

# Sentence initial lexical bundles in Chinese and New Zealand PhD theses in the discipline of General and Applied Linguistics

Liang Li<sup>1</sup>, Margaret Franken<sup>2</sup>, Shaoqun Wu<sup>1</sup>

<sup>1</sup>University of Waikato | <sup>2</sup>Ministry of Education

Lexical bundles are recurrent multiword combinations and often function as discourse building blocks. Lexical bundles have been analysed in university students' writing to detect linguistic errors, measure writing competence, and investigate the divergence between L1 and L2 writing. Few studies, however, have focused on the high-stakes genre of PhD thesis and investigated the bundle productions of the same genre within the same level and discipline. This paper compares sentence initial lexical bundles in the corpora of English theses written by Chinese and New Zealand PhD students in the discipline of General and Applied Linguistics. Forty-six bundles from a Chinese corpus and forty-two bundles from a New Zealand corpus were generated. Among them, 94% of sentence initial bundles were identified as metadiscursive bundles. Chinese and New Zealand doctoral students showed considerably different preferences in their bundle selection. The paper examines the possible impact of these preferences and suggests there is a need to extend the metadiscourse knowledge of doctoral students in terms of lexical bundles.

**Keywords:** corpus analysis, lexical bundles, metadiscourse, L2 academic writing, PhD thesis, disciplines

## 1. Introduction

Lexical bundles are recurrent multiword combinations of three or more words (Biber et al., 1999) generated automatically from a corpus (Biber, 2015). Examples of lexical bundles are *the fact that*, *in the case of*, and *it should be noted that*. Lexical bundles play an important role in signifying fluency, accuracy, and idiomaticity in academic writing. These recurrent multiword combinations are important discourse building blocks. The repertoire of lexical bundles allows writers to expand their texts from ready-made lexical chunks instead of individual words (Coxhead & Byrd, 2007). The use of lexical bundles, that is, shared conventional expressions in a specific discourse, expresses writer's community-recognized voice and indicates writer's membership of the target discourse community (Hyland, 2012; Wray, 2002). More importantly, lexical bundles as structurally open functional blocks (Biber & Barbieri, 2007) often act as discourse frames where writers can start from and weave their communicative contents into these "points of fixation" (Dechert, 1984, p. 223).

An increasing number of lexical bundle studies have emerged after earlier bundle studies by Biber and his colleagues (e.g., Biber et al., 2003, 2004; Biber et al., 1999). Among them, many are in the area of L2 academic writing. Different bundles have been identified in relation to writer groups and discourse functions carried by these bundles are examined in almost all bundle studies in this area. For example, Ädel and Erman (2012) compared undergraduate essays between the first and second year L1 English student writing and the first to fourth term L1 Swedish student writing. Chen and Baker (2010) explored lexical bundles in the excerpts of published academic texts, and L1 English and L1 Chinese student essays. Hyland (2008) compared research articles with PhD dissertations and master's theses written mainly by L1 Cantonese/Chinese students. Wei and Lei (2011) investigated research

articles and L1 Chinese PhD dissertations. In these studies, Biber and his colleagues' taxonomy (Biber et al., 2003, 2004) (i.e., referential bundles, discourse organizer, and stance bundles) or Hyland's (2008) framework (i.e., research-oriented, text-oriented, and participant-oriented bundles) was adopted to analyse the functions of the lexical bundles. Despite the differences between these two function models, both Biber et al.'s (2003, 2004) taxonomy and Hyland's (2008) framework include interpersonal bundles (stance bundles in Biber's taxonomy and participant-oriented bundles in Hyland's framework) to facilitate writer-reader interaction and organisational bundles (discourse organizers in Biber's taxonomy and text-oriented bundles in Hyland's framework) to indicate the structure of text. Therefore, although the above studies used different models, we can generalise the similarities between their results: a smaller percentage of interpersonal bundles and a larger percentage of organisational bundles are reported in L1 Swedish undergraduate essays (Ådel & Erman, 2012) and L1 Chinese student essays (Chen & Baker, 2010), master's theses (Hyland, 2008) and doctoral dissertations (Wei & Lei, 2011). However, language background is often not the only explanatory variable of these earlier studies. Confounding variables such as genres and levels of study could all affect the results as the linguistic features of writing can be largely determined by its situational context, communicative purposes and target audiences (Biber & Conrad, 2019).

Many recent studies tend to minimize the influences of other factors during comparison. They examined the bundle productions of the same genre written by the same cohort, although the corpora in some studies covered texts from different disciplines. For example, Pérez-Llantada (2014), Pan et al. (2016), and Esfandiari and Barbary (2017) focused on published research articles written by L1 Spanish, L1 Chinese, and L1 Persian professional writers respectively, using the corresponding L1 English articles as their reference corpus. Bychkovska and Lee (2017) investigated argumentative essays at the undergraduate level, although the years of study are different between their L1 English and L1 Chinese groups. Lu and Deng (2019) examined doctoral dissertation abstracts of L1 Chinese students from Tsinghua University compared to those of L1 English students from the Massachusetts Institute of Technology. One of the interesting findings is a wide range of stance bundles found in the L1 Chinese writing by Bychkovska and Lee (2017), Lu and Deng (2019), and Pan et al. (2016). Stance bundles were less frequent in L2 writing in many previous studies and remain an unfavourable category in articles written by L1 Spanish and L1 Persian professional writers (Esfandiari & Barbary, 2017; Pérez-Llantada, 2014). Another difference between these studies and previous ones is that they further examined the subcategories of functions (e.g., frame bundles and resultative bundles) with a focus on the use of individual bundles. Framing bundles, such as *in the form/case/context of*, were found to be underused in research articles, argumentative essays, and doctoral dissertation abstracts produced by L1 Chinese writers (Bychkovska & Lee, 2017; Lu & Deng, 2019; Pan et al., 2016). Regarding resultative bundles, the preferred pattern of Chinese students was *(the) result(s) show(s/ed)/indicate(d) that (the)*, while two alternative patterns *I/we find/found/show that the* and *have been found/shown to* (or *is/was/were found/shown to be*) were favoured by L1 students in abstract writing (Lu & Deng, 2019).

Few studies have explored lexical bundles in L2 PhD thesis writing, a high-stakes genre for students (Hyland, 2008), and compared it with the same genre within the same level and discipline. This study intends to fill this gap, comparing the lexical bundles used by L1 Chinese and New Zealand PhD students. Our purpose of using a New Zealand thesis corpus as a reference corpus is to identify the features of bundles used in L1 Chinese thesis writing. As the examination of New Zealand theses involves both domestic (i.e., NZ and Australian) and international examiners, New Zealand theses can represent the international practices to

some extent. The purpose of our comparison is not to examine Chinese PhD students' deviations, that is, linguistic deficiency from native-speaker norms. This can be attributed to the following two factors. First, PhD theses are the final products of advanced research students and after many years of learning and practice these apprentice writers should have already become emerging researchers who are competent in academic writing and are able to present their voices or join the conversations in their academic discourse communities. The constant feedback received from their supervisors, peer students, journal reviewers and any other researchers in their fields further enhances the quality of their language production. Second, native speakers also require "prolonged formal education" to familiarize themselves with the conventions of academic writing and to obtain their writing competence, as "academic English is no one's first language" (Hyland, 2019, p. 19). In addition, some theses in the New Zealand corpus might not be written by L1 English writers. It is difficult to identify L1 theses, although we tried to exclude L2 theses by writer names and thesis titles. Therefore, this study intends to shed light on the different academic writing practices or preferences in Chinese and New Zealand contexts by comparing PhD theses. Lexical bundles as recurrent multiword expressions are chosen as the focus of this study because these shared combinations are more likely to reveal the shared language features and communicative purposes of the studied genre (Cortes, 2015; Hyland, 2008). The following research questions are used to focus our analysis:

- 1) What are the most frequent 4-word sentence initial bundles in the Chinese and the New Zealand thesis corpus?
- 2) What are the discourse functions served by the bundles in each corpus?
- 3) What are the similarities and differences in the selection of specific bundles between the Chinese and the New Zealand PhD students?

## 2. Methodology

### 2.1. Corpus collection

We collected PhD theses submitted from 2000 to 2013 in the discipline of General and Applied Linguistics from a number of universities in both China and New Zealand. The Chinese PhD theses were randomly selected and downloaded from one of the most prominent and accessible academic databases in China: Wanfang Data Knowledge Service Platform (<http://www.wanfangdata.com.cn/>). The New Zealand PhD theses were downloaded from the university library websites and only open-access New Zealand theses were collected for this study. The New Zealand theses that appeared to have been written by L2 (particularly Chinese) authors were carefully excluded based on author names and thesis titles as the writing of these students was more likely to reflect the features of other learning contexts. Only the main part of the theses (i.e., introduction, body, and conclusion) was included in the corpora, while other sections such as acknowledgements, the table of contents, abstracts, references, and appendixes were removed. Both corpora (see Table 1) are the same size, so the raw frequencies were used in this paper for comparison. The Chinese PhD theses on average are 30% shorter than the New Zealand theses, but similarities such as genre, discipline, and level of writing ensure broad comparability of these two corpora.

**Table 1.** Corpus collection

	CH PhD	NZ PhD
<b>Number of universities</b>	12	5
<b>Number of theses</b>	67	46
<b>Total words</b>	c.3,800,000	c.3,800,000
<b>Average length</b>	57,232 words	82,609 words

## 2.2. Bundle identification

The focus of this study is on sentence initial bundles as sentence initial bundles serve as the departure point of messages to locate and orient the clauses (e.g., *It should be noted*), while non-initial ones complete clauses or provide additional information (e.g., *the extent to which*) (Cortes, 2013). FLAX (<http://flax.nzdl.org>), a self-access language analysis and learning system documented in Witten et al. (2013), was used in this study to automatically generate lexical bundles. There are two differences between FLAX and the programmes or programme settings used in many other bundle studies (e.g., Biber et al., 2003; Cortes, 2004; Hyland, 2008). 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 generates both uninterrupted word sequences and sequences containing a non-sentence-boundary punctuation mark (mostly commas).

The key criteria for bundle generation are the length of word combinations, the frequency threshold, and the breadth of distribution. As in most previous studies, 4-word bundles were identified as target bundles because they are of a sufficient length to present productive grammatical structures (Cortes, 2004; Hyland, 2008). 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 sequences that occurred over 5 times per million words (i.e., 19 times in each corpus) and across more than 5 texts were included.

## 2.3. Bundle analysis

Biber and his colleagues' taxonomy (Biber et al., 2003, 2004) and Hyland's (2008) framework have been employed in a wide range of bundle studies. Biber et al.'s (2003, 2004) taxonomy was developed from both written and spoken data with categories indicating features of spoken language (e.g., personal epistemic bundles, desire bundles), which was hard to apply to our written data. Based on the classification of Biber et al. (2003, 2004) and Halliday's (1994) linguistic metafunctions, Hyland formed his framework to reflect the unique features of research-focused writing. This framework was the starting point of our PhD thesis analysis. However, during the analysis, we found Hyland's other framework, the interactive and interactional model of metadiscourse (2005a, 2005b, 2018) was a better fit for the majority of our bundle data. His metadiscourse model is also developed from Halliday's (1994) work so it is similar to Hyland's bundle framework. At the same time, this model includes a more detailed classification under Hyland's bundle categories: text-oriented and participant-oriented bundles. On the other hand, the analysis of metadiscourse, as Ädel and Mauranen (2010) argue, often extends beyond pre-determined small search terms and covers larger chunks. Therefore, we attempted to adapt the metadiscourse model to interpret our bundle data. The use of the metadiscourse model could possibly reveal the connections between the two concepts (e.g., the bundle *It should be noted* can be regarded as a directive device) and enable academic writers to view the generated sentence initial bundles as metadiscourse devices, the devices used to facilitate readers to decode the texts.

The first two authors worked independently to code the data. Every bundle was examined against its contexts and the primary function was assigned to each bundle. As a few bundles serve both interactive and interactional functions, it is difficult to identify their primary functions. Two of them, *In this chapter/section I*, contained both frame markers *In this chapter/section* and the first-person pronoun *I*. The other three, *As/It can be seen and Look at the following*, functioned as both endophoric markers to refer to other parts of the text and directives to guide the reader. These five bundles were regarded as both interactive and interactional bundles in this study. This categorization is in line with the inclusion of bi-dimensional metadiscourse devices in Bouziri's (2021) recently-developed tripartite metadiscourse model, which distinguishes organising metadiscourse, involving/evaluative metadiscourse, and bi-dimensional metadiscourse. The inter-coder reliability was 86% and the disputed cases were resolved in the discussion.

### 3. Sentence initial bundles in thesis writing

#### 3.1. Bundle distribution

We generated 46 sentence initial bundles, totalling 2,071 occurrences (mean = 45.02, standard deviation = 39.32) in our Chinese PhD corpus and 42 bundles, altogether 1,418 occurrences (mean = 33.76, standard deviation = 19.51) in our New Zealand corpus. The Chinese PhD students relied on sentence initial bundles to a greater extent than their New Zealand counterparts ( $P < 0.05$ ). This finding is consistent with the findings of many comparative studies between L1 and L2 writing, such as Bychkovska and Lee (2017), Lu and Deng (2019), Pan et al. (2016). The heavier reliance on sentence initial bundles of L1 Chinese writers might be a result of their conscious avoidance of self-coined expressions (Bychkovska & Lee, 2017) or a result of their sentence-building strategy to write more efficiently from these semi-fixed blocks, their *islands of reliability* (Dechert, 1984). Altogether 15 bundles (i.e., 34%) were shared between the two corpora and most of them were interactive bundles. These shared bundles reflected the shared rhetoric needs between these two groups of PhD students for their thesis writing, which will be discussed under each metadiscourse category. The Appendix lists all the generated bundles along with their raw frequencies and the shared ones were marked in bold.

Table 2 presents the distribution of sentence initial bundles in each metadiscourse category. The majority of sentence initial bundles fell into the interactive subset, while a few bundles acted as interactional devices. There were no bundles indicating shared knowledge, reader pronouns, personal asides and questions as in Hyland's (2005b) interactional model. The eight bundles categorized as *others* can be regarded as research-oriented bundles in Hyland's (2008) bundle framework. They are not the focus of our study as five of them (*In the present study*, *The present study is*, *All of the participant*, *There was a/no significant*) are content-specific and the other three (*At the end/time/beginning of*) are time indicators external to the discourse. In other words, they are not metadiscourse devices (Hyland, 2018).

The total number of bundle occurrences in each category is listed in Table 2 but we chose not to focus on quantitative comparisons as it is difficult to compare the over/underuse of metadiscourse devices from a bundle perspective. Most bundle-generation programmes generate lexical bundles up to a punctuation mark. FLAX identifies sentence boundaries but ignores punctuation markers (mainly commas) within sentences. The different generation approach resulted in a number of sentence initial chunks with internal commas (e.g., *As a result, the*; *To sum up, the*; *In this section, I*; *In other words, the*; *For example, in the*). This

occurred more often in our Chinese PhD corpus. We manually removed these chunks from our final lists as lexical bundles are “uninterrupted sequences of words” (Biber et al., 2003, p. 74). But the metadiscourse devices embedded in such chunks suggest the potential pitfalls of a quantitative comparison: both shorter commonly used metadiscourse expressions (e.g., *as a result, in other words, for example*) and word sequences containing punctuations (*In this section, I* vs. *In this section I*) were excluded. At the same time, the inconsistency between researchers in categorizing bundles and multifunctionality of many bundles add to the difficulties in classifying bundles (Ädel & Erman, 2012). Therefore, we focused on bundle selections of the two PhD groups rather than testing the differences statistically.

**Table 2.** Distribution of sentence initial bundles

Categories		Examples	CH PhD Occurrences	NZ PhD Occurrences
Interactive	Transition bundles	<i>On the other hand</i>	515	304
	Frame bundles	<i>In the case of</i>	450	312
	Code gloss bundles	<i>That is to say</i>	243	44
	Endophoric bundles	<i>As discussed in Chapter</i>	395	232
Total			1603	892
Interactional	Attitude bundles	<i>It is important to</i>	81	116
	Hedge bundles	<i>It is possible that</i>	52	147
	Booster bundles	<i>It is clear that</i>	169	49
	Self-mention bundles	<i>In this chapter I</i>	0	52
	Directive bundles	<i>It should be noted</i>	192	76
Total			494	440
Others		<i>All of the participants</i>	95	190

### 3.2. Interactive bundles

Interactive bundles acted as discourse frames to create a sense of cohesive and coherence flow and to guide the readers throughout texts (Hyland, 2018). In the two PhD corpora, we identified transition bundles, frame bundles, code gloss bundles, and endophoric bundles as follows.

#### 3.2.1. Transition bundles

Transition bundles are expressions highlighting internal relations between units of texts, either comparison, addition or consequence. The majority of transition bundles (i.e., 5/6) were shared between Chinese and New Zealand PhD corpora. These were *On the other hand*, *On the one hand*, *In addition to the*, *At the same time*, and *As a result of*. The only unshared bundle in the Chinese corpus was *In spite of the* and in the New Zealand corpus was *In contrast to the*.

The bundle *On the other hand* was the most frequent bundle in both corpora, shared between the two corpora but used differently by the two groups of students. More sentence initial *On the other hand* bundles were identified in the Chinese student corpus (246 vs. 108) and in the New Zealand student writing dataset this pattern often immediately followed the

subject of the sentence (subjects were underlined in the following two examples), as in examples (1) and (2).

- (1) *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 mid twenties. 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)*
- (2) *By claiming that her view is in line with the already accepted research findings, the writer on the one hand shows her expertise (sic) and thus expresses confidence in the validity of her finding by evaluating it as consistent. **On the other hand**, the room for negotiation of claims is deliberately limited by imposing attitudes, interpretations, and assessments of truth-value, and by predisposing the reader towards certain references. (CH PhD)*

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 and Bizup (2014) advise writers to start their sentences with subjects/topics and to choose familiar and related concepts as subjects of their sentences to form “a strong topic string” (p. 74). In examples (1) and (2), the New Zealand PhD student began his or her sentence with the short simple noun phrase *The women*, 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*, formed a set of related concepts. In other words, the New Zealand student relied on the noun phrase *The women* together with the transition marker *on the other hand* to create a sense of cohesive and coherent flow. The Chinese PhD student, however, solely depended on the transition markers *on the one hand* and *on the other hand* to connect sentences, with *the writer* as the subject and topic of the first sentence; *the room for negotiation of claims*, a new and unrelated concept as the topic of the second sentence.

### 3.2.2. Frame bundles

Frame bundles function as signposts, signalling the boundaries of arguments (e.g., *In this chapter I*), introducing/sequencing the stages of texts (e.g., *The first of these*) or specifying limiting conditions (e.g., *In the case of*). According to these three functions, frame bundles were further classified into boundary bundles, stage bundles, and condition bundles. The New Zealand PhD students used two boundary bundles (*In this chapter/section I*), three stage bundles (*There are a number (of)*, *The first of these*, *This is followed by*), and six condition bundles. Four of the condition bundles (*In the case of*, *On the basis of*, *In terms of the*, *With regard to the*) were shared with the Chinese students and two (*For the purposes of*, *In the context of*) were unique ones. No boundary bundles and stage bundles were generated from the Chinese PhD writing dataset. However, a close examination showed that the Chinese PhD students also used recurrent boundary and stage markers but many of their sentence initial ones were cut off by commas (e.g., *In this section*, *I/the*, *To sum up*, *the*), and cannot be regarded as 4-word bundles. Condition bundles were popular among the Chinese students. Besides the four shared ones, another six (*As far as the [..... is/are concerned]*, *From the*

*perspective of, With respect to the, For the sake of, In the field of, When it comes to*) were used at the beginning of sentences. The Chinese students tended to specify the frames before presenting their statements. The popularity of frame bundles among Chinese PhDs is divergent with the finding of Pan et al. (2016) and one of the reasons might be the cut-off of the commas.

### 3.2.3. Code gloss bundles

Hyland (2007) distinguishes two functions of code gloss markers: reformulation and exemplification. All the code gloss bundles in our study functioned to reformulate or rephrase the previous message and no example was introduced by the bundles. No code gloss bundle was shared between the two groups. The Chinese students employed five code gloss bundles (*That is to say, To be more specific, To put it another, This suggests that the, This means that the*) to elaborate their statements. Two of them, *That is to say* and *To put it another (way)*, conveyed a sense of equivalence between the preceding and succeeding text. The equivalent, probably simpler or more exact information was introduced by the bundles to enhance the reader's knowledge construction. The bundle *To be more specific* supplied the reader with more detailed information. The other two bundles *This suggests/means that the* expanded the previous statement with an explanation.

The New Zealand student chose another two different bundles *This is not to/a*, not to elaborate but to modify the anaphoric argument, to reject the possible alternative interpretation (3).

- (3) *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)*

### 3.2.4. 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, and so forth (e.g., *As shown in Table*). Three bundles (*As can be seen, The results of the, The analysis of the*) were shared. Besides these three, the major difference between the two groups of students was the New Zealand students deployed shell nouns and demonstrative *this* in their bundles (e.g., *The results/purpose/aim of this*) to specify what they referred their readers to and to closely link their texts. Shell nouns are abstract nouns that carry little or no meaning but operate to encapsulate the meaning from the surrounding clauses or phrases (Aktas & Cortes, 2008). The research-related shell nouns identified in our sentence initial bundles facilitated the student writers to semantically characterize and conceptualize their research process and outcomes, and at the same time, linked to the preceding or succeeding discourse as cohesive devices. As shown in example (4), *The results of this* introduced the different effects of the Deductive and Inductive groups on the comprehension tasks:

- (4) ***The results of this** study show that there were greater gains for students in the Deductive group than for those in the Inductive group on comprehension tasks. (NZ PhD)*



Shell nouns were also used by the Chinese students, but they employed a number of multi-reference bundles (e.g., *It can be seen*, *As is shown in*, *The following is/are a/an/some*) to steer their readers to a fairly wide variety of items including sections of texts, examples, or figures (e.g., *The following is a/some discussion/examples/diagram*). These more generic bundles did not allow writers to characterize/conceptualize the referred text and did not provide the readers with useful information to predict what is referred to. In Li et al. (2018), we reported our Chinese postgraduate participants' little awareness of the functions of shell nouns even though these nouns were spotted in their writing. They resorted to the shell nouns as temporary choices because they could not find a better sentence structure at the time of writing, or they deliberately selected different nouns of a similar meaning to avoid word repetition.

### 3.3. Interactional bundles

Interactional bundles acted to foster solidarity between the writer and the reader by means of expressing the writer's community-recognized personality and inviting the reader to jointly interpret the presented research information (Hyland, 2018). Not many bundles were identified as interactional devices, but they played important discourse functions as attitude markers, hedges, boosters, self-mention markers, and directives.

#### 3.3.1. Attitude bundles

Attitude bundles express one's subjective evaluation of one's arguments or personal feelings towards one's research-related experiences. Most sentence initial bundles in this category (e.g., *It is important/necessary/interesting/difficult to*) were part of anticipatory-*it* clause used to depersonalize the writer's attitudes and to emphasize the statements placed at the end of sentences (Biber et al., 1999; Groom, 2005; Hewings & Hewings, 2002). Although the grammatical structure was the same, the Chinese and New Zealand PhD students selected different adjectives to serve different functions. The only shared bundle is the *important* bundle. Besides the *important* bundle, two different personal feeling bundles, *It is interesting/difficult to*, were generated from the New Zealand corpus. The bundle *It is interesting to*, mainly used with infinitive verb *note* and its *that*-clause, showcased the writer's enthusiasm about their findings, as in (5):

- (5) *It is interesting to note that both groups produced more gestures than head-movement, more head-movement than facial expression/gaze and more facial expression/gaze than body-movement.* (NZ PhD)

The *difficult* bundle, on the other hand, was usually followed by a range of activity verbs (e.g., *compare*, *distinguish*, *identify*, *explain*, and *envision*) to indicate different obstacles or challenges encountered during selecting, comparing, interpreting, and evaluating data in the process of research, as the example (6) demonstrates:

- (6) *It is difficult to distinguish individuals who are more likely than others to have brought these innovations into their community.* (NZ PhD)

Given the presence of personal feeling bundles in the New Zealand theses, there was no personal feeling bundle found in the Chinese student texts. These student researchers, unlike New Zealand students, appeared to be conservative in revealing their feelings. As Yang

(2013) argues the general belief among Chinese scientists is that “scientific discourse should be impersonal, serious, objective, and factual” and it is “inappropriate and non-authoritative” to express personal feelings in research writing (p. 30). This is also supported by the interview finding of Li et al. (2018), in which their postgraduate participant considered academic writing as “statements of objective facts rather than subjective personal arguments” (p. 11).

Two other bundles *It is necessary to* and *It is hoped that* were favoured by the Chinese students. The use of *necessary* bundle indicated the students’ keenness to justify their own decisions on research and had a stronger orientation towards research activities (e.g., *conduct, provide, introduce, find out, and point out*) as in commenting on what the researcher should do or has done, as in (7):

- (7) ***It is necessary to*** *conduct more extensive studies to corroborate the above findings, and to clarify on what conditions recasts can function as a facilitator of interlanguage development.* (CH PhD)

The bundle *It is hoped that* often occurred in the introduction chapter of the Chinese PhD theses. The mental attitude verb *hope* (Pearson, 2020) was used to indicate the student researcher’s subjective expectation of the potential values of his or her study (8):

- (8) ***It is hoped that*** *the findings of the present study will add to the body of literature currently available by examining the role of feedback or a combination of feedback types in ESL/EFL writing, and that the study will pave the way for subsequent research.* (CH PhD)

### 3.3.2. Hedge bundles

Hedge bundles address the writers’ uncertainty and express their cautiousness towards making claims or statements. No hedge bundle was shared between the two corpora. The New Zealand students used four bundles *It is possible that*, *It is also possible*, *It is not clear*, and *It may be that*. Three of them were anticipatory-*it* clauses embedded with possibility adjectives. The bundle *It is possible that*, as the second most popular bundle in the New Zealand corpus, was used in a wide variety of contexts to predict research findings or contradictory findings (9), to suggest alternative approaches (10), or to provide possible pedagogical implications (11):

- (9) ***It is possible that*** *the results of the studies may have been different if another television program had been utilized.* (NZ PhD)
- (10) ***It is possible that*** *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)
- (11) ***It is possible that*** *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)

No *possible* bundle was found in the Chinese PhD corpus. Alternatively, the Chinese students showed their reliance on verbs to tone down their voices with the use of *It seems that the* and *It is argued that*. This extends Yang's (2013) finding of the different weight placed on epistemic adjectives and lexical verbs between her English-medium research article corpus and her Chinese-authored English-medium research article corpus. The Chinese PhD writers in our study also relied on reporting verbs rather than adjectives to express their levels of certainty.

### 3.3.3. *Booster bundles*

Booster bundles express the writers' certainty towards their propositions and are used to silence the alternative voices. By manipulating the weight of hedges and boosters, the writer balances "objective information, subjective evaluation and interpersonal evaluation" (Hyland, 2005b, p. 180).

The bundle *It is clear that* was the only booster bundle shared between the New Zealand and Chinese PhD students. Another booster bundle in the New Zealand student writing was *The fact that the*. Booster bundles occurred in the Chinese student writing were *It is true/obvious that*, *As a matter of (fact)*, *The results showed that*, and *It should be pointed (out)*.

The extensive and diverse use of the abstract noun *fact* between the two corpora deserved a close examination. In the New Zealand student bundle *The fact that the*, the noun *fact* was used as a shell noun (Aktas & Cortes, 2008) or metadiscursive noun (Jiang & Hyland, 2017). It enclosed its meaning from the succeeding appositive *that*-clause and performed both interactional and interactive functions: semantically characterising the statement in its *that*-clause as unarguable objective evidence and linking to the succeeding statement, as illustrated in example (12):

- (12) ***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)

The Chinese students also used the same word *fact* to emphasize the objectivity of their statements: their bundle *As a matter of* was part of the idiom *As a matter of fact*. As shown in example (13), this *fact*-embedded idiom introduced the objective information *a generally similar pattern has been identified from the data* to further support its preceding statement:

- (13) ***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)

A quick search of the two bundles (*The fact that the* and *As a matter of fact*) in the BAWE corpus<sup>1</sup> (Nesi & Gardner, 2012), a corpus of good standard university student assignments

---

<sup>1</sup> The British Academic Written English (BAWE) corpus 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 (RES000-23-0800).

collected from universities in the UK, showed a substantial difference in their occurrences: 47.14 per million for the former and 1.56 per million for the latter. In comparison to *As a matter of fact*, *The fact that the* occurs far more frequently in academic writing. As to L2 writing, Chen and Baker (2010) and Ädel and Erman (2012) found *The fact that the* was less commonly used in Chinese and Swedish student writing. Ädel and Erman (2012) consider *The fact that the* as a “complex structure” and suggest it is likely to be learned later (Ädel & Erman, 2012, p. 88). In our study, the Chinese PhD students’ use of the idiom *As a matter of fact* might be their alternative option of this complex structure.

### 3.3.4. Self-mention bundles

Two bundles embedded with the first-person pronoun *I*, *In this chapter/section I*, were generated from the New Zealand PhD writing and the PhD writers acted as a discourse guide to introduce or summarize the main points in a particular chapter or section. There was no self-mention bundle in the Chinese PhD corpus, but first-person pronouns *I* and *we* also appeared in the Chinese PhD corpus. The reason is that the Chinese PhD writers used a comma after *in this chapter/section*, which cut off their lexical chunks (e.g., *In this chapter/section, we*; *In this section, I*). By comparing these chunks with the self-mention bundles in the New Zealand corpus, we can see that the Chinese students tended to use editorial *we* (Ädel, 2006) for self-reference. Hyland (2001) and Kuo (1999) interpret this feature as a way of reducing personal attribution and claiming less individualised authorial identity through the use of the plural form. The Chinese students may not feel comfortable or confident constructing a strong personal authorial identity as emerging researchers.

### 3.3.5. Directive bundles

Directive bundles express a writer’s guidance to help the readers see or interpret the texts in the writer expected way. Two bundles were shared by Chinese and New Zealand PhD students: one mental verb *note* bundle (*It should be noted*) and one sense verb *see* bundle (*As can be seen*). The use of *should* in the *note* bundle claims the higher authority of the writer by focusing the reader’s attention on a particular point. The bundle *As can be seen* was usually followed by the preposition *from* or *in*, steering the reader to a number of sources including tables, figures, examples, and results.

Besides these two bundles, the Chinese PhD students used another three sense verb bundles *It can be seen*, *We can see that*, and *Look at the following*. Both *see* bundles were used to announce the writer’s conclusion from the data or literature and invite the reader to view it through the writer’s eyes. Both bundles were followed by a *that*-clause to state the conclusion. The bundle *Look at the following* was used in the Chinese PhD corpus to introduce the example(s) as in *Look at the following example(s)*.

## 4. Discussion

The present study compares the sentence initial bundles in New Zealand L1 and Chinese L2 thesis writing using Hyland’s model of metadiscourse (2005a, 2005b, 2018). One-third of the bundles are shared between the two corpora (e.g., *On the other hand*, *The results of the*, and *As can be seen*) and a wider range of bundles are unique to the Chinese or New Zealand thesis writing (e.g., *The following is a*, *It is interesting to*, and *As a matter of*). Most sentence

initial bundles are interactive bundles, framing the texts; while a small number of bundles act as interactional devices, facilitating writer-reader interactions.

Unlike many previous studies, we did not use inferential statistics to infer the differences in bundle distribution between the two corpora because we were not able to estimate the use (either underuse or overuse) of metadiscourse devices from a bundle perspective. Only recurrent multiword combinations occurring at the beginning of sentences were generated as the sentence initial bundles in our study, which excluded many other metadiscourse devices such as individual words (e.g., *also*, *surprisingly*) or shorter word combinations (e.g., *defined as*, *tend to*) as presented 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 suggested in Cortes's (2013) lexical bundle study. The focus on sentence initial position also left out the resources used in other parts of sentences (e.g., *the extent to which*). Moreover, other factors such as the variation in thesis length (57,232 vs. 82,609 words) and the use of punctuation (e.g., *In this section*, *I*) affected the results of quantitative comparisons.

We chose to closely examine individual bundles within their contexts with regard to their positions within sentences, any specific functions served by bundles, and word selections or structural preferences. The most frequent bundle in both corpora, transition bundle *On the other hand*, more often occurs as the first element of sentences in the Chinese student writing and New Zealand students tend to start their sentences with subjects/topics, followed by this bundle, relying on related concepts to create a sense of flow. Frame bundles prove to be a general category and bundles in this category can be further categorized into boundary bundles (e.g., *In this chapter I*), stage bundles (e.g., *The first of these*), and condition bundles (e.g., *In the case of*). The code gloss bundles in the Chinese corpus (e.g., *That is to say*, *To be more specific*, *This suggests that the*) are used to elaborate the anaphoric statements, while the ones in the New Zealand corpus (*This is not to/a*) act to narrow down the possible alternative interpretations. The endophoric bundles of the New Zealand corpus (e.g., *The results/purpose/aim of this*) are featured by the use of shell nouns and demonstrative *this* to specify the references, while a number of multi-reference bundles (e.g., *It can be seen*, *As is shown in*, *The following is/are a/an/some*) are employed by the Chinese students. Besides shell nouns, the two groups of students indicate their different preferences of adjectives or verbs in expressing their stance: the New Zealand students use *It is interesting/difficult to* to reveal their feelings and *It is possible that* to hedge their claims and Chinese students select a range of verb-embedded bundles such as *It is hoped/argued that*, *The results showed that*, and *It should be pointed (out)* to express their attitudes and certainty towards the arguments. The different weight placed on the first-person pronouns *I* and *we* possibly shows their degree of comfort or confidence in voicing their authorial identity. Even the same word *fact* is used differently – as a shell noun in the New Zealand student bundle *The fact that the* and as part of an idiom in the Chinese student bundle *As a matter of (fact)*.

Two more differences between this study and many previous ones are the focus on sentence initial bundles and the use of a metadiscourse model. The focus on sentence initial bundles might be the reason that drives us to choose a metadiscourse model as writers tend to orient or set the scene at the beginning of their sentences. The use of a metadiscourse model allowed us to explore metadiscourse devices by taking a bottom-up corpus-driven approach, which starts with lexical bundles, the word sequences identified solely on the basis of distribution criteria (Biber, 2015). This “completely corpus-driven” (Granger, 2018, p. 192) approach, in which lexical bundles are viewed as components of metadiscourse resources, can potentially add to previous metadiscourse studies by verifying existing researcher-generated metadiscourse lists and functions, and by including bundle-based metadiscourse items and unearthing new functional categories.

## 5. Conclusion and implications

The bundle comparison between New Zealand L1 and Chinese L2 thesis writing improves our understanding of the research genre (i.e., PhD thesis), suggests how the two groups of students realize their interaction with their readers by means of lexical bundles, and examines the possibility of analysing lexical bundles from the perspective of metadiscourse. The PhD thesis, as a book-length work, requires a large number of interactive bundles to signpost the reader throughout the long journey; as a high-stakes genre, it is intrinsically objective and authoritative with much fewer interactional bundles. Both New Zealand and Chinese PhD students have successfully presented their research to their target audiences in two different contexts. However, compared to the New Zealand counterparts, as L2 writers and in the Chinese academic community, the Chinese PhDs tend to rely more on set phrases (e.g., to link sentences, to frame statements), be limited by a smaller range of expressions (e.g., the use of multi-reference bundles), and act more conservatively on revealing subjective feelings and authorial presence (e.g., the lack of personal feeling bundles and little use of *I*). The avoidance of self-coined expressions, the belief in the objectivity of scientific discourse, and an underdeveloped authorial voice might all contribute to their bundle use.

According to Biber and Conrad (2019), linguistic choices are associated with the situational context of use. The Chinese doctoral students' bundle selection might not hamper the communication within the Chinese discourse community, but could possibly cause confusion for a wider audience internationally and limit future publication possibilities. New Zealand theses, as a representative of international practices, can be used for their reference. The bundle lists generated from this study, together with metadiscourse functions, can be resorted as thesis writing resources. Sentence initial bundles, particularly the ones with shell nouns and anticipatory-*it* structure, can be introduced as sentence starting strategies to Chinese PhD students. Personal feeling bundles and self-mention bundles can also be highlighted to deliberately raise Chinese emerging researchers' attention to the importance of authorial presence in academic writing. The access to the corpus-based language learning tool that affords L2 students bundle lists with frequency-based displays, multiple contexts, and typographical salience (Franken, 2014) allows student writers to act as language researchers to interact with any in-built specific corpus, to explore the metafunctions of sentence initial bundles, and to choose appropriate ones to realize their communication purposes.

The present study only focuses on four-word sentence initial bundles in PhD theses. Bundles produced by doctoral students of various lengths and at other parts of sentences deserve equal attention considering the crucial roles of these recurrent multiword combinations in facilitating language production. Our study did not consider the influence of research paradigms (i.e., quantitative, qualitative, and mixed methods research) on bundle production and future studies can compare L1 and L2 bundles in relation to different paradigms.

## References

- Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. 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. <https://doi.org/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.  
<https://doi.org/10.1016/j.jeap.2008.02.002>
- Biber, D. (2015). Corpus-based and corpus-driven analysis of language variation and use. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 193-224). Oxford University Press.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.  
<https://doi.org/10.1016/j.esp.2006.08.003>
- Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2 ed.). Cambridge University Press. <https://doi.org/10.1017/9781108686136>
- 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*. 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.  
<https://doi.org/10.1093/applin/25.3.371>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
- Bouziri, B. (2021). A tripartite interpersonal model for investigating metadiscourse in academic lectures. *Applied Linguistics*, 42(5), 970-989.  
<https://doi.org/10.1093/applin/amab001>
- Bychkovska, T., & Lee, J. J. (2017). *At the same time*: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, 30(Supplement C), 38-52. <https://doi.org/10.1016/j.jeap.2017.10.008>
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423.  
<https://doi.org/10.1016/j.esp.2003.12.001>
- Cortes, V. (2013). *The purpose of this study is to*: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, 12(1), 33-43. <https://doi.org/10.1016/j.jeap.2012.11.002>
- Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum. In V. Cortes & E. Csomay (Eds.), *Corpus-based research in applied linguistics: Studies in honor of Doug Biber* (Vol. 66, pp. 197-216). John Benjamins Publishing Company.  
<https://doi.org/10.1075/scl.66>
- 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.  
<https://doi.org/10.1016/j.jslw.2007.07.002>
- Dechert, H. W. (1984). Second language production: Six hypotheses. In H. W. Dechert, D. Mohle, & M. Raupach (Eds.), *Second language productions* (pp. 211-230). Gunter Narr Verlag.
- Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes*, 29, 21-42.  
<https://doi.org/10.1016/j.jeap.2017.09.002>

- Franken, M. (2014). The nature and scope of student search strategies in using a web derived corpus for writing. *The Language Learning Journal*, 42 (1), 85-102.  
<https://doi.org/10.1080/09571736.2012.678013>
- Granger, S. (2018). Tracking the third code: A cross-linguistic corpus-driven approach to metadiscursive markers. In A. Čermáková & M. Mahlberg (Eds.), *The Corpus Linguistics Discourse: In honour of Wolfgang Teubert* (pp. 185-204). Benjamins.
- Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes*, 4(3), 257-277.  
<https://doi.org/10.1016/j.jeap.2005.03.002>
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold.
- Hewings, M., & Hewings, A. (2002). "It is interesting to note that ...": A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes*, 21, 367-383.
- Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, 20(3), 207-226. [https://doi.org/10.1016/S0889-4906\(00\)00012-0](https://doi.org/10.1016/S0889-4906(00)00012-0)
- Hyland, K. (2005a). *Metadiscourse: Exploring interaction in writing*. Continuum.
- Hyland, K. (2005b). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192.
- Hyland, K. (2007). Applying a gloss: Exemplifying and reformulating in academic discourse. *Applied Linguistics*, 28(2), 266-285.
- Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62.  
<https://doi.org/10.1111/j.1473-4192.2008.00178.x>
- Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32, 150-169. <https://doi.org/http://dx.doi.org/10.1017/S0267190512000037>
- Hyland, K. (2018). *Metadiscourse: Exploring interaction in writing* (1 ed.). Bloomsbury Academic. <https://doi.org/10.5040/9781350063617>
- Jiang, F., & Hyland, K. (2017). Metadiscursive nouns: Interaction and cohesion in abstract moves. *English for Specific Purposes*, 46, 1-14.  
<https://doi.org/10.1016/j.esp.2016.11.001>
- Kuo, C.-H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes*, 18(2), 121-138. [https://doi.org/10.1016/S0889-4906\(97\)00058-6](https://doi.org/10.1016/S0889-4906(97)00058-6)
- Li, L., Franken, M., & Wu, S. (2018). Chinese postgraduates' explanation of the sources of sentence initial bundles in their thesis writing. *RELC Journal*, 50(1), 37-52.  
<https://doi.org/10.1177/0033688217750641>
- Lu, X., & Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. *Journal of English for Academic Purposes*, 39, 21-36.  
<https://doi.org/10.1016/j.jeap.2019.03.008>
- Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge University Press.
- Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60-71.  
<https://doi.org/10.1016/j.jeap.2015.11.003>



- Pearson, H. (2020). Attitude verbs. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmerman (Eds.), *The Wiley Blackwell companion to semantics*. John Wiley & Sons Inc.
- 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. <https://doi.org/10.1016/j.jeap.2014.01.002>
- Vande Kopple, W. J. (1989). *Clear and coherent prose: A functional approach*. Scott, Foresman and Company.
- Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC Journal*, 42(2), 155-166. <https://doi.org/10.1177/0033688211407295>
- Williams, J. M., & Bizup, J. (2014). *Style: Lessons in clarity and grace* (11th ed.). Pearson.
- Witten, I. H., Wu, S., Li, L., & Whisler, J. L. (2013). *The book of FLAX: A new approach to computer-assisted language learning* (2nd ed.). University of Waikato.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- 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. <https://doi.org/10.1016/j.pragma.2013.01.008>

## **Appendix. Sentence initial bundles in Chinese and NZ PhD corpora**

Data related to this study can be found at

[https://osf.io/u5edc/?view\\_only=9a33dbdf8cf14ef5acef8f23e6f19e85](https://osf.io/u5edc/?view_only=9a33dbdf8cf14ef5acef8f23e6f19e85).

## **Address for correspondence**

Liang Li  
Te Kura Toi School of Arts  
University of Waikato  
Private Bag 3105  
Hamilton 3240  
New Zealand  
[liangl@waikato.ac.nz](mailto:liangl@waikato.ac.nz)

## **Co-author details**

Margaret Franken  
Ministry of Education  
Mātauranga House  
33 Bowen Street  
Wellington 6011  
New Zealand  
[margaret.franken@gmail.com](mailto:margaret.franken@gmail.com)

Shaoqun Wu  
School of Computing & Mathematical Sciences  
University of Waikato  
Private Bag 3105  
Hamilton 3240

New Zealand  
shaoqun.wu@waikato.ac.nz