Chinese Postgraduates Explanation of the Sources of Sentence Initial Bundles in their Thesis Writing

Liang Li

University of Waikato, New Zealand

Margaret Franken

University of Waikato, New Zealand

Shaoqun Wu

University of Waikato, New Zealand

Abstract

Lexical bundles, recurrent multiword combinations in a register, are extremely common and important discourse building blocks in academic writing. An increasing number of studies have investigated lexical bundles in academic writing in recent years, but few studies have explored L2 learners' interpretations of their own bundle production, particularly sentence initial bundle production. Investigating the sources that have appeared to influence learners' choices and knowledge of bundles is important as it complements what we know about the structural and functional features of lexical bundles and provides useful first-hand information for second language writing pedagogy. The present study interviewed five Chinese postgraduate students to probe possible reasons for their use of the typical sentence initial bundles identified in the self-built Chinese Masters and PhD thesis corpora. The interviews revealed diverse explanations including interlingual transfer, classroom learning, noticing in reading, a lack of rhetorical confidence, and misunderstanding of rhetorical conventions. The results suggest the need for raising students' awareness of the common sentence starters in postgraduate academic writing, increasing their confidence as student writers, familiarizing them with rhetorical conventions, and incorporating effective corpus-based tools into pedagogical practices.

Keywords

Introduction

Multiword combinations have been recognized as an essential aspect of vocabulary knowledge and an important focus for supporting language production (Lewis, 2008; Nattinger and DeCarrico, 1992; Sinclair, 1991; Nation, 2013). Lexical bundles are recurrent multiword combinations of three or more words, identified empirically on the frequency of co-occurrence and distribution across texts (Biber et al., 1999). Bundles, according to their positions in sentences, are divided into sentence initial bundles and non-initial bundles, with the former occurring at the beginning of sentences and the latter, at the second part of sentences.

Many scholars have investigated the discrepancies between L2 (second language) students and L1 (first language) or advanced writers in their bundles choices (e.g. Ädel and Erman, 2012; Chen and Baker, 2010; Hyland, 2008a; Wei and Lei, 2011; Allen, 2009; Staples et al., 2013). The research on lexical bundles has been very text focussed. 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 student bundle choices (e.g. Cortes, 2004; Hyland, 2008a; Wei and Lei, 2011; Allen, 2009; Qin, 2014; Paquot, 2013; Ruan, 2016). 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 suggest the reasons for the 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 and Lei, 2011), learning experience (Hyland, 2008a; Wei and Lei, 2011; Ruan, 2016), cultural preference (Wei and Lei, 2011; Hyland, 2008a), 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. Ruan (2016) suggests the main reason for the increase of discourse organizers in the undergraduate writing is English for Academic Purposes (EAP) coursework. Hyland (2008a), and Wei and Lei (2011) attribute impersonality in Chinese student writing to the teaching materials and practices, and to cultural preferences. Hyland (2008a) considers confidence as another factor for the absence of stance bundles in dissertations written by PhD students who are natives of Hong Kong. Allen (2009) and Qin (2014) suggest the length of student writing possibly affects the number of text organizers (e.g. *in the next section*). Allen (2009) also acknowledges the role of linguistic transfer, and as an example

cites the bundle *it can be said* in his Japanese student corpus, as the result of interlingual transfer from the Japanese expression *to iwareru*. Qin (2014) adds audience awareness as another reason for student bundle production by arguing that raising students' awareness will possibly help them to produce clearer and more consistent writing.

Corpus linguists can postulate reasons and make hypotheses, but evidence of their interpretations cannot be obtained from corpus-based analysis, and can only be collected using other methods, such as interviews. This is because 'corpus data does not interpret itself' (Baker, 2006: 18). There is a need to interrogate text users to 'understand how and why language users make the choices they do when they speak/write' (Hyland, 2011: 106).

The present study, therefore, intends to address this gap by interviewing a group of Chinese postgraduates to elicit and identify these L2 learners' interpretations of their own bundle choices, particularly selections of sentence initial bundles in thesis writing. The research question posed was:

What reasons do Chinese postgraduates give for their sentence initial bundle choices

in their own thesis writing?

It is important to determine the sources of learner bundles as this can complement the existing knowledge gained from the bundle studies and provide useful first-hand information for English as a second language (ESL) or English as a foreign language (EFL) writing pedagogy.

Methodology

Interview Data Collection

This is a follow-up study of a corpus-based bundle analysis between Chinese L2 and New Zealand L1 postgraduate thesis writing (i.e. Li, 2016), in which typical sentence initial bundles in Chinese student thesis writing were identified by comparing four self-built corpora: a Chinese Masters thesis corpus, a New Zealand Masters thesis corpus, a Chinese PhD thesis corpus and a New Zealand PhD thesis corpus. These four corpora contain theses submitted from 2000 to 2013 in the discipline of General and Applied Linguistics. One-on-one interviews were conducted on the basis of Li's (2016) bundle analysis. Five Chinese postgraduates, one Masters student and four PhD students, studying social science at the University of Waikato were recruited as interviewees because the academic writing in social science bears many similarities (Hyland, 2008b). We would have liked more interviewees, particularly Masters students, to have had a better representation and a wider coverage of the corpus findings. We would have liked to recruit interviewees within the same discipline of our corpus data, but it proved to be extremely difficult to recruit interviewees at our university. However, the shared written work of our interviewees is comparatively long, 10,160 words on average, and our interviewees, as postgraduates, all have many years of experience in learning and teaching writing.

Therefore, it is possible to identify a number of overlapping bundles and common features in their writing and to obtain broad interpretations from them.

The original drafts with no editing from supervisors or other language tutors were collected. The related expressions between corpus findings and interviewees' writing were manually identified. These expressions completely or partially overlap with the generated sentence initial bundles (e.g. *In order to make* and *In order to make*, *Last but not least* and *The last but not least*), convey the same meaning with them (e.g. *There is no doubt* and *It is undoubted that*), or bear the same feature as the use of bundles in the Chinese Masters and PhD thesis corpora (e.g. *underuse of shell noun bundles*). Table 1 presents the bundles in the Chinese postgraduate thesis corpora and the expressions in the interviewees' writing. These related expressions will be illustrated with contextualized sentences from both the Chinese postgraduate thesis corpora and interviewees' texts in the next section. The reasons for choosing them will also be provided in the form of interview transcripts from the Chinese interviewees.

Table 1. Related Expressions from Interviewees' Writing.

Bundles in Chinese Student Corpora	Expressions in Chinese Interviewees' Writing
(In order) to make/get/find	*In order to make, to-infinitive phrases*
There is no doubt	*It is undoubted that*
It is obvious/clear that	*Clearly*
Last but not least	*The last but not least*
Underuse of shell noun bundles, e.g. *The use of the*	Only a few shell noun phrases identified, e.g. *The complaints from my colleagues and the results of the meetings, the definition of old, The common measurement, The identity of old age*
With the development of	*With the development of*
So it is necessary	sentence initial conjunctions
Underuse of *note* bundles, e.g. *It should be noted*	The use of *It is interesting to find*, instead of *It is interesting to note*
Underuse of attitude bundles, e.g. *It is interesting to*	Only a few expressions contained *interesting*, e.g. *the interesting is*
Underuse of *I* bundles, e.g. *In this chapter I*	Only one *I* identified

Semi-structured interviews, as an approach to 'enter into the other person's perspective' (Patton, 2015: 426) and 'a balance between structure and openness'(Gillham, 2005: 79), were conducted on the use of particular bundles in the interviewees' writing to gain these interviewees' perspectives on and learning experiences of these bundles. Each interview took approximately one

hour. The interviews consisted of two parts. Part 1 was concerned with the background of our interviewees in relation to L2 academic writing, for example, their disciplines, length of English-medium study, and relevant language teaching or learning experience. Part 2 was largely based on the identified expressions in their own writing. Interviews focussed on the reasons and possible sources of their selections. Table 2 provides an overview of the interviewees' information and experiences, which were considered closely relevant to the current research.

Table 2. Overview of Five Chinese Interviewees.

Interviewee	Z	A	J	W	V
Age	40+	30+	30+	25+	25+
Gender	Male	Male	Female	Female	Female
Level	PhD	PhD	PhD	PhD	Masters
Discipline	Applied Linguistics	Knowledge Management	Tourism Management	Management Communication	Applied Linguistics
Months of English-Medium Study	3	32	16	6	16
Particular Language Teaching & Learning Experience	14 years of lecturing at foreign language department of a Chinese university			half-a-year experience of writing English correspondence	4 years of English language teaching experience, teaching Cambridge English, New Concept English; having attempted TOEFL and IELTS for several times

The current learning context of these five interviewees is different from that of the Chinese writers of the thesis corpora studied in by Li (2016). These interviewees 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' (2012: 353); therefore, the expectations of institutions and supervisors will possibly influence their students' writing practices. The local influence should be taken into consideration

while interpreting the interview data. However, the five interviewees all received their primary, secondary and undergraduate education in mainland China. Except for V, the other interviewees had completed their Masters degrees in China. The years of formal education in China had schooled them into the expectations of the Chinese context and introduced them to 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 interviewees' writing.

Interview Data Analysis

All interviews were conducted in Chinese, the L1 of both the interviewees and Li, the author who conducted the interviews. The interview data were transcribed by Li and checked by the interviewees for clarification before analysis (Patton, 2015). The interviews were translated into English and if some words were untranslatable, transliteration – the original Chinese words with their closest English meanings given in brackets – was adopted (Halai, 2007). Around 10% of all the translation was double-checked by Wu, whose L1 is also Chinese.

Coding and thematic analysis (Braun and Clarke, 2006; Coffey and Atkinson, 1996) were used to analyse the interview data. A proportion of about 20% of the data was independently coded by Li and Franken, and the inter-coder reliability was 78%, a comparatively satisfactory result according to Miles and Huberman (1994). Disputed cases on coding were resolved in discussion. Then Li coded the rest of the data, and codes and themes were double-checked and refined by Franken and Wu. Table 3 presents a brief overview of the interview data with the identified themes, interview excerpts and number of occurrences in each theme.

Table 3. Coding of the Interview Data.

Themes	Excerpts	Number of Occurrences
interlingual transfer	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*).	11
classroom learning	My teacher told us cohesion and coherence were important in English and these conjunctions were fairly important.	22

noticing in reading	I have never noticed the position of *in order to* in my reading.	13
a lack of rhetorical confidence	I do not want to replace *find* with *note* and I am not comfortable to use *note*.	4
misunderstanding of rhetorical conventions	Academic writing should be objective and scientific, so I try not to use first person pronouns.	6

Reasons for Chinese Students' Sentence Initial Bundle Choices

Five reasons were given by the interviewees for the different bundle selections in the Chinese student writing.[1] These reasons were interlingual transfer, classroom learning, noticing in reading, a lack of rhetorical confidence, and misunderstanding of rhetorical conventions. Among them, classroom teaching is the most significant reason revealed by the interviewees, followed by noticing in reading and interlingual transfer (see Table 3).

Interlingual transfer refers to the transfer from the interviewees' L1, 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 the Chinese students follow word sequences in Chinese sentences while constructing English ones. A typical example is the use of (*in order*) *to*-infinitive bundles at the beginning of sentences to indicate the pre-conditions of the main clauses:[2]

> 1. ***In order to make*** *the interviewees 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).

In contrast, it appeared from the analysis of the examples in Li (2016) that the New Zealand students usually employed these bundles in the second part of their sentences to add complementary information to their main clauses, as in:

> 2. *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).

In the following interview transcripts, both interviewees A and V attributed their use of sentence initial (*in order*) *to*-infinitive phrases to the transfer of Chinese writing practices, 'Chinese sentences usually start with the indications of purposes' (Tables 4 and 5).

Table 4. V's Interview on Her Use of *To-phrase* Fragment.

Text	Interpretation
In order to make *the interviews operating smoothly, some questions were prepared beforehand as prompts for interviews* (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)

Table 5. A's Interview on His Use of *To-phrase* Fragment.

Text	Interpretation
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* (A)	This is my writing habit, maybe learned from my Chinese writing. Chinese sentences usually start with the indications of purposes (A)

Literal transfer means word for word translation. The bundle *There is no doubt* was used by the Chinese students in the thesis corpora to indicate their complete certainty. According to the interviewee V, the direct Chinese translation led to her use of the expression *It is undoubted that*, which has the same meaning as *There is no doubt* (Table 6). She translated this literally from its corresponding Chinese expression 毫无疑问 without considering the difference in the degree of centainty.

Table 6. V's Interview on Her Use of *Undoubted.*

Text	Interpretation
It is undoubted that *learning English is necessary for elder immigrants in English speaking countries* (V)	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).

Semantic transfer was found when the Chinese students chose English bundles according to the meanings of equivalent Chinese words. An example of semantic transfer was the preference of the bundle *It is obvious that* to *It is clear that*. As the interviewee Z reported, this was the result of Chinese students' judgment between 明显 and 清楚: '清楚 reflects a higher degree of certainty' (Table 7).

Table 7. Z's Interview on His Use of *Clear* and *Obvious.*

Text	Interpretation
Clearly*, trilingual education is not simply the mere extension of bilingualism ...* (Z)	I also like to use *it is obvious that*, but for *clear*, I like to use *clearly*. Chinese do not like to use *it is clear that* because we prefer 明显 (*apparent, easy to notice or understand*) to 清楚 (*certain, impossible to doubt*) in Chinese. We think *obvious* means 明显 and *clear* 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 (Z).

Classroom teaching is another important factor reported as contributing to the Chinese students' bundle use. Teachers have been reported to emphasize or even overemphasize certain language features while other salient features of academic English have been overlooked. The sentence initial bundle *Last but not least* was found as a general sequence signpost without any specific reference in the Chinese Masters writing corpus. As the interviewee V stated, *The last but not least* was introduced as an effective strategy to achieve cohesion and coherence in English writing and 'Chinese teachers put efforts on teaching conjunctions in English writing classes' (Table 8).

Table 8. V's Interview on Her Use of *The Last but Not Least.*

Text	Interpretation
The last but not least, *I will analyse some possible suggestions for future research* (V).	I learned this phrase in China. We do not use so many conjunctions 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). |

At the same time, shell nouns (e.g. *fact*, *problem*, *approach*) (Aktas and Cortes, 2008) and shell noun bundles (e.g. *The results of the*, *The purpose of this*) were underused in the corpus data and their linking power was largely overlooked in the writing pedagogy. During the interviews, it was found that the Chinese interviewees had little knowledge of these nouns and noun phrases, although they employed a few shell nouns in their texts. The interviewee Z considered his use of the noun phrase *The complaints from my colleagues and the results of the meetings* as an inferior and temporary choice because it resulted in a long subject (Table 9). The interviewee V was more conscious of the need to avoid word repetition, as a result of her learning and testing experiences, rather than the characterization and linking functions performed by her selected shell nouns *definition*, *measurement* and *identity* (Table 10), although her frequent change of words may interrupt the flow of the text.

Table 9. Z's Interview on His Use of Noun Phrase.

Text	Interpretation
***The complaints from my colleagues and the results of the meetings** often linger in my mind* (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 10. V's Interview on Her Use of Noun Phrase.

Text	Interpretation
*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* (V).	I am changing the nouns in this paragraph to avoid repetition. These words are the same meaning (V). My teacher suggested me not to repeat words and 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).

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' (2004: 417). It is interesting to note that familiarity is a necessary precondition to students' noticing (i.e. conscious learning during reading). Familiar language items, for example, the typical Chinese student bundle *with the development of* seem salient to the interviewee Z and have been consciously noticed in his course book and English newspaper (Table 11). In contrast, unfamiliar items or unknown features like the position of (*in order*) *to*-infinitive phrase fragments have gained little attention from the interviewee V (Table 12). Both their limited L2 processing ability and the shortage of awareness-raising tasks are likely to have contributed to their lack of noticing.

Table 11. Z's Interview on His Use of *With the Development Of.*

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

Table 12. V's Interview on Her Use of *To-phrase* Fragment.

Text	Interpretation
In order to make the interviews operating smoothly, some questions were prepared beforehand as prompts for interviews (V).	I have never noticed the position of *in order to* in my reading. Nobody has ever picked up my sentence initial (*in order*) *to*- infinitive phrases as a mistake (V).

Rhetorical confidence has been raised as another determinant of the 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: xiv). For example, our corpus analysis revealed that the Chinese students preferred to begin their sentences with conjunction bundles (e.g. *So it is necessary*) to immediately illustrate how they were related to the preceding sentences whereas the New Zealand students were more likely to rely on the topics/subjects of sentences to create a sense

of cohesive and coherent flow by placing conjunctions in the first part rather than exactly at the beginning of their sentences. The interviewee V provided a reason for the frequent sentence initial *however* in her writing: she dared not take the risk of placing it after the subject of the sentence because she was not familiar with placing it in this position: 'I'm afraid I will make a mistake' (Table 13).

Table 13. V's Interview on Her Use of *However.*

Text	Interpretation
However, *the definition of old varies from one society to another* (V). **However**, *ageing or old age does not mean dull or stagnant* (V).	I am used to putting *however* at the beginning of my sentences. It has been years. I am confident with it (V). I'm afraid I will make a mistake if I put *however* after the subject of the sentence, so I try not to use it in this way (V).

The Chinese students also seem to be highly conscious of their identity as student researchers and apprentice writers. They appeared more comfortable with expressing their attitudes towards their own research procedures rather than influencing the evaluations of their readers. One example of this is the underuse of *note* bundles. In the case of the interviewee J, she chose *It is interesting to find*, rather than *It is interesting to note*, a salient bundle in the New Zealand students' writing, to express personal stance from the writer's perspective. When the bundle *it is interesting to note was suggested to her*, she responded as follows (Table 14).

Table 14. J's Interview on Her Use of *Note.*

Text	Interpretation
It is interesting to find *that there is wide networking coordination within the DMO in Zhu Jiayu and other stakeholders* (J).	I do not want to replace *find* with *note* and I am not comfortable to use *note*. 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).

A few bundles in the Chinese students' thesis writing might be the results of their misunderstanding of rhetorical conventions. The popular attitude bundles in the New Zealand student theses (e.g. *It is interesting to*) were almost absent in the Chinese students' writing. In Table 15, the interviewee V revealed the reason for her rare usage of this bundle. She regarded academic writing as statements of objective facts rather than subjective personal arguments: 'Academic writing should be neutral'. Therefore, she tried to hide her personal feelings and appeared reluctant to reveal her existence as a writer.

Table 15. V's Interview on Her Use of *Interesting*.

Text	Interpretation
*However, **the interesting is**, although some other interviewees agreed that memory might decline with aging, they believed it was not the main reason which could stop English learning* (V).	I rarely use *interesting* 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).

In addition, the representation of the Chinese students' 'authorial identity' (Hyland, 2002) has largely been hidden by the underuse of the first-person pronoun *I* bundles. The same as the corpus data, only one *I* was found in the interviewees' work (Table 16), and during the interview W indicated her preference of *the researcher* as a result of her understanding of academic writing as being 'objective and scientific'.

Table 16. W's Interview on Her Use of *I*.

Text	Interpretation
*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* (W).	Strictly speaking, I should use *the researcher* here. Academic writing should be objective and scientific, so I try not to use first person pronouns (W).

Pedagogical Implications of Sources of Bundles

With regard to the reasons for bundle selections reported by the Chinese interviewees, corpus-based language learning is proposed here as an effective approach to compensate for insufficient usage by L2 students and teachers. They do not have L1 writers' intuition and may not have sufficient language resources to consult. This approach, together with other teaching and learning approaches, is suggested to achieve the following four recommendations:

1.   Raise Chinese students' awareness of the discrepancies between L1 and L2 bundles.

2.   Emphasize bundle noticing in academic reading and writing.

3.   Increase Chinese students' confidence as student writers.

4.   Familiarize Chinese students with rhetorical conventions.

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.

Recommendation 1: Raise Chinese students' awareness of the discrepancies between L1 and L2 bundles

The present study supports Paquot's proposal for L1 transfer effects on lexical bundles: 'EFL learners bring knowledge of the L1 lexicon to the writing task in the foreign language, including preferred collocations and lexicogrammatical patterns of words, as well as their stylistic or register specificities, discourse functions and frequency of use' ((2013) 411). In the earlier stages of language learning, transference is likely to be a productive and helpful strategy (Hoey, 2005). However, at later stages awareness-raising activities, targeting the discrepancies between L1 and L2 bundles, are clearly needed, if the writer intends to reach a wider audience from different language backgrounds. Awareness-raising activities can help learners to identify their bundle production in terms of word order, literal meaning and semantic association.

Recommendation 2: Emphasize 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 the use of the bundle. 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 bundles they were familiar with and ignore the unfamiliar ones. In order to direct students' attention to bundles they are unfamiliar with, it is necessary to enhance the input (Sharwood Smith, 1993; Sharwood Smith, 1991) 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 *in order to* 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, Flexible Language Acquisition System, see http://flax.nzdl.org) 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 with frequency-

based displays, multiple contexts and typographical salience, and can be easily identified while reading. During the writing process, EAP teachers can use discourse focussed 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 sentence starters.

Recommendation 3: Increase Chinese students' confidence as student writers

Chinese students, as non-native writers of English, 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 categorized into different metadiscourse functions (Li et al., 2017) or moves and steps (Cortes, 2013) for example, or teachers can require students to collect useful bundles before they start writing. With support, students can expand their writing from these bundles, which serve as 'islands of reliability', in other words, 'points of fixation, anchoring ground to start from and return to' (Dechert, 1984: 223). 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.

During writing, Chinese students seem less comfortable and confident than New Zealand students to cognitively involve their readers and they tend to resist the use of *note* bundles to direct their readers. Instead, they feel more confident to choose words (e.g. *find*) to describe their own research procedures from the writer's perspective. Supervisors can raise students' awareness of their current apprentice identity and encourage them to engage in 'legitimate peripheral participation' (Flowerdew, 2000: 131) so that students can have their voices heard by means of communicating with authority. A parallel corpus, for example a New Zealand thesis corpus at the same level in the same discipline, can be used to show Chinese students the practices of L1 writers (e.g. the use of *note* bundles) to increase their confidence in writer-reader interaction.

Recommendation 4: Familiarize Chinese students with rhetorical conventions

The current research reveals Chinese students' possible misinterpretations of rhetorical conventions in terms of sentence initial bundles. The absence of attitude bundles (e.g. *It is interesting to*) and *I* bundles is due to Chinese students' misunderstanding of rhetorical conventions. 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). Bundles and context sentences from a relevant corpus can be used to demonstrate these conventions. The reason for using bundles is that lexical bundles are extremely common expressions (Biber and Barbieri, 2007) and the use of bundles represents the common practices in a certain discourse. At the same time, bundles can serve as useful resources adopted to follow these conventions.

Conclusion

An increasing number of studies have been conducted in the area of lexical bundles in L2 writing, but few studies have investigated the reasons for learner bundle choices. This research addressed this gap by interviewing five Chinese postgraduate students about their own bundle production. The research revealed five reasons for the different bundle selections in Chinese student writing as interlingual transfer, classroom learning, noticing in reading, a lack of rhetorical confidence, and misunderstanding of rhetorical conventions. The findings of this study with writer self-reported reasons have not only verified and complemented researchers' predictions in the previous bundle studies, but also highlighted the top three reasons for learner discrepancy as classroom learning, noticing in reading and interlingual transfer. On the basis of these reasons, different teaching and learning approaches, particularly a corpus-based language learning approach, can be used to compensate for insufficient inputs of L2 students and teachers and to resolve the misunderstandings about rhetorical conventions.

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 as the corpus data, or even from the same participants. Another approach is to gather supervisors' and examiners' evaluations of students' bundle selection to investigate the relationship between bundle selection and target audiences or quality of writing.

Acknowledgements

Funding

Notes

1.	In this article, the word *interviewee* is used to refer to the five Chinese interview participants and the word *student(s)* is used to refer to the student writer(s) of the thesis corpora.

2.	All examples are from the four self-built thesis corpora and the writing of the Chinese interviewees. They are the original texts of the students with spelling, grammatical, lexical and punctuation mistakes unedited.

References

Ädel A and 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: 81-92.

Aktas RN and Cortes V. (2008) Shell nouns as cohesive devices in published and ESL student writing. *Journal of English for Academic Purposes* 7: 3-14.

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.

Baker P. (2006) *Using corpora in discourse analysis,* London, United Kingdom: Continuum.

Biber D and Barbieri F. (2007) Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26: 263-286.

Biber D, Johansson S, Leech G, et al. (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.

Braun V and Clarke V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology* 3: 77-101.

Chen Y-H and Baker P. (2010) Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology* 14: 30-49.

Coffey A and Atkinson P. (1996) *Making sense of qualitative data: Complementary research strategies,* Thousand Oaks, CA: Sage.

Cohen AD. (1983) Reformulating second-language compositions: A potential source of input for the learner. ERIC database (ED 228866).

Cortes V. (2004) Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23: 397-423.

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: 33-43.

Coxhead A. (2012) Academic vocabulary, writing and English for academic purposes: Perspectives from second language learners. *RELC Journal* 43: 137-145.

Dechert HW. (1984) Second language production: Six hypotheses. In: Dechert HW, Mohle D and Raupach M (eds) *Second language productions.* Tübingen, Germany: Gunter Narr Verlag, 211-230.

Flowerdew J. (2000) Discourse community, legitimate peripheral participation, and the nonnative-English-speaking scholar. *TESOL Quarterly* 34: 127-150.

Gillham B. (2005) *Research interviewing: The range of techniques,* Maidenhead, United Kingdom: Open University Press.

Halai N. (2007) Making use of bilingual interview data: Some experiences from the field. *The Qualitative Report* 12: 344-355.

Hoey M. (2005) *Lexical priming: A new theory of words and language,* New York, NY: Routledge.

Hyland K. (2002) Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics* 34: 1091-1112.

Hyland K. (2008a) Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics* 18: 41-62.

Hyland K. (2008b) As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27: 4-21.

Hyland K. (2011) Looking though corpora into writing practices. In: Viana V, Zyngier S and Barnbrook G (eds) *Perspectives on corpus linguistics.* Amsterdam, Netherlands: John Bejamins Publishing Company, 99-113.

Leech G. (1998) Preface. In: Granger S (ed) *Learner English on computer.* London, United Kingdom: Longman, xiv-xx.

Lewis M. (2008) *Implementing the lexical approach: Putting theory into practice,* London, United Kingdom: Heinle Cengage Learning.

Li L. (2016) Sentence initial bundles in L2 thesis writing: A comparative study of Chinese L2 and New Zealand L1 postgraduates' writing. Hamilton, New Zealand: University of Waikato.

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

Li T and Wharton S. (2012) Metadiscourse repertoire of L1 Mandarin undergraduates writing in English: A cross-contextual, cross-disciplnary study. *Journal of English for Academic Purposes* 11: 345-356.

Miles MB and Huberman AM. (1994) *Qualitative data analysis: An expanded sourcebook* Thousand Oaks, CA: Sage.

Nation ISP. (2013) *Learning vocabulary in another language,* Cambridge, United Kingdom: Cambridge University Press.

Nattinger JR and DeCarrico JS. (1992) *Lexical phrases and language teaching,* Oxford, United Kingdom: Oxford University Press.

Paquot M. (2013) Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics* 18: 391-417.

Patton MQ. (2015) *Qualitative research & evaluation methods : Integrating theory and practice,* Thousand Oaks, CA: Sage.

Qin J. (2014) Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System* 42: 220-231.

Ruan Z. (2016) Lexical bundles in Chinese undergraduate academic writing at an English medium university. *RELC Journal*: 1-14.

Schmidt RW. (1990) The role of consciousness in second language learning. *Applied Linguistics* 11: 129-158.

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: 118-132.

Sharwood Smith M. (1993) Input enhancement in instructed SLA: Theoretical bases. *SSLA* 15: 165-179.

Sinclair J. (1991) *Corpus, concordance, collocation,* Oxford, United Kingdom: Oxford University Press.

Staples S, Egbert J, Biber D, et al. (2013) Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes* 12: 214-225.

Wei Y and Lei L. (2011) Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC Journal* 42: 155-166.

Wu S, Franken M and Witten IH. (2010) Supporting collocation learning with a digital library. *Computer Assisted Language Learning* 23: 87-110.