Lexical Bundles in Applied Linguistics: Variations across Postgraduate Genres

Hassan Jalali

Isfahan University of Technology
h.jalali@cc.iut.ac.ir

Abstract

As building blocks of coherent discourse, lexical bundles, frequent word combinations that commonly occur in different registers, have attracted the attention of researchers in corpus linguistics in the last decade. The importance of these word clusters lies in their often necessary functional contribution to the development of evolving discourse. While most previous studies of bundles have been mainly concerned with variations in the use of these word sequences across different registers (e.g., conversation, classroom teaching, and lectures) and a number of disciplines (e.g., history and biology), postgraduate genres have not been the subject of adequate rigorous analysis. This corpus-based study explored possible differences and\or similarities between two students' genres in terms of the variety, structure, and function of lexical bundles. Through two corpora of master theses and doctoral dissertations in one single discipline of applied linguistics, the study showed that not only was there a large intradisciplinary difference between the two genres in the range of bundles employed, but also there were some striking differences in the total frequency and function of these word combinations. Some implications for academic writing instruction have also been discussed.

Keywords: Lexical bundles, master theses, doctoral dissertations, corpus linguistics, applied linguistics

Introduction

As a particular and relatively new category of word combinations, lexical bundles, also known as clusters and chunks (Hyland, 2008b), were defined by Biber, Johansson, Leech, Conrad, and Finegan (1999) in their innovative and extensive treatment of English grammar. They defined lexical bundles as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status" (p. 990). More importantly, they

referred to frequency as the most salient and defining characteristic of bundles; in order for a word combination (e.g., *on the other hand, at the same time, in the case that*, etc.) to count as a bundle, it must occur at least twenty times in a corpus made of one million words and at least five different texts to guard against idiosyncratic or repetitive uses. Fixedness in form (e.g., *on the basis of* not **on a basis of**) and non-idiomatic meaning (e.g., the meaning of a four-word bundle like *in the presence of* is almost easily retrievable form the meaning of its individual parts) are among other properties of bundles.

Along with the structural grouping of bundles (see Table 1), Biber et al. (1999) also compared their uses and distributions across two registers: conversation and academic prose. However, it is not so much just their pervasive presence in the language that has made bundles a topic of high interest especially in recent corpus-based studies, but rather their often necessary functional contribution to the coherence and organization of different texts, either spoken or written (Cortes, 2004, 2008; Biber, Conrad & Cortes, 2003; Biber, Conrad & Cortes, 2004; Hyland, 2008a, 2008b; Jeblonaki, 2009, 2010; Jalali, 2009; Jalali, EslamiRasekh & TavangarRizi, 2008, 2009; Jalali & Ghayoomi, 2010). As building blocks of coherent discourse, these word clusters can serve such a wide range of discursive functions as organization of discourse, expression of stance, and reference to textual or external entities. Lexical bundles in some different disciplines and registers have been classified structurally (Biber et al, 1999; Biber et al., 2004; Biber, 2006b; author, 2008) as well as functionally (Cortes 2001, Biber et al., 2003 ; Biber et al., 2004; Biber, 2006b; Biber & Barbieri, 2007; Hyland, 2008a, 2008b; Jalali, 2009; Jalali et al., 2008, 2009; Jalali & Ghayoomi, 2010). Some of these studies are reviewed here.

Table 1

Most common patterns of 4-word bundles in academic writing (Biber et al., 1999, pp. 997–1025)

Structure	Examples
Noun phrase + of	the end of the, the nature of the, the beginning of the, a large number of
Other noun phrases	the fact that the, one of the most, the extent to which
Prepositional phrase + of	at the end of, as a result of, on the basis of, in the context of
Other prepositional phrases	on the other hand, at the same time, in the present study, with respect to the
Passive + prep phrase fragment	is shown in figure, is based on the, is defined as the, can be found in
Anticipatory it + verb/adj	it is important to, it is possible that, it was found that, it should be noted
Be + noun/adjectival phrase	is the same as, is a matter of, is due to the, be the result of
Others	as shown in figure, should be noted that, is likely to be, as well as the

Among studies focusing on disciplinary variations in the use of these word clusters, Cortes (2002, 2004) found that lexical bundles were used much more frequently in biology research articles than in history, which is a soft field. Her study also showed some major structural and at the same time, some few functional differences between these two disciplines in their uses of bundles. Similarly, Hyland (2008a), working on a large corpus of academic writing, came up with the finding that different disciplines drew on different ranges and types of bundles in their respective discourses. Furthermore, his study found more similarities between cognate fields (i.e., electrical engineering and microbiology on the one hand, and business studies and applied linguistics on the other hand). There were also structural and functional differences between disciplines.

In the studies of variations across registers, Biber et al. (1999) compared conversation and academic prose, while Biber et al. (2004) worked on two other registers: classroom teaching and textbooks. Collectively, these studies showed that the number of lexical bundles in classroom teaching was almost twice more than that of conversation and

around four times more than that of textbooks and academic prose. The strong use of clusters in the classroom teaching was attributed to the heavy reliance of this register on both 'oral' and 'literate' bundles. More extensively, Biber and Barbieri (2007) investigated the use of bundles in a wider range of university registers. They reported the differential pervasiveness of bundles in different university registers, the heavier reliance of written non-academic registers on bundles, and some other differences between registers in structural and functional types of bundles.

With regard to possible generic variations in the use of bundles, there is only one study: Hyland (2008b). Comparing three corpora of master theses, doctoral dissertations, and research articles in four different disciplines, Hyland (2008b) showed that these three different genres relied on different kinds and numbers of bundles with master theses employing bundles more than dissertations and much more than research articles. Structurally, his study showed that unlike research articles, bundles in student genres were more phrasal than clausal. Using his own functional classification of bundles, Hyland (2008b) also demonstrated that bundles in master theses were heavily research-oriented, while bundles in research articles were for the most part text-oriented. Bundles in doctoral dissertations were more similar to research articles in being more text-oriented and less research-oriented, but the bundles in the latter were more participant-oriented (expressing writer's attitude toward the text, content, and the readers) than the former. The study concluded that less proficient and confident writers might rely on formulaic expressions more. The problem with this study, though, was that academic genres in each discipline were not explored separately.

Interestingly, to present researcher's knowledge, just one study has focused specifically on examining possible variations in the use of bundles across different degrees of writing expertise: Cortes (2002, see also Cortes, 2004). Cortes examined students' use of bundles in their essays in two disciplines of history and biology at three levels: undergraduate low level, undergraduate upper level, and graduate. Overall, her study showed that many lexical bundles favored and used by experts in these two fields, as examined by their respective research articles, were never or quite rarely used by either group of students. It was just in the case of biology students that a gradual pattern of development was seen in their use of bundles.

It seems that there is a paucity of studies that would focus on the study of bundles within one single disciplinary area especially with an aim to describe and explain possible differences and/or similarities between high-stake postgraduate genres. Master theses and doctoral dissertations are the most highly valued academic genres created by postgraduate students, who are for the most part still novices in any given disciplinary area. These two students' genres are of particular importance since "They carry the burden of assessment and determine future life chances, but with different expectations for particular forms of argument, cohesion, and reader engagement" (Hyland, 2008b, p. 50).

To address possible intradisciplinary variations in the use of lexical bundles, this study chose to investigate, compare, and contrast the variety, frequency, structure, and function of these word clusters in master theses and doctoral dissertations representing postgraduate writing in one disciplinary area, namely, applied linguistics. The main reason for choosing applied linguistics as the discipline of interest was similar to what Ruiying and Allison (2003, p. 366) said: "Besides being still relatively under-researched, applied linguistics is of particular interest for pedagogic reasons, because raising awareness of genre features becomes directly relevant as part of its disciplinary content as well".

The Study

This study, therefore, was set out to address the following questions:

1. What are the most frequent four-word lexical bundles in two postgraduate genres (master theses and doctoral dissertations) of the discipline of applied linguistics?

2. How can the lexical bundles identified in each of these genres be classified structurally and functionally?

3. To what extent is there evidence to support similarity and\or contrast between the two postgraduate genres in terms of variety, frequency, form, and function of lexical bundles?

Corpora and Text Analysis Programs

Two corpora of doctoral dissertations and master theses, all written by Iranian L1 Persian graduate students of applied linguistics at a state

university in Iran, were collected for the purpose of this study. As Table 2 shows, two relatively small corpora were used in this study to represent the two postgraduate genres. The number of master texts was almost twice more than that of doctoral dissertations although the size of the two corpora was not so much different. Only the main parts of each thesis or dissertation entered each of the corpora, and other parts (e.g., title, headings, acknowledgements, tables, figures, graphs, references, appendices, etc.) were removed from the texts.

Table 2

Master theses and doctoral dissertations corpora word count

Students' genres	Number of texts	Number of words
Master theses	22	441033
Doctoral dissertations	12	476922
Total	34	917955

Two computer programs were used in this study in order to explore lexical bundles, their frequencies, and the number of texts in which they had been used: Antconc 3.2.1w (Anthony, 2007), and Wordsmith (Scott, 2008). The former was used for the identification of lexical bundles and concordancing while the latter was only used to find the number of texts within which each bundle had been used. Each of these programs is described more below.

Antconc 3.2.1.w, developed by Anthony (2007), is a free computer concordancing program used in this study to first identify lexical bundles and then find their frequencies. Among its useful features and tools, this text analysis computer program has a tool by which it can identify lexical bundles of different lengths and frequencies in small or large corpora. By giving it a set of commonly key words with which clusters and bundles usually collocate like articles, prepositions, anticipatory *it*, modals, etc., and deciding on the minimum optimal frequency and the required number of words in clusters, this program can find and display all lexical bundles in corpora of different sizes with their actual frequencies. The concordancer also makes it possible to see clusters in actual textual context within which they had originally been used. In this study, like some other previous studies of lexical bundles (e.g., Cortes, 2002, 2004, 2006, 2008; Biber et al., 2004; Biber & Barbieri, 2007; Hyland, 2008a,

2008b; Jalali, 2009; Jalali et al., 2008, 2009, Jalali & Ghayoomi, 2010), only four-word combinations or bundles were investigated. This was because in comparison to five-word bundles, four-word bundles are much more frequent and also in comparison to three-word bundles, they serve a wider range of functions.

The other computer program, Wordsmith tools 5 has been developed by Scott (2008). It is in many ways similar to Antconc 3.2.1.w, but it can count and display the number of files with which a given bundle is used. So when all candidate lexical bundles were identified by the first computer program, each of them was again searched on Wordsmith tools 5 to find the number of texts. Only those four- word combinations could count as lexical bundles that had been used in at least five different texts no matter how frequent they were. This was to guard against all possible idiosyncratic and\or repetitive uses of the same bundle in the same text by the same writer or writers.

Normalization Procedure
Since the two corpora used in this study were not parallel in size, a normalization procedure had to be employed in order to make it possible to explore lexical bundles and compare the two genres in terms of overall frequency of bundles used. While the limitations of such a procedure have already been pointed out (e.g., Biber & Barbieri, 2007; Cortes, 2002, 2004, 2008), as corpora used in this study were not large enough, the use of such a normalization procedure was inevitable. Of course, despite the limitations, this procedure had already been used in some previous corpus-based studies of lexical bundles like Biber et al. (2004), Biber and Barbieri (2007), and Hyland (2008a, 2008b). This procedure is described more below.

First, two different frequency cut-offs were decided for exploration of lexical bundles in the two genres. As in this study the more conservative frequency cut-off of twenty in one million was adopted, based on the size of the two corpora, two frequency cut-offs of nine and ten were chosen for master theses and doctoral dissertations respectively with the additional requirement that these raw frequencies be realized in at least five different texts as found by the second computer program. Then, the actual total frequency of candidate bundles in each of the genres was multiplied by one million and the result was divided by the actual size of each corpus used in the study.

Structural and Functional Analysis of Lexical Bundles

All lexical bundles identified in each of the two genres were classified structurally using the most widely-used structural taxonomy of bundles developed in Biber et al. (1999) (see Table 1). As for functions, Hyland's functional taxonomy of bundles in academic writing (2008a, 2008b) was used as an initial framework for classification of bundles. While there were some other functional taxonomies of bundles (e.g., Biber et al., 1999; Biber et al., 2003, 2004; Cortes, 2002, 2004, 2008), this taxonomy was used since it was specifically based on academic writing while other classification schemes covered a wider range of registers. However, it seems that as the corpora become more specific (e.g., texts in one disciplinary area rather than different areas), there is a need for the development of further sub-categories to cater for more special functions of bundles in a given discipline (Cortes, 2002). Table 3 represents the extended taxonomy of bundles used in this study.

The three major functions of bundles correspond to Halliday's (1994) tripartite metafucntions of language. Research-oriented bundles serve a more ideational role of encoding activities, experiences, and practices in the real world; text-oriented bundles serve a more textual function of organizing and connecting different parts of discourse; and finally, participant-oriented bundles play a more interpersonal role by establishing interactions between writers and readers (Thompson, 2001; Hyland, 2008a, 2008b). In the case of research-oriented bundles, four new sub-categories were added: study-focusing bundles, statistical bundles, goal-oriented bundles, and discipline-bound bundles (this last sub-category was used instead of 'topic' in the original taxonomy). In the case of text-oriented bundles, only the sub-category 'rephrasing bundles' was added. Finally, in the case of participant-oriented bundles, while engagement features were taken as one of the sub-categories just as that of original taxonomy, different stance features were divided into seven different sub-categories: attitude markers, epistemic-certain, epistemic-uncertain, epistemic-impersonal, intention, ability, and interrogative. Each of these sub-categories will be described and explained in functional analysis and comparison of bundles in the two postgraduate genres.

Results

Lexical Bundles in Master Theses and Doctoral Dissertations: Varieties and Frequencies

Probably, the most surprising finding of this corpus-based study was related to the relatively large difference between the two academic genres under investigation in terms of the range and overall frequency of bundles employed. As can be seen from Table 4, in the corpus of doctoral dissertations there were no more than 141bundles while in the corpus of master theses there were 255 different lexical bundles. So the range of lexical bundles in master theses was 114 more than that of doctoral dissertations. To put generic differences in the variety of bundles used in terms of percentages, it can be said that the range of bundles used in master theses was approximately 45% more than that of doctoral dissertations. The results obtained by the use of normalization procedure also showed that the overall frequency of lexical bundles in master theses was considerably more than that of doctoral dissertations (6860, and 11514, respectively). This means that the total use of lexical bundles in master theses was around 68% percent more than that of doctoral dissertations. It seems; therefore, that master theses go beyond all norms not only in the variety of bundles used but also in the overall use of these word combinations.

Lexical Bundles Used in Both Doctoral Dissertations and Master Theses

There were 83 different lexical bundles used in both doctoral dissertations and master theses. *On the other hand*, *the results of the*, *in terms of the*, and *significant difference between the* were some instances of these bundles in students' genres. In terms of variety, this means that 59% of bundles in doctoral dissertations were also used in master theses while only 32% of all bundles in master theses were used in doctoral dissertations.

Table 3

An extended functional classification of lexical bundles in academic writing (based on Hyland, 2008a, 2008b)

Major functions	Sub-categories	Examples
Research-oriented bundles	Location (time\place)	at the same time, at the end of, at the beginning of
	study-focusing	in the present study, in the current study, in this study is
	quantification	a wide variety of, a large number of, a great deal of
	statistical	the correlation between the, the mean scores of
	procedure	through the use of, the use of the, by the use of
	description	the structure of the, the analysis of the, the quality of the
	goal-oriented	the purpose of the, for the purpose of,
	discipline-bound	as a foreign language, English as second, native speakers of English
Text-oriented bundles	transition signals	on the other hand, as well as the, on the one hand
	resultative signals	on the basis of, as a result of, the influence of the
	structuring signals	as shown in table, in the following chapter
	framing signals	in terms of the, with regard to the, with respect to the
	rephrasing signals	that is to say, in other words the, this means that the
Participant-oriented bundles	attitude markers	it is important to, it is difficult to, it is clear that
	epistemic-certain	that there is a, to the fact that, the fact that they
	epistemic-uncertain	it is possible to, may be due to, it is possible that
	epistemic-impersonal	seems to be a, it is believed that
	intention	in order to find, to find out the
	ability	can be used to, can be found in
	interrogative	to see whether or, to see if the
	engagement	should be noted that, can be seen as

Table 4

Variety and overall use of lexical bundles in postgraduate genres

Genres	Doctoral dissertations	Master theses
Number of bundles	141	255
Actual frequency	3272	5078
Normalized frequency (in one million)	**6860**	**11514**

However, there was a slight difference in terms of the overall frequencies of these shared bundles in students' genres. Through the use of the normalization procedure, it was found that the overall frequency of these shared bundles in doctoral dissertations was 4950 (in one million) and in master theses, 5485. So, it seems that the overall use of shared bundles in master theses was around 10% more than that of doctoral dissertations.

Lexical Bundles in Doctoral Dissertations not Used in Master Theses
There were 58 different bundles in doctoral dissertations not used in master theses. *The performance of the*, *a significant difference between*, *that is to say*, *to the fact that*, and *in the first place* were some examples of these bundles. So, it can be estimated that 41% of bundles in doctoral dissertations were not used in master theses.

Structural Description and Comparison of Lexical Bundles in Postgraduate Genres
The differences and similarities between the two students' academic genres in terms of structural groups of bundles were explored by comparing the variety of bundles used in each structural group and also their respective percentages for their overall use. As Table 5 shows, in terms of variety of bundles used in each structural class, given the higher variety of bundles in master theses, not surprisingly, master theses outnumbered doctoral dissertations in the range of bundles employed for most structural groups. More specifically, except for the seventh structural class of bundles (be+ noun\adjectival phrases), which was used to the same extent and the least in both genres, in other groups, master theses employed more bundles than doctoral dissertations. Furthermore, the relatively scarce use of anticipatory *it* bundles and passive structures (sixth and seven structural groups of bundles) by students (Jalali et al., 2009) could be first attributed to the absence of these structures in some

languages including students' first language, Persian, and more importantly, the strong association of the former with more overt expressions of stance as will be explained in the functional comparison of bundles.

Table 5

Structural comparison of bundles (doctoral dissertations\master theses)

Structures	No of bundles	Overall frequency (in one million)	Percentage (%)
Noun phrase+ of	33\58	1778\2905	25.92\25.23
Other noun phrases	12\28	610\1288	8.90\11.19
Prepositional phrase+ of	35\43	1673\1802	24.38\15.66
Other prepositional phrases	27\54	1503\2765	21.92\24
Passive+ prepositional phrase fragment	6\7	157\301	2.30\2.62
Anticipatory it+ verb\adjective	4\6	145\215	2.10\1.88
Be +noun\adjectival phrase	4\4	124\197	1.80\1.72
Others	20\55	870\2041	12.68\17.70
Total	141\255	6860\11514	100\100

In the case of phrasal bundles, the two academic genres were quite similar in the extent to which they relied on noun phrase and prepositional phrase bundles, lending support to the findings of previous studies (e.g., Biber et al., 1999; Cortes, 2002, 2004) that academic writing, unlike some registers like conversation and classroom teaching, could best be characterized as depending for the most part on phrasal rather than clausal bundles. As can be seen from Table 5, around 76 to 81 % of bundles in the two genres were in the first four phrasal groups (in terms of overall rate of occurrence). However, there were also some differences between the genres in their use of these phrasal groups of bundles.

To begin with the total use, the overall occurrence of phrasal bundles in doctoral dissertations was around 5% more than that of master theses. One of the biggest generic differences in the case of phrasal bundles referred to the overall use of lexical bundles made of prepositional phrases with *of*. As Table 5 can show, the overall use of this structural class of bundles in doctoral dissertations was much more than that of

master theses. This could be because of the more textual nature of doctoral dissertations as such prepositional phrases can lend themselves well to framing, scaffolding, and circumscribing the arguments (Hyland, 2008a, 2008b). On the other hand, the use of other noun phrases and other prepositional phrases in master theses was comparatively more frequent than that of doctoral dissertations. In the case of the three clausal groups of bundles (passive+ prepositional phrase fragments, anticipatory *it* bundles, and *be*+ noun\adjectival phrase), there did not seem to be any significant difference between the two students' genres except their relatively higher variety in master theses.

Functional Description and Comparison of Bundles in Postgraduate Genres

Table 6 shows the varieties, overall frequencies (normalized in one million), and percentages of lexical bundles in terms of the three major functional categories used in this study. As for similarities, the two genres made a much heavier use of research-oriented and text-oriented bundles than participant-oriented bundles. More specifically, more than 85 percent of all bundles in the two genres were research and text-oriented. In terms of the variety of bundles used in each major functional category, as expected, master theses outweighed doctoral dissertations in the range of all bundles employed. As can be seen, the variety of research-oriented bundles in master theses was twice more than that of doctoral dissertations. In the case of the second functional category, text-oriented bundles, the difference between the two genres in terms of variety was less than the previous category, but, interestingly, the overall use of text-oriented bundles in doctoral dissertations was much more than that of master theses, showcasing the more textual nature of this genre. With regard to participant-oriented bundles, which reflect different attitudinal, interpersonal, and affective meanings, the variety in master theses was almost twice more than that of doctoral dissertations although the difference between the genres in their overall use of such bundles was quite small. Therefore**,** unlike few other studies (e.g., Hyland, 2008a, 2008b), master students' use of participant-oriented bundles to express different interactional and evaluative meanings did not seem to be less than that of doctoral students. In the next three parts, functional differences are discussed in more details in terms of different sub-categories.

Table 6

Functional comparison of bundles in the two postgraduate genres (doctoral dissertations\master theses)

Categories	Number of bundles	Frequency (normalized)	Percentage (%)
Research-oriented	51\130	2300\5735	33.53\49.81
Text-oriented	67\85	3692\4115	53.83\35.73
Participant-oriented	23\40	868\1664	12.64\14.46
Total	141\255	6860\11514	100\100

Functional description and comparison of postgraduate genres in terms of research-oriented bundles

As said in the previous part, the overall use of research-oriented bundles in master theses was more than that of doctoral dissertations. As can be seen from Table 7, eight different sub-categories were employed in this study in order to classify all research-oriented bundles in the two corpora. Table 7 can show that there were both similarities and differences between the two genres with regard to sub-categories of research-oriented bundles.

Table 7

Comparison of genres in terms of research-oriented sub-categories (doctoral dissertations\master theses)

Sub-categories	Number	Frequency (normalized)	Percentage (of all bundles)
Location (time\place)	6\7	277\342	4.03\2.97
Study-focusing	7\23	312\1093	4.55\9.50
Quantification	11\16	306\558	4.46\4.85
Statistical bundles	3\16	331\739	4.84\6.41
Procedure	4\11	206\478	3\4.15
Description	8\27	384\1152	5.60\10
Goal-oriented	2\3	75\124	1.10\1.08
Discipline-bound	10\27	409\1249	5.95\10.85
Total	**51\130**	**2300\5735**	**33.53\49.81**

To begin with the first sub-category, location (marking time and place), although the variety of such bundles in the two genres was not

very different, the overall use of them in doctoral dissertations was more than that of master theses. *At the same time*, *in the context of*, *at the end of*, and *in the course of* were some examples of locative bundles used in both master theses and doctoral dissertations. The following examples can showcase the use these bundles in doctoral dissertations and master theses, respectively:

(1) It might as well be admitted that collocations are both indispensable and *at the same time* problematic for foreign language learners and they therefore should play an important role in second language acquisition (SLA), especially for adult learners.

(2) *At the end of* Khordad, 25 out of 100 students produced all 42 vocabulary items correctly in the first task.

Study-focusing bundles were a new sub-category of research-oriented bundles developed in this study to individualize and mark a given study or research. Many times the word 'study' was one of the main constituents of such bundles (e.g., *in the present study*). As can be seen from Table 7, there was a larger difference between master theses and doctoral dissertations in the case of study-focusing bundles. The variety of such bundles in master theses was much more than that of doctoral dissertations (23, 7, respectively). *Of the present study* and *in the present study* were two of the most frequent bundles of this type in both postgraduate genres. The heavy use of study-focusing bundles in master theses could be attributed to a high tendency among students at this level to refer to and perhaps, overemphasize their own studies. The following two examples can illustrate the use of some bundles of this type in doctoral dissertations and master theses, respectively:

(3) Altogether six general patterns of collocation of prepositions were recognized and investigated *in the present study*.

 (4) Regarding the second reason recognized *in the present study* (overload of memory), it seems that human beings, through experience, develop intuitive criteria to predict the limit of working memory span in relation to the load of a specific task.

Quantifying bundles encode number, degree, amount, or variety of their following nominal entities (e.g., *a wide range of*, *a great deal of*). Table 7 shows that although the variety of such bundles in master theses

was more than that of doctoral dissertations, in terms of overall use, there was no significant difference between the two genres. *One of the most*, *in a variety of*, and *a great deal of* are some of the most frequent quantifying bundles in doctoral dissertations while *one of the most*, *in each of the*, and *as one of the* were the three most frequent bundles of this type in master theses.

Statistical bundles, another sub-category used in this study and some other previous studies of bundles (e.g., Cortes, 2002), referred more to statistical notions and terminology (e.g., *the correlation between the*). One can find that master theses employed a wider variety of such bundles than doctoral dissertations. The preponderance of such bundles in master theses could be attributed to the more research-oriented and displaying nature of discourse in this genre (Hyland, 2008b). *Significant difference between the*, *the mean scores of*, and *difference between the mean* were some of the most frequent statistical bundles in master theses.

Comparing the two academic genres in terms of the fifth sub-category of research-oriented bundles, procedure, which stands for bundles that show the way or instrument through which something is done, reflects once more the heavier use of such research-oriented bundles by students at the master's level. *In the use of*, *the use of the*, and *the role of the* were some of procedural bundles used in both genres:

(5) The results of statistical analysis of ANOVA (one-way) for the first research hypothesis indicate that the DDL approach makes a difference in <u>*the use of the*</u> collocation of prepositions. (corpus of doctoral dissertations)

One of the largest differences between master theses and doctoral dissertations was in the use of descriptive bundles. As can be seen from Table 7, the variety of descriptive bundles in master theses was three times more than that of doctoral dissertations. The overall use of such bundles also showed that students at the master's level relied more on this group than doctoral students. The following two examples can showcase the use of some of these bundles by students at the master's level and doctoral students, respectively:

(6) Table 4.8 illustrates <u>*the findings of the*</u> chi-square statistics related to the second null hypothesis.

(7) It means that the post test 1, being communicative and qualitative, significantly improved *the performance of the* experimental group.

The difference between the two academic genres in their use of research-oriented bundles was the least in the case of goal-oriented bundles. As can be seen, in both genres, there were few bundles of this type. The overall use also did not show any considerable generic differences. *The purpose of the*, *for the purpose of*, and *the purposes of the* were the only three bundles of this type used in master theses or doctoral dissertations.

Discipline–bound bundles, used in this and some other studies (e.g., Cortes, 2002), are specialized discipline-specific word combinations used exclusively and in varying degrees by members of a given disciplinary community. As can be seen from Table 7, the use of such bundles in master theses was considerably heavier than that of doctoral dissertations. While *as a foreign language*, *English as a foreign*, *in the target language*, and *in a foreign language* were some of the most frequent bundles of this type in master theses, *native speakers of English*, *in the target language*, and *in second language acquisition* were the frequent clusters in doctoral dissertations:

Functional description and comparison of postgraduate genres in terms of text-oriented bundles

The overall use of text-oriented bundles in doctoral dissertations, as said before, was noticeably more than that of master theses, supporting the more argumentative and norm-developing nature of this genre (Hyland, 2008b). There were some other differences between the two academic genres in terms of sub-categories of text-oriented bundles. To begin with transition signals, which mark additive or contrastive relations between prior and coming discourse (e.g., *as well as the*, *on the other hand*), one can notice a large difference between doctoral dissertations and master theses: although the variety of such bundles in each of the genres was the same, the overall frequencies and percentages indicated that students' use of such bundles at the master's level was almost twice less than that of doctoral students (see Table 8). Interestingly, *on the other hand*, *as well as the*, and *on the one hand* were the three most frequent bundles of this type in both genres.

Table 8

Comparison of genres in terms of text-oriented sub-categories (doctoral dissertations\master theses)

Sub-categories	Number	Frequency (normalized)	Percentage % (of all bundles)
Transition signals	6\6	574\438	8.38\3.80
Resultative signals	17\24	1132\1537	16.5\13.35
Structuring signals	0\13	0\399	0\3.46
Framing signals	41\41	1818\1719	26.5\14.92
Rephrasing signals	3\1	168\22	2.45\0.20
Total	**67\85**	**3692\4115**	**53.83\35.73**

Although the variety of bundles serving as resultative signals (showing cause-effect relationships) in master theses was more than that of doctoral students, the overall use of such bundles indicated that doctoral students seemed to rely on these bundles again more than students at the master's level. The following examples can illustrate the function of these bundles in their actual contexts of use:

(8) *The results of this* analysis helped to specify the extent to which the monolingual and bilingual participants are sensitive to superiority and subjacency movement constraints exerted on multiple wh-questions in complex constructions. (Corpus of doctoral dissertations)

(9) The results of this study showed that English majors used learning strategies with high to medium frequency, and that the highest rank (79.6%) was for metacognitive strategies while the lowest (63%) was for compensation strategies. (Corpus of master theses)

Structuring signals are used to announce different text stages and signpost different parts of the evolving text (e.g., *in the next chapter*, *are shown in table*). Surprisingly, there were no bundles in doctoral dissertations which would serve as structuring signals while this group of text-oriented bundles was found to be used in master theses. The absence of structuring signals in doctoral dissertations could be partly attributed to the relatively fewer number of texts used in this corpus. *As can be seen*, *will be discussed in*, *will be presented in*, and *can be seen in* were some of the most frequent structuring signals in master theses:

(10) <u>*As can be seen*</u> in figure 4.3 the answers for group one are illustrated in the first columns, while the answers for the second group are displayed in the second columns and the third columns represent the third group answers.

Framing signals were the most widely-used sub-category of text-oriented bundles in the two genres. Bundles used in this way served to condition propositional meaning by setting restricting conditions on arguments (Hyland, 2008b) (e.g., *in the case of*, *from the perspective of*). As can be seen from Table 8, the variety of bundles serving such a function in students' genres was the same, but the overall use of such bundles in master theses was noticeably less than that of doctoral dissertations. *In terms of the*, *with respect to the*, and *in the case of* were some of the most frequent bundles of this kind in doctoral dissertations while bundles *in the case of*, *in the process of*, and *of the three groups* were found to be more frequent in master theses.

The last sub-category of text-oriented bundles, rephrasing signals, which was added in this study to identify those lexical bundles which served a more reformulatory and explanatory role in discourse, did not show any significant difference between genres except its comparatively higher use in doctoral dissertations. While *in the sense that* was the only bundle of this kind in master theses, in doctoral dissertations there were two other bundles as well: *that is to say* and *is to say that*.

Functional description and comparison of postgraduate genres in terms of participant-oriented bundles

As said before, participant-oriented bundles play a more interpersonal role by reflecting different kinds of epistemic, attitudinal, and interactional meanings. Many of these meanings have also already been studied under such varied labels as 'metadiscourse elements' (Hyland, 1999, 2000, 2004 2005; Hyland and Tse, 2004) and 'stance expressions' (Biber et al., 1999; Biber, 2006a, 2006b). While in Hyland's functional taxonomy (2008a, 2008b), bundles serving such functions were only classified into two broad categories of stance and engagement features, in this study, while retaining engagement features as one of the sub-categories, different stance meanings were differentiated on the basis of analysis of bundles in their actual contexts of use and some other prior studies of stance (Cortes, 2002, 2004, Biber et al., 2004; Biber, 2006a; Biber & Barbieri, 2007; Jalali et al., 2009).

As can be seen from Table 9, in both genres, participant-oriented bundles were the least used in comparison to the previous two categories. In fact, just less than fifteen percent of all word combinations served more as participant-oriented bundles in the two genres. The use of such bundles in master theses, both in terms of variety and overall use, also seemed to be more than those of doctoral dissertations. There were some other generic differences in terms of each sub-category of participant-oriented bundles to be discussed below.

Table 9

Comparison of genres in terms of participant-oriented sub-categories (doctoral dissertation\master theses)

Sub-categories	Number	Frequency (normalized)	Percentage (of all bundles)
Attitude markers	5\3	155\86	2.27\0.75
Epistemic-certain	8\7	392\493	5.72\4.27
Epistemic-uncertain	3\4	107\129	1.55\1.12
Epistemic-impersonal	3\7	88\270	1.28\2.35
Intention	3\12	105\419	1.52\3.65
Ability	0\1	0\63	0\0.55
Interrogative	0\4	0\129	0\1.12
Engagement	1\2	21\75	0.3\0.65
Total	**23\40**	**868\1664**	**12.64\14.46**

To begin with bundles serving as attitude markers, one can find a relatively large difference between doctoral dissertations and master theses especially in their overall use of these bundles. As can be seen from Table 9, doctoral students employed attitude markers at least three times more than students at the master's level. While an attitude marker like *it is important to* was found to be used in both of the genres to varying degrees, there were some bundles like *it is necessary to* and *should be mentioned that* that were only used by doctoral students. The relatively absence of bundles acting as attitude markers in master theses could be partly accounted for by referring to students' incipient growing disciplinary identity at this level (Jalali et al., 2009). The use of some attitude markers (e.g., *it is necessary to*) implies the voice of a disciplinary expert who in one way or another directly judges and comments on the value and status of propositional meaning following

such bundles. It seems that students at the master's level are not yet confident enough to overtly signal their presence in the text through such bundles:

(11) This is not the place to change this, but in order to be able to discuss the question of whether parameter (re)setting is possible in SLA, *it is necessary to* spell out at least the core of PT and to identify some empirical facts which might corroborate or which would constitute counterevidence to hypotheses of this theory. (Corpus of doctoral dissertations)

(12) Finally, it *should be mentioned that* almost the same significant differences were found across the six groups on the oral and written translation of the mentioned three types of object wh-questions. (Corpus of doctoral dissertations)

The use of those lexical bundles conveying an epistemic-certain meaning turned the pendulum again toward doctoral dissertations. Although the total frequency of such bundles in master theses was even more than doctoral dissertations, the overall use showed that the use of such bundles in doctoral dissertations was more than that of master theses:

(13) A comparison of the frequency and percentages of errors made by the participants reveals *that there is a* tendency among Iranian EFL students to use their L1 collocational patterns into L2 settings. (Corpus of doctoral dissertations)

(14) This general agreement cannot, however, ignore *the fact that the* role of awareness in second language acquisition has generally been challenging. (Corpus of master theses)

Bundles conveying an epistemic-uncertain meaning, also referred to as hedging devices in some previous literature (e.g., Hyland, 1996), show a kind of tentativeness and uncertainty toward the following propositions (e.g., *it is possible to*, *may be due to*). There did not seem to be any significant difference between the two academic genres in their use of these bundles. It can be seen that the use of such bundles was generally not very frequent either in doctoral dissertations or master theses.

Bundles conveying an epistemic-impersonal meaning usually have passive agentless structures (e.g., *it was found that*, *they were asked to*) by which writers can encode an argument as a fact without identifying or committing themselves to it. In other words, by using such bundles, writers can give a factual status to a proposition without attributing the argument to any one including themselves. As can be seen from Table 9, the use of such bundles in master theses was more than that of doctoral dissertations. There were some bundles like *they were asked to* and *it is believed that* that were only used by students at the master's level.

In the case of bundles encoding intentionality (e.g., *in order to find*, *to find out the*, *in order to investigate*), there was a relatively large difference between master theses and doctoral dissertations. In terms of frequency and variety of such bundles, it can be seen that master's students' use of such bundles was four times more than that of doctoral dissertations. In terms of overall use; however, students at the master's level used bundles expressing intention just twice more than doctoral students.

Bundles expressing ability were one of the least used in both corpora. There was only one bundle of this type in the whole corpus of master theses (to *be able to*) although the frequency of this bundle in this genre was relatively high. Interestingly, doctoral dissertations did not embark even on this bundle to encode ability. Bundles serving a more interrogative role were also only used in master theses (e.g., *to see if the*, *to see whether or*).

Bundles serving as engagement features specifically address the reader (Thompson, 2001) and play a direct interactional relationship between the writers and imagined readers (Hyland, 2005, 2008a, 2008b). As can be seen from Table 9, the use of bundles serving such a function in both doctoral dissertations and master theses was quite rare. While in the corpus of doctoral dissertations, only the bundle *should be noted that* served such a function, in the corpus of master theses, *it should be noted* was also found. The scarce use of such bundles in students' genres could be attributed to the scope of the readership for each of these genres and writers' audience sensitivity. Unlike research articles, doctoral dissertations and master theses do not have a wide readership, so writers may not be very careful in their managing of readers and bringing them into the text (Jalali et al., 2009). On the other hand, students at the master's or doctoral level may not have a clear sense of the audience and

not think so much about the readers, their expectations, reactions, and processing difficulties.

Discussion

Students' abundant use of bundles, especially at the master's level, in their writing could be taken as the most surprising result found in this study given the finding of the previous research that showed (e.g., Cortes, 2004, 2006) generally, students, whether native, non-native, graduate, or undergraduate, as novices to any particular disciplinary community, tended to rely less on bundles in the development of their discourses. In fact, it has often been the failure to use such multi-word sequences that identifies students as novices and newcomers to a community (Haswell, 1991; Hyland, 2008a). It seems that postgraduate students both at the master's and doctoral levels tend to use lexical bundles for a wide variety of discursive functions. Less proficient language users and generally those who have not yet established themselves well in the discourse community of experts may need to rely more on multi-word expressions like lexical bundles (Hyland, 2008b).

More specifically, unlike postgraduate students, research article writers as experts in any given disciplinary area may just need to draw upon a specific and limited set of bundles in the development of their discourses. Such expert writers may rely on other linguistic resources like specialized vocabulary, diverse word choices, conjunctions, discourse markers, and manipulation of syntactic devices as well as bundles to develop their arguments (Jalali et al., 2009). Postgraduate students, on the other hand, may not have access to all these different resources. Although it is yet far from clear whether or not lexical bundles have a formulaic status (Biber & Barbieri, 2007), the abundant use of these word combinations by postgraduate students can suggest that generally less proficient and expedient members of a discourse community may need to rely more on these expressions since they could act as easier and more accessible short-cuts through which they can develop their discourses. The findings of this study also run counter to claims made by some previous researchers (e.g., Yorio, 1989; Cortes, 2004, 2006) that generally formulaic sequences are difficult to acquire.

One other explanation that may provide a partial account for the heavy use of bundles in postgraduate writing could be provided by referring to a notion of limited vs. extended discourses. Unlike research

article, which is an academic genre characterized by compactness of information, less repetition and redundancy, and limitations in space, theses and dissertations are extended discourses which allow student writers to go to as much length and detail as they need to serve their more demonstrative and less argumentative purpose of displaying their knowledge and familiarity with research practices (Hyland, 2008b; Milton, 1999).

One other important thing about lexical bundles that this study was able to discover and demonstrate quite well was showcasing the undeniable influence of corpus size and the variety of texts used on the range and types of bundles recognized in each of academic genres. While Hyland (2008b) found that in his corpus of doctoral dissertations, there were 95 different bundles, in this study, more than 140 bundles were identified in this genre. On the other hand, in the case of master theses, Hyland discovered 149 bundles while the number of bundles in master theses in this study was 255. These differences could be attributed mostly to the corpus size, and also partly to the effect of first language and culture. Overall, it can be postulated that variations within genres of a single discipline are generally much more than those between the same genres but across different disciplines.

Another finding of this study which needs explanation refers to the observed wide discrepancy between master theses and doctoral dissertations in the variety and overall frequency of lexical bundles. This study showed that the overall use of lexical bundles in master theses was much more than that of doctoral dissertations. While part of this large gap between the two postgraduate genres could be attributed to differences in the kind of genre (Hyland, 2008a), the potential influence of other factors like second language proficiency, rhetorical awareness, students' knowledge of genre expectations, the topics and areas of research, students' prior reading experience, supervisors' preferences, and students' degree of familiarity and their expertise in the use of clusters cannot easily be underestimated. Overall, it seems that doctoral students are more similar to research article writers in their less reliance on bundles and also their more dependence on text-oriented bundles to develop their discourses.

Some previous studies of bundles have shown that these multi-word sequences could be strong register and discipline discriminators (Biber et al., 1999; Cortes, 2002, 2004; Biber et al., 2004; Hyland, 2008a), i.e.,

each register or discipline draws on a distinctive set of bundles. This study showed that lexical bundles could also be relatively strong genre discriminators as well since the two academic genres under investigations in this study embarked on specific and almost different sets of bundles in their discourses. More importantly, it could be claimed that just as research articles in different discipline rely on different set of bundles (Cortes, 2002, 2004; Hyland, 2008a, 2008b), students' postgraduate genres in different disciplinary areas could be distinguished, among other means, by their different phrase logical preferences in their choice of bundles. However, further studies of lexical bundles in postgraduate genres of different disciplines, writers with different L1s and cultural communities are needed to shed more light on the extent to which master theses and doctoral dissertations are similar and\or different across different disciplinary communities.

Conclusion

Although there are already some models on how to introduce students to different word combinations (e.g., Nattinger & DeCarrico, 1992; Lewis, 1997; Willis, 2003), the findings of this study call for a more increased pedagogical focus on different multi-word sequences like lexical bundles (Neely & Cortes, 2009). The findings can also stress a more genre-focused EAP (English for academic purposes) especially in advanced writing courses, where students are helped to prepare themselves to join the community of research article writers (Byrd & Coxhead, 2010; Pang, 2010). It is important for students to come to the awareness that if not many, some of their preferred word preferences may be frowned upon when they use them in their submissions.

Developing instructional packages especially "corpus-enhanced disciplinary writing courses" (Cortes, 2006) through which lexical bundles, their distributions across different genres and registers, as well as their functions would be introduced to students could be one of the main frontiers in EFL\ESL writing courses. This study showed that different academic writers in different disciplines and genres drew on different lexical bundles to develop their arguments and persuade the readers. It is important especially for EAP course designers to be well aware of this and expose students to those clusters that they will likely need to use in their target genres. The use of noticing (Schmidt, 1990; Cortes, 2004, 2006), conscious raising tasks (Lewis, 2000a, 2000b), clusters lists, and concordances (Hyland, 2008a) could be some of the

means by which students could come to a better understanding of these word combinations.

References

Anthony, L. (2007). *Antconc 3.2.1: Freeware corpus analysis toolkit.* Retrieved from http://www.antlab.sci.waseda.ac.jp/

Biber, D. (2006a). *University language*: *A corpus-based study of spoken and written registers*. Amsterdam: Benjamin.

Biber, D. (2006b). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, *5* (2), 97-116.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26* (3), 263-286.

Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: An initial taxonomy. In A. Wilson, P. Rayson, & T. McEnery (Eds.), *Corpus linguistics by the lune: A festschrift for Geoffrey Leech* (pp. 71-92). Frankfurt: Peter Lang.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at …: lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25* (3), 371–405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson.

Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, *5* (5), 31-64.

Cortes, V. (2001). *Lexical bundles in context*: *A new taxonomy* (Unpublished manuscript). Northern Arizona University, Arizona.

Cortes, V. (2002). *Lexical bundles in academic writing in history and biology* (Unpublished doctoral dissertation). Northern Arizona University, Arizona.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23* (4), 397–423.

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example form a writing intensive history class. *Linguistics and Education*, *17* (4), 391-406.

Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, *3* (1), 43-57.

Halliday, M. A. K. (1994). *Functions of language* (2nd Ed.). London: Arnold.

Haswell, R. (1991). *Gaining ground in college writing: Tales of development and interpretation*. Dallas: Southern Methodist University Press.

Hyland, K. (1996). Writing without conviction? Hedging in scientific research articles. *Applied Linguistics*, *17* (4), 433-454.

Hyland, K. (1999). Talking to students: Metadiscourse in introductory coursebooks. *English for Specific Purposes*, *18* (1), 3-26.

Hyland, K. (2000). *Disciplinary discourses: Social interaction in academic writing*. London: Longman.

Hyland, K. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, *13* (2), 133–151.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, *7* (2), 173–192.

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* (1), 41-62.

Hyland, K., & Tse, P. (2004). Disciplinary interactions: A reappraisal. *Applied linguistics*, *25* (2), 156-177.

Jalali, H. (2009). *Lexical bundles in applied linguistics: Variations within a single discipline* (Unpublished doctoral dissertation). Isfahan University, Isfahan.

Jalali, H., Eslami Rasekh, A., & Tavangar Rizi, M. (2008). Lexical bundles and intradisciplinary variation: The case of applied linguistics. *Iranian Journal of Language Studies*, *2* (4), 447-484.

Jalali, H., Eslami Rasekh, A., Tavangar Rizi, M. (2009). Anticipatory 'it' lexical bundles: A comparative study of student and published writing in applied linguistics. *Iranian Journal of Language Studies*, *3* (2), 177-194.

Jalali, H., & Ghayoomi, S. (2010). A comparative qualitative study of lexical bundles in three academic genres of applied linguistics. *Modern Journal of Applied Linguistics*, *2* (4), 323-333.

Jeblonaki, R. (2009). "IN THE LIGHT OF": A corpus-based analysis of lexical bundles in two EU-related registers. *WoPaLP*, *3*, 1-27.

Jeblonaki, R. (2010). English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes, 29* (4), 253–267.

Lewis, M. (1997). *Implementing the lexical approach. Hove:* Language Teaching Publications.

Lewis, M. (2000a). Learning in the lexical approach. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 155-185). Hove: Language Teaching Publications.

Lewis, M. (2000b). Materials and resources for teaching collocation. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 186-204). Hove: Language Teaching Publications.

Milton, J. (1999). Lexical Thickets and electronic gateways: Making text accessible by novice writers. In C. N. Candlin & K. Hyland (Eds.),

Writing: Texts, processes, and practices (pp. 221-243). London: Longman.

Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching.* Oxford: Oxford University Press.

Neely, E., & Cortes, V. (2009). A little bit about: Analysing and teaching lexical bundles in academic lectures. *Language Value*, *1* (1), 17-38.

Pang, W. (2010). Lexical bundles and the construction of an academic voice: A pedagogical perspective. *Asian EFL Journal*, *47*, 1-13.

Ruiying, Y., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes*, *22* (4), 365-385.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11* (2), 129-158.

Scott, M. (2008). *Wordsmith Tools 5*. Oxford: Oxford University Press.

Thompson, G. (2001). Interaction in academic writing: Learning to argue with the reader. *Applied Linguistics*, *22* (1), 58-78.

Willis, D. (2003). *Rules, patterns, and words: Grammar and lexis in English language teaching*. Cambridge: Cambridge University Press.

Yorio, C. (1989). Idomaticity as an indicator of second language proficiency. In K. Hyltenstam & K. Obler (Eds.), *Bilingualism across the lifespan* (pp. 55–72). Cambridge: Cambridge University Press.