadays. (CH MA)*

(35) *It is obvious that with the statistical algorithm, all those sequences that are recurrent in a corpus, though unacceptable to native speaker norms, can also be identified. (CH PhD)*

As noted by the informants J, Z and V (Tables 79-81), the primary reason for Chinese students to choose *obvious* instead of *clear* was that these two words conveyed different meanings. The meaning of *obvious* corresponds to the Chinese conventional expression 明显 (easy to understand), which describes the simple process of understanding. The Chinese equivalence of *clear* is 清楚 (certain), which refers to a high degree of certainty. The equivalent Chinese meanings largely affected the Chinese students' word selection because they interpreted the meanings of these two English words through their Chinese equivalences.

The informant J considered the reader's background knowledge, so she used *obvious* to describe the information that is easy to understand. The informant Z added that *clear* and *obvious* also varied in degree of certainty and many Chinese students dared not to use *clear* to argue with absolute certainty and they felt safer to use the less certain word *obvious*. As can be seen from the above *clear* (28, 29) and *obvious* examples (34, 35), the Chinese students were more likely to deploy *It is clear/obvious that* bundles to present objective facts instead of subjective arguments; while the New Zealand students chose *It is clear that* to strengthen their positions (30, 31). The informant V articulated her preference for the academic word *obvious* rather than the high frequency *clear* to show her sophisticated knowledge of academic vocabulary.

**Table 79. J's interview on her use of *clear* and *obvious***

Text	Interpretation
<i>Obviously, even though the superior government offers great help of funding to the DMO, it still lacks of tourism talents who are the key factor of the long-term development of rural tourism in Zhu Jiayu.</i> (J)	<i>Obvious</i> means 明显, and <i>clear</i> means 清楚. What I want to highlight here is the fact is very obvious and they lack tourism talents. <i>Obvious</i> describes something on the surface or something easy to understand. It is not that easy to achieve the degree of <i>clear</i> . I think the reader can easily understand this point here, which does not require any background knowledge. (J)

**Table 80. Z's interview on his use of *clear* and *obvious***

Text	Interpretation
<i>Clearly, trilingual education is not simply the mere extension of bilingualism but a complicated process involving sociolinguistic, cognitive and psycholinguistic aspects and a product of economic globalization, cultural pluralism and mobility and migration of human beings.</i> (Z)	I think it has been widely proved that trilingual education is a complicated process. This must be correct, just like the truth, so I used <i>clearly</i> . (Z) I also like to use <i>it is obvious that</i> , but for <i>clear</i> , I like to use <i>clearly</i> . (Z) Chinese do not like to use <i>it is clear that</i> because we prefer 明显 ( <i>apparent, easy to notice or understand</i> ) to 清楚 ( <i>certain, impossible to doubt</i> ) in Chinese. We think <i>obvious</i> means 明显 and <i>clear</i> refers to 清楚. In fact, these two words are somewhat different according to their Chinese translations: 清楚 reflects a higher degree of certainty. Let me use a metaphor: <i>obvious</i> refers to the stuff floating on the water, which is obvious to see; but <i>clear</i> means clear to see even the bottom of the water. The word <i>clear</i> indicates a very high degree of certainty, so many Chinese students dare not to use it. (Z)

**Table 81. V's interview on her use of *clear* and *obvious***

Text	Interpretation
<i>From the statistics above, it is obvious that elder immigrants account for a great proportion of the population in New Zealand society, and their needs and well-being should not be neglected.</i> (V) <i>It is obvious that the participants in both focus-groups have good relationship with each other and they felt free to talk with each other.</i> (V)	The word <i>clear</i> is not as good as <i>obvious</i> . Whenever I want to express the Chinese meaning 明显 ( <i>apparent, easy to notice or understand</i> ), <i>It is obvious that</i> pops up in my mind. The word <i>clear</i> means 清楚 ( <i>certain</i> ) 清澈 ( <i>limpid</i> ) and 干净 ( <i>clean</i> ) rather than 明显. (V) The word <i>clear</i> is too common, which is not as academic as <i>obvious</i> . (V)

*It is true that* is another interesting bundle, which was used by the Chinese PhD students together with *It is clear that* and *It is obvious that* and had the same structure and close meaning with the other two bundles, but performed a different function. Rather than posing a clear idea, more than two-thirds of *It is true that* bundles were followed by *but* or *however*-clauses, to express a concessive relation, which assumed the conceded proposition that the reader might raise, and established the writer's position through challenging the hypothetical proposition (36).

(36) *It is true that most dictionary user guides have adopted headings, subheadings, bold face and so on to indicate its inherent structure, but how many readers are patient enough to browse in entirety a user guide or aware enough to compare the size of one typeface with that of another? (CH PhD)*

In accordance with this study, Pang (2010) identified the same bundle *it is true that* from his EAP (English for academic purposes) course book corpus, which also acts as a concessive bundle. It is possible that learning materials contribute to the prevalence of this bundle in the Chinese student writing.

The *show* bundles in the Chinese corpora presented the research results (*The following table shows*) or the interpretations of the results (e.g. *The results showed/show that*) as objective hard facts (37). The *found* bundle (*It is found that*) described the research results as the existing facts discovered by the researcher (38). The *believe* bundle (*It is believed that*) articulated the researcher's subjective inferences with great confidence (39).

(37) *The results showed that recasts had positive developmental effects for more advanced learners even though recasts were usually not repeated and rarely elicited MO from the learners. (CH PhD)*

(38) *It is found that learners with more aptitude can not only imitate the foreign sounds correctly but can also distinguish one sound from another while those with less aptitude can not. (CH MA)*

(39) *It is believed that the students who received such strategies training will have a better understanding of the reading process in terms of how they read*

*and whether or not they need or are able, to improve their reading strategies by themselves. (CH MA)*

The informant J explained her use of *believe* in writing (Table 82). She used *believe* to highlight her argument as a universally recognised fact, although the argument was merely derived from her specific research and was not supported in the literature.

**Table 82. J's interview on her use of *believe***

Text	Interpretation
<i>On the one hand, networking relationships, or informal social network (also known as Guanxi in China), are <b>believed</b> to be used predominantly by the DMO in China society within the traditional collectivism system. (J)</i>	Here I want to express the meaning of <i>universally believed with no doubt</i> . This is obtained from my first-hand research. I think this point is very important, but I have not searched the literature. (J)

The bundle *As a matter of* in the Chinese student texts was part of the idiom *As a matter of fact*. Like the New Zealand bundle *The fact that the*, this bundles also emphasised the objectiveness of the statement provided by the writer (40). Both bundles contained the same core word *fact*. However, the New Zealand students used it as a shell noun, while the Chinese postgraduates deployed a *fact*-embedded idiomatic phrase.

(40) *As **a matter of fact**, a careful examination of the frequency order identified in the corpus data and the difficulty order obtained in the elicitation measure reveals a generally similar pattern. (CH PhD)*

The use of *should* in the bundle *It should be pointed (out)* highlighted the necessity for the writer to raise the point articulated in the *that*-clause (41).

(41) ***It should be pointed out** that English proficiency here actually refers to the self-perceived L2 proficiency instead of their true English proficiency level. (CH PhD)*

*There is no doubt*, occurring in the masters corpus, was followed by an appositive *that*-clause, indicating the writer's complete certainty (42).

(42) *There is no doubt that collocations can pose daunting problems to foreign language users and learners. (CH MA)*

The informant V's interpretation (Table 83) identified the source of this bundle: the interlingual transfer of the Chinese expression 毫无疑问.

**Table 83. V's interview on her use of *undoubted***

Text	Interpretation
<i>It is undoubted that learning English is necessary for elder immigrants in English speaking countries. (V)</i>	I am absolutely certain with this argument and I want to express my certainty here. This expression might come from the equivalent Chinese phrase 毫无疑问. (V) I have not thought this expression is so strong while writing, It is not that strong in our Chinese mind or if it is the corresponding Chinese expression. (V)

Booster bundles were used extensively in the Chinese students' writing and it was common to see over generalisation with the use of booster bundles. As in the following example (43), it was one possible explanation from the writer that the source of the error was the negative interlingual transfer from the acquired language (i.e. Chinese verb *zuo*) to the target language (i.e. English verb *do*). Therefore, the *obvious* bundle was too strong to use here and proper caution should be taken while making inferences.

(43) *It is completely a convention in English to talk about make a decision rather than do a decision, although any speaker of English will also understand the latter unconventional expression. It is obvious that the use of 'do a decision' is the result of L1 transfer and both have the meaning of zuo in Chinese. (CH MA)*

Another example is the use of *There is no doubt* (44). The writer's doubtless idea in the first sentence, *communicative language teaching has greatly enhance English teaching in China*, conflicted with his or her statement in the second sentence, *this*

---

*innovation (i.e. communicative language teaching) does not seem to bring about significant improvement in Chinese students' communicative competence.*

(44) ***There is no doubt*** that the development of communicative language teaching has a profound effect on both methodology and syllabus design and has greatly enhanced English teaching in China. However, with more than ten years' effort in adopting communicative language teaching, this innovation does not seem to bring about significant improvement in Chinese students' communicative competence and it seems to suggest that a communicative approach may have its limitations. (CH MA)

#### **8.4 Self-mention bundles**

Self-mention resources include subject or object first-person pronouns and first-person possessive adjectives and pronouns (i.e. *I, we, me, us, my, our, mine* and *ours*). Self-mentions explicitly establish the writer's presence. As Hyland (2001b) argues, "self-mention can help construct an intelligent, credible and engaging colleague, by presenting an authorial self firmly established in the norms of the discipline and reflecting an appropriate degree of confidence and authority" (p. 216). First-person pronouns, *I* and *we*, were the only self-mention devices found within the sentence initial bundles of this study, which acted as discourse guides to introduce or summarise the main points in a particular section or chapter. See examples 45 and 46.

(45) ***In this section, we*** will address this issue primarily from three perspectives: *achieving native-like selection and native-like fluency (Section 3.2.1), supporting social interaction (Section 3.2.2) and facilitating language development (Section 3.2.3).* (CH PhD)

(46) ***In this chapter I*** have described and analysed the general communication patterns in this team, identified the communication challenges they faced and have begun to examine the discursive strategies members of the team had at their disposal to manage the occurrences of miscommunication and problematic talk that inevitably arose in their daily working lives. (NZ PhD)

The use of self-mention pronouns was considerably different between the Chinese and New Zealand students (Table 84). Only one overlapping bundle, *In this section, I*, occurred between the two PhD corpora. Apart from this bundle, the New Zealand students showed their preference to single first-person pronoun *I*, as in the bundles *In this chapter/section I*, to construct their authority and express their confidence as an emerging researcher. The Chinese students extensively used the plural first-person pronoun *we* in *In this chapter/section, we*.

**Table 84. Self-mention bundles in the New Zealand and Chinese corpora**

	NZ	Chinese
<i>I</i> bundles	<b>In this section, I (0, 5)</b> In this chapter I (7, 8); In this section I (0, 6)	<b>In this section, I (0, 7)</b>
<i>we</i> bundles		In this chapter, we (7, 9); In this section, we (0, 7)

According to Ädel (2006), metadiscursive *we* can function as inclusive authorial *we* and exclusive *we*. The former refers to both the writer and the reader and creates writer-reader solidarity. This can be seen from the bundle *As we all know*, which will be discussed later in the section of shared knowledge. The latter is comprised of collective *we* and editorial *we*. Collective *we* refers to co-authors of the writing. Editorial *we* is the most unusual but interesting type, which is used by single writer for self-reference. All the *wes* of the Chinese bundles functioned as editorial *we*, although the Chinese theses were all single-authored texts. Quirk, Greenbaum, Leech, and Svartvik (1985) explain the motivation of using editorial *we* as a “desire to avoid *I*, which may be felt to be somewhat egotistical” (p. 350). Hyland (2001b) and Kuo (1999) interpret this phenomenon as an intention to reduce personal attributions and to obtain authority from the plural form. This may also reflect the slower pace of the development of English academic writing in China, where the traditional “author-evacuated” view of academic prose is possibly still prevailing (Geertz, 1988, p. 9). The collective pronoun *we* used here seems to downplay the presence of writers.

Lack of personal voice was also found in the interview informants' writing. Three informants, Z, A and V, explained the reasons for the absence and these emerging researchers' voices were largely shut down by teacher expectation (Z, V), view of rhetorical convention (A) and personal confidence (V).

I will not use *I* or *we* in academic writing. I still remember my writing teacher requires us not to use them because they are too subjective. (Z)

We are not allowed to use *I* or *we* in our writing, particularly in the finding chapter. It is all about my findings, so I need to stay neutral and my findings should be objective. (A)

Our teachers expect us not to use *I* or *we* in academic writing because nobody cares who you are and what you say. Whenever I use *I* or *we*, the teachers will change it into another pattern. No *I* or *we* is allowed in academic writing. I also agree to avoid *I* or *we* because I am not qualified to write things like *I suggest* or *I argue*. (V)

One *I* was found in W's work (Table 85), but during the interview she indicated her preference for *the researcher*. Like the other informants, W also learned from her Chinese teacher that academic writing should be objective and first person pronouns should be excluded from the text.

**Table 85. W's interview on her use of *I***

Text	Interpretation
<p><i>Following the 3 dimensions, I will use transitivity in ideational meta-function of language in Halliday's (1985) systemic-functional grammar and lexical classification (p.129) as analysis tools and interpret how discourse is produced, then the selected interviewing discourse will be will closely scrutinized in relation to the dominant ideology of the time when they are produced. (W)</i></p>	<p>Strictly speaking, I should use <i>the researcher</i> here. Academic writing should be objective and scientific, so I try not to use first person pronouns. (W) I learned this point from my teacher in China when I studied my masters degree. (W)</p>

The frequency of self-mention pronouns was greatly different across the masters and PhD levels and the PhD students employed far more self-mention bundles. This result supports Hyland's (2004b) finding. This is because more experienced student researchers are likely to have more confidence in presenting their writer identity, while less experienced researchers tend to believe that self-mention conflicts with



objectivity and formality of academic writing (Hyland, 2004b). This is true of both New Zealand and Chinese students' writing.

### 8.5 Directive bundles

Directive bundles are the most popular strategy to engage readers, which function as signposts to guide readers throughout arguments (Hyland, 2001a, 2005b). Hyland (2002b) classifies directives according to the principal forms of the directed activities and divides these activities as textual acts, physical acts and cognitive acts.

Textual acts steer readers to another part of the text or to another text.

Physical acts instruct readers to perform some action in a research or real world situation.

Cognitive acts guide readers through a line of reasoning or to understand a point in a certain way.

The Chinese students, especially the PhDs, used more directive bundles opposed to the New Zealand students. As listed in Table 86, three shared bundles were *As can be seen*, *It can be seen* and *It should be noted*. The most significant difference between the Chinese and New Zealand writing was the use of verbs: the Chinese students used sense verbs *see* and *look at* to direct readers mostly to textual acts, whereas the New Zealand students employed mental verb *note* to guide readers to cognitive acts. There was no bundle referring to physical acts in this study.

**Table 86. Directive bundles in the New Zealand and Chinese corpora**

	NZ	Chinese
<i>see/look at</i> bundles	<b>As can be seen (13, 9); It can be seen (5, 0)</b>	<b>As can be seen (5, 14); It can be seen (13, 15)</b> Look at the following (0, 11); We can see from (5, 0); We can see that (0, 5)
<i>note</i> bundles	<b>It should be noted (10, 11) It must be noted (5, 0)</b>	<b>It should be noted (0, 14)</b>

This supports Hyland's (2002b) study on directives: the most imposing directives, which direct to cognitive acts, are least used by the Hong Kong L2 students, who tend to use directives to guide their readers through research procedures or to steer their attention to non-linear information of their texts (e.g. tables, examples and

appendices). In other words, the degree of risk during communication largely affected the Chinese students' writing: they tended to choose words or lexicogrammatical patterns that performed low-risk functions and were not confident to take high risks.

### 8.5.1 Shared directive bundles

Three overlapping bundles between the New Zealand and Chinese corpora were *As can be seen*, *It can be seen* and *It should be noted*. The first two bundles have been discussed in the Section 7.3.1 Shared endophoric bundles because they act as both endophoric and directive bundles. This section will only focus on the last bundle.

*It should be noted*, followed by *that*-clause, rarely occurred in the Chinese masters corpus but was used frequently in the other three corpora. The use of *should* carries strong connotations of unequal power, claiming the higher authority of the writer by forcefully focusing the reader's attention on a particular point (Hyland, 2001a), in this study, the research background (47), limitation of the research (48) or research finding (49).

- (47) ***It should be noted*** that during the current study, *Summarise-Pair-Share* is used to prompt readers to summarise text and to talk about the use of this strategy, rather than to teach them how to summarise. (NZ MA)
- (48) ***It should be noted that***, for methodological reasons, the analysis in this section relates only to references to singular persons. (NZ PhD)
- (49) ***It should be noted that*** the percentage of the word families produced at this level was lower than that at both the 3,000 (24.5% lower) and the 5,000 (2.3% lower) word levels, while in the receptive vocabulary test, the percentage of the academic words they knew was the highest among the four word levels. (CH PhD)

The Chinese masters students relied greatly on a range of similar patterns to arouse the readers' attention, consisting of *It is worth noting that*, *It should be notice that* and *We should notice that*. From the informant V's Chinese translation of these patterns 值得关注, 应该注意 and 我们应该注意 (Table 87), we can see her

awareness of the minor difference between the word *note* and *notice*: *note* means deserve attention and *notice* means become aware.

**Table 87. V's interview on her use of *note* and *notice***

Text	Interpretation
<p><i>However, it is worth noting that sometimes the family's support could be a barrier for English language learning. Some elder participants were not eager to study English in New Zealand because they have family to support them to deal with affairs which need English.</i> (V)</p>	<p><i>It is worth noting that</i> means 值得关注 (it deserves attention) (V)</p>
<p><i>It should be noticed that most of the studies concerned about the comparison between younger learners and elder learners, but few of them compared the starting age among the single group of elder learners.</i> (V)</p>	<p><i>It should be noticed that</i> means 应该注意 (it should gain awareness), and <i>notice</i> means 注意 (become aware). (V)</p>
<p><i>However, we should notice that using Chinese to help learning English has a disadvantage, if there is any inaccurate and misleading translation between these two languages or if it is difficult to express the original content in another language, it would be very difficult for elder people who have limited English ability to discover.</i> (V)</p>	<p><i>We should notice that</i> means 我们应该注意 (we should become aware of it), and Chinese like to use <i>we</i> to refer to the reader and writer. (V)</p>

### 8.5.2 Directive bundles in the New Zealand students' writing

As a deviation from *It should be noted*, *It must be noted* only existed in the New Zealand masters corpus, which was also followed by *that*-clause and performed a fairly similar function (50). However, according to Salager-Meyer's (1992) scale of certainty: must-should-would-can-may-might (p. 93), *must* bundles carry a stronger sense of obligation, as in the following example:

- (50) *It must be noted that the interviews with all participants were transcribed while out in the research field, in order to afford all participants an opportunity to comment or withdraw any statements they made.* (NZ MA)

### 8.5.3 Directive bundles in the Chinese students' writing

Unlike the New Zealand student writing, *note* bundles are not popular among the Chinese students, especially at the masters level. As the informant J stated, the Chinese students may not feel comfortable and confident to actively involve their readers (Table 88). J used *find* instead of *note* in her original writing to express personal stance from the writer's perspective. Being suggested the bundle *it is interesting to note*, she responded as follows.

**Table 88. J's interview on her use of *note***

Text	Interpretation
<i>It is interesting to find that there is wide networking coordination within the DMO in Zhu Jiayu and other stakeholders.</i> (J)	I do not want to replace <i>find</i> with <i>note</i> and I am not comfortable to use <i>note</i> . If my reader agrees with me and finds this interesting, it is interesting to them; otherwise, it is not. I am not willing to forcefully involve my readers and require them to pay attention to this point. Instead, I want to tell them this is my finding. I think my readers should have their freedom. If they think this is an interesting finding, they will note this point. If they do not think so, then they do not share the same opinion with me. (J) If I choose <i>note</i> , does it mean I am more confident with my finding? Along with comfort, confidence is also important in writing. (J)

Other directive bundles in the Chinese writing were *Look at the following* and *We can see from/that*. The bundle *Look at the following* was used in the Chinese PhD corpus to introduce example(s) as *Look at the following example(s)*. The verb phrase *look at* occurs relatively frequently in spoken rather than written register, 499.7 versus 108.8 cases per million words in the BNCweb, which explains the low frequency in another three thesis corpora.

*We can see from/that*, only occurred in the Chinese student corpora. The plural pronoun *we* was used as an inclusive pronoun to create writer-reader solidarity, inviting the reader to interpret the succeeding research result through the writer's eyes. As shown in examples 51 and 52, the only difference between *We can see*

*from* and *We can see that* was that the former also introduced the media (e.g. table, figure and data) to the reader with a *from*-phrase, inserted between the main clause *We can see* and subordinate *that*-clause.

(51) *We can see from table 4.4 and 4.5 that the use of the restricted collocations is similar to the results of free collocations. (CH MA)*

(52) *We can see that under the framework of SFL, abstract nouns play an important role in the construal of human experience. (CH PhD)*

According to Wen, Ding, and Wang (2003), the use of first-person pronoun *we* in the Chinese students' writing is regarded as a feature of English spoken discourse. In this case, these two bundles were less formal than the two shared *see* bundles, *As/It can be seen*.

The prevalence of *see* in the Chinese writing may also be related to its frequency. In the written texts of the BNCweb, *see* occurs nearly ten times as many as *note* (1013 compared to 109 per million words). As the informant A noted, the frequent encounters in his reading contributed to his preference of *see* (Table 89).

**Table 89. A's interview on his use of *see***

Text	Interpretation
<i>See Figure 1. (A)</i>	I do not know any other way. I have seen this pattern many times during my reading. It is easy to use. (A)
<i>See Figure 3. (A)</i>	
<i>See one example from Figure 5. (A)</i>	

## 8.6 Shared knowledge bundles

There was one shared knowledge bundle, *As we all know*. This occurred in the Chinese masters corpus with relatively high frequency (i.e. 11 times per million words) to introduce the writer-recommended "truth" (53-55).

(53) *As we all know, English is a widely-used language in the world. (CH MA)*

(54) *As we all know language learning is an abstract process which requires a large amount of memorization, so it is not an easy task to learn it well for second language speakers. (CH MA)*

---

(55) *As we all know*, when students do not know the proper English expressions in writing, they usually turn to literal translation, no matter whether the translation can be accepted in English. (CH MA)

As indicated in the informant V's interview (Table 90), the plural first-person pronoun *we* not only included the writer and reader but also encompassed the whole discourse community; therefore, the use of this bundle allowed her to attribute her statement to common knowledge and no reference was needed to support the argument.

**Table 90. V's interview on her use of *As we all know***

Text	Interpretation
<i>As we all know</i> , the linguistic distance between Chinese and English is huge. (V)	Here, I want to express everybody knows it and it is a common knowledge. If I had not used <i>As we all know</i> , I should have referenced this proposition. I really cannot find the reference and nobody studies the linguistic distance between English and Chinese. (V)

The informant W distinguished phenomena from concepts, and she felt more confident to use *as we all know* to modify phenomena. She also pointed out the sources of this phrase: English newspapers, reading comprehension exercises or secondary school textbooks.

I feel confident to use *as we all know* or *as it is known to all* to modify social phenomena rather than concepts. These phrases appear in English newspapers, reading comprehension exercises or secondary school textbooks. (W)

## 8.7 Summary

This chapter has focused on the frequency, structure and function of interactional bundles in the Chinese and New Zealand theses. The differences between the Chinese and New Zealand writing and between the masters and PhD texts were analysed on the basis of Hyland's model of writer-reader interaction.

---

### 8.7.1 Differences between the Chinese and New Zealand writing

The Chinese and New Zealand theses were considerably different in the use of interactional bundles, which were attitude marker bundles, hedge/booster bundles, self-mention bundles, directive bundles and shared knowledge bundles. Compared to the New Zealand postgraduates, the Chinese students were more conservative in expressing their personal attitudes towards certain statements or research-related experiences, which can be seen from the absence of *It is interesting/difficult to* bundles in the Chinese high-frequency bundle lists. As to hedge and booster bundles, the New Zealand students tended to soften their assertions and expressed their tentativeness, extensively using the bundles such as *It is possible that* and *It may be that*, which is also regarded as a feature of academic writing in the social science disciplines (Hyland, 2004b). The Chinese students, on the other hand, preferred to strengthen their arguments through booster bundles (e.g. *It is clear/obvious that*) and lack the awareness of cautiously weighting their claims and negotiating with their readers. Another major difference was the use of self-mention pronouns: the Chinese students intended to create a multiple-author representation by means of plural pronoun *we*. However, the New Zealand students preferred to choose single pronoun *I* to establish their own authority as emerging researchers.

In addition, the Chinese students' writing showed the traces of negative transfer from Chinese and influence from spoken English. For example, *As we all know* was probably attribute to an equivalent expression 众所周知 in Chinese. Other typical lexical bundles in the Chinese students' writing, such as *Look at the following*, appeared to be prevalent in informal or spoken contexts rather than academic genres.

### 8.7.2 Differences between the masters and PhD writing

The greatest differences between the masters and PhD writing were in the use of attitude markers and self-mentions. The most significant difference in the use of attitude markers between the masters and PhD writing was the function of *important* bundles. The masters writing seemed to be research-focused with a greater proportion of *important* bundles modifying research actions, while the PhD writing appeared to be claim-focused with more *important* bundles reinforcing argumentation. Self-mention bundles were another category that differed between

---

the masters and PhD writing. The masters students showed their lack of confidence in exposing their writer identity and a lack of knowledge of changing rhetorical conventions with a comparatively low frequency of self-mention bundles.

The following chapter is the last chapter of this thesis, the discussion and conclusion chapter. I will summarise the discrepancies in the use of sentence initial bundles and reasons for Chinese students' bundle choices. At the same time, I will address the limitations of the present study, provide suggestions for future research and highlight the theoretical, methodological and pedagogical implications of this bundle research.





---

## Chapter 9 Discussion and conclusion

Corpora and corpus linguistics have attracted increasing attention in recent years, as different types of corpora have been built and various corpus linguistic approaches have been used to generate word lists or conduct discourse analysis. In the area of academic discourse analysis, corpus linguistic approaches have been used to investigate languages, registers, genres, disciplines and learner interlanguage. In terms of learner language, particularly learner language in academic writing, contrastive interlanguage analysis (i.e. the CIA<sup>2</sup> approach) (Granger, 2015) has been adopted in the areas of both lexical bundle research and metadiscourse analysis, two major areas beyond the word level. Lexical bundle research compares the use of bundles between L1 and L2 writing with regard to frequency, structure and function. However, many studies have ignored the differences between genres, overlooked the influences of academic competence, conflated sentence initial and non-initial bundles, and failed to take into account learners' perspectives. Metadiscourse research compares the number of pre-determined metadiscourse devices or devices in a specific metadiscourse subcategory (e.g. hedges) between L1 and L2 texts. However, most studies have used Hyland's (2005a) list of pre-determined items, mainly individual words. It is sometimes insufficient to determine the function of text by means of single words. This top-down approach with pre-determined research items is also likely to miss some salient features of academic language. Moreover, few lexical bundle and metadiscourse studies, if any, have included learner perspectives on their own lexical choices.

To fill the gaps in the previous research, this study identified discrepancies in the use of sentence initial bundles in thesis writing between Chinese L2 postgraduates and New Zealand L1 postgraduates. Four collections 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. This study compared both masters and PhD theses because comparing the same genre at the same level is likely to eliminate influencing factors such as communicative goals, text lengths and writer identities. This study focused primarily on sentence initial bundles

---

because they function differently from non-initial ones. This study examined lexical bundles as metadiscourse devices to better inform writing pedagogy. In addition, this study explored the underlying reasons for the Chinese postgraduates' typical bundle choices.

In this chapter, I will first and briefly summarise the principal findings in relation to the research questions. Then I will discuss the implications of this study in relation to theory, methodology and pedagogy. Finally, I will examine the limitations of the current study and provide suggestions to future research.

## **9.1 Discrepancies and reasons of discrepancies**

This study takes the CIA approach. It explores the discrepancies between sentence initial bundles produced by Chinese L2 and New Zealand L1 students and compares bundles between masters and PhD levels. The comparison of Chinese L2 bundle production at different academic levels can be regarded as a dimension of learner variables in the CIA model. This study focuses on four research questions with regard to discrepancies in bundle use in frequency, structure and function, and reasons for these discrepancies. I summarise the key findings below.

### **9.1.1 Discrepancies in frequency**

This section addresses the first research question:

*What are the frequencies of four-word sentence initial bundles in the Chinese and New Zealand masters and PhD corpora? Are there any differences between Chinese and New Zealand postgraduates, or between masters and PhD levels of study in the use of sentence initial bundles?*

With respect to this question, I have obtained the following four findings:

1. Chinese postgraduates, particularly masters students, rely more heavily on sentence initial bundles. This echoes previous bundle studies on Chinese student writing in which masters level students use the most bundles (e.g. Hyland, 2008a; Wei & Lei, 2011; Xu, 2012). This is also consistent with previous metadiscourse research on L2 learner writing in which L2 writers

---

generally deploy more metadiscourse devices than native writers. (e.g. Ädel, 2006; Cao & Wang, 2009; Heng & Tan, 2010).

2. It is also interesting to note that New Zealand postgraduates also follow the same trend as New Zealand masters student writing appears to rely more on bundles. Few studies have been conducted to investigate native students' thesis writing and their language development, so it is difficult to generalise this finding to a wider context.
3. Both groups of PhD students (i.e. Chinese PhDs and New Zealand PhDs) share more bundles than their masters counterparts, although they use fewer bundles in total. This further affirms Qin's (2014) finding of the correlation between academic levels and the use of target bundles of native writing: the number of overlapping bundles steadily increases from masters to PhD level.
4. Both groups of Chinese postgraduates (i.e. Chinese masters and Chinese PhDs) show their stronger preference for interactive bundles in comparison to New Zealand postgraduates, who deploy a higher percentage of interactional bundles. It is difficult to compare this finding with previous metadiscourse research because a bottom-up (i.e. more corpus-based), instead of a top-down (i.e. more discourse-analytic) approach, has been taken in the present study; and also because the conflicting research results were reported in the former studies (e.g. Cao & Wang, 2009; Heng & Tan, 2010). It is, however, possible to compare this finding with the findings of bundle research because of the strong correlation between lexical bundle taxonomies and metadiscourse models. The finding of the present study is in line with the findings from studies in bundle research in which discourse organisers or text-oriented bundles were found more popular in L2 student writing, and stance bundles or participant-oriented bundles were more pervasive in native student corpora (e.g. Ädel & Erman, 2012; Chen & Baker, 2010; Hyland, 2008a; Wei & Lei, 2011; Xu, 2012).

### **9.1.2 Discrepancies in structure**

This section deals with the second research question:

---

*What are the salient structures of these bundles in the Chinese and New Zealand corpora? Are there any differences between Chinese and New Zealand postgraduates, or between masters and PhD levels of study in the distribution of structures?*

The structural analysis has been informed by the previous work of Biber and his colleagues (Biber et al., 2004; Biber et al., 1999) and Chen and Baker (2014). Five structural categories were identified and three new patterns were created with regard to the sentence initial bundles in this study: NP-based, PP-based, VP-based (including two patterns *active or passive verb + noun/preposition phrase fragment* and *(in order) + to-clause fragment*), clause-based (including *anticipatory it-clause fragment*, and three created patterns *there be-clause fragment*, *noun phrase + verb phrase fragment* and *conjunction + clause fragment*) and other bundles. NP-based bundles, PP-based bundles, anticipatory-*it* bundles, and *noun phrase + verb phrase fragment* bundles were generally found to be dominant across all four corpora. This is slightly different from other lexical bundle research on academic prose, in which NP-based bundles, PP-based bundles, anticipatory-*it* bundles, and passive verb bundles have been found pervasive. The reason for the inclusion of *noun phrase + verb phrase fragment* bundles and exclusion of passive verb bundles is the different object of this study: primarily sentence initial bundles.

Discrepancies in the use of bundles between Chinese and New Zealand students were examined within each structural category:

1. NP-based bundles, mostly shell noun bundles (e.g. *The results of the*), are largely underused in Chinese postgraduate writing, even at the PhD level. Neither masters nor doctoral informants are seemingly aware of the important functions of these shell noun bundles. Only two bundles were identified in the Chinese PhD corpus (*The results of the*, *The analysis of the*). In contrast, Wei and Lei (2011) report a similar number of NP-based bundles between their Chinese PhD corpus and journal article corpus. The different quality of PhD theses between my corpus and their corpus may contribute to the conflicting results. The current PhD corpus consists of theses from 12 universities including both top and common ones. In contrast,

their PhD corpus merely contains theses from top universities. In this sense, the results of this study are likely to be more representative of Chinese PhD thesis writing.

2. PP-based bundles, especially the bundles indicating logical relations (e.g. *As a result of*, *On the other hand*, *In addition to the*), occur more frequently in Chinese student writing. Both Chinese masters and PhDs prefer to begin their sentences with prepositional phrases. This supports Hyland's (2008a) finding for masters and PhD thesis writing. A closer examination of bundle patterns further reveals that Chinese masters, compared with the other three student groups, appear less inclined to use *of*-phrase PP-based bundles (e.g. *In the case of*, *In terms of the*, *As a result of*).
3. VP-based sentence initial bundles only exist in Chinese student writing. This differs from the previous studies, in which passive verb bundles (e.g. *is based on the*) have been recognised as an important feature of academic writing (e.g. Ädel & Erman, 2012; Chen & Baker, 2010; Hyland, 2008a; Wei & Lei, 2011). One possible explanation is the different focus of the present study, sentence initial bundles, and New Zealand students prefer to position VP-based bundles in the second part of their sentences. Another reason is both Chinese masters and PhDs prefer to use (*in order*) *to* bundles (e.g. *To sum up*, *the*, *In order to make*). This finding extends the research of Chen and Baker (2010) on Chinese undergraduate writing to postgraduate level. Chinese tertiary students' preference for (*in order*) *to* bundles exists regardless of levels and they particularly prefer to use this type of bundle as sentence starters.
4. Clause-based bundles consist of four patterns: anticipatory it-clause fragment, there be-clause fragment, noun phrase + verb phrase fragment and conjunction + clause fragment. Chinese masters and PhDs deploy fewer anticipatory-*it* bundles than native writers, in this case, New Zealand postgraduates. Like Ädel and Erman's (2012) study, few *there be* constructions appear in Chinese masters and PhD writing. Chinese postgraduates, however, incorporate more noun phrase + verb phrase fragment and conjunction + clause fragment bundles in their theses. The preference for the former type of bundles indicates a lack of NP-based

---

bundles, in other words, nominal phrases, in Chinese student writing and the prevalence of the latter type of bundles shows their high reliance on conjunctions.

The discrepancies between levels of study (i.e. masters level compared to PhD level) mostly confirm the findings of Xu (2012) and Hyland (2008a). There are more research-related NP-based bundles in both masters corpora (i.e. Chinese masters and New Zealand masters), and more PP-based bundles and anticipatory-*it* bundles in both groups of PhD students' writing (i.e. Chinese PhDs and New Zealand PhDs).

### 9.1.3 Discrepancies in function

This section answers the third research question:

*What are the metadiscourse functions of these bundles in the Chinese and New Zealand corpora? Are there any differences between Chinese and New Zealand postgraduates, or between masters and PhD levels of study in the distribution of functions?*

The functional analysis is based on Hyland's interactive (2005a) and interactional (2005c) model of metadiscourse. No interactive bundle was identified as evidential, but two new subcategories — condition bundles and introduction bundles — were created within the category of interactive bundles. As for interactional bundles, no bundle was found as personal asides, questions or embedding reader pronouns. Therefore, interactive bundles in this study comprise transition bundles, frame bundles, endophoric bundles, code glosses bundles, condition bundles and introduction bundles. Interactional bundles, on the other hand, consist of attitude bundles, hedge bundles, booster bundles, self-mention bundles, directive bundles and shared knowledge bundles. Interactive and interactional bundles have been examined respectively in this study. A few sentence initial bundles act as both interactive and interactional devices, and they have been classified in both categories.

Most lexical bundles studies put great effort into overall comparisons rather than comparisons within each subcategory (e.g. comparisons of epistemic stance

---

bundles), or on quantitative analysis rather than context-based qualitative analysis. Metadiscourse research, on the other hand, covers both general and specific comparisons, and involves both quantitative and qualitative approaches. However, many metadiscourse studies focus on interactional instead of interactive metadiscourse resources. Interactional devices such as hedges, boosters, self-mentions and directives have been examined extensively, but little attention has been devoted to interactive devices.

This study took a quantitative as well as qualitative approach, and investigated both interactive and interactional bundles. Seven major findings are summarised below:

For interactive bundles

1. Chinese students showed their preference to use transition bundles embedded with transition markers or one-word conjunctions (e.g. *on the other hand, However, it is not*), rather than noun phrases to start their sentences.
2. Chinese students were found to lack knowledge of endophoric bundles, particularly shell noun bundles, an effective strategy to achieve cohesion and coherence.
3. Chinese students were observed to employ a wider variety of condition bundles and many of them only appear in the Chinese corpora (e.g. *As far as the, In this way, the, When it comes to, With the development of, In order to make*).
4. Chinese students in the present study tended to use bundles without specific references, for example, *The first one is*. Compare this to *The first of these* in New Zealand writing.
5. Both Chinese and New Zealand PhD students in this study paid more attention to linking, modifying and elaborating their ideas, using more transition bundles (e.g. *In addition to the*), condition bundles (e.g. *In the case of*) and section-level frame bundles (e.g. *In this section, I*).

For interactional bundles



- 
6. Chinese students were found to prefer to use booster (e.g. *It is obvious that*) rather than hedge bundles (e.g. *It is possible that*). This confirms many previous studies (e.g. Gilquin & Paquot, 2008; Hyland & Milton, 1997; Yang, 2013).
  7. Chinese students appeared to be more reluctant to express their personal feelings, to reveal their writer identity and to cognitively involve their readers, with a relatively low use of attitude bundles (e.g. *It is interesting to*), self-mention bundles (e.g. *In this chapter, I*) and directive bundles of cognitive acts (e.g. *It should be noted that*).
  8. Both Chinese and New Zealand PhD students in this study were more cautious to express their attitude and less reluctant to indicate their writer identity, using fewer attitude bundles (e.g. *It is important to*) but more self-mention bundles (e.g. *In this section, I*).

In addition, a range of typical bundles has been found in Chinese student texts. Some examples are *In a word, the, To be more specific, To put it another (way), From the perspective of, As far as the (... is/are concerned), When it comes to, As we all know, and Look at the following*. The possible reasons for these bundles will be discussed in the next section.

#### **9.1.4 Reasons for discrepancies**

This section considers the last research question:

*What reasons do Chinese postgraduates give for their sentence initial bundle choices in their thesis writing?*

In order to answer this question, six Chinese postgraduates were interviewed in regard to their identical or partially identical language production to the typical sentence initial bundles in the two Chinese corpora. Seven reasons have been provided for the different bundle selections in Chinese student writing, which include:

1. Interlingual transfer
2. Classroom learning
3. Noticing in reading

- 
4. A lack of rhetorical confidence
  5. Misunderstanding of rhetorical conventions
  6. Limited word knowledge
  7. Learner strategies

Interlingual transfer refers to the transfer from Chinese. According to the interviewees' interpretations, interlingual transfer involves word order transfer, literal transfer and semantic transfer for bundle production. Word order transfer occurs when Chinese students follow word sequences in Chinese sentences while constructing English sentences. Examples are the use of *By means of*, *In order to make*, and other *to*-phrase fragments at the beginning of sentences. Literal transfer means word for word translation, for example, when the direct Chinese translation of 毫无疑问,从这个方面来讲 and 当提到老年人的时候 leads to the use of *There is no doubt, from the perspective of* and *when elder is mentioned*. Semantic transfer is found when Chinese students choose English bundles according to the meanings of equivalent Chinese words. A typical example of semantic transfer is the preference for *It is obvious that* to *It is clear that*, which is the result of Chinese students' judgement between 明显 and 清楚. Interlingual transfer facilitates these Chinese writers with their language production. However, during the transfer, Chinese postgraduates fail to notice the variations in pragmatics between the source language and target language. As in the example *There is no doubt* and 毫无疑问, the English one conveys a much higher degree of certainty than the Chinese equivalent.

Classroom teaching is another important factor contributing to Chinese students' bundle use. Teachers have been reported to emphasise or even overemphasise certain language features, while many salient features of academic English have been overlooked. In the context of EFL teaching in mainland China, single-word conjunctions (e.g. *however, therefore*) and sequence markers (e.g. *last but not least*) are introduced as a strategy to achieve cohesion and coherence in English writing. At the same time, the linking power of shell nouns (e.g. *fact, problem, approach*) and shell noun bundles (e.g. *The results of the, The purpose of this*) has been overlooked. Students are encouraged to avoid word repetition in order to show the richness of their vocabulary, but the linking function of repeated words and the

---

distinction between words, particularly between synonyms (e.g. *important* compared to *necessary*, *indicate* compared to *show*) are largely neglected. Students have learnt the anticipatory-*it* pattern as a specific grammatical structure, but they rarely have the ready-made anticipatory-*it* bundle repertoire (e.g. *It is interesting/difficult to*) to draw on during academic writing. Students are familiar with a range of reference-free expressions (e.g. *One of the most*, *As we all know*) to turn their arguments into universal truth or common knowledge. However, the crucial role of reading in writing has been marginalised, and reading and writing have mostly been taught as two isolated skills.

Noticing is an essential prerequisite for bundle learning. However, as Cortes (2004) reports, “even though students might have frequently encountered these expressions in their academic reading, simple exposure to the frequent use of lexical bundles in published academic writing does not result in the acquisition of these expressions by university students” (p. 417). It is interesting to note that familiarity is a necessary precondition for students’ noticing (i.e. conscious learning during reading). Familiar language items, such as conjunctions, sequence markers and some fixed expressions *for example*, *with the development of* and *as we all know* seem salient to students, and have been consciously noticed. In contrast, unfamiliar items or unknown features like the bundles *in the case of* and *in terms of*, and the position of (*in order*) *to*-infinitive phrase fragments have gained little attention from these students. Both their limited L2 processing ability and the shortage of awareness-raising tasks are likely to have contributed to their lack of noticing.

Rhetorical confidence has been raised as another determinant of Chinese students’ bundle selection. Students may resort to avoidance strategies, “failing to exploit the full range of the target language’s expressive possibilities” (Leech, 1998, p. xiv). They tend to put the conjunction *however* all at the beginning of their sentences and dare not to take the risk to place it after the subject of the sentence, an unfamiliar position for them. They use the high-frequency and transparent word *way* as their lexical teddy bear (Hasselgren, 1994), and avoid choosing a more appropriate but unfamiliar one such as *style*, *method* or *approach*. Students also seem to be highly conscious of their identity as student researchers and apprentice writers. They appear more comfortable to express their attitudes towards their own research

---

procedures rather than influence the evaluation of their readers. They employ more *It is important to + verb + object* patterns and *necessary* bundles (e.g. *It is necessary to*), and underuse *It is important to + verb + that-clause* patterns and *important* bundles (e.g. *It is important to*). They are more likely to steer their readers to textual acts rather than cognitive acts, and low-level instead of high-level cognitive acts. They use *look at* and *see* bundles (e.g. *Look at the following*, *We can see from*) to guide their readers to examples, tables, figures and data, but avoid *note* bundles (e.g. *It should be noted*) to draw their readers' attention to a particular statement. They prefer *obvious* to *clear* bundles to present objective facts instead of subjective arguments, to describe the information that is easy to understand and to indicate a low degree of certainty. They intend to create a multiple-author representation by means of *we* bundles (e.g. *In this chapter, we*), and are reluctant to choose *I* bundles (e.g. *In this chapter, I*) to establish their own authority as emerging researchers.

A few bundles in Chinese students' writing are results of their misunderstanding of rhetorical conventions. Chinese students are more likely to regard academic writing as statements of objective fact rather than subjective personal arguments. Therefore, they try to hide their personal feelings, lack reader-awareness, and appear reluctant to reveal their existence as writers. This seems to explain why the popular personal feeling bundles in New Zealand student theses (e.g. *It is interesting/difficult to*) are absent in Chinese student writing. Moreover, different weight is put on hedge and booster bundles. New Zealand students tend to soften their assertions and express their tentativeness, extensively using hedge bundles such as *It is possible that* and *It may be that*, which is regarded as a feature of academic writing in the social science disciplines (Hyland, 2004b). Chinese students, on the other hand, prefer to strengthen their arguments through booster bundles (e.g. *There is no doubt*) and lack awareness of how to more cautiously weight their claims and negotiate with their readers. In addition, Chinese students' existence as writers has largely been hidden by the absence of first-person pronoun *I* bundles or by their use of collective pronoun *we* bundles.

Word knowledge consists of form, meaning and use (Nation, 2013). In this study, these Chinese postgraduates appear to lack word knowledge in the areas of collocations, grammatical patterns and register constraints. The inappropriate use

---

of *necessary* with mental verbs (e.g. *It is necessary to note that*) reflects their insufficient learning of collocations, particularly collocations between synonyms (e.g. *important* compared to *necessary*). The absence of *interesting* and *difficult* bundles indicates their incomplete knowledge of grammatical patterns. On the one hand, they may not have a ready-made bundle repertoire to draw from (e.g. *It is interesting to note that*). On the other hand, the incorrect guidance they may have received from some writing books (e.g. the book written by Xiaoyi Shen) or teachers discourages their use of certain patterns. The preference for less formal *seem* bundles (e.g. *It seems that the*) reflects their limited knowledge in regard to registers, in other words, where to use the bundle. In this case, the synonym *appear* is a more formal word than *seem*, and the bundle *It would appear that* is more appropriate to use in academic writing.

Learner strategies refer to the strategies consciously adopted by L2 learners to achieve their particular purposes, which include strategies used to overcome limited language proficiency, to avoid the trouble of reference searching, and to balance their long sentences. The first type has also been referred to as a type of communication strategy in the literature, that is, Tarone's (1980) approximation strategy or Færch and Kasper's (1984) substitution strategy. Chinese students tend to use a word or phrase of a close meaning to substitute the intended word or phrase, using *way* for *style, method* or *approach*, or *in this way* for *from this perspective*. In the case of reference searching, Chinese students may employ set expressions such as *one of the most* and *as we all know*, if they are reluctant to search for the literature or cannot find the references. An example of sentence balancing is that Chinese students may place the long modification phrases (e.g. *by means of, in order to*) at the beginning of sentences, if they consider their sentences are overloaded.

## **9.2 Limitations and suggestions for future research**

The limitations of this study mainly rise from the nature of the corpora, the approach to the bundle analysis and the selection of the interview informants. I will address these limitations and provide suggestions for future research in the section below.

---

### 9.2.1 Limitations

The Chinese and New Zealand masters and PhD corpora were self-built particularly for this study and there is no validation corpus-based analysis for this research, although some findings have been validated from the interview data. Moreover, the corpora for this study were built from postgraduate theses in the discipline of general and applied linguistics. Caution should be taken while interpreting and generalising the findings of this study to a broader context. The research results and identified bundles may not be transferrable to other genres (e.g. journal articles), other disciplines (e.g. computer science) or other academic levels (e.g. undergraduate theses).

It should be noted that 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 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 (1-3). The study of these bundles complements the findings of this study and is equally important to extend learners' bundle knowledge.

(1) *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)*

- 
- (2) *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)*
- (3) *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)*

Furthermore, the cut-off frequency is set relatively high to limit the number of bundles to a manageable size. This manipulation probably results in the exclusion of a number of low frequency but valuable metadiscourse items.

As to the selection of the interview informants, the most convincing interpretations would be presented if these informants could be selected directly from the writers of the corpus data. If this is not possible, more convincing results could alternatively be achieved through selecting the informants who have the same education background with the corpus writers (e.g. composing their theses in Chinese universities, studying general or applied linguistics). However, due to the constraints of the research, it was impossible to collect interview data from writers of corpus texts and mainland China. Although the interviews feature the overlapping expressions between these informants' writing and the sentence initial bundles of the corpus data, the divergent learning contexts between these informants and the thesis writers in China should be borne in mind while interpreting the interview data.

The postgraduates (i.e. corpus writers) in China are writing for different audiences. Although these Chinese postgraduates choose different sentence initial bundles from their English speaking counterparts, it may not be regarded as inefficient and ineffective writing in their own context. This is because their target readers, their supervisors, thesis examiners or other general readers, come from the same linguistic and cultural background as them, and are less likely to find the students' language anomalous or regard these typical sentence initial bundles as an obstacle to their understanding. However, these Chinese students need to craft their writing if they intend to have their work published in English and accessed by wider audiences from different backgrounds.

---

### 9.2.2 Suggestions

With regard to the above limitations in terms of corpus building, bundle identification and informant recruitment, future bundle research is greatly needed to explore the features of non-native writers' bundle selections and the recurrent patterns in the texts of (proficient) native writers. In the present study, four comparative thesis corpora were built within one discipline and comparisons have been made between Chinese L2 and New Zealand L1 postgraduates, and between masters and PhDs. Broader comparisons can be drawn in the future studies with ready-made or self-built corpora comprising other genres, in diverse disciplines, or across different academic levels. In this study, possible reasons for typical bundle choices of Chinese postgraduates were explored solely from the interview data. In order to obtain more convincing evidence, there is a need to build comparative corpora of L1 and L2 writing, in other words, to integrate the CIA (contrastive interlanguage analysis) and CA (contrastive analysis) approaches. As Granger (1996) argues that "CA data helps analysts to formulate predictions about interlanguage which can be checked against CIA data . . . . Conversely, CIA results can only be reliably interpreted as being evidence of transfer if supported by clear CA descriptions" (p. 46).

Besides comparative corpora of L1 and L2 (or IL) writing, a set of learner corpora of different L1 but the same IL could also be compiled considering Jarvis's (2000) framework of transfer studies. This suggested process involves comparisons between learner corpora of the same L1 and IL to identify the common IL features (type 1), between learner corpora of different L1 but the same IL to exclude developmental and universal factors (type 2), and between L1 and IL corpora of learners to determine effects of L1 influence (type 3):

1. intra-L1-group homogeneity in learners' IL performance,
2. inter-L1-group heterogeneity in learners' IL performance, and
3. intra-L1-group congruity between learners' L1 and IL performance. (Jarvis, 2000, pp. 253-255)



---

In the same vein, the study of interlanguage development would greatly benefit from longitudinal corpus study of learner language, ideally the language data produced by the same groups of learners over time (Paquot, 2012).

The present study investigates four-word sentence initial bundles. Bundles of various lengths deserve equal attention considering the crucial roles of these recurrent multiword combinations in facilitating learner language production. Bundles can be examined in relation to their positions within sentences (e.g. sentence non-initial bundles) or even paragraphs (e.g. paragraph initial bundles) (P. Baker, personal communication, October 25, 2014). Bundles can also be analysed in regard to moves in writing to show features of various types of texts and to provide more specific language resources for learners (e.g. Cortes, 2013).

Future research can gain better insights into the development and sources of learner language with the interview data collected in the same or similar contexts of the corpus data, or even from the same participants. Researchers can also choose to conduct longitudinal case studies, tracking L2 learners' acquisition of lexical bundles in ESL/EFL or English-speaking contexts. In this way, a richer picture will be created documenting the development of learners' lexical repertoire and the sources of their acquisition. A pioneering study in this area is Li and Schmitt (2009), which explores a Chinese masters student's improvement in the area of lexical phrases over an academic year and her self-reported explicit and implicit sources for this improvement. Another approach is to gather supervisors' and examiners' evaluations on students' bundle selections to investigate the correlations between bundle selections and target audiences or quality of writing. Interview data could also be collected to verify the claims of New Zealand native writers' bundle selections.

### **9.3 Implications**

Despite the limitations, this study has important implications for future research and pedagogy. I will present these implications in terms of theory, methodology and particularly pedagogy with reference to previous arguments and current development of corpus-based tools (e.g. FLAX).

### 9.3.1 Theoretical implications

The existing metadiscourse models are mostly developed through top-down approaches, with pre-determined metadiscourse devices, largely individual words. This study takes a bottom-up corpus-based approach and extends metadiscourse analysis to the bundle-based four-word units, as presented in Table 91. 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 bundle-driven metadiscourse categorisation confirms most of the categories of Hyland's (2005a, 2005c) metadiscourse model and develops the model by adding another two categories, namely condition bundles and introduction bundles. 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 a current understanding of metadiscourse devices and functions.

**Table 91. 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 text	<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 /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>

---

### 9.3.2 Methodological implications

This research has three implications for methodology in that it distinguishes between sentence initial and non-initial bundles, combines lexical bundle research (i.e. bottom-up approach) with metadiscourse analysis (i.e. top-down approach), and supplements corpus-based analysis with interviews.

The study distinguishes sentence initial and non-initial bundles because these two types of bundles pose different challenges to learners and perform different functions in sentences. Sentence initial bundles are considered more challenging for learners. They need to consider at least three factors — reader expectation, sequence of information, and cohesion and coherence — while starting a sentence. The study of these bundles can provide learners with a range of sentence starters, better inform learners about the various functions these bundles perform, and raise learners' awareness of crucial factors they need to consider while writing a sentence.

Very recently Granger (2014) called for the combination of lexical bundle research and metadiscourse analysis:

Languages have been shown to differ markedly in their use of metadiscourse. However, hardly any studies rely on lexical bundles and use a truly corpus-driven methodology. This is a pity as lexical bundles are an efficient way of accessing the longer stretches of discourse which are often used to express metadiscourse and have so far been largely neglected. (Granger, 2014, p. 59)

This is consistent with the approach of the current research, which has worked to fill this gap by exploring the possibility of combining these two approaches theoretically and empirically. The findings of this study have shown that this combination is an effective and productive way to investigate written discourse and to provide directly applicable resources for writing pedagogy. Lexical bundles, as the units of analysis, can stand alone somewhat from their contexts, which allows for comparatively easier identification of functions. The use of the lexical bundle approach, in other words, a bottom-up approach, also verifies and expands the existing knowledge of metadiscourse, as discussed above. On the other hand, the use of the metadiscourse model extends the application of bundle analysis in

---

pedagogy, which allows learners to access lexical bundles as devices in interpersonal communication to manage information flow or to mediate writer-reader interaction.

The use of interviews in this study has informed the interpretation of the corpus data from writer's perspective. Corpus data does not explain itself. Corpus linguists can postulate reasons and make hypotheses, but evidence of their interpretations can only be collected using other methods, such as the interviews used in this study, and those suggested in Section 9.2.2.

### **9.3.3 Pedagogical implications**

A small number of studies have focused on the teaching of lexical bundles in academic writing (e.g. Cortes, 2006; Eriksson, 2012; Jones & Haywood, 2004). Jones and Haywood (2004) selected a list of target bundles on the basis of Biber and his colleagues' (1999) academic bundle lists. They then asked their non-native students in a pre-sessional EAP (English for Academic Purposes) course to analyse the grammatical structures and discourse functions of these bundles during reading, and to use these bundles in writing. Cortes (2006) introduced a group of target bundles identified in a history journal article corpus to a class of history majors, who were native speakers of English, and who attended a writing-intensive history course. She used different types of bundle exercises (e.g. filling in the blanks, multiple choice, inappropriate use correction) to enhance student learning. Eriksson (2012) taught bundles to non-native doctoral students of biochemistry and biotechnology. He based his bundle selection on two self-compiled corpora: a journal article corpus in the same fields and a corpus of participating doctoral students' writing. He also drew from three published bundle lists: Hyland's (2008b) two bundle lists in engineering and technology, and Simpson-Vlach and Ellis's (2010) Academic Formulas List. These were used to identify the underused but important bundles in the PhD student writing. Learning activities were also designed to help these students to understand the functions of bundles and encourage them to employ bundles in their writing.

No significant improvement in bundle production has been found in these studies, although the students' awareness of bundles has been raised. Various issues such

---

as time constraints, teacher support and activity design were considered as influencing factors.

The current bundle analysis suggests that teaching lexical bundles should not be confined to a list of fixed multiword chunks (mostly 3-5 words), either retrieved from a related corpus (or two comparative corpora), or selected from previously-generated bundle lists. Instead, both the implications of bundle studies and a variety of language resources based on the same frameworks should be part of the pedagogy. In this section, I will explain these two points in regard to the findings of the current research, which include teaching and learning recommendations, and the use of corpus-based tools for bundle teaching and learning.

#### *9.3.3.1 Teaching and learning recommendations*

Language is formulaic to a great degree. As formulaic multiword combinations, lexical bundles can act as points of fixation to facilitate writing construction, and therefore deserve special attention in academic writing pedagogy. With reference to the discrepancies between Chinese and New Zealand student bundle production and the reasons reported by the Chinese postgraduates, a range of teaching and learning approaches are suggested to address the following recommendations:

1. Equip Chinese students with bundles used by advanced native writers,
2. Emphasise bundle noticing in academic reading and writing,
3. Increase Chinese students' confidence as student writers,
4. Familiarise Chinese students with rhetorical conventions, and
5. Expand Chinese students' word knowledge of multiword combinations.

Most of these recommendations refer to Chinese students because they are the subjects of this study; however, these recommendations may also apply to other L2 learner groups. FLAX, as a corpus-based language learning tool, will be used as an example to illustrate the potential of corpus-based tools because of my familiarity with it. Alongside FLAX, many corpus-based tools are available in the market: some are free resources for learners (e.g. AntConc, Compleat Lexical Tutor, BYU-BNC, COCA, WebCorp and SKELL) and some are commercial tools (e.g. WordSmith Tools, Collocate and Sketch Engine). These tools (including FLAX)

---

support similar search functions, and have their own strengths and limitations. Learners can choose to use any of them according to their needs and preferences.

**Recommendation 1: Equip Chinese students with bundles used by advanced native writers**

The bundles used by advanced native writers represent both good practices of academic writing (e.g. the use of shell noun bundles) and common practices of target academic communities (e.g. the use of personal feeling bundles). The use of these bundles can be arguably labelled as native norms. The different bundles (e.g. *It is obvious that* and *It is clear that*) and the different use of the same bundles (e.g. the different positions of *in order to* bundles) identified in this study are likely to increase Chinese students' knowledge of good writing practices and raise Chinese students' awareness of the differences between their practices in the academic discourse community of mainland China and the practices of other communities, such as those of New Zealand universities. Therefore, these bundles should be incorporated into writing pedagogy no matter whether this is in an EFL (e.g. mainland China) or an ESL (e.g. New Zealand) context of teaching.

With the bundle knowledge, Chinese students can adopt the bundles of advanced native writers to write more effectively in English. The present study has not only generated function-based bundle lists used by New Zealand postgraduates, but has also highlighted unique bundles in New Zealand postgraduate writing for quick reference. These lists, together with the typical bundles highlighted in the current research, can be used to compile teaching resources, such as academic or thesis writing handbooks for advanced Chinese learners. These lists can also be introduced to students in class under the topics of cohesion and coherence, exemplification and reformulation, modification and certainty, stance and engagement, writer identity, etc. Therefore, students not only supported to understand the requirements and conventions of academic writing, but also have the bundles at hand to meet these requirements and follow the conventions. EAP teachers can also use search tools (e.g. FLAX) to present students with relevant and accessible bundle resources for any specific writing tasks. The task-based bundle lists complement and are more relevant than the existing general lists of formulaic

---

sequences such as the lexical bundle lists extracted from various spoken and written registers (e.g. textbooks and class sessions) by Biber and his colleagues (Biber & Barbieri, 2007; Biber et al., 2003, 2004; Biber et al., 1999), and Simpson and Ellis's (2010) Academic Formulas List (including the Core, Spoken and Written sublists) generated from different genres. At the same time, teacher self-generated bundle lists effectively address some of the challenges discussed in Byrd and Coxhead (2010), which are little knowledge of generating process of bundle lists published in research reports, difficulty in choosing the length of lexical bundles to teach, and lack of information on use in context of bundles in published lists. Teachers will understand the generating process better and can self-manipulate the length of bundles involved in the generation process. In addition, teacher-created bundle lists allow students to access the context in which bundles are used, and to learn when and how to use these bundles.

With awareness of the differences, Chinese students are more likely to perceive writing as a community-based practice, in which different target audiences require different communication approaches. The typical bundles used in their theses may not hinder the effectiveness of communication between themselves as writers and their supervisors, examiners or other readers coming from the same discourse community in China. The completion and publication of these theses have already proved this. These bundles, however, could possibly cause confusion for their wider audience from other communities and limit future publication possibilities, so they need to become aware of their bundle selections and choose culturally appropriate ones, if they intend to reach a wider or different audience. The present study uncovers the divergent use of sentence initial bundles between Chinese L2 and New Zealand L1 postgraduates in their thesis writing. More discrepancies can be revealed through manipulating variables, such as types of bundles (e.g. sentence non-initial bundles, paragraph-initial bundles), cultural backgrounds of writers (e.g. French learners, British learners), levels of proficiency (e.g. secondary school level, undergraduate level), and genres (e.g. narratives, research reports). Comparative corpora can therefore be built to fulfil different pedagogical purposes.

**Recommendation 2: Emphasise bundle noticing in academic reading and writing**

---

Nation (2013) outlines three cognitive processes for vocabulary learning: noticing, retrieval and creative use, which could or should be transferrable to bundle learning. Noticing, known as consciousness in Schmidt (1990), refers to seeing a bundle as unfamiliar and attending to it. Noticing is a determining factor in bundle learning. However, simple exposure to lexical bundles does not guarantee bundle noticing. Students were found to habitually pay attention to their familiar bundles and ignore the unfamiliar ones. In order to direct students' attention to their unfamiliar bundles, it is necessary to enhance the input (Sharwood Smith, 1991, 1993) of these bundles in academic reading and writing. During the reading process, EAP Teachers can ask students to highlight the bundles within texts, negotiate the appropriateness of the bundles (e.g. position of transition bundles), explain the meanings of the bundles, or classify the bundles into different function categories. The bundle search functions of some corpus-based tools (e.g. FLAX) allow students to view typographically-highlighted bundles and to access bundles within their function-based categories. Teachers can build reading materials into a corpus-based tool, so that language chunks, such as bundles, will be perceptually salient and can be easily identified while reading. During the writing process, EAP teachers can use discourse focused techniques like reformulation (Cohen, 1983) to rewrite students' sentences, preserving their ideas but replacing the inappropriate sentence starters with sentence initial bundles for example. Students' noticing of bundles can be enhanced by comparing the differences between their original writing with the reformulated one.

### **Recommendation 3: Increase Chinese students' confidence as student writers**

Chinese students, as non-native writers, are often conservative and avoid adopting unfamiliar bundles in their writing to minimize the risk of making errors. If the risk can be reduced, students should become more confident to try unfamiliar bundles, so that these bundles can be gradually acquired. EAP teachers and textbook writers could provide students with a set of target bundles categorised into different metadiscourse functions, or teachers can require students to collect useful bundles before they start writing. With the support, students can expand their writing from these "islands of reliability" (Dechert, 1984). If possible, the corpus-based bundle learning approach can be applied during bundle production, which is in line with



Nation’s (2013) retrieval and creative use theories, and aligns with Wu, Franken and Witten’s (2010) argument on collocation learning. In the context of this research, retrieval refers to the recall process of any previously met bundle, which will be enhanced when learners negotiate the use of an unfamiliar bundle (e.g. *from the perspective of*) through searching its content word (e.g. *perspective*), structure (e.g. *preposition + perspective*) and multiple contexts, as illustrated in Figure 5. Creative use means the use of a previously met bundle in another context, which can also be enriched by the application of corpus-based tools. Creative use is achieved when learners negotiate the appropriateness of an unfamiliar bundle through a range of contexts and incorporate the target bundle in their writing. Creative use is regarded as the most effective process in retention of vocabulary knowledge (Nation, 2013) including bundle knowledge, as it is only when students feel confident to take the risk and deploy the target bundle in their productive language, that they can learn the bundle.

verb + preposition + perspective	told from the perspective of	190	put into perspective	13
	written from the perspective of	187	played from the perspective of	12
	sung from the perspective of	92	depending on perspective	10
	seen from the perspective of	44	considered from the perspective of	9
	viewed from the perspective of	32	narrated from the perspective of	7
	viewed from the perspective of	32	>>> more	
noun + preposition + perspective	<ul style="list-style-type: none"> <li>• (Biographical book of Arien and his wife <b>viewed from the perspective of</b> Arien's son.</li> <li>• A critical account of Popper's thought, <b>viewed from the perspective of</b> contemporary analytic philosophy.</li> <li>• When <b>viewed from the perspective of</b> the planet the quasi-satellite will appear to travel in an oblong retrograde loop around the planet.</li> <li>• Generally called "sci-fi noir," it is "a view of what the 1980s might have looked at <b>viewed from the perspective of</b> a 1940s filmmaker."</li> <li>• S.J.A. Turney's 2009 debut, <i>Marius' Mules</i>, based on Julius Caesar's invasion of Gaul, <b>viewed from the perspective of</b> a Legionary commander.</li> <li>• Stressing the animals is not desirable, but the concept of and measurement of stress must be <b>viewed from the perspective of</b> the animal using the scientific method.</li> </ul>			20
	<ul style="list-style-type: none"> <li>• Blood-Dark Track: A Family History (Granta Books) by Joseph O'Neill, contains a detailed and evocative history of the city, <b>viewed from the perspective of</b> a Christian Syrian family long resident in Mersin.</li> <li>• Warcraft III, like Blizzard's previous title StarCraft, allows for single and multiplayer replays to be recorded and viewed, allowing a game to be played at slower and faster speeds and <b>viewed from the perspective of</b> all players.</li> <li>• The term "perfective aspect" is sometimes used to refer to the perfect, which describes a state resulting from a previous action (also described as a previous action with relevance to a particular time, or a previous action <b>viewed from the perspective of</b> a later time).</li> <li>• Spanish and Ancient Greek, for example, have a perfect (not the same as the perfective), which refers to a state resulting from a previous action (also described as a previous action with relevance to a particular time, or a previous action <b>viewed from the perspective of</b> a later time).</li> </ul>			19
noun + of + perspective	<ul style="list-style-type: none"> <li>• Blood-Dark Track: A Family History (Granta Books) by Joseph O'Neill, contains a detailed and evocative history of the city, <b>viewed from the perspective of</b> a Christian Syrian family long resident in Mersin.</li> <li>• Warcraft III, like Blizzard's previous title StarCraft, allows for single and multiplayer replays to be recorded and viewed, allowing a game to be played at slower and faster speeds and <b>viewed from the perspective of</b> all players.</li> <li>• The term "perfective aspect" is sometimes used to refer to the perfect, which describes a state resulting from a previous action (also described as a previous action with relevance to a particular time, or a previous action <b>viewed from the perspective of</b> a later time).</li> <li>• Spanish and Ancient Greek, for example, have a perfect (not the same as the perfective), which refers to a state resulting from a previous action (also described as a previous action with relevance to a particular time, or a previous action <b>viewed from the perspective of</b> a later time).</li> </ul>			14
				13
verb + perspective	viewed from the perspective of	1	represent the perspective of	4
	viewed from the Greek-Arcadian perspective	1	reflects the perspective of	4
	changed her perspective	6	having the perspective of	4
	control perspective	4	uses the perspective of	4
	>>> more			
perspective + to + verb	perspective to create	11	third-person perspective to interact with	4
	perspective to enhance	5	perspective to depict	2
	perspective to give	4	perspective to more accurately represent	2
	perspective to make	4	perspective to properly judge	2
	perspective to understand	4	perspective to mediate	2
	>>> more			
adjective + preposition + perspective	important from the perspective	5	playable from the perspective	3
	subject to perspective distortions	3	wrong from an intelligence perspective	3
	due to perspective	3	robust to perspective transformations	3

Figure 5. Search for *perspective* bundle in FLAX

---

Chinese students, in comparison with New Zealand students, feel less comfortable and confident to cognitively involve their readers and to establish their individual identity as emerging researchers. Supervisors can raise students' awareness of their current apprentice identity and encourage them to engage in "legitimate peripheral participation" (Flowerdew, 2000, p. 131), so that students can have their voices heard by means of communicating with the authority and establishing their own identity. The parallel New Zealand thesis corpora can be used here to show students the practices of New Zealand L1 writers (e.g. the use of *note* bundles and *I* bundles) to increase their confidence in writer-reader interaction.

#### **Recommendation 4: Familiarise Chinese students with rhetorical conventions**

The absence of personal feeling bundles, hedge bundles and *I* bundles are also due to Chinese students' misunderstanding of rhetorical conventions. The current research reveals these conventions and highlights Chinese students' misinterpretations in terms of sentence initial bundles. Teachers can directly explain these conventions to students. Or, if applicable, they could invite university lecturers of different subject areas to discuss the conventions (Coxhead, 2012). With this study, it is easy to focus on the key points. Teachers can illustrate these conventions with the bundles and context sentences from a relevant corpus. The reason for using bundles is these strings always occur with high frequencies, which represent the common practices in a certain discourse. At the same time, bundles can serve as useful resources adopted to follow these conventions.

#### **Recommendation 5: Expand Chinese students' word knowledge of multiword combinations**

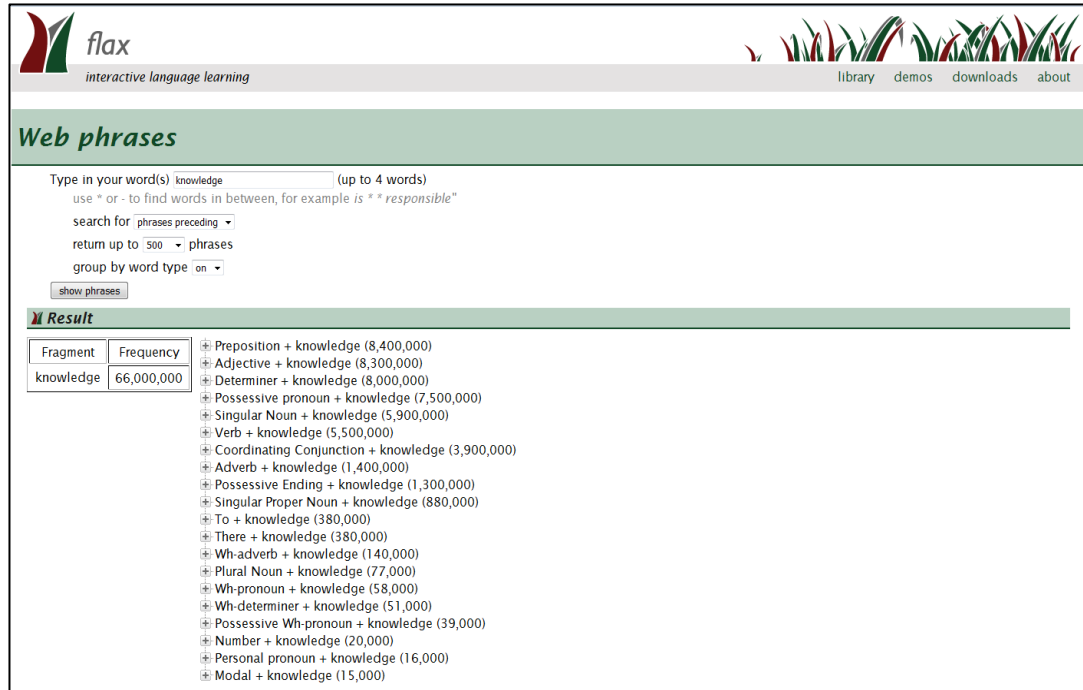
With regard to word knowledge, this study reveals Chinese students' limited knowledge of multiword combinations such as collocations (e.g. incorrectly collocating *necessary* with *note*) and lexico-grammatical patterns (e.g. the lack of *interesting* bundle knowledge), and their little knowledge of registers (e.g. the preference for less formal *seem* bundles). Coxhead (2012) suggests in-class discussion with L2 writers appropriate word use and register. Traditional resources (e.g. textbooks and dictionaries) are familiar to teachers and students, which can be referred to during writing, although they are often limited in the size of multiword

combinations presented. Corpus-based learning approaches greatly exceed traditional resources in the number of multiword combinations provided and the embedded search functions. For example, *Learning Collocations* and *Web Phrases* collections in FLAX automatically extract collocations from build-in corpora (e.g. BNC, BAWE and Wikipedia) and the Web (Figures 6 & 7). Target collocations are structurally grouped according to their frequency of occurrence, and typographically highlighted within their original contexts, and richly linked with related collocations, topics, definitions based on internal as well as external sources. With the help of corpus-based tools, students are more likely to find their unfamiliar expressions and can learn to use these expressions in a given register. The British Academic Written English Corpus (BAWE) as an example, containing 2860 highly graded student assignments (6M words) (Nesi & Gardner, 2012), supports the learning of lexico-grammatical patterns. Details will be presented in the next section.

The screenshot shows the FLAX Learning Collocations interface. At the top, there is a search bar with 'knowledge' entered and a dropdown menu for 'contemporary English (Wikipedia)'. Below the search bar, there are family words: 'know', 'knew', 'knowing', 'known', 'knows', 'unknown', 'knowledgeable'. The main content area is titled 'used as a noun' and displays a table of collocations grouped by structure.

Structure	Collocation	Frequency	Collocation	Frequency
adjective + knowledge	scientific knowledge	951	common knowledge	465
	extensive knowledge	599	human knowledge	418
	public knowledge	517	technical knowledge	332
	prior knowledge	513	intimate knowledge	331
	new knowledge	477	general knowledge	325
noun + of + knowledge	lack of knowledge	538	knowledge of the world	125
	body of knowledge	302	level of knowledge	121
	source of knowledge	168	dissemination of knowledge	120
	theory of knowledge	155	knowledge of the language	119
	state of knowledge	126	pursuit of knowledge	105
knowledge + noun	knowledge base	564	knowledge sharing	116
	knowledge management	324	knowledge transfer	114
	knowledge representation	147	knowledge workers	105
	background knowledge	132	advance knowledge	101
	knowledge economy	129	prior knowledge	90
verb + knowledge	have knowledge of	212	requires knowledge of	76
	share knowledge	101	denied knowledge of	70
	working knowledge of	91	gain knowledge of	59

**Figure 6. Search of *knowledge* in *Learning Collocations* collection in FLAX**



The screenshot shows the FLAX web interface. At the top, there is a logo for 'flax interactive language learning' and navigation links for 'library', 'demos', 'downloads', and 'about'. The main heading is 'Web phrases'. Below this, there is a search input field containing the word 'knowledge' and a note '(up to 4 words)'. Below the search field, there are options for 'search for' (phrases preceding), 'return up to' (500), 'group by word type' (on), and a 'show phrases' button. The search results are displayed under a 'Result' heading in a table with two columns: 'Fragment' and 'Frequency'. The first row shows 'knowledge' with a frequency of 66,000,000. To the right of the table, there is a list of grammatical categories and their frequencies, each with a plus sign icon.

Fragment	Frequency
knowledge	66,000,000

- + Preposition + knowledge (8,400,000)
- + Adjective + knowledge (8,300,000)
- + Determiner + knowledge (8,000,000)
- + Possessive pronoun + knowledge (7,500,000)
- + Singular Noun + knowledge (5,900,000)
- + Verb + knowledge (5,500,000)
- + Coordinating Conjunction + knowledge (3,900,000)
- + Adverb + knowledge (1,400,000)
- + Possessive Ending + knowledge (1,300,000)
- + Singular Proper Noun + knowledge (880,000)
- + To + knowledge (380,000)
- + There + knowledge (380,000)
- + Wh-adverb + knowledge (140,000)
- + Plural Noun + knowledge (77,000)
- + Wh-pronoun + knowledge (58,000)
- + Wh-determiner + knowledge (51,000)
- + Possessive Wh-pronoun + knowledge (39,000)
- + Number + knowledge (20,000)
- + Personal pronoun + knowledge (16,000)
- + Modal + knowledge (15,000)

**Figure 7. Search of *knowledge* in *Web Phrases* collection in FLAX**

### 9.3.3.2 Use of corpus-based tools for bundle teaching and learning

The use of corpus-based tools refers to the hands-on corpus searches for language learning, that is, data-driven learning (DDL), a term coined by Johns (1991), indicating the idea of students as language researchers. As for sentence initial bundle learning, students can combine built-in bundle lists and hands-on corpus searches, supported by corpus-based tools, for example, FLAX (Wu, 2010). The bundle lists used in this study can be viewed within FLAX, as shown in Figure 8. With the help of FLAX, students can access multiple contexts in which these bundles appear by clicking on them (Figure 9). The application of FLAX affords L2 students a certain amount of sentence initial bundles of thesis writing with frequency-based displays, multiple contexts and typographical salience (Franken, 2012). The access to the corpus-based language learning tool, FLAX or tools like it, allows students to act as language researchers to learn to interact with the corpora, to explore the metafunctions of sentence initial bundles and to choose appropriate ones for writing.

The screenshot shows the 'flax interactive language learning' website. The header includes navigation links: library, demos, downloads, about, and Login. The main heading is 'NZ PhD theses'. Below this, there are tabs for 'About Collection', 'Search', 'Browse by Title', 'Collocations', 'Wordlist', and 'LexicalBundles'. A 'My Cherry Basket' icon is also present. The 'LexicalBundles' tab is active, displaying a list of sentence initial bundles with their respective counts:

- At the beginning | In the middle
- On the other hand, (108)
- It is possible that (75)
- In the case of (68)
- At the same time, (68)
- It is important to (63)
- As discussed in Chapter (63)
- At the end of (56)
- In addition to the (54)
- The results of the (44)
- In other words, the (40)
- It should be noted (40)
- As can be seen (36)
- All of the participants (33)
- It is interesting to (32)
- In this chapter I (31)
- For the purposes of (29)
- There are a number (29)
- In the context of (28)
- As a result of (28)

**Figure 8. Sentence initial bundles in the New Zealand PhD thesis corpus**

The screenshot shows the same website as Figure 8, but with the 'LexicalBundles' tab selected. The bundle 'It is important to (63)' is highlighted. Below the list, several context sentences are provided, each starting with 'It is important to' followed by a blue link to the bundle name:

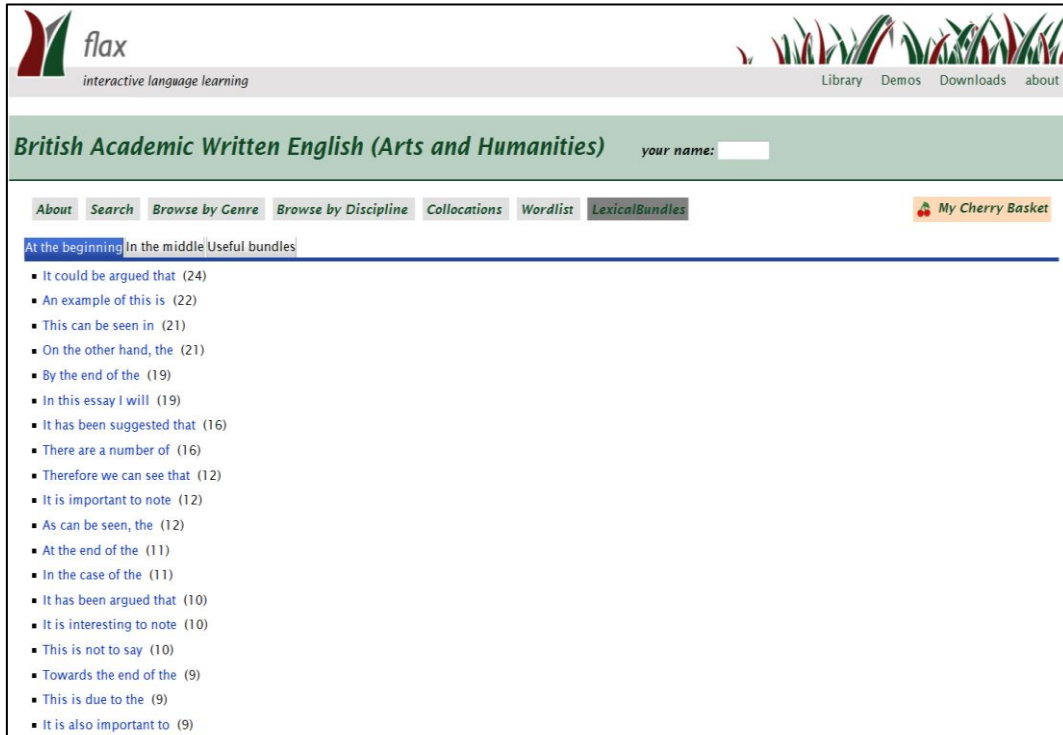
- It is important to** note that while I have divided the analysis description into the various processes used for each research question, the processes were recursive and mutually informative across research questions.
- It is important to** point out that there is a difference between the percentage of copied text as identified by the textual analysis2 and similarity index produced by turnitin.com.
- It is important to** note, however, that the chapters which discuss the concepts of family, literacy, family literacy and wellbeing simply point to the epistemological underpinnings and consistency in the perspectives argued for.
- It is important to** remember that Pacific people came to New Zealand in the hope of a better quality of life (Meleisea & Schoeffel, 1998).
- It is important to** note, though, that for Maori, family and community is often the same (Robinson & Williams, 2001) and that for Pacific peoples living in New Zealand, community life is often based around their churches (Meleisea & Schoeffel, 1998), thus there are culturally-preferred starting points for community.
- It is important to** criticalists that the voices of people who fare less well in the world are heard, so attention is paid by critical researchers to seeking out these voices and creating space for them, and most importantly, to ensuring that they are represented authentically.
- It is important to** view language policy as not only the explicit, written, overt, de jure, official, and "top-down" decision-making about language, but also the implicit unwritten, covert, de facto, grass-roots, and unofficial ideas and assumptions, which can influence the outcomes of policy-making just as emphatically and definitively as the more explicit decisions (Schiffman,
- It is important to** note that Shohamy has primarily used her framework to explore hidden political agendas and the way in which governments and

**Figure 9. Context sentences of the bundles *It is important to***

This combination of bundle lists and corpus searches has also been applied to other build-in corpora of FLAX, such as the BAWE collection<sup>4</sup>, a collection of high-quality student assignments of British universities, to satisfy different writing needs and to cater for a wider variety of writers. Figure 10 is the list of sentence initial bundles in Arts and Humanities collection of BAWE, including *It could be argued that*, *An example of this is*, *This can be seen in*, *On the other hand*, *the* and *By the end of the*. Each bundle contains five-words, another common length for lexical bundles. (The length of bundles can be manipulated within FLAX.) As shown in Figure 11, the sentence initial bundles in this collection have been manually categorised according to their metadiscourse functions to reduce students' search load. The terminology in metadiscourse models such as Hyland's (2005a, 2005c) model is developed for discourse analysis and some terms (e.g. *directives*, *endophoric markers*, *hedges* and *boosters*) are too complex for teachers and students to understand. Therefore, I renamed those categories with plain language, for example, *to instruct readers*, *to refer to information in other parts of text*, and *to express certainty or uncertainty*. The combination of bundle lists and corpus searches efficiently transfers the results of corpus-based analysis into pedagogy. At the same time, it greatly decreases students' search load and students no longer need to interpret a considerable number of concordances.

---

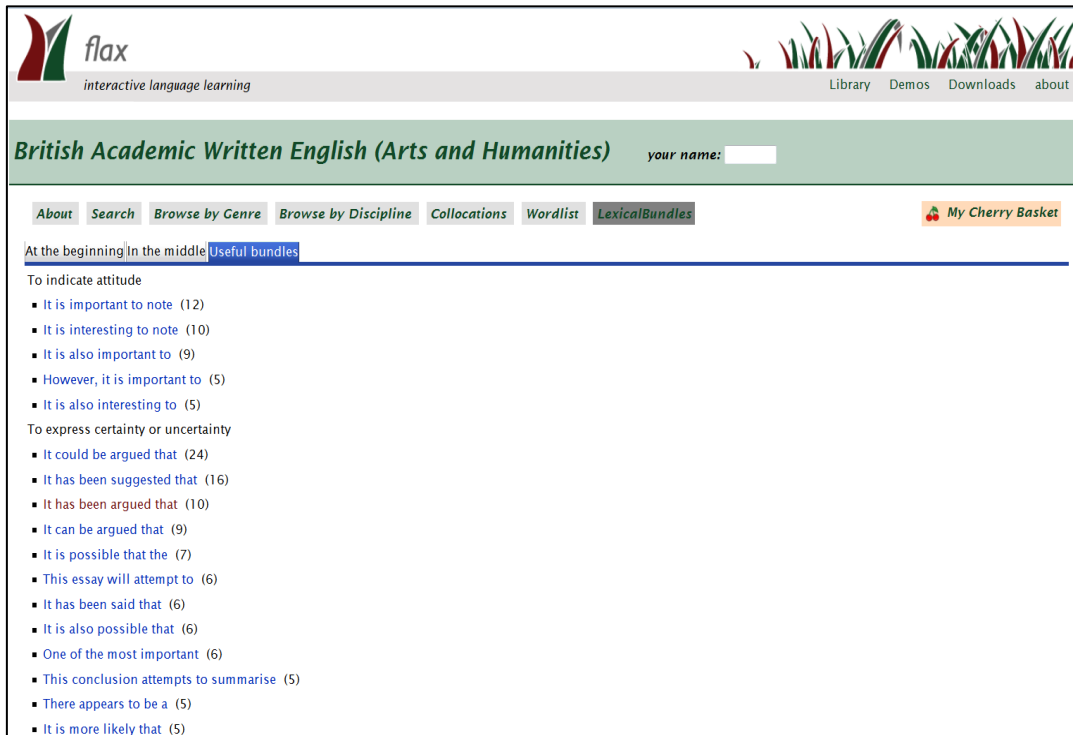
<sup>4</sup> The examples in this part come from the British Academic Written English (BAWE) corpus, which was developed at the Universities of Warwick, Reading and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics [previously called CELTE], Warwick), Paul Thompson (formerly of the Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800).



The screenshot shows the 'flax interactive language learning' website. The main heading is 'British Academic Written English (Arts and Humanities)' with a 'your name:' input field. A navigation bar includes 'About', 'Search', 'Browse by Genre', 'Browse by Discipline', 'Collocations', 'Wordlist', and 'LexicalBundles'. A 'My Cherry Basket' icon is also present. The 'LexicalBundles' section is active, displaying a list of sentence initial bundles under the heading 'At the beginning | In the middle | Useful bundles'.

- It could be argued that (24)
- An example of this is (22)
- This can be seen in (21)
- On the other hand, the (21)
- By the end of the (19)
- In this essay I will (19)
- It has been suggested that (16)
- There are a number of (16)
- Therefore we can see that (12)
- It is important to note (12)
- As can be seen, the (12)
- At the end of the (11)
- In the case of the (11)
- It has been argued that (10)
- It is interesting to note (10)
- This is not to say (10)
- Towards the end of the (9)
- This is due to the (9)
- It is also important to (9)

**Figure 10. Sentence initial bundles in BAWE**



The screenshot shows the 'flax interactive language learning' website. The main heading is 'British Academic Written English (Arts and Humanities)' with a 'your name:' input field. A navigation bar includes 'About', 'Search', 'Browse by Genre', 'Browse by Discipline', 'Collocations', 'Wordlist', and 'LexicalBundles'. A 'My Cherry Basket' icon is also present. The 'LexicalBundles' section is active, displaying a list of function-based sentence initial bundles under the heading 'At the beginning | In the middle | Useful bundles'.

To indicate attitude

- It is important to note (12)
- It is interesting to note (10)
- It is also important to (9)
- However, it is important to (5)
- It is also interesting to (5)

To express certainty or uncertainty

- It could be argued that (24)
- It has been suggested that (16)
- It has been argued that (10)
- It can be argued that (9)
- It is possible that the (7)
- This essay will attempt to (6)
- It has been said that (6)
- It is also possible that (6)
- One of the most important (6)
- This conclusion attempts to summarise (5)
- There appears to be a (5)
- It is more likely that (5)

**Figure 11. Function-based sentence initial bundle list in BAWE**

Another function of FLAX, the *Search for sentences* function with its *Group by pattern* option, can be used to extend students' bundle knowledge. As illustrated in Figure 12, all the sentences containing *important* in Arts and Humanities collection are grouped into two columns by word position — near the beginning (321 sentences) or in the middle (285 sentences). Each line represents a general pattern, but rather than using grammatical terminology a single concrete example is shown. The top pattern *It is important to + verb* has also been generated as a useful bundle in this study. Figure 13 displays the sentences with different verbs (e.g. *note*, *consider*, *remember*, *examine* and *recognise*) deployed to fill in the *verb* slot. Unlike bundle lists, this function provides students with a variety of language patterns. Therefore, teachers' concerns about students' repetitive use of a limited number of bundles can be addressed.

The screenshot shows the FLAX website interface. At the top, there is a logo for 'flax interactive language learning' and navigation links for 'Library', 'Demos', 'Downloads', and 'about'. Below this is a header for 'British Academic Written English (Arts and Humanities)' with a 'your name:' field. A navigation menu includes 'About', 'Search', 'Browse by Genre', 'Browse by Discipline', 'Collocations', 'Wordlist', 'LexicalBundles', and 'My Cherry Basket'. The main content area is titled 'Search Words in Collection' and shows a search for 'sentences' that contain the word 'important'. Below the search bar, it says 'You may be interested in importance, importantly, unimportant'. The search results are displayed as 'Search Result: 606 patterns found'. Under 'Group by pattern', there are two columns: 'At the beginning (321)' and 'In the middle (285)'. The 'At the beginning' column lists the following patterns:

- It is important to note (81)
- The important aspects of (16)
- Bodily positioning is an important part of (10)
- Advertising played an important role in (9)
- It is important for (9)
- Therefore the careful consideration are necessarily important in (8)
- Another important aspect to consider (7)
- Important factors in (7)
- Slautati was important in (6)
- It is important that (6)
- The most important opponent (5)
- Another important factor that (4)

**Figure 12. Sentences containing *important* at the beginning, grouped by pattern**



The screenshot shows the 'flax' website interface. At the top, there is a navigation bar with 'Library', 'Demos', 'Downloads', and 'about'. Below this is a header for 'British Academic Written English (Arts and Humanities)' with a 'your name:' field. A secondary navigation bar includes 'About', 'Search', 'Browse by Genre', 'Browse by Discipline', 'Collocations', 'Wordlist', 'LexicalBundles', and 'My Cherry Basket'. The main content area is titled 'Search Words in Collection' and shows a search for 'sentences' containing the word 'important'. Below the search bar, it indicates 'Search Result: 606 patterns found'. The results are grouped by pattern, with 'At the beginning (321)' selected. A list of sentences is displayed, each starting with 'It is important to' followed by a verb and a count in parentheses. The sentences are:
 

- It is important to note (81)
- It is important to note that Allende gives Esteban the most intimate narrative role in the book. (21)
- It is important to consider the 'delayed effect' between input and output. (21)
- It is important to remember that Alba is not the only narrator in the novel. (8)
- It is important to examine why this was the case. (3)
- Therefore, it is important to look at who is addressing the child when considering the effects of CDS. 5.2. (3)
- It is important to recognise that in many ways they were quite similar to modern humans. (3)
- It is important to analyse why it was so appealing to this sector of society. (2)
- It is important to recognize that these are not temporally distinct moments. (2)
- It is also important to take into account not only whether Ranke's method was scientific but whether this term could also be applied to his historical outlook in general. (2)
- It is important to notice, what materials is the object made of and what is the workload involved in the production of such saddle and if it was the same or different in the past. (2)

**Figure 13. Sentences with the same pattern *It is important to + verb***

## 9.4 Concluding remarks

My PhD journey into lexical bundles has led me to explore the exciting fields of discourse analysis, corpus linguistics and data-driven learning, to join the fascinating conversations with scholars and fellow students (particularly corpus linguists), and to experiment with various corpus-based tools and corpus-based approaches. I have sought to unify knowledge of metadiscourse into corpus-based analysis, to combine text analysis with interviews, and to integrate the outcomes of corpus linguistics into writing pedagogy. The potential of information and computing technology (ICT) in language study and language learning has not yet been fully recognised, realised and exploited, and many valuable and interesting topics are waiting for teachers and researchers to explore.

---

## References

- Abdi, R. (2002). Interpersonal metadiscourse: An indicator of interaction and identify. *Discourse Studies*, 4(2), 139-145.
- Ackermann, K., & Chen, Y.-H. (2013). Developing the Academic Collocation List (ACL) — A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235-247.  
doi:10.1016/j.jeap.2013.08.002
- Ädel, A. (2010). *Just to give you kind of a map of where we are going: A taxonomy of metadiscourse in spoken and written academic English.* *Nordic Journal of English Studies*, 9(2), 69-97.
- 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
- Allen, D. (2009). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education*, 1, 105-127.
- Altenberg, B. (1993). Recurrent word combinations in spoken English. In J. M. D'Arcy (Ed.), *Proceedings of the Fifth Nordic Association for English Studies Conference* (pp. 17-27). Reykjavik, Iceland: University of Iceland.
- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 101-122). Oxford, United Kingdom: Oxford University Press.
- Baker, P. (2006). *Using corpora in discourse analysis.* London, United Kingdom: Continuum.

- 
- Baker, P. (2010). *Sociolinguistics and corpus linguistics*. Edinburgh, United Kingdom: Edinburgh University Press.
- Benson, M. (1990). Collocations and general-purpose dictionaries. *International Journal of Lexicography*, 3(1), 23-34.
- Benson, M., Benson, E., & Ilson, R. (2010). *The BBI dictionary of English word combinations* (3rd ed.). Amsterdam, Netherlands: John Benjamins Publishing Company.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge, United Kingdom: Cambridge University Press.
- Biber, D. (2006). *University Language: A corpus-based study of spoken and written registers*. Philadelphia, PA: John Benjamins Publishing Company.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3), 275-275. doi:10.1075/ijcl.14.3.08bib
- 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. (2009). *Register, genre, and style*. Cambridge, United Kingdom: Cambridge University Press.
- 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., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly*, 36(1), 9-48. doi:10.2307/3588359
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2-20. doi:10.1016/j.jeap.2010.01.001

- 
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, *45*(1), 5-35.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London, United Kingdom: Longman.
- Block, D. (2003). *The social turn in second language acquisition*. Washington, DC: Georgetown University Press.
- Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum*, *1*, 1-14.
- Bolinger, D., & Sears, D. A. (1981). *Aspects of language* (3rd ed.). New York, NY: Harcourt Brace Jovanovich.
- Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, *21*(1), 37-54.  
doi:10.1017/S0958344009000068
- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, *60*(3), 534-572. doi:10.1111/j.1467-9922.2010.00566.x
- Boulton, A. (2012). Beyond concordancing: Multiple affordances of corpora in university language degrees. *Languages, Cultures and Virtual Communities*, *34*, 33-38.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101.  
doi:10.1191/1478088706qp063oa
- Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics*, *36*(1), 1-22.  
doi:10.1093/applin/amt018
- Brown, J. D. (2014). *Mixed Methods Research for TESOL*. Edinburgh, the UK: Edinburgh University Press.
- Bunton, D. (1999). The use of higher level metatext in Ph.D theses. *English for Specific Purposes*, *18*, Supplement 1(0), S41-S56. doi:10.1016/S0889-4906(98)00022-2

- 
- 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
- Butterfield, J. (2009). *Damp squid: The English language laid bare*. New York, NY: Oxford University Press.
- Byrd, P., & Coxhead, A. (2010). *On the other hand: Lexical bundles in academic writing and in the teaching of EAP*. *University of Sydney Papers in TESOL*, 5, 31-64.
- Cambridge University Press. (2015). *Cambridge Learner Corpus*. Retrieved from [http://www.cambridge.org/gb/elt/catalogue/subject/custom/item3646603/Cambridge-English-Corpus-Cambridge-Learner-Corpus/?site\\_locale=en\\_GB](http://www.cambridge.org/gb/elt/catalogue/subject/custom/item3646603/Cambridge-English-Corpus-Cambridge-Learner-Corpus/?site_locale=en_GB)
- Cao, F., & Hu, G. (2014). Interactive metadiscourse in research articles: A comparative study of paradigmatic and disciplinary influences. *Journal of Pragmatics*, 66, 15-31. doi:10.1016/j.pragma.2014.02.007
- 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.
- Chambers, A., & O'Sullivan, Í. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL*, 16(1), 158-172. doi:10.1017/S0958344004001211
- Chan, T.-p., & Liou, H.-C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning*, 18(3), 231-250. doi:10.1080/09588220500185769
- Chang, J.-Y. (2014). The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL : the Journal of EUROCALL*, 26(2), 243-259. doi:10.1017/S0958344014000056
- Charles, M., Pecorari, D., & Hunston, S. (2009). Introduction: Exploring the interface between corpus linguistics and discourse analysis. In M. Charles, D. Pecorari & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 1-10). London, United Kingdom: Continuum.

- 
- Chen, H.-J. H. (2011). Developing and evaluating a web-based collocation retrieval tool for EFL students and teachers. *Computer Assisted Language Learning*, 24(1), 59-76. doi:10.1080/09588221.2010.526945
- Chen, L. (2010). An investigation of lexical bundles in ESP textbooks and electrical engineering introductory textbooks. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 107-125). London, United Kingdom: Continuum International Publishing Group.
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.
- Chen, Y.-H., & Baker, P. (2014). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics* doi:10.1093/applin/amu065
- Cohen, A. D. (1983). Reformulating second-language compositions: A potential source of input for the learner. from ERIC database (ED 228866)
- 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. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, 17, 391-406.
- 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
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238. doi:10.2307/3587951
- Coxhead, A. (2012). Academic vocabulary, writing and English for academic purposes: Perspectives from second language learners. *RELC Journal*, 43(1), 137-145. doi:10.1177/0033688212439323

- 
- Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing, 16*(3), 129-147. doi:10.1016/j.jslw.2007.07.002
- 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.
- Dahl, T. (2004). Textual metadiscourse in research articles: A marker of national culture or of academic discipline? *Journal of Pragmatics, 36*(10), 1807-1825. doi:10.1016/j.pragma.2004.05.004
- Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning, 28*(2), 130-144. doi:10.1080/09588221.2013.803982
- 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.
- Durrant, P. (2015). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics, 1*-30. doi:10.1093/applin/amv011
- Eriksson, A. (2012). Pedagogical perspectives on bundles: Teaching bundles to doctoral students of biochemistry. In J. Thomas & A. Boulton (Eds.), *Input, process and product: Development in teaching and language corpora* (pp. 195-211). Brno, Czech Republic: Masaryk University Press.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *TEXT, 20*(1), 29-62.
- Færch, C., & Kasper, G. (1984). Two ways of defining communication strategies. *Language Learning, 34*(1), 45-63.
- Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods, 5*(1), 1-11.
- Firth, J. R. (1957). Modes of meaning. In J. R. Firth (Ed.), *Papers in linguistics 1934-1951* (pp. 190-215). London, United Kingdom: Oxford University Press.

- 
- 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.
- Flowerdew, J. (2000). Discourse community, legitimate peripheral participation, and the nonnative-English-speaking scholar. *TESOL Quarterly*, 34(1), 127-150. doi:10.2307/3588099
- Flowerdew, J. (2003). Signalling nouns in discourse. *English for Specific Purposes*, 22(4), 329-346. doi:10.1016/S0889-4906(02)00017-0
- Flowerdew, L. (2001). The exploitation of small learner corpora in EAP materials design. In M. Ghadessy, A. Henry & R. L. Roseberry (Eds.), *Small corpus studies and ELT theory and practice* (pp. 363-379). Amsterdam, Netherlands: John Benjamins Publishing Company.
- Flowerdew, L. (2014). Which unit for linguistic analysis of ESP corpora of written text? In M. Gotti & D. S. Giannoni (Eds.), *Corpus analysis for descriptive and pedagogical purposes* (pp. 25-41). Bern, Switzerland: Peter Lang.
- Francis, G. (1986). *Anaphoric nouns*. Birmingham, United Kingdom: English Language Research, University of Birmingham.
- Franken, M. (2012). The nature and scope of student search strategies in using a web derived corpus for writing. *The Language Learning Journal*, 1-18. doi:10.1080/09571736.2012.678013
- Fu, X., & Hyland, K. (2014). Interaction in two journalistic genres: A study of interactional metadiscourse. *English Text Construction*, 7(1), 122-144. doi:10.1075/etc.7.1.05fu
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305-327. doi:10.1093/applin/amt015
- Geertz, C. (1988). *Works and lives: The anthropologist as author*. Cambridge, United Kingdom: Polity Press.
- Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL*, 26(2), 225-242. doi:10.1017/S0958344014000044



- 
- Gillaerts, P., & Velde, F. V. d. (2010). Interactional metadiscourse in research article abstracts. *Journal of English for Academic Purposes*, 9(2), 128-139. doi:10.1016/j.jeap.2010.02.004
- Gillham, B. (2005). *Research interviewing: The range of techniques*. Maidenhead, United Kingdom: Open University Press.
- Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6(4), 319-335. doi:10.1016/j.jeap.2007.09.007
- Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction*, 1(1), 41-61.
- Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson (Eds.), *Language in contrast: Papers from a symposium on text-based cross-linguistic studies* (pp. 37-51). Lund, U.K.: Lund University Press.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145-160). New York, NY: Oxford University Press.
- 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
- Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, 1(1), 7-24. doi:10.1075/ijlcr.1.1.01gra
- Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *International Corpus of Learner English Version 2: Handbook and CD-ROM*. Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.
- Gray, B., & Biber, D. (2013). Corpus approaches to the study of discourse. In K. Hyland & B. Paltridge (Eds.), *The Bloomsbury companion to discourse analysis* (pp. 138-152). London, United Kingdom: Bloomsbury Academic.
- Halai, N. (2007). Making use of bilingual interview data: Some experiences from the field. *The Qualitative Report*, 12(3), 344-355.

- 
- Halliday, M. A. K. (1985). *An introduction to functional grammar*. London, United Kingdom: Arnold.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London, United Kingdom: Edward Arnold.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London, United Kingdom: Longman.
- 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.
- Herbel-Eisenmann, B., & Wagner, D. (2010). Appraising lexical bundles in mathematics classroom discourse: Obligation and choice. *Educational Studies in Mathematics*, 75(1), 43-63.
- Hill, J., & Lewis, M. (1997). *LTP dictionary of selected collocations*. Hove, United Kingdom: Language Teaching Publications.
- Hinkel, E. (2001). Matters of cohesion in L2 academic texts. *Applied Language Learning*, 12, 111-132.
- Hinkel, E. (2002). *Second language writer's text: Linguistic and rhetorical features*. NJ: Lawrence Erlbaum Associates.
- 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
- Hu, G., & Cao, F. (2011). Hedging and boosting in abstracts of applied linguistics articles: A comparative study of English- and Chinese-medium journals. *Journal of Pragmatics*, 43(11), 2795-2809. doi:10.1016/j.pragma.2011.04.007
- Huat, C. M. (2012). Learner corpora and second language acquisition. In K. Hyland, C. M. Huat & M. Handford (Eds.), *Corpus applications in applied linguistics* (pp. 191-207). London, United Kingdom: Continuum International Publishing Group.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge, United Kingdom: Cambridge University Press.
- Hyland, K. (2001a). Bringing in the reader: Addressee features in academic articles. *Written Communication*, 18(4), 549-574.
- Hyland, K. (2001b). 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. (2002a). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*, 34(8), 1091-1112. doi:10.1016/S0378-2166(02)00035-8
- Hyland, K. (2002b). Directives: Argument and engagement in academic writing. *Applied Linguistics*, 23(2), 215-239. doi:10.1093/applin/23.2.215
- Hyland, K. (2004a). *Disciplinary discourse: Social interactions in academic writing*. Ann Arbor, MI: University of Michigan Press.
- Hyland, K. (2004b). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, 13(2), 133-151. doi:10.1016/j.jslw.2004.02.001
- 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. (2007a). Applying a gloss: Exemplifying and reformulating in academic discourse. *Applied Linguistics*, 28(2), 266-285.
- Hyland, K. (2007b). Different strokes for different folks: Disciplinary variation in academic writing. In K. Fløttum (Ed.), *Language and discipline perspectives on academic discourse* (pp. 89-108). Newcastle, United Kingdom: Cambridge Scholars Publishing.
- 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. (2008c). Persuasion, interaction and the construction of knowledge: Representing self and others in research writing. *International Journal of English Studies*, 8(2), 1-23.
- Hyland, K. (2009). *Academic discourse: English in a global context*. London, United Kingdom: Continuum.
- Hyland, K. (2010). Constructing proximity: Relating to readers in popular and professional science. *Journal of English for Academic Purposes*, 9(2), 116-127. doi:10.1016/j.jeap.2010.02.003
- Hyland, K. (2011). Looking though corpora into writing practices. In V. Viana, S. Zyngier & G. Barnbrook (Eds.), *Perspectives on corpus linguistics* (pp. 99-113). Amsterdam, Netherlands: John Benjamins Publishing Company.
- Hyland, K. (2012). Corpora and academic discourse. In K. Hyland, C. M. Huat & M. Handford (Eds.), *Corpus applications in applied linguistics* (pp. 30-46). London, United Kingdom: Continuum.
- 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
- Hyland, K., & Tse, P. (2005). Evaluative that constructions: Signalling stance in research abstracts. *Functions of Language*, 12(1), 39-63. doi:10.1075/fol.12.1.03hyl

- 
- Intaraprawat, P., & Steffensen, M. S. (1995). The use of metadiscourse in good and poor ESL essays. *Journal of Second Language Writing*, 4(3), 253-272. doi:10.1016/1060-3743(95)90012-8
- Ivanič, R. (1991). Nouns in search of a context: A study of nouns with both open- and closed-system characteristics. *International Review of Applied Linguistics*, 29, 93-114. doi:10.1515/iral.1991.29.2.93
- Jablonkai, R. (2010). English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes*, 29(4), 253-267. doi:10.1016/j.esp.2010.04.006
- Jakobson, R. (1980). *The framework of language*. Michigan, MI: Oxford Publishing Limited.
- Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning*, 50(2), 245-309. doi:10.1111/0023-8333.00118
- Jiang, F., & Hyland, K. (2015). 'The fact that': Stance nouns in disciplinary writing. *Discourse Studies*, 1-22. doi:10.1177/1461445615590719
- Jiang, H. (2009). 汉英学术语篇中语码注解标记使用情况对比分析 [The contrastive analysis of the use of code glosses in Chinese and English academic discourse]. *外语学刊 [Foreign Language Research]*, 5, 88-91.
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *English Language Research Journal*, 4, 1-16.
- Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 269-300). Philadelphia, PA: John Benjamins Publishing Company.
- Kaneyasu, M. (2012). *From frequency to formulaicity: Morphemic bundles and semi-fixed constructions in Japanese spoken discourse* (Doctoral dissertation). University of California, Los Angeles, CA. Retrieved from <https://escholarship.org/uc/item/1zp613xj#page-1>
- Karabacak, E., & Qin, J. (2013). Comparison of Lexical Bundles used by Turkish, Chinese, and American University Students. *Procedia — Social and Behavioral Sciences*, 70(0), 622-628. doi:10.1016/j.sbspro.2013.01.101

- 
- Khedri, M., Heng, C. S., & Ebrahimi, S. F. (2013). An exploration of interactive metadiscourse markers in academic research article abstracts in two disciplines. *Discourse Studies*, 15(3), 319-331.  
doi:10.1177/1461445613480588
- Kilgarriff, A. (2009). *Corpora in the classroom without scaring the students*. Paper presented at the 18th International Symposium on English Teaching, Taipei, Taiwan. Retrieved from <https://www.kilgarriff.co.uk/.../2009-K-ETA-Taiwan-scaring.doc>
- Kilgarriff, A., & Grefenstette, G. (2003). Introduction to the special issue on the web as corpus. *Computational Linguistics*, 29(3), 333-347.  
doi:10.1162/089120103322711569
- Kim, L. C., & Lim, J. M.-H. (2013). Metadiscourse in English and Chinese research article introductions. *Discourse Studies*, 15(2), 129-146.
- Kim, Y. (2009). Korean lexical bundles in conversation and academic texts. *Corpora*, 4(2), 135-165. doi:10.3366/E1749503209000288
- Koutsantoni, D. (2006). Rhetorical strategies in engineering research articles and research theses: Advanced academic literacy and relations of power. *Journal of English for Academic Purposes*, 5(1), 19-36.  
doi:10.1016/j.jeap.2005.11.002
- 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
- Lautamatti, L. (1978). Observations on the development of the topic in simplified discourse. In V. Kohonen & N. E. Enkvist (Eds.), *Text linguistics, cognitive learning and language teaching* (pp. 71-104). Turku, Finland: University of Turku.
- Lee, D. Y. W. (2001). Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning and Technology*, 5(3), 37-72.
- Leech, G. (1998). Preface. In S. Granger (Ed.), *Learner English on computer* (pp. xiv-xx). London, United Kingdom: Longman.

- 
- Leńko-Szymańska, A., & Boulton, A. (2015). *Multiple affordances of language corpora for data-driven learning*. Amsterdam, Netherlands: John Benjamins Publishing Company.
- Lewin, B. A. (2005). Hedging: An exploratory study of authors' and readers' identification of 'toning down' in scientific texts. *Journal of English for Academic Purposes*, 4(2), 163-178. doi:10.1016/j.jeap.2004.08.001
- Lewis, M. (2008). *Implementing the lexical approach: Putting theory into practice*. London, United Kingdom: Heinle Cengage Learning.
- Li, C. (2004). *A corpus-based analysis of collocation errors in writing by Chinese English learners* (Master's thesis). University of Electronic Science and Technology of China, Chengdu, China. Retrieved from [http://d.wanfangdata.com.cn/Thesis\\_W003526.aspx](http://d.wanfangdata.com.cn/Thesis_W003526.aspx)
- Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18, 85-102.
- Li, T., & Wharton, S. (2012). Metadiscourse repertoire of L1 Mandarin undergraduates writing in English: A cross-contextual, cross-disciplinary study. *Journal of English for Academic Purposes*, 11, 345-356.
- Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31(1), 25-35. doi:10.1016/j.esp.2011.07.002
- Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System*, 33(4), 623-636. doi:10.1016/j.system.2005.02.002
- Marandi, S. (2003). Metadiscourse in Persian/English master's thesis: A contrastive study. *IJAL*, 6(2), 23-42.
- Marco, M. J. L. (2011). Exploring atypical verb+noun combinations in learner technical writing. *International Journal of English Studies*, 11(2), 77-95.
- 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
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge, United Kingdom: Cambridge University Press.

- 
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London, United Kingdom: Routledge.
- Molino, A. (2010). Personal and impersonal authorial references: A contrastive study of English and Italian linguistics research articles. *Journal of English for Academic Purposes*, 9(2), 86-101.  
doi:10.1016/j.jeap.2010.02.007
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, United Kingdom: Cambridge University Press.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, United Kingdom: Cambridge University Press.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford, United Kingdom: Oxford University Press.
- Neely, E., & Cortes, V. (2009). A little bit about: Analyzing and teaching lexical bundles in academic lectures. *Language Value*, 1(1), 17-38.
- Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, 11(3), 283-283. doi:10.1075/ijcl.11.3.04nes
- Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge, United Kingdom: Cambridge University Press.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.
- Nesselhauf, N. (2004). What are collocations? In D. J. Allerton, N. Nesselhauf & P. Skandera (Eds.), *Phraseological units: Basic concepts and their application* (pp. 1-21). Basel, Switzerland: Schwabe.
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge, United Kingdom: Cambridge University Press.
- O'Sullivan, Í., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15(1), 49-68. doi:10.1016/j.jslw.2006.01.002



- 
- Pang, P. (2009). A study on the use of four-word lexical bundles in argumentative essays by Chinese English-majors: A comparative study based on WECCL and LOCNESS. *CELEA Journal*, 32(3), 25-45.
- Pang, W. (2010). Lexical bundles and the construction of an academic voice: A pedagogical perspective. *Asian EFL Journal*, 47, 1-13.
- Paquot, M. (2012). *Academic vocabulary in learner writing: From extraction to analysis*. London, United Kingdom: Continuum.
- Paquot, M. (2013). Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics*, 18(3), 391-417. doi:10.1075/ijcl.18.3.06paq
- Parkinson, D., & Francis, B. (2007). *Oxford idioms dictionary for learners of English* (2nd ed.). Oxford, United Kingdom: Oxford University Press.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). London, United Kingdom: Longman.
- Pecorari, D. (2009). Formulaic language in Biology: A topic-specific investigation. In M. Charles, D. Pecorari & S. Hunstan (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 91-104). London, United Kingdom: Continuum.
- Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, 14, 84-94. doi:10.1016/j.jeap.2014.01.002
- Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System*, 42(1), 220-231. doi:10.1016/j.system.2013.12.003
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London, United Kingdom: Longman.
- Rundell, M. (2010). *Macmillan collocations dictionary for learners of English*. London, United Kingdom: Macmillan Education.
- Salager-Meyer, F. (1992). A text-type and move analysis study of verb tense and modality distribution in medical English abstracts. *English for Specific Purposes*, 11(2), 93-113.

- 
- Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing*. Amsterdam, Netherlands: John Benjamins Publishing Company.
- Schiffrin, D., Tannen, D., & Hamilton, H. E. (2001). *The handbook of discourse analysis*. Malden, MA: Blackwell Publishers.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing, and use*. Philadelphia, PA: John Benjamins Publishing Company.
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1-22). Philadelphia, PA: John Benjamins Publishing Company.
- Schnur, E. (2014). Phraseological signaling of discourse organization in academic lectures: A comparison of lexical bundles in authentic lectures and EAP listening materials. *Yearbook of Phraseology*, 5(1), 95-122.  
doi:10.1515/phras-2014-0005
- Sharwood Smith, M. (1991). Speaking to many minds: on the relevance of different types of language information for the L2 learner. *Second Language Research*, 7(2), 118-132. doi:10.1177/026765839100700204
- Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *SSLA*, 15, 165-179. doi:10.1017/S0272263100011943
- Shih, R. H.-H. (2000). Collocation deficiency in a learner corpus of English: From an overuse perspective. In A. Ikeya & M. Kawamori (Eds.), *Proceedings of the 14th Pacific Asia Conference on Language, information and Computation* (pp. 281-288). Tokyo, Japan: PACLIC 14 Organizing Committee.
- Siefring, J. (2004). *Oxford dictionary of idioms* (2nd ed.). Oxford, United Kingdom: Oxford University Press.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, United Kingdom: Oxford University Press.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing

- 
- section. *Journal of English for Academic Purposes*, 12(3), 214-225.  
doi:10.1016/j.jeap.2013.05.002
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. London, United Kingdom: Blackwell.
- Suomela-Salmi, E., & Dervin, F. (2009). *Cross-linguistic and cross-cultural perspectives on academic discourse*. Philadelphia, PA: John Benjamins Publishing Company.
- Tarone, E. (1980). Communication strategies: Foreign talk, and repair in interlanguage. *Language Learning*, 30(2), 417-431.
- 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.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam, Netherlands: John Benjamins Publishing Company.
- Tracy-Ventura, N., Cortes, V., & Biber, D. (2007). Lexical bundles in speech and writing. In G. Parodi (Ed.), *Working with Spanish corpora* (pp. 217-231). London, United Kingdom: Continuum International Publishing.
- Vande Kopple, W. J. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication*, 36(1), 82-93.  
doi:10.2307/357609
- Vande Kopple, W. J. (1989). *Clear and coherent prose: A functional approach*. Glenview, IL: Scott, Foresman and Company.
- 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
- Walter, E. (2006). *Cambridge idioms dictionary* (2nd ed.). Cambridge, United Kingdom: Cambridge University Press.
- Wang, H., & Zhou, X. (2009). 中国英语学习者动名词搭配行为的发展特点研究—语料库驱动的研究方法 [A study on the verb-noun collocational behavior of Chinese EFL learners at three developmental stages: A corpus driven approach]. *外语学刊* [*Foreign Language Research*], 6, 59-62.

- 
- Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC Journal*, 42(2), 155-166.  
doi:10.1177/0033688211407295
- Wen, Q., Ding, Y., & Wang, W. (2003). 中国大学生英语书面语中的口语化倾向：高水平英语学习者语料对比分析 [Features of oral style in English compositions of advanced Chinese EFL learners : An exploratory study by contrastive learner corpus analysis]. *外语教学与研究 [Foreign Language Teaching and Research]*, 35(4), 268-274.
- Wen, Q., Liang, M., & Yan, X. (2008). *中国学生英语口语笔语语料库 (2.0 版) [Spoken and written English corpus of Chinese learners] (version 2.0)*. Beijing, China: Foreign Language Teaching and Research Press.
- West, M. (1953). *A general service list of English words*. London, United Kingdom: Longman.
- Williams, J. M. (1981). *Style: Ten lessons in clarity and grace* (1st ed.). Boston, MA: Scott, Foresman.
- Williams, J. M. (2003). *Style: Ten lessons in clarity and grace* (7th ed.). New York, NY: Addison-Wesley.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, United Kingdom: Cambridge University Press.
- 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.

- 
- Xu, F. (2011). 中国学生英语学术写作中身份语块的语料库研究 [A corpus-based study of identity constructions in Chinese students' English academic writing]. *中国英语教育 [English Education in China]*, 3, 1-9.
- Xu, F. (2012). 中国学习者英语学术词块的使用及发展特征研究 [The use and developmental features of lexical bundles in Chinese learners' English academic writing]. *外语与外语教学 [Foreign Languages and Their Teaching]*, 4, 51-56. doi:10.13564/j.cnki.issn.1672-9382.2012.04.013
- Xu, H., & Gong, S. (2006). 元语篇手段的使用与语篇质量相关度的实证研究 [An investigation into the correlation between use of metadiscourse markers and writing quality]. *现代外语 [Modern Foreign Languages]*, 29(1), 54-61.
- 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. doi:10.1016/j.linged.2011.09.004
- 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
- Yeh, Y., Li, Y.-H., & Liou, H.-C. (2007). Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning*, 20(2), 131-152. doi:10.1080/09588220701331451
- Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning and Technology*, 12(2), 31-48.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257-283. doi:10.1016/j.jslw.2004.06.002

## Appendix A: Adaptations of Biber and his colleagues' taxonomy

Subcategory	Criterion	Examples	Category	Study
<b>Inferential bundles</b>	identify a logical relationship	<i>on the basis of</i>	discourse organizers	Biber, Conrad & Cortes, 2003
	indicate inference	<i>on the basis of, as a result of</i>	discourse organizers	Cortes, 2004
	make inference	<i>as a result of, in view of the, this is due to</i>	discourse organizers	Chen & Baker, 2010
	reflect relationships between prior and coming discourse		discourse organizers	Adel & Erman, 2012
<b>Contrast /Comparison bundles</b>	identify a logical relationship	<i>on the other hand</i>	discourse organizers	Biber, Conrad & Cortes, 2003
	indicate comparison/contrast	<i>on the other hand, in contrast to the</i>	discourse organizers	Cortes, 2004
<b>Frame bundles</b>	identify the textual conditions	<i>on the basis of, in the case of</i>	discourse organizers	Biber, Conrad & Cortes, 2003
	identify textual conditions	<i>in the case of, in the context of, the nature of the, the extent to which</i>	discourse organizers	Cortes, 2004
	specify a given attribute or condition	<i>in the context of, the nature of the</i>	referential expressions	Chen & Baker, 2010
	make direct reference to physical or abstract entities or to the textual context itself		referential expressions	Adel & Erman, 2012

	specification of attributes	<i>in the context of, the nature of the</i>	referential expressions	Cortes, 2013
<b>Quantifying bundles</b>	introduce quantities or amounts and statistical expressions	<i>a wide range of, in the number of, one of the most, not significantly different from</i>	referential expressions	Cortes, 2004
	expressions related to anything potentially measurable as quantifying bundles such as size, number, amount or extent	<i>a wide range of, in a number of, the extent to which</i>	referential expressions	Chen & Baker, 2010
	make direct reference to physical or abstract entities or to the textual context itself		referential expressions	Adel & Erman, 2012
<b>Focus bundles</b>	focus on the noun phrase following the bundle as especially important	<i>those of you who, that's one of the, one of the things</i>	referential expressions	Biber, Conrad & Cortes, 2004
	identification/focus bundles	<i>one of the most, one of the major</i>	referential expressions	Cortes, 2013
	qualify a certain element of the succeeding discourse in terms of its importance or degree of difficulty	<i>it is important to, it is difficult to</i>	discourse organizers	Cortes, 2004
	preview, emphasize or summarize the main point	<i>that's one of the, one of the things</i>	discourse organizers	Biber & Barbier, 2007

---

identify the focus that the writer is making	<i>one of the most, there would be no</i>	discourse organizers	Chen & Baker, 2010
reflect relationships between prior and coming discourse		discourse organizers	Adel & Erman, 2012

---





---

## Appendix B: Ädel's (2006) taxonomy of personal metadiscourse

---

Category	Function	Examples
<b>Metatext</b>		
<b>Code</b>	<p><b>Defining</b> explicitly comments on how to interpret terminology.</p> <p><b>Saying</b> involves general <i>verba dicendi</i> such as say, speak, talk or writer, in which the fact that something is being communicated is foregrounded.</p>	<p><i>What do we mean by . . . then?</i></p> <p><i>We have to consider our definition of . . .</i></p> <p><i>What I am saying is . . .</i></p> <p><i>A question I ask myself is . . .</i></p>
<b>Text: focus on structure of essay</b>	<p><b>Introducing the topic</b> gives explicit proclamations of what the text is going to be about, which facilitates the processing of the subsequent text for the reader.</p> <p><b>Focussing</b> refers to a topic that has already been introduced in the text: announces that the topic is in focus again or it narrows down.</p> <p><b>Concluding</b> is used to conclude a topic.</p> <p><b>Exemplifying</b> explicitly introduces an example.</p> <p><b>Reminding</b> points backwards in the discourse to something that has been said before.</p> <p><b>Adding</b> overtly states that a piece of information or an argument is being added to existing one(s).</p> <p><b>Arguing</b> stresses the discourse act being performed in addition to expressing an opinion or viewpoint. Verbs used are performatives.</p> <p><b>Contextualising</b> contains traces of the production of the text or comments on (the condition of) the situation of writing.</p>	<p><i>In the course of this essay, we shall attempt to analyse whether . . .</i></p> <p><i>I will discuss . . .</i></p> <p><i>Now I come to the next idea which I presented in the beginning . . .</i></p> <p><i>I will only discuss the opponents of . . .</i></p> <p><i>In conclusion, I would say that . . .</i></p> <p><i>As an example of . . . , we can look at . . .</i></p> <p><i>If we take . . . as an example</i></p> <p><i>As I mentioned earlier, . . .</i></p> <p><i>As we have seen, . . .</i></p> <p><i>I would like to add that . . .</i></p> <p><i>The . . . which I argue for is . . .</i></p> <p><i>I have chosen this subject because . . .</i></p> <p><i>I could go on much longer, but . . .</i></p>
<b>Writer-reader interaction</b>		
<b>Participant: focus on writer and/or</b>	<p><b>Anticipating the Reader's Reaction</b> pays special attention to predicting the reader's reaction to</p>	<p><i>I do realise that all this may sound . . .</i></p>

---

---

<b>reader of current text</b>	<p>what is said, e.g. by explicitly attributing statements to the reader as possible objections or counterarguments conceived by him.</p> <p><b>Clarifying</b> marks a desire to clarify matters for the reader; motivated by a wish to avoid misinterpretation. Negative statements are common.</p> <p><b>Aligning perspectives</b> takes it for granted that the reader takes the writer's perspective. The reader's agreement is presupposed.</p> <p><b>Imagining Scenarios</b> is a 'picture this' type of encouragement that (often politely) asks the reader to see something from a specific perspective. It allows writers to make examples vivid and pertinent to the reader.</p> <p><b>Hypothesising about the Reader</b> makes guesses about the reader and his knowledge or attitudes.</p> <p><b>Appealing to the Reader</b> attempts to influence the reader by emotional appeal. The writer persona conveys her attitude with the aim of correcting or entreating the reader.</p>	<p><i>You probably never heard of . . . before either</i></p> <p><i>I am not saying . . . , I am merely pointing out that . . . By this I do not mean that . . .</i></p> <p><i>If we [consider/compare] . . . , we [can/will] [understand/see] . . .</i></p> <p><i>If you consider . . . , you can perhaps imagine . . . Think back to when you were . . .</i></p> <p><i>You have probably heard people say that . . .</i></p> <p><i>I hope that now the reader has understood . . . In order for . . . , you and I must keep our minds open</i></p>
-------------------------------	---	--

---

*Note.* Adapted from *Metadiscourse in L1 and L2 English* (pp.60-61), by A. Ädel, 2006, Amsterdam, Netherlands: John Benjamins Publishing Co. Reprinted with permission.

## Appendix C: Bundles identified in the four postgraduate corpora

CH MA		CH PhD		NZ MA		NZ PhD	
On the other hand,	62	On the other hand,	65	On the other hand,	34	On the other hand,	28
That is to say,	51	In other words, the	39	It is important to	26	It is possible that	20
At the same time,	38	That is to say,	37	The results of the	23	In the case of	18
The results of the	29	On the one hand,	27	It is possible that	14	At the same time,	18
In the process of	28	The results of the	24	In the case of	14	It is important to	17
On the basis of	22	In the case of	20	The results of this	13	As discussed in Chapter	17
With the development of	21	In the present study,	19	As can be seen	13	At the end of	15
In other words, the	20	At the same time,	18	It is interesting to	13	In addition to the	14
In the present study,	20	On the basis of	18	The purpose of this	12	The results of the	12
In this chapter, the	19	In this sense, the	17	As a result of	12	In other words, the	11
At the end of	17	In terms of the	17	At the same time,	12	It should be noted	11
On the one hand,	17	In this way, the	16	The majority of the	11	As can be seen	9
In order to make	16	In addition to the	15	At the time of	10	It is interesting to	8
As a result, the	15	It can be seen	15	In addition to the	10	In this chapter I	8
In this way, the	13	As far as the	14	It should be noted	10	For the purposes of	8
In this study, the	13	It should be noted	14	That is to say,	10	There are a number	8
It can be seen	13	As can be seen	14	By the end of	9	The fact that the	7
At the beginning of	13	From the perspective of	12	In other words, the	9	In the context of	7
It is obvious that	12	As a result, the	11	The chapter concludes with	9	As a result of	7
During the process of	12	With regard to the	11	At the end of	9	At the time of	7
The purpose of this	12	Look at the following	11	For the purpose of	9	It is not clear	7
As a matter of	11	It is clear that	10	In the current study,	9	There was a significant	7

As we all know,	11	The following is a	9	The aim of the	8	There was no significant	7
The present study is	11	To sum up, the	9	In the current study	8	The purpose of this	7
With the help of	10	To be more specific,	9	In this study the	8	On the one hand,	6
Based on the above	10	In this chapter, we	9	There appears to be	8	It may be that	6
As is shown in	10	With respect to the	8	There are a number	8	This is not to	6
It is suggested that	9	For the sake of	8	The fact that the	8	In terms of the	6
The results show that	9	In this case, the	8	It may be that	8	There were no significant	6
The purpose of the	8	It is true that	8	In addition to this,	7	The first of these	6
From the perspective of	8	It is hoped that	8	It is important that	7	The results of this	6
In the light of	8	In this section, we	7	There was no significant	7	At the beginning of	6
To sum up, the	8	It is important to	7	However, it is important	7	In contrast to the	6
In order to get	8	It is obvious that	7	In spite of the	7	It is also possible	6
The following table shows	8	It seems that the	7	In this chapter I	7	The analysis of the	6
One of the most	8	As shown in Table	7	This is because the	7	The aim of this	6
The result of the	8	As a matter of	7	The purpose of the	6	In this section I	6
In addition to the	8	The analysis of the	7	The findings of this	6	With regard to the	6
As far as the	7	In this section, I	7	In contrast to the	6	It is difficult to	6
It is clear that	7	As is shown in	7	In other words the	6	It is clear that	6
It is believed that	7	In spite of the	7	The next chapter will	6	On the basis of	5
It is found that	7	In this section, the	7	This chapter presents the	6	In this section, I	5
The results showed that	7	It is necessary to	6	This suggests that the	6	This is not a	5
Therefore, it is necessary	7	It should be pointed	6	The analysis of the	6	This is followed by	5
However, it is not	7	It is argued that	6	The aim of this	6		
Last but not least,	7	The results showed that	6	The results from the	6		
For example, in the	7	To put it another	6	As a result, the	6		

In this chapter, we	7	The present study is	6	At the same time	6
In this section, the	7	This suggests that the	6	It is difficult to	6
The main purpose of	7	In the field of	6	It is interesting that	6
In other words, they	7	For example, in the	6	This is not to	6
It is important to	7	This means that the	6	As discussed in Chapter	6
It is necessary to	7	In other words, it	6	The limitations of the	5
The results indicate that	6	The following is an	6	The findings of the	5
The thesis consists of	6	When it comes to	6	The use of the	5
When it comes to	6	As a result of	5	In terms of the	5
In a word, the	6	In other words, they	5	In this section the	5
In this part, the	6	We can see that	5	It was important to	5
In order to find	6	The following are some	5	It is possible to	5
First of all, the	6	However, it should be	5	It would appear that	5
As one of the	6			It can be seen	5
With regard to the	6			It must be noted	5
There is no doubt	6			This chapter describes the	5
The first one is	6				
The following are some	6				
That is to say	6				
From the above table,	5				
This thesis consists of	5				
The following is a	5				
As shown in Table	5				
In the course of	5				
In view of the	5				
As a result, it	5				
It is evident that	5				
It is hoped that	5				

---

We can see	5
from	
It means that the	5
The following is	5
the	
As can be seen	5
So it is	5
necessary	

---

## Appendix D: 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						
<b>Introduction bundles</b>	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 E: 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						

## Appendix F: Ethical approval

Dean's Office  
Faculty of Education  
Te Kōwhiri Wānanga  
The University of Waikato  
Private Bag 3100  
Hamilton, New Zealand

Phone +64 7 838 4300  
[www.waikato.ac.nz](http://www.waikato.ac.nz)



THE UNIVERSITY OF  
**WAIKATO**  
Te Kōwhiri Wānanga o Waikato

### MEMORANDUM

**To:** Liang Li  
**Cc:** Dr Margaret Franken  
**From:** Associate Professor Garry Falloun  
Chairperson, Research Ethics Committee  
**Date:** 26 February 2014  
**Subject:** Supervised Postgraduate Research – Application for Extension of Ethical Approval  
(FD1102/12)

Thank you for your application for an extension of ethical approval for the project:

**Corpus-based collocation learning for L2 writing: Strategies, perspectives and efficacy\***

It is noted that you wish to include up to six Chinese master and PhD level students in your research. Thank you for providing copies of the revised participant information sheet and consent form.

I am pleased to advise that your application has received ethical approval.

Please note that researchers are asked to consult with the Faculty's Research Ethics Committee in the first instance if any further changes to the approved research design are proposed.

The Committee continues to wish you all the best with your research.

Associate Professor Garry Falloun  
Chairperson  
Faculty of Education Research Ethics Committee

\* During the interval of the ethical approval and submission, the title of this thesis has been changed.



---

## Appendix G: Interview questions

1. How many years have you been learning English? When did you come to study abroad? What do you study?
2. I have noticed you have used ..... Was this a careful choice or did you just do this automatically?
3. Why did you choose the one you did? Was this a good choice do you think?
4. Did you consider other options to express this? What were they?
5. Here are some suggestions ..... What do you think?
6. What are the sources of the chosen sentence initial bundles?
7. Is there anything else you want to talk about?

### Overlapped expressions from one participant's writing

<b>Interactive markers</b>	<b>Expressions</b>
Transition markers	<i>On the one hand, ... on the other hand, ...</i>
Frame markers	<i>The last but not least,</i>
Endophoric markers	
Code glosses	<i>In other words,; To be specific,</i>
Condition markers	<i>In order to make; With the development of</i>
Introduction markers	
<b>Interactional markers</b>	<b>Expressions</b>
Attitude markers	<i>it is necessary to; the interesting is,</i>
Hedges	<i>one of the most</i>
Boosters	<i>it is obvious that; It is undoubted that</i>
Self-mentions	<i>I</i>
Directives	<i>It should be noticed that; it is worth noting that</i>
Shared knowledge	<i>As we all know,</i>





## Appendix H: The 50 most frequent sentence initial bundles in each corpus

CH MA	CH PhD	NZ MA	NZ PhD
On the other hand,	On the other hand,	On the other hand,	On the other hand,
That is to say,	In other words, the	It is important to	It is possible that
At the same time,	That is to say,	The results of the	In the case of
The results of the	On the one hand,	It is possible that	At the same time,
In the process of	The results of the	In the case of	It is important to
On the basis of	In the case of	The results of this	As discussed in Chapter
With the development of	In the present study,	As can be seen	At the end of
In other words, the	At the same time,	It is interesting to	In addition to the
In the present study,	On the basis of	As a result of	The results of the
In this chapter, the	In this sense, the	The purpose of this	In other words, the
At the end of	In terms of the	At the same time,	It should be noted
On the one hand,	In this way, the	The majority of the	As can be seen
In order to make	In addition to the	That is to say,	It is interesting to
As a result, the	It can be seen	In addition to the	In this chapter I
In this way, the	As far as the	It should be noted	For the purposes of
It can be seen	As can be seen	At the time of	There are a number
In this study, the	It should be noted	In other words, the	In the context of
At the beginning of	From the perspective of	The chapter concludes with	As a result of
It is obvious that	As a result, the	By the end of	The fact that the
During the process of	With regard to the	In the current study,	At the time of
The purpose of this	Look at the following	At the end of	There was a significant
As we all know,	It is clear that	For the purpose of	It is not clear
As a matter of	The following is a	In the current study	There was no significant
The present study is	To sum up, the	There appears to be	The purpose of this
Based on the above	To be more specific,	There are a number	This is not to
With the help of	In this chapter, we	The aim of the	On the one hand,
As is shown in	With respect to the	In this study the	It may be that
It is suggested that	In this case, the	The fact that the	In terms of the
The results show that	For the sake of	It may be that	The first of these
The purpose of the	It is true that	However, it is important	The results of this
From the perspective of	It is hoped that	It is important that	In contrast to the

To sum up, the	It is important to	There was no significant	There were no significant
In the light of	As shown in Table	In addition to this,	It is also possible
The following table shows	In this section, we	This is because the	At the beginning of
In order to get	It is obvious that	In spite of the	In this section I
One of the most	It seems that the	In this chapter I	It is difficult to
In addition to the	As a matter of	In contrast to the	The analysis of the
The result of the	The analysis of the	The next chapter will	It is clear that
Therefore, it is necessary	In this section, I	The purpose of the	With regard to the
It is believed that	As is shown in	In other words the	The aim of this
It is found that	In spite of the	This chapter presents the	On the basis of
The results showed that	In this section, the	This suggests that the	This is not a
It is clear that	The results showed that	The findings of this	In this section, I
As far as the	It is necessary to	This is not to	This is followed by
However, it is not	It should be pointed	The results from the	For example, in the
Last but not least,	It is argued that	It is difficult to	The majority of the
For example, in the	To put it another	As a result, the	It can be seen
In this chapter, we	The present study is	As discussed in Chapter	As shown in Table
In this section, the	This suggests that the	It is interesting that	During the course of
It is important to	For example, in the	The analysis of the	In other words, it