

Journal of English Language
Teaching and Learning
No. 11, 2013

The Use of Lexical Bundles in Native and Non-native Post-graduate Writing: The Case of Applied Linguistics MA Theses

Zahra Amirian

Assistant Professor University of Isfahan

Somaye Ketabi

MA. Student at TEFL, University of Isfahan

Hamed Eshaghi

MA. Student at General linguistics, University of Tehran

Abstract

Connor et al. (2008) mention “specifying textual requirements of genres” (p.12) as one of the reasons which have motivated researchers in the analysis of writing. Members of each genre should be able to produce and retrieve these textual requirements appropriately to be considered communicatively proficient. One of the textual requirements of genres is regularities of specific forms and content. Lexical bundles are one of the features which play significant role in building genres’ regularities. Many researchers have tried to define academic writing with resort to the lexical bundles employed in it. Advanced and high intermediate L2 students’ pieces of writing and also post-graduate writing have been analyzed in different aspects. However, the important element in the analysis of post-graduate writing has always been the differences between genres across disciplines. In other words, in investigating lexical bundles in different genres, researchers have not focused on the issue of “nativity of the writer. To be exact, they consider native and non-native writing to share the same features. By considering this gap in lexical bundles studies, the present paper is an attempt to explore the nature of lexical bundles in native and non-native post-graduate students’ writing. In order to do so, a corpus of about one-million words from Iranian students’ applied linguistics theses is compared with a corpus of the same size from native English students’ applied linguistics theses. The results show significant differences in the frequency of lexical bundles used by native and Iranian students and also in structural and functional patterns used.

Keywords: Lexical bundles, Academic writing, L2 writing, Corpus analysis.

تاریخ وصول: ۹۲/۳/۵، تاریخ تصویب: ۹۲/۵/۳۰

*-E-mail: samk1989.sk@gmail.com

Introduction

University students face different challenges when they enter academic world. One of the great challenges is to get their articles published, and in order to do so they should elaborate their ideas like experts. The academic genre, like other genres, has its special words and structures. Using special formulaic language, which is defined by the field students are writing in, is one of the ways to be successful in the academic world. Applying especial prefabricated patterns shows that the author is a member of the academic group, and her/his writing is considered commendable by experts.

Native speakers produce formulaic language more than non-native speakers (Schmitt and Carter, 2004). The production of these pieces of language is automatic for native speakers, but non-native speakers usually produce non-existent chunks affected by their first language and also the academic instruction they had (Howarth, 1998). Even the learners who are advanced in English grammar and vocabulary might fail in using formulaic language. As Ädel and Erman (2012, p. 81) say “the degree of proficiency correlates significantly with the proportion and/or types of formulaic language used.”

It is not yet clear that how frequent these patterns should be in academic writing, but studies have shown different frequencies of prefabricated patterns in the writing of novices and experts (Ädel and Erman, 2012; Cortes, 2006; Jalali, 2008). The focus in recent studies on formulaic language is lexical bundles.

Lexical bundles

Recurrent patterns have been studied under different names such as lexical phrases and formulas (Biber et al., 2004), but Biber and his colleagues were the first scholars who introduced “lexical bundles” and defined them as “recurrent expressions regardless of their idiomaticity, and regardless of their structural status” (Biber et al., 1999). After this introduction and analyzing more corpora, some criteria were set for the expressions to be counted as lexical bundles. The first one is frequency; a group of words is considered as lexical bundles if they occur at least 20 times in a corpus of one-million

words. Of course, this cut-off point is relative and is defined by the register and the mode of the corpus under study. For example, as Biber (Ibid.) showed, lexical bundles are more frequent in spoken language, so the frequency criterion in spoken language is different from written language. Additionally, for a phrase to be considered as a lexical bundle, it should be seen in five different texts as to be sure the occurrence is not because of a writer's specific style.

Two significant features about lexical bundles are their high frequency and their role in discourse construction. After examining different texts and registers, Biber et al. (2004) and Hyland (2008b) concluded that the high frequency of lexical bundles is not by chance and needs explanation. The explanation is provided by bundles' discourse functions. As Biber et al. (2004, p. 400) said "they [lexical bundles] are important building blocks of discourse, associated with basic communicative functions." They can be used as discourse organizers (e.g. *if you look at, know a little about*) or referential phrases (e.g. *at the end of*) to express new information (Biber et al., 2004; Biber and Barbiery, 2007). As Biber et al. (2004) claimed:

These sequences of words can be regarded as structural 'frames', followed by a 'slot'. The frame functions as a kind of discourse anchor for the 'new' information in the slot, telling the listener/reader how to interpret that information with respect to stance, discourse organization, or referential status. (p. 399)

Lexical bundles in L1 and L2 writing

Biber and Barbiery (2007) mention that since lexical bundles are highly frequent, it is supposed that they could be acquired naturally and there is no need to teach them. But they assert that discourse functions of lexical bundles should be taught. Following this assumption, several researchers conducted corpus-based studies to investigate differences between articles and pieces of writing of experts and those of advanced students (Ädel and Erman, 2012; Cortes, 2006; Jalali, 2008). There have been differences in their writing from the perspective of lexical bundles. Ädel and Erman concluded that advanced learners of English used more limited sets of

bundles than native students; however, the corpus size was small in their study. In Cortes's study (2006), articles of history in three journals were compared to the writing of students, and based on differences, some lesson plans were designed to be taught. In another study, Jalali (2008) examined lexical bundles used in articles published in linguistics journals and linguistics MA theses and PhD dissertations of Iranian students, and found that there were major differences in the frequency of the lexical bundles employed. But he did not focus on the role of English as a second language in his study; his main concern was the differences between two genres (research articles and post-graduate writing).

In his comprehensive research, Hyland (2008a, b) studied MA theses, PhD dissertations, and articles of different majors (applied linguistics, biology, electrical engineering, and business studies) and found specific patterns for each genre. Nevertheless, in spite of the fact that the students were EFL learners (Chinese students) in Hyland's study, the aim was not the differences in L1 and L2 pieces of writing, but structural and functional categories in different disciplines were important.

Overview of the present study

As it can be seen, the important element in the analysis of post-graduate writing has usually been the differences between disciplines (biology vs. applied linguistics) or genres (e.g. articles vs. theses). In other words, the effect of native or non-native academic setting has hardly ever been investigated in post-graduate writing. Post-graduate writing is considered to be special since the students are proficient in English and they have studied many articles and books written by experts in order to write their theses and dissertations. So, native and non-native post-graduate students are both familiar with academic expressions. The point is that the degree to which the setting can affect their writing has not investigated yet.

By considering the gap in the studies of lexical bundles in post-graduate writing, this study is an attempt to explore the nature of lexical bundles in applied linguistics MA theses of native and non-native students. The corpora used in this study are composed of about two million words of MA theses of native and Iranian students which

is the optimal size to investigate lexical bundles. This was not true for previous studies in which the number of words was less than one million in MA theses.

Native students' theses were gathered by the use of Edinburgh research archive, and Iranian students' theses were collected by the help of post-graduates in University of Isfahan and University of Tehran. The writer hopes that the results of this study shed more light on the nature of lexical bundles in L1 and L2 post-graduate writing.

Corpora

The data in the present study consist of two corpora of applied linguistics MA theses of Iranian students (which were written in non-native setting by Iranian students and edited by Iranian supervisors) and native students' applied linguistics MA theses (which were written in native setting by English students and edited by English supervisors). The first corpus was composed of 46 theses and the second one consists of 48 theses:

Table 1. *Corpora Word Count*

	Total Words
applied linguistics MA theses of Iranian students	5,740
applied linguistics MA theses of native Students	9,690

Method

Four-word bundles were chosen to be investigated in this study since they are more frequent than five-word clusters and present more range of functions and structures than three-word bundles (Hyland, 2008a). The frequency cut-off point is set at 20 times per million words and the least occurrence in 5 texts. AntConc 3.2.1 (Anthony, 2007) was used to find lexical bundles. The frequencies and functional and structural categories of each corpus was first identified and then compared.

Results

Considerable differences were seen in the frequency of bundles in post-graduate writing of native and Iranian students. Table 2 presents the frequency in each corpus:

Table 2. Bundle Frequency in Corpora

Genres	No. of Texts	No. of Total Words	No. of Different Lexical Bundles	No. of Total Cases
Applied linguistics MA theses of Iranian students	46	1,035,740	211	8465
Applied linguistics MA theses of native Students	48	1,059,690	61	2073

The number of different bundles used in native students' theses is much less than the frequency of clusters in Iranian students' theses. The reason for this overuse might be the instruction they had in their writing classes. They are usually taught that in order to be considered as a part of academic community, they should use formulaic language. However, because of "the lack of rich input" (Schmitt and Carter, 2004, p. 13), overuse, underuse, or misuse of these expressions are common in L2 writing.

There are also some differences in the most frequent lexical bundles used in two corpora. Table 3 represents these differences:

Table 3. *Thirty Most Frequent Four-word Bundles in Corpora*

	<i>Native Students' Linguistics MA Theses</i>		<i>Iranian Students' Linguistics MA Theses</i>	
1	the results of the	251	on the other hand	110
2	on the other hand	246	in the case of	84
3	in the case of	145	the University of Edinburgh	76
4	in the target language	131	as well as the	75
5	of the present study	107	in the context of	59
6	the meaning of the	102	it is important to	57
7	one of the most	100	at the end of	51
8	at the same time	96	the end of the	48
9	significant difference between the	86	as a result of	47
10	in the field of	83	I would like to	47
11	in the process of	81	the rest of the	47
12	at the end of	78	for the purposes of	44
13	is one of the	78	can be found in	41
14	as well as the	77	the structure of the	41
15	in the target text	77	an example of a	38
16	on the basis of	77	that there is a	38
17	the results of this	76	the use of the	38
18	in the present study	73	in relation to the	36
19	in other words the	73	should be noted that	35
20	in the use of	71	can be seen in	34
21	results of this study	68	in terms of the	32
22	to the fact that	68	the purposes of this	32
23	there is a significant	66	in the form of	31
24	the end of the	66	it should be noted	30
25	as a result of	63	the total number of	30
26	as a foreign language	62	at the same time	29
27	of the target language	62	at the University of	29
28	the findings of the	62	can be used to	29
29	used in this study	61	in the present study	29
30	the analysis of the	60	that there is no	29

Some of the bundles which were used the most in native students' writing are not common in the writing of Iranian students (compared with other bundles) such as "as a result of." Some bundles such as

“the rest of the” and “I would like to” are shared in two corpora, but they are among the least frequent bundles in Iranian writing. All in all, it was found that 34 bundles were shared in two corpora, but the frequencies are not the same in native and Iranian writing.

Structural categorization of lexical bundles

Biber et al’s taxonomy (1999) is chosen to classify bundles structurally. However, the researchers in the present study modified the definition of “other” group (shown in Table 4). In Biber et al’s taxonomy, lexical bundles such as “by the fact that” and “in a way that”, which consist of a prepositional phrase and a word of another phrase, are in “other prepositional phrases” group, but in this study they are in “other” group (bundles which do not fit into other categories and are not complete phrases). Moreover, there are separate groups named “noun phrase” and “prepositional phrase” which represent complete structures of NPs and PPs.

Table4. *Most Common Patterns of Four-word Bundles in Academic Writing*

Structure	Examples
NP + of-phrase fragments	<i>the results of the, the use of the, the end of the, the findings of the</i>
n phrase with other post modifier fragments	<i>participants in the, the relationship between the, English as a second</i>
	<i>its of this study, findings of this study, purpose of this study</i>
PP + of-phrase fragments	<i>in the case of, in the form of, at the end of, as a result of</i>
r PP fragments	<i>respect to the, in English as a, in relation to the</i>
	<i>he other hand, in the target language, in the same way</i>
copulatory it + VP/AdjP	<i>necessary to, it is important to, it is possible to</i>
ive + PP fragment	<i>l in this study, used in order to</i>
Copula be + NP/AdjP	<i>is a significant difference, is a kind of</i>
ould like to	<i>ould like to, in other words the, of language learning the</i>

(Biber et al., 1999, pp. 997-1025)

Structural comparison of four-word lexical bundles in two corpora

Structural grouping of bundles in Iranian and native students' theses is shown in Table 5. It is evident in Table 5 that both groups of students were more interested in using phrasal clusters (NP, NP + of-phrase fragments/other post-modifier fragments, PP, PP + of-phrase fragments, other PP fragments) than clausal bundles (it + VP/AdjP, passive + PP fragments, be + NP/AdjP). The most employed bundles in Iranian post-graduate writing are those without complete and unified structure which are labeled as "others". Clusters like "as well as the" and "in other words the" are among the most used bundles in Iranian students' post-graduate writing. The second mostly used clusters are prepositional phrases. One of the most frequent four-word bundles in Iranian students' writing is "on the other hand" which a prepositional phrase is. Phrases with prepositions as their heads such as PP with embedded *of*-phrase fragments, other PP fragments, and PP were employed more than phrases with nouns as their heads which are NP, Noun phrase with other post modifier fragments, NP + of-phrase fragments (38.51% vs. 30.53%). The least frequent four-word lexical bundles are "it + VP/AdjP".

Table 5. *Structural Comparison of Lexical Bundles in Linguistics MA Theses of Iranian and Native Students*

Structure	Frequency of Different Categories	Frequency of Total Cases	Percentage (%)
NP + of-phrase fragments	34\18	1340\620	16.21\29.9
Noun phrase with other post modifier fragments	17\0	660\0	8.30\0
NP	13\0	479\0	6.02\0
PP with embedded of-phrase fragments	27\12	1376\424	16.64\20.68
Other PP fragments	7\1	245\36	3.08\1.57
PP	27\4	1553\149	18.79\7.26
Anticipatory it + VP/AdjP	6\7	150\193	1.88\9.41
Passive + PP fragment	11\4	314\132	3.94\6.43
Copula be + NP/AdjP	7\4	267\89	3.35\4.34
Others	62\11	1881\430	22.75\20.74
Total	211\61	8265\2073	100\100

*numbers to the left of the slash are for Iranian students and those to the right are for native students

Noun phrases with *of*-phrase fragments were used more than other structural groups in native students' writing. But they did not employ noun phrases with other post modifier fragments. For example, "one of the most" –which is among the most common clusters in Iranian students' master theses– was not used at all in native students' post-graduate writing. Moreover, native students did not use lexical bundles with complete NP structures in their writing. Phrases with nouns as their heads were almost as common as prepositional phrases (29.9 vs. 29.69) in native students' theses. The least frequent bundles in native students' theses are prepositional phrase fragments (without *of*).

There are some similarities and differences in structural patterns of two corpora. Talking about similarities, one can say that both groups of students use phrasal structures more than clausal ones. Another similarity is in the use of "Be + NP/AdjP" and PP fragments without *of* labeled as "other PP fragments"; both Iranian and native students did not employ these group very much.

Nevertheless, differences are seen more than similarities in comparing structural patterns of lexical bundles in Iranian and native students' post-graduate writing. One of the dissimilarities is found in employing clausal and phrasal bundles. Although both groups used more phrasal bundles than clausal bundles, the percentage of clausal clusters in native students' writing is twice more than its percentage in Iranian students' writing (Figure 1). Among clausal clusters in Iranian students' theses (and also among all structural groups), bundles with anticipatory *it* are the least frequent ones; however, this is not the case for native students' theses. Moreover, Iranian students were more inclined to use complete prepositional phrases without *of* than native students were. Iranian students were also more interested in employing two structural groups with noun phrases (complete noun phrases and noun phrases with other post modifiers). Instead, native students used noun phrases with *of*-phrase fragments more.

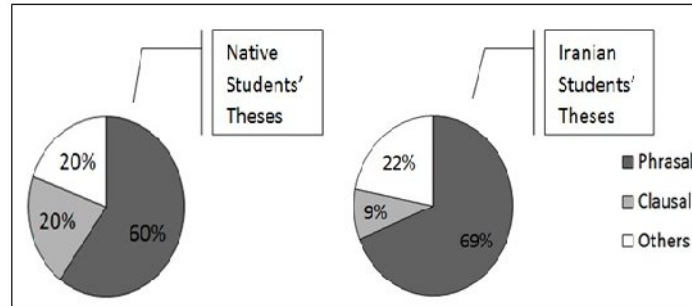


Figure 1. Phrasal and Clausal Bundles in Two corpora

Functional categorization of lexical bundles

Hyland's grouping (2008a), which was originally developed from Biber et al.'s study (2004), has been used to functionally organize four-word bundles in this study. This classification is preferred to Biber et al.'s (2004) since it is specifically designed for academic writing. Table 6 presents functional categorization of four-word bundles based on Hyland's model (2008a).

Table 6: Functional Categorization of Four-word Lexical Bundles in Academic Writing (Hyland, 2008a)

Major Functions	Sub-categories	Examples
<i>Research-oriented</i> Help writers to structure their activities and experiences of the real world	Location	<i>at the beginning of, at the same time, in the present study</i>
	Procedure	<i>the use of the, the role of the, the purpose of the, the operation of the</i>
	Quantification	<i>the magnitude of the, a wide range of, one of the most</i>
	Description	<i>the structure of the, the size of the, the surface of the</i>
	Topic	<i>in the Hong Kong, the currency board system</i>
<i>Text-oriented</i> Concerned with the organization of the text and its meaning as a message or	Transition signals	<i>on the other hand, in addition to the, in contrast to the</i>
	Resultative signals	<i>as a result of, it was found that, these results suggest that</i>

<i>argument</i>	Structuring signals <i>text-reflexive markers which organize stretches of discourse or direct reader elsewhere in text</i>	<i>in the present study, in the next section, as shown in figure</i>
	Framing signals <i>situate arguments by specifying limiting conditions</i>	<i>in the case of, with respect to the, on the basis of, in the presence of</i>
<i>Participant-oriented These are focused on the writer or reader of the text</i>	Stance features <i>convey the writer's attitudes and evaluations</i>	<i>are likely to be, may be due to, it is possible that</i>
	Engagement features <i>address readers directly</i>	<i>it should be noted that, as can be seen</i>

Functional comparison of four-word lexical bundles in the two corpora

Table 7 shows the functional grouping of bundles in this study. The difference between the frequencies of three categories in native speakers' writing is not like Iranian students' writing. In other words, the difference between the percentages of three categories frequencies is about 5 and 10 in native students' writing, but this is not the case for Iranian students' writing. The difference between the most and the least used categories is about 52 percent. It can be said that lexical bundles are distributed more equally in three functional groups in native students' writing.

More than half of four-word lexical bundles are in research oriented category in post-graduate writing of Iranian students. In order to prove that they are proficient enough in explaining and elaborating on the research processes, MA students mostly use research-oriented bundles (Hyland, 2008b). This statement is true for Iranian students' writing, but native students are more interested to organize their texts by using text-oriented bundles.

Among research oriented bundles, Iranian students are mostly interested in using clusters which represent the field of research (topic). These bundles are made of words and expressions which are directly and specifically related to the subject under study (Hyland, 2008a). A contextualized example of these bundles is shown in sentence 1:

1) *A brief look at four decades of language teaching and learning literature reveals that any language teaching method that once made its way to the field paid special attention to vocabulary learning as one of the major building blocks of a language.*

Table 7. *Functional Categorization of Lexical Bundles in Linguistics MA Theses of Iranian and Native Students*

Function	Frequency of Different Categories	Frequency of Total Cases	Percentage(%)
Research-oriented	124\25	4718\	57.08\35.69
• Location	16\6	783\208	9.47\10.14
• Procedure	33\5	1132\141	13.69\6.87
• Quantification	15\3	618\62	7.47\3.02
• Description	28\8	912\201	11.03\9.8
• Topic	32\3	1273\128	15.40\6.17
Text-oriented	70\19	3078\809	37.24\39.46
• Transition signals	28\4	1098\246	13.28\12
• Resultative signals	8\2	590\75	7.13\3.65
• Structuring signals	16\5	566\165	6.84\8.04
• Framing signals	18\8	824\323	9.96\15.57
Participant-oriented	17\17	469\524	5.67\25.56
• Stance features	11\12	305\340	3.68\16.58
• Engagement features	6\5	164\184	1.98\8.97
Total	211\61	8265\2073	100\100

**numbers to the left of the slash are for Iranian students and those to the right are for native students*

However, native students mostly used text-oriented bundles which are “concerned with the organization of the text and its meaning as a message or argument” (Hyland, 2008a, p. 13). Framing signals are the most common subcategory of text-oriented bundles in native students’

theses. Sentence 2 is a contextualized example of these bundles in native students writing:

2) In the context of machine translation, these models are expected to enrich the existing surface word translation model with additional linguistic feature models.

The least frequent clusters in both groups are participant-oriented ones which bring writers' views into the study and talk to readers directly (Hyland, 2008a). Among this category of lexical bundles, stance features which deal with writer's attitudes were employed more than engagement features:

3) It is important to inquire into the issue of the role of L1 in an L2 composing task, since the use of L1 seems to be a natural and frequent cognitive behavior in a bilingual mind engaging in an L2 task.

The great difference in functional patterns of these two corpora is in the use of participant-oriented bundles. The frequency of participant-oriented bundles in native students' writing is five times more than their frequency in Iranian students' writing. This finding is in line with the results of Hyland's (2008a) study. The avoidance of participant-oriented bundles by non-native MA students may be because of the fact they were not interested to take the risk and support their claims by using expressions which bring their own ideas into the research.

Comparing functional patterns of lexical bundles in native and non-native students' post-graduate writing shows that bundles are more equally distributed in three functional groups in native students' writing. The difference in the frequencies of the mostly used bundles in Iranian students' writing (research-oriented bundles) and the least ones (participant-oriented bundles) is more than 50 percent. In other words, Iranian students have used research-oriented bundles ten times more than participant-oriented bundles. On the other hand, the

difference between the most and the least employed categories in MA theses of native students is about 10 percent. So, it can be said that native students tried to make use of all types of bundles, and there is more variety in their writing. But Iranian students relied on one type of bundles more than the other two groups.

If the percentage of each category is considered, one can say that Iranian students relied on research-oriented bundles more than the other functional groups. However, this is not true in Native students' writing. They mostly made use of text-oriented bundles, but the difference between text-oriented and research-oriented bundles (the second mostly used bundles) is slight (about 5 percent). However, Iranian and native students relied on text-oriented bundles almost equally as the percentages in two categories are about 38 and 39 percent respectively. Moreover, the difference in the use of research-oriented bundles is about 21 percent. Almost the same difference is seen in participant oriented bundles. By considering these differences and the definition of functional categories of bundles, one can conclude that Iranian students try to make their theses as objective as possible by more mentioning their activities and experiences of the real world and less referring to their own ideas or their readers.

The mostly used subcategory of research-oriented bundles in Iranian students' writing is those related to the field of study (topic-related bundles: 15.72%). This type of bundles is not very common in native students' theses; just about 5 percent of all the clusters are about the subject under study. By considering this analysis, one can conclude that Iranian students try to bring unity in their writing by employing more subject-related expressions and clusters. The least frequent bundles in research-oriented category in both Iranian and Native students' writing are those related to quantity of the elements in the study. Moreover, native students are most inclined to use location-related bundles in their writing, but this type of bundles are at the fourth place in Persian students' theses.

The most frequent text-oriented bundles' subcategory in Iranian students' writing is transition signals which are the second highly-used bundles in native students' writing. Native students employed framing signals more than other signals. In addition, the least frequently used signals in native students' writing are resultative

signals, but the percentage of their frequency in Iranian students' writing is twice more than its frequency in the other corpus.

Native students were more interested in employing Participant-oriented bundles. In subcategories of this type of lexical bundle, both groups of students were more inclined to use stance features which express writer's evaluation rather than engagement features which deal with readers.

Discussion

The results of this study again prove that lexical bundles are constructing elements in academic discourse. The importance of them is proved by their high frequency in the two corpora used in this study.

However, what is remarkable about this piece of research is that for the first time post-graduate writing of native and non-native speakers of English is compared. It is true that previous studies such as Jalali (2008) and Hyland (2008a) investigated lexical bundles in post-graduate writing, but the focus in those studies was not whether the students are native speakers of English or not. For example, the master theses in Hyland's study were written by Chinese students, and the purpose of the study was to explore lexical bundles in different disciplines. In Jalali's study, Iranian students' post-graduate writing was compared with research articles whose writers were from different countries, so the aim was not to investigate the effect of first language and native or non-native academic setting on using lexical bundles. In another investigation, Hyland (2008b) analyzed master theses, PhD dissertations, and research articles, but again the focus was not the differences between native and non-native writers.

The effect of academic setting will be evident if Hyland's study (2008a, b) is compared with the present study. The writers in Hyland's study were Chinese students (EFL students); however, the frequency of bundles used in Iranian students' theses were almost twice more than those in Chinese students' theses. The reason for this variation might be the non-native setting and the effect of first language. But the results should be interpreted with caution, as the corpora in

Hyland's study were composed of research articles as well as MA theses and PhD dissertations.

The results of different frequencies of functional categories in native students' writing are more compatible with the frequencies of bundles in Chinese students' post graduate writing in Hyland's study (2008a). The mostly used functional category in both groups is "text-oriented bundles". Hyland (2008b) also analyzed MA theses separately and found that MA students are more interested in talking about the methodology they have employed in their writing. This is true for Iranian students in the present study, too. Comparing with the results of Jalali's study (2008), one can find that Iranian students use research bundles more than other categories and participant-oriented clusters are the least frequent ones. But, the percentages are different in two groups. It may be due to the fact that the corpus in Jalali's analysis included PhD dissertation, too. PhD students are more inclined to bring their ideas into their research by using participant-oriented bundles. As Hyland (2008a) mentioned, this observation may be because PhD students' more academic experience.

It was interesting that structural comparison of bundles revealed no compatibility with previous studies. The mostly used bundles in Jalali's study were found to be "prepositional phrase + of" in two groups (post-graduate writing and research articles), and the most frequent clusters in Hyland's study were "other prepositional phrase". Comparing with the present study, in which the most frequent lexical bundles in native students' post-graduate writing were found to be "noun phrase + of" and the mostly used clusters in Iranian students' post-graduate writing were those without complete structure (labeled as "others" in Biber's taxonomy), one observes no similarity. It is true that in all studies of academic genres phrasal bundles were more frequent than clausal ones (and compatibility is also observed in subcategories of clausal structures), but similarities in subcategories of phrasal structures were rarely seen. The reason might be because of vague structural classification of lexical bundles. Biber et al. (1999) were the first scholars who structurally classified lexical bundles by analyzing a large corpus of spoken and written genres. Twelve structural categories in academic writing were specified in the original work of Biber and his colleagues. However, in the works of Hyland

(2008b) and Jalali (2008), not all of the categories were used. Other researchers such as Chen and Baker (2012) have developed their own taxonomy. This lack of agreement may result in different categorization which cannot clearly identify the specific characteristics of each genre.

Another point should be mentioned about structural categorization of Biber and his colleagues (1999): in the classification of prepositional phrases and noun phrases, it is supposed that they cannot occur as complete structures such as “results of this study” or “at the same time”. It is true that Biber has put complete PP structures under the forth category (other prepositional phrase) ; however, the bundles under this category have different structures; some of them are complete PP phrases, such as “at the same time”, some of them need just one word to become complete structures, such as “of English as a”, and some of them are complete PP structures plus a word from another phrase, such as “of this study was” and “in other words the”. The last mentioned category is not considered as PP fragment by the researchers in this study, as they are a mixture of two phrases. Some examples will explain this statement:

- *In other words the difference between formal grammar and functional grammar is that the generation of rule-governed sentences is not the aim of the functional grammar but rather the production of rule-governed sentences is the means to coherent communication*

In this example, “the” belongs to the next phrase which is a noun phrase, and together they can make a sentence.

- *As the aim of this study was measuring the improvements of young learners in learning English vocabulary after a certain period of learning and the best method for improving communication skills in English for EFL young learners, the questions in these final exams were divided into vocabulary and communicative questions.*

Again, the last word in the lexical bundle belongs to the next phrase which is TP is this example.

If the researchers exactly define the structural categorization they use in their studies, agreement will be met in comparing similar genres, and structural definition of each genre will be more precise.

Conclusion

Investigating formulaic language in the writing of students has been the area of interest for many researchers, but the purpose in the present study was filling the missing gap in lexical bundles studies which is the difference in the use of lexical bundles in native and non-native students' post-graduate writing. The results show considerable differences in the frequency, structures, and functions of lexical bundles in native and Iranian post-graduate writing.

Summary of the results

What is significant in the findings of this study is that Iranian post-graduate students rely on lexical bundles more than native post-graduate students and even more than Chinese students in Hyland's study (2008b). There are also major differences in the use of different functional and structural categories.

Similar to previous studies on exploring lexical bundles in academic writing, the present investigation revealed students' more interest in using clausal bundles in their theses. Despite this general finding, the structural patterns in this study did not show marked similarities with the previous investigations. This might be due to the vague and various structural classifications of lexical bundles in different studies.

Functional patterns in native students' post-graduate writing indicate more variety, as native students employed all categories with almost the same frequency. In addition, Iranian students show more concern about mentioning the procedures of completing their research by using more research-oriented bundles. While native students try to organize their writing by using more text-oriented bundles. By considering the infrequency of participant-oriented bundles in Iranian students' writing, one can conclude that Iranian students try to be as

objective as possible by talking more about real world's procedures and less about their ideas in their theses.

Pedagogical implications

Iranian students' writing, unlike previously studied corpora such as Hyland (2008a,b) showed great differences compared with the writing of native students in frequency and structural and functional categories of lexical bundles. The overuse of one category and underuse of another can be a sign of unfamiliarity with the range and function of these bundles. So, EAP (English for Academic Purposes) experts and teachers should consider the importance of lexical bundles in academic writing and try to bring variety to students' writing by exposing students to different lexical bundles and their use.

Explicit teaching of these bundles, as Cortes (2006) showed, does not have significant effect on students' future writing. In other words, presenting a list of expressions (even with contextualized examples) and doing some fill-in-the-blank tasks are not effective. However, if the students are presented with experts' articles and books, and if the structural analysis of texts is done by the students themselves, there might be the possibility that students acquire these bundles and employ them in their writing. In the corpus used in this study, it was observed that Iranian MA students employed a bulk of lexical bundles; almost twice more than other non-native authors. It shows that the students believe using these clusters present them as being academically rich, but the overuse of them may result in lack of creativity in their writing. University students should care for creativity as well as accuracy in academic writing. Overemphasis on each of these elements will result in an article which is not accepted by experts.

So, writing classes need some modifications. It is common that students' essays and pieces of writing are collected every session and reviewed by the professor, but analyzing experts' writing is not so usual in these courses in Iran. Exposure to the usage of these bundles in articles should be done in organized session alongside the writing assignments. Therefore, not only can the students compare their writing with their peers, but also they can refer to a rich resource of

academic writing and know more about the road they should go along to become professional writers.

Limitations of the study

There were some unavoidable limitations in this study. The first one is about the size of the corpora. The number of available theses which were written by native students was limited. Second, the theses were gathered from the students of limited number of universities. This is again because of the lack of resources.

Suggestions for further research

An important result of this study is the ideas and questions it can create in readers' minds. One of these ideas can be exploring the effect of exposure to lexical bundles on the writing of L2 students in a longitudinal study. Another idea can be about investigating the patterns of lexical bundles in PhD dissertations which are supposed to be more professional.

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, 84-91.
- Altenberg, B. (1998). On the phraseology of spoken English. In J. D'Arcy (Ed.), *Proceedings of the Fifth Nordic Association for English studies Conference*. University of Iceland, Reykjaví, Iceland.
- Anthony, L. (2007). Antconc 3.2.1: Freeware corpus analysis toolkit. [on-line]. Available: <http://www.antlab.sci.waseda.ac.jp/>
- Biber, D, & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263-286.
- Biber, D, Conrad, S & Cortes, V. (2004). If you look at ...: lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371-405.
- Biber, D, Johansson, S, Leech, G, Conrad S, & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson.
- Chen, Y.-H., and Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and 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, 397-423.
- Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example form a writing intensive history class. *Linguistics and Education*, 17, 391-406.
- Firth, J. R. (1964). *Papers in linguistics*. London and New York: Oxford University Press.
- Howarth (1998). The phraseology of learner's academic writing. In A. Cowie (Ed.), *Phraseology* (pp.161-186). Oxford: Claredon Press.

- Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, 4-21.
- Hyland, K. (2008b). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18, 41-62.
- Jalali, H., Eslami Rasesh, A., and Tavangar, M. (2008). Lexical bundles and interdisciplinary variation: the case of applied linguistics. *Iranian Journal of Language Studies*, 2(4), 447-484.
- Wray, A., & Perkins, R. M. (2000). The functions of formulaic language: An integrated model. *Language and Communication*, 20, 1-28
- Schmitt, N., and Carter, R. (2004). Formulaic sequences in action. In N. Schmitt, *Formulaic sequences* (pp.1-22). Amsterdam: John Benjamins.
- Wray, A., and Perkins, R. M. (2000). The functions of formulaic language: An integrated model. *Language and Communication*, 20, 1-28.

Appendix A: Lexical bundles in linguistics MA theses of Iranian students

	<i>Bundles</i>	<i>Frequen cy</i>		<i>Bundles</i>	<i>Frequency</i>
1	the results of the	251	28	the findings of the	62
2	on the other hand	246	29	used in this study	61
3	in the case of	145	30	the analysis of the	60
4	in the target language	131	31	the findings of this	60
5	of the present study	107	32	results of the study	60
6	the meaning of the	102	33	findings of this study	59
7	one of the most	100	34	in the form of	59
8	at the same time	96	35	a significant difference between	58
9	significant difference between the	86	36	is a significant difference	56
10	in the field of	83	37	that there is a	55
11	in the process of	81	38	the fact that the	53
12	at the end of	78	39	the mean scores of	53
13	is one of the	78	40	for the purpose of	52
14	as well as the	77	41	in terms of the	52
15	in the target text	77	42	of the source language	52
16	on the basis of	77	43	of the source text	50
17	the results of this	76	44	the participants in the	50
18	in the present study	73	45	a great deal of	48
19	in other words the	73	46	of this study was	47
20	in the use of	71	47	English as a foreign	46
21	results of this study	68	48	the difference between the	46
22	to the fact that	68	49	on the one hand	45
23	there is a significant	66	50	findings of the study	44
24	the end of the	66	51	in the area of	44
25	as a result of	63	52	in the target sentence	44
26	as a foreign language	62	53	the extent to which	44
27	of the target language	62	54	in this study was	42

	<i>Bundles</i>	<i>Frequen cy</i>		<i>Bundles</i>	<i>Freque ncy</i>
55	as one of the	41	84	of the study the	33
56	in the control group	41	85	the present study is	33
57	mean scores of the	40	86	the scores of the	33
58	of English as a	40	87	according to the text	32
59	should be noted that	40	88	English as a second	32
60	the beginning of the	40	89	of the participants in	32
61	the reliability of the	40	90	the content of the	32
62	to be able to	40	91	the nature of the	32
63	at level of significance	39	92	the other hand the	32
64	in a second language	39	93	the relationship between the	32
65	by the use of	38	94	to find out the	32
66	second or foreign language	38	95	with respect to the	32
67	in a foreign language	37	96	can be concluded that	31
68	in the sense that	37	97	in each of the	31
69	the purpose of the	37	98	in order to be	31
70	there is no difference	37	99	in the context of	31
71	in this study were	36	100	is based on the	31
72	significant at the level	36	101	the meaning of a	31
73	at the beginning of	35	102	the use of a	31
74	between the two groups	35	103	the use of the	31
75	can be used to	35	104	they were asked to	31
76	in order to find	35	105	in second language learning	30
77	is significant at the	35	106	in such a way	30
78	of this study is	35	107	in this study the	30
79	on the part of	35	108	No significant difference between	30
80	due to the fact	34	109	order to find out	30
81	in terms of their	34	110	Language teaching and learning	29
82	as a means of	33	111	structure of the English	29
83	in a way that	33	112	the structure of the	29

	<i>Bundles</i>	<i>Frequency</i>		<i>Bundles</i>	<i>Frequency</i>
113	will be discussed in	29	141	participants were asked to	25
114	as a second language	28	142	purpose of this study	25
115	can be said that	28	143	significant difference in the	25
116	in the above example	28	144	such a way that	25
117	is an example of	28	145	the one hand and	25
118	it should be noted	28	146	the present study was	25
119	language learning and teaching	28	147	the rest of the	25
120	of the fact that	28	148	to find out whether	25
121	the effect of the	28	149	a second or foreign	24
122	difference between the two	27	150	and at the same	24
123	I would like to	27	151	for the sake of	24
124	of the results of	27	152	in relation to the	24
125	of the two languages	27	153	is referred to as	24
126	participants in this study	27	154	of the most important	24
127	the basis of the	27	155	of the study was	24
128	between the performances of	26	156	the results showed that	24
129	can be regarded as	26	157	the target language and	24
130	in order to make	26	158	this study was to	24
131	is no difference between	26	159	to the use of	24
132	it can be concluded	26	160	a large number of	23
133	it seems that the	26	161	can be considered as	23
134	it was found that	26	162	in this study is	23
135	of the text and	26	163	is a kind of	23
136	the differences between the	26	164	it is necessary to	23
137	through the use of	26	165	the characteristics of the	23
138	in English as a	25	166	the degree to which	23
139	in the course of	25	167	the part of the	23
140	in the source text	25	168	the quality of the	23

	<i>Bundles</i>	<i>Frequency</i>		<i>Bundles</i>	<i>Frequency</i>
169	the role of the	23	196	used to refer to	21
170	the study will be	23	197	whether or not the	21
171	analysis of the data	22	198	as can be seen	20
172	as it was mentioned	22	199	in other words they	20
173	can be seen in	22	200	in the same way	20
174	no difference between the	22	201	investigate the effect of	20
175	of the target text	22	202	mean score of the	20
176	purpose of the study	22	203	on the role of	20
177	seems to be a	22	204	one of the main	20
178	the form of the	22	205	that the difference between	20
179	an important role in	21	206	that there is no	20
180	at the level of	21	207	the means of the	20
181	difference is significant at	21	208	the total number of	20
182	in other words it	21	209	there any difference between	20
183	in the one way	21	210	to find out if	20
184	in the source language	21	211	to make sense of	20
185	is concerned with the	21			
186	is the result of	21			
187	it can be said	21			
188	learners of English as	21			
189	of language learning and	21			
190	of second language acquisition	21			
191	of the study will	21			
192	the importance of the	21			
193	to be the most	21			
194	to take part in	21			
195	used in order to	21			

Appendix B: Lexical bundles in linguistics MA theses of Native students

	<i>Bundles</i>	<i>Frequency</i>		<i>Bundles</i>	<i>Frequency</i>
1	on the other hand	110	31	would like to thank	29
2	in the case of	84	32	can be seen as	28
3	the University of Edinburgh	76	33	the extent to which	28
4	as well as the	75	34	the fact that the	28
5	in the context of	59	35	the nature of the	28
6	it is important to	57	36	the results of the	28
7	at the end of	51	37	at the time of	27
8	the end of the	48	38	it is possible that	27
9	as a result of	47	39	it is possible to	27
10	I would like to	47	40	more likely to be	27
11	the rest of the	47	41	as part of the	26
12	for the purposes of	44	42	for each of the	26
13	can be found in	41	43	on the basis of	26
14	the structure of the	41	44	in the same way	25
15	an example of a	38	45	at the beginning of	24
16	that there is a	38	46	is likely to be	24
17	the use of the	38	47	it is difficult to	24
18	in relation to the	36	48	the use of a	23
19	should be noted that	35	49	to be able to	23
20	can be seen in	34	50	to refer to the	23
21	in terms of the	32	51	a great deal of	22
22	the purposes of this	32	52	are more likely to	22
23	in the form of	31	53	are a number of	21
24	It should be noted	30	54	the size of the	21
25	the total number of	30	56	a large number of	20
26	at the same time	29	57	a wide range of	20
27	at the University of	29	58	the context of the	20
28	can be used to	29	59	the length of the	20
29	in the present study	29	60	the meaning of the	20
30	that there is no	29	61	would need to be	20

Appendix C: Shared bundles in two corpora

	Shared Bundles	Frequency in MA thesis of Iranian Students	Frequency in MA thesis of native Students
1	the results of the	251	28
2	on the other hand	246	110
3	in the case of	145	83
4	the meaning of the	102	20
5	at the same time	96	29
6	at the end of	78	51
7	as well as the	77	75
8	on the basis of	77	26
9	in the present study	73	29
10	the end of the	66	48
11	as a result of	63	47
12	in the form of	59	31
13	that there is a	55	38
14	the fact that the	53	28
15	in terms of the	52	32
16	a great deal of	48	22
17	the extent to which	44	28
18	should be noted that	40	35
19	to be able to	40	23
20	at the beginning of	35	24
21	can be used to	35	29
22	the nature of the	32	28
23	in the context of	31	59
24	the use of a	31	23
25	the use of the	31	38
26	the structure of the	29	41
27	it should be noted	28	30
28	I would like to	27	47
29	the rest of the	25	47
30	in relation to the	24	36
31	a large number of	23	20
32	can be seen in	22	34
33	that there is no	20	29
34	the total number of	20	30