ACADEMIC WRITING REVISITED: A PHRASEOLOGICAL ANALYSIS OF APPLIED LINGUISTICS HIGH-STAKE GENRES FROM THE PERSPECTIVE OF LEXICAL BUNDLES

Hassan Jalali *
Assistant Professor
Isfahan University of Technology
h.jalali@cc.iut.ac.ir

Gholam Reza Zarei
Associate Professor
Isfahan University of Technology
grzarei@cc.iut.ac.ir

Abstract

Lexical bundles are frequent word combinations that commonly appear in different registers. They have been the subject of much research in the area of corpus linguistics during the last decade. While most previous studies of bundles have mainly focused on variations in the use of these word combinations across different registers and a number of disciplines, not much research has been done to explore some high-stakes written academic genres of one single disciplinary area. This more qualitative study aimed at finding the way in which target bundles in the discipline of applied linguistics, as identified in research articles, were used by two groups of EFL postgraduate students (master-level and doctoral students) as novice discourse community members in the same discipline. Surprisingly enough, the study, contrary to some findings of the previous research, found that in many cases, postgraduate students were able to use target bundles as published writers did. The study, therefore, revealed little if any difference between the three groups of writers in their actual use of lexical bundles. Notwithstanding this, there were some remarkable discrepancies between the three groups with regard to some structural and functional classes of bundles.

Keywords: lexical bundles, research articles, doctoral dissertations, master theses, applied linguistics

1. Introduction

Lexical bundles, also known as clusters and chunks (Hyland, 2008b), are a particular and relatively new category of word combinations with a possibly formulaic status (see Biber & Barbieri, 2007). These word combinations were introduced and 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; to explain further, 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 ten times in a corpus made of one million words with the additional requirement that this rate of occurrence be realized in at least five different texts to guard against idiosyncratic or repetitive uses. Fixedness of 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.

Lexical bundles have also been classified structurally in some disciplines and registers (Biber et al, 1999; Biber, 2006a; Biber & Conrad, 1999; Biber, Conrad & Cortes, 2004). However, it is not only their pervasive presence in the language that has made bundles a topic of high interest, especially in recent corpus-based studies, but rather it is often their noteworthy functional contribution to the development, coherence and organization of different texts, either spoken or written, (Biber et al, 2004; Cortes, 2004; Hyland, 2008a, 2008b) that has attracted the researchers in the field. There have been, therefore, a number of studies that have developed functional classifications of these word combinations (e.g., Biber, Conrad &Cortes, 2003; Biber et al, 2004; Biber & Barbieri, 2007; Cortes 2001, 2002, 2004, 2008; Hyland, 2008a, 2008b). It has been shown quite well that as building blocks of coherent discourse (Hyland, 2008a, 2008b), 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. It seems that the frequent use of lexical bundles signals competent

language use within a register such that learning conventions of a register use may entail learning how to use certain fixed phrases (Hyland, 2008a, 2008b). However, significant questions remain about the extent to which fixed phrases are unique to particular registers and the extent to which expert disciplinary writing differs from that of novices with respect to the use of discipline-specific frequent word combinations.

An important feature of bundles is their variability across different genres. Biber (2006a), for instance, discovered that the spoken genre of classroom teaching uses about twice as many different bundles as conversations and about four times as many as textbooks. He suggests that this extremely high density can be explained by the fact that teaching draws heavily on both oral and written genres. He also found that the bundles are required to do very different jobs in the two genres, with classroom talk comprising much higher proportions of discourse organisers (*going to talk about*, *it has to do with*) and stance bundles (*I don't know if*, *I want you to*) than textbooks. Similarly, Cortes (2004) found systematic differences between genres, with bundles typical of published academic prose being far less common in writing by second language students. In fact, it is often a failure to use native-like formulaic sequences which identifies students as outsiders and there is a general consensus that formulaic sequences are difficult for L2 learners to acquire (Yorio, 1989).

According to Cortes (2004), generally there are two different approaches for the identification of frequent word combinations. The first approach relies on a group of expressions selected prior to the study (Nattinger & DeCarrico, 1992), which are considered significant because they can be recognized as familiar by native speakers of the language or because they have been found to be frequent in the related literature. The second approach employs a search tool that explores lexical co-occurrences of different length, at different cut-off frequency points (Biber, Johansson, Leech, Conrad & Finegan, 1999). This study then takes up the second approach and uses an already developed tool to deal with the issue in question. Some of the studies, those related to lexical bundles, are reviewed here.

2. Literature Review

Since 1999, a number of corpus-based and mostly comparative studies have been specifically launched to explore possible differences and/or similarities in the use of bundles between different disciplinary fields, registers, different degrees of writing expertise, and genres. Among studies focusing on disciplinary variations in the use of word clusters, Cortes (2002, 2004) found that research articles in biology as a kind of hard field employed bundles much more than those of history, which is a soft field. Her study also showed some major structural and a few functional differences between these two disciplines in their uses of bundles. Similarly, Hyland (2008a), through a relatively large corpus of academic writing, showed 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).

In studies addressing 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. These two 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. As another more extensive research, Biber and Barbieri (2007) investigated the use of bundles in a wider range of university registers. This study pinpointed the differential pervasiveness of bundles in different university registers, the heavier reliance of written non-academic registers (e.g. course syllabi) on bundles, and some other differences between registers in structural and functional types of bundles.

Interestingly, few studies have specifically focused on examining possible variations in the use of bundles across different degrees of writing expertise, especially in EFL settings (see also Levy, 2003; Cortes, 2002, 2004). Among these, 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 level.

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 groups of students.

With regard to possible generic variations in the use of bundles, among few studies done, we can mention Hyland (2008b). Comparing three corpora of master theses, doctoral dissertations, and research articles in four different disciplines, this study 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. This study also showed that unlike research articles, bundles in student genres were structurally more phrasal than clausal. It was also demonstrated that bundles in master theses were heavily research-oriented (describing the world, facts, and activities), while bundles in research articles were for the most part text-oriented (organizing and connecting different parts of the discourse). Bundles in doctoral dissertations were more similar to research articles in being more text-oriented and less research-oriented, but at the same time, the bundles in the latter were more participant-oriented (expressing writers' attitudes toward the texts, content , and the readers) than the former. The study concluded that less proficient and confident writers might rely on formulaic expressions more. The major problem with this study, though, was that academic genres in each discipline were not explored separately and a relatively small number of research articles (30) were used for the exploration of bundles in each discipline.

The review of the above research shows that there have been very few studies focusing on the study of bundles within one single disciplinary area especially with an aim to describe and explain possible differences and/or similarities between published experts and novice postgraduates in the use of these word combinations in their respective high-stakes genres: research articles, doctoral dissertations, and master theses. Research article represents the most prestigious genre of the academy and is "the principal site of disciplinary knowledge-making" (Hyland, 2008b, p.46), and the means by which experts in a given field construct, shape, and interpret disciplinary knowledge and understanding. Doctoral dissertations and master theses, on the other hand, are the most highly valued academic genres created by

postgraduate students, who are for the most part still novices in a given disciplinary area. Both doctoral dissertations and master theses are important because "They carry the burden of assessment and determine future life chances, but with different expectations for particular forms of argument, cohesion, and reader engagement" (ibid. p.47).

It might be assumed that lexical bundles are simple expressions that can be acquired easily in the natural course of language learning. If this is the case, then advanced language learners, such as university students at postgraduate levels, should get to master the use of these expressions in their language communities. Making an attempt to test this hypothesis and following on several studies which have compared the use of different linguistic features in the written production of published authors and university students (Conrad, 1996; Hewings & Hewings, 2002), the present research employed a corpus-based approach to examine lexical bundles in published and student academic writing in applied linguistics.

More specifically, this study chose to first identify lexical bundles in research articles, known as target bundles (Cortes, 2002, 2004), and then investigate the frequency, form, and function of such bundles in the two postgraduate genres of applied linguistics: master theses and doctoral dissertations. The rationale behind this is that the use of typical lexical bundles in applied linguistics published writing is part of disciplinary expertise expected from postgraduate students. Furthermore, The main reason for choosing applied linguistics as the discipline of interest is similar to what Ruiying and Allison (2003) say "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" (p. 366).

The paper proceeds with the description of corpora, text analysis programs, and data analysis methods used in the study. In the results section, lexical bundles identified in research articles are classified both structurally and functionally and then the study shows the extent to which target bundles, as identified in published writing, are used in the two students' genres. In the discussion section, an attempt is made to interpret and explain

the findings. Finally, the paper ends with two separate sections for pedagogical implications and conclusion.

Built up on the rationale stated above, this study addresses the following questions:

1. What are the most frequent four-word lexical bundles in published writing in applied linguistics?

2. How can these lexical bundles (called target bundles) be classified structurally and functionally?

3. Are these target bundles used by postgraduate students in applied linguistics master theses and doctoral dissertations? If so, how frequently are these bundles used?

4. Are the functions students try to convey with those bundles similar to those conveyed by published authors?

3. Method

3.1 Corpora

Three corpora were used in this study. The first two ones consisted of doctoral dissertations and master theses in applied linguistics that all had been written by Iranian (L1 Persian) postgraduate students of applied linguistics in Iran during 2004-2008. All of these theses and dissertations, which had been written in English, were reflecting the relevance to English language education. As Table 1 shows, the number of master texts used was almost twice more than that of doctoral dissertations although the size of the two students' corpora was not so very different. Only the main parts of each thesis or dissertation were incorporated in each of the corpora, and other parts (e.g. titles, headings, acknowledgements, tables, figures, graphs, references, appendices, etc.) were removed from the texts. The third corpus, the corpus of research articles, was much larger than students' corpora. To cater for corpus representativeness and size, 201 articles from seven different journals in applied linguistics were taken, covering a number of volumes during 2006-2007. Decision about the choice of the journals in the field was based mostly on experts' opinions, previous corpus-based studies on applied linguistics (e.g. Ruiying, & Allison, 2003, 2004), practical access

to electronic files of articles, and relevance to English language education (see Table 2).

Table 1. 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

Table 2. Research articles corpus word count

Journals	Number of texts	Number of words
Applied Linguistics	29	240212
English Language Teaching	45	151506
English for Specific Purposes	37	250576
English for Academic Purposes	20	125236
Second Language Writing	14	108663
Linguistics and Education	11	94614
System	45	247156
Total	201	1217963

3.2 Text analysis programs

Two computer programs were used in this study in order to explore lexical bundles, their frequencies, the number of texts in which they had been used, and their actual contexts of use: AntConc 3.2.1w (Anthony, 2007), and Wordsmith tools 5 (Scott, 2008). The former was used for the identification of lexical bundles and concordancing while the latter was only employed to find the number of texts within which each bundle had been used. AntConc 3.2.1 is a free text analysis computer program that can identify word combinations, clusters, or lexical bundles of different lengths and frequencies in small or large corpora. By giving it a set of common key words with which clusters and bundles usually collocate like articles (e.g. *the*), prepositions (e.g., *of, in, on, at*, etc.) anticipatory *it*, modals (e.g., *can, should*), etc., and deciding on the minimum optimal frequency (e.g. twenty in a corpus of one million words) and the required number of words in clusters (i.e. three, four, five, or six), 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 each of the clusters in their actual textual context. The only problem with AntConc 3.2.1 is that it cannot display the number of texts within which a given lexical bundle had been used. This problem was tackled by employing another text analysis program: Wordsmith tools 5. This computer program, developed by Scott (2008), is in many ways similar to AntConc 3.2.1.w, but it can count and display the number of files, and hence the number of texts in which a given bundle had been used.

In this study, like some other previous studies of lexical bundles (e.g. Biber et al, 2004; Biber & Barbieri, 2007; Cortes, 2002, 2004, 2006; Hyland, 2008a, 2008b), 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 they serve a wider range of functions than three-word bundles, which are for the most part too frequent to be managed in a study of this kind (Hyland, 2008a, 2008b).

3.3 Data analysis

Given the relatively large difference in the size of the corpora and the possible unreliability of employing a normalization procedure (Biber & Barbieri, 2007; Cortes, 2002, 2004, 2008), a more qualitative approach, similar to that used by Cortes (2004), was followed in this study. It must be noted that this study is qualitative in the sense that lexical bundles identified in research articles and known as target bundles were taken as some kind of yardstick by which theses and dissertations were evaluated in terms of their use of these word sequences as part of accepted disciplinary practices in the field. In other words, the first part of the study tried to elicit a number of word sequences which could be recognized as bundles. This was achieved through definitional and qualitative criteria suggested in the literature. The findings in the first part were subsequently drawn on as the evaluative procedure for the establishment of disciplinary nature of the corpora used. This is further described here.

First, the corpus of research articles was explored to identify candidate bundles in the published writing in applied linguistics. In this study, as the more conservative frequency cut-off of twenty in one million was adopted,

to find lexical bundles in research articles corpus, which was larger than one million (see Table 2), four-word clusters had to occur at least twenty-five times and in five different texts to count as bundles. Bundles identified in this way were regarded as target bundles and the other two corpora were examined to see how they were used by postgraduate students and whether the functions for which they were used corresponded to those of published authors.

All lexical bundles identified in research articles were classified structurally using the most widely-used structural taxonomy of bundles developed by Biber et al. (1999) (see Table 3). As for functions, Hyland's functional taxonomy of bundles in academic writing (2008a, 2008b) was used as an initial framework for the classification of bundles. While there were some other functional taxonomies of bundles (e.g. Biber et al., 1999; Biber et al., 2003, 2004), this taxonomy was used since it was specifically based on academic writing.

Table 3. 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

While this taxonomy was used as the preliminary framework for the classification of bundles, the analysis of bundles in their actual contexts of use made some modifications in this taxonomy quite necessary. These modifications involved the addition of some sub-categories to the first and third major functional groups of bundles, including research-oriented and participant-oriented bundles. It must be noted that both authors analyzed the functions of bundles in their actual contexts of use and in most cases they came up with the same decisions. In cases they had different views regarding the functions of bundles, they discussed things together initial reaching 100 % agreement.

The three major functions of bundles corresponded 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 the 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 (Hyland, 2008a, 2008b; Thompson, 2001). In the case of research-oriented bundles, three new sub-categories were added: study-focusing bundles (e.g., *the purpose of this*), goal-oriented bundles (e.g. *the goal of this*), and discipline-bound bundles (e.g., *English as a foreign*). With regard to text-oriented bundles, only the last sub-category of rephrasing bundles (e.g. *in other words*) was added. Finally, in the case of participant-oriented bundles, while engagement features were taken as one of the sub-categories as that of original taxonomy, different stance features were divided in to four different sub-categories: attitude markers, epistemic-certain, epistemic-uncertain, and intention. Each of these sub-categories will be described more in the functional analysis of bundles.

4. Results & Discussion

4.1 Lexical bundles in research articles in applied linguistics

The results showed that in the corpus of research articles in applied linguistics, which was representative of published writing in this disciplinary area, there were 121 lexical bundles with *on the other hand*, *in the case of*, *at the same time*, *the extent to which*, and *the end of the* being some of the

most frequent. Table 4 shows the variety and overall frequency of bundles in research articles.

Table 4. Variety and overall use of bundles in research articles

Genres	Research articles
Number of bundles	121
Actual frequency	5697

Structurally, as can be seen in Table 5, the following structural patterns were found regarding the use of lexical bundles in applied linguistics published writing:

1. More than 80% of all bundles in research articles were phrasal rather than clausal (either noun phrases or prepositional phrases), thereby strongly lending support to the findings of some previous studies, such as Biber et al. (1999) and Cortes (2002, 2004) that academic writing, unlike some other registers such as conversation and classroom teaching, depends for the most part on phrasal rather than clausal bundles. This means that published academics are drawing on phrasal bundles in their development and organization of discourse.

Table 5. Overall Structural Description of Lexical Bundles in Research
Articles

Structures	No of bundles	Percentage (%)
Noun phrase+ of	32	23.45
Other noun phrases	12	10.05
Prepositional phrase+ of	34	30.47
Other prepositional phrases	19	19
Passive+ prepositional phrase fragment	4	2.40
Anticipatory it+ verb\adjective	8	5.60
Be +noun\adjectival phrase	2	1.22
Others	10	7.81
Total	121	100

2. Around half of all bundles in research articles were prepositional phrases (prepositional phrases with *of* as well as other prepositional phrases). This means that this structural class of bundles is the most frequent among all categories. As will be explained later, this group of bundles can serve a wide variety of functions in the development of the evolving discourse.

3. Noun phrases with *of* and other noun phrases comprised almost one-third of all bundles.

4. Only less than ten percent of all bundles in research articles were clausal.

5. In terms of the range and variety of a particular structural group of bundles, noun phrases and prepositional phrases with *of* were the most widely-employed (32 and 34, respectively).

6. Among clausal bundles in research articles, those beginning with anticipatory *it* were preferred more than the other two groups (passive+ prepositional fragments, be+ noun\adjectival phrase). This can be attributed to different discursive functions that such *it* clauses serve in the academic discourse, as discussed below.

4.2 Functional description of lexical bundles in research articles

Table 6 shows the overall use of lexical bundles in research articles in terms of the three major functional categories used in this study based on the functional taxonomy of lexical bundles developed by Hyland (2008a, 2008b). Table 7 also represents the functional classification of all target bundles identified in published writing. These functions are discussed more details below.

Table 6. Functional description of lexical bundles in research articles in applied linguistics

Categories	Number of bundles	Percentage %
Research-oriented	58	45.17
Text-oriented	44	41.95
Participant-oriented	19	12.88
Total	121	100

4.2.1 Research-oriented bundles in applied linguistics published writing

Overall, the results showed that a large number of bundles in research articles were mostly employed to encode experiences, activities, and events in the real world (research-oriented bundles). These were clusters that focused more on the external relations in the world describing time and place relations (examples 1 and 2), size and magnitude (examples 3 and 4), the study itself (example 5) and research procedures (example 6):

(1) At the time of the research, a third of the senior secondary cohort were international students, the overwhelming number being from China.

(2) In this paper I will focus on "pragmatic" resources, but *within the context of* the interpretations and customs of those involved.

(3) The development of Hull's SAC involved *a wide range of* stakeholders: academic staff, Head of Department, Dean, technical and secretarial staff and the open learning adviser to ensure it reflected *a wide range of* pedagogic as well as technical, architectural and administrative needs.

(4) With regard to these various activities, however, the most note-worthy point was that all eight regarded watching English movies and television series as *one of the most* effective ways of improving their English.

(5) Second, one may assume that multiple choice and open ended tasks were selected because the students *in the current study* would be familiar with them and this may have reduced anxiety that could have been introduced by the inclusion of unfamiliar task types in a test.

(6) Unlike *the use of the* past simple which may not show any connection between a past action and the present, *the use of the* present perfect shows that an immediate past action has a result or an effect in the present.

There were some research-oriented bundles which were more descriptive in nature, representing some features, abstract or concrete, in the physical or mental world. The more general ones were labeled under the category of 'description' (examples 7 and 8) as used in Hyland's functional taxonomy (2008a, 2008b), and more specific ones were put in the last category, discipline-bound bundles, which seemed to be unique to the field of applied linguistics (examples 9 and 10), describing those word sequences that are almost exclusively used by people in the field:

(7) If the existing rules do not allow the listener to decompose the utterance, he guesses *the meaning of the* utterance from the given environmental cues, and incorporates this mapping into his rule repertoire.

(8) *The analysis of the* teachers' assertions on the in-class activities that they assign to the pupils might reveal whether these activities are affected by their views about the concept of communicative competence.

Table 7. Functional classification of target bundles

Major functions	Sub-categories	Bundles
Research-oriented bundles	location (time\place)	at the same time, the end of the, in the context of, at the end of, at the beginning of the, the beginning of the, at the time of, in the course of, at the university of, the context of the, the time of the, within the context of
	study-focusing	in the present study, of the present study, in the current study, participants in this study
	Quantification	a wide range of, one of the most, a small number of, as part of the, per cent of the, is one of the, a great deal of, the majority of the, the total number of, for each of the, in each of the, the rest of the, in a number of, as one of the, a large number of, the number of words
	Procedure	the use of the, in the use of, the role of the, through the use of, the use of a , that the use of, on the use of, of the use of, to the use of
	Description	the meaning of the, the analysis of the, the language of the, in the experimental group, the form of a, the development of the, an understanding of the, the source of the
	goal-oriented	the purpose of the, for the purposes of
	discipline-bound	as a second language, English as a second, in the target language, as a foreign language, native speakers of English, English as a foreign, Both L1 and L2
Text-oriented bundles	transition signals	on the other hand, as well as the, on the one hand, in addition to the, as well as to
	resultative signals	on the basis of, as a result of, the results of the, the results of this, the basis of the, a result of the
	structuring signals	can be seen in, as can be seen, as shown in table, are shown in table
	framing signals	in the case of, the extent to which, in terms of the, the ways in which, on the part of, in the form of, the nature of the, in the process of, in the field of, the part of the, the way in which,

Table 7. (cont.) Functional classification of target bundles

Text-oriented bundles	framing signals	in terms of their, in relations to the, the students in the, with respect to the, at the level of, with regard to the, of English as a, to the development of , of the target language, in the construction of, the relationship between the, in the light of, from the perspective of, the degree to which, the ways in which the, the case of the , in the area of
	rephrasing signals	in the sense that
Participant-oriented bundles	attitude markers	it is important to, it is difficult to, it is necessary to, it is clear that, it is interesting to
	epistemic-certain	the fact that the, to the fact that, that there is a, by the fact that
	epistemic-uncertain	it is possible that, it is possible to, can be used to
	Intention	to be able to
	engagement	it should be noted, is important to note, should be noted that, can be seen as, be seen as a, interesting to note that

(9) At the same time, a sizeable PBL component was added to an already established *English as a Second* Language (ESL) Course for Medicine to help first-year students develop PBL discussion skills.

(10) Ellis (1999) argues that it is in this stage that systematic variation begins to set in (while learners use the forms categorically) before they proceed to a possibly eventual stage called 'completion' where different forms are employed systematically in accordance with rules *in the target language*.

4.2.2 Text-oriented bundles in applied linguistics published writing

A relatively large number of bundles in research articles had a more discursive function of marking the relationship between prior and coming discourse (text-oriented bundles), lending support to Hyland's position (2008a, 2008b) that lexical bundles in research articles serve a textual function to a large extent. This also shows the characteristic interpretative nature of published academic writing, where claims need to be mitigated. This also pinpoints the evaluative patterns of argument in applied linguistics

as a soft discipline, where persuasion is more explicitly interpretative and less empiricist (Hyland, 2004). So while claims are often based on the observations of real-world phenomena, knowledge is made by plausible reasoning rather than the results speaking themselves. Text-oriented bundles are heavily used to provide familiar and shorthand ways of engaging with the literature, connecting ideas, directing readers around the text, and specifying limitations.

These bundles were used diversely to act as transition markers (examples 11 and 12), to show causative relations (example 13), to mark text stages (example 14 and 15), to interpret, limit, and specify the textual conditions (examples 16 and 17), and to rephrase the preceding arguments (Hyland, 2007) (example 18):

(11) Pre-task planning directs learners' attention to the conveyance of message that is reflected in greater fluency and lexical variety. Online planning, *on the other hand*, encourages learners' attention to grammatical accuracy but results in reduced fluency.

(12) In an outsourced context where the CSR is a non-native English speaker (NNES) the cultural *as well as the* linguistic demands of the interaction must be significant.

(13) What seems strikingly different here is both speakers' persistence to clarify and double check ambiguities and potential misunderstandings that may arise *as a result of* their limited competence in English as well as the nature of the task.

(14) As *can be seen in* Table 4, with the highest minimum score of 48.00 on the FLCAS, the band 1 (the least proficient) students had the lowest maximum score of 157.00 as well.

(15) *As shown in Table* 5, both the L1 and L2 writers used about five individual paraphrases per summary, and, for both groups, about 45% of an average summary was made up of paraphrases.

(16) In contrast, when an item is presented after some time has elapsed and after some intervening items have been shown, full processing will be necessary, as the previous presentation will not be as readily available as *in the case of* massed presentations.

(17) The current study attempts to develop preliminary understandings of _the extent to which_ subject teachers perceive themselves to be responsible for supporting Chinese international students as they meet the challenges of the exam-oriented senior secondary school in Australia.

(18) It is inaccurate _in the sense that_ it ignores the distributional behavior of 59.5% of the most frequently used adjectives, which can appear in both positions.

4.2.3 Participant-oriented bundles in applied linguistics published writing

Table 7 also shows that participant-oriented bundles, which are used to express different stance meanings and encode engagement features (Hyland, 2008a, 2008b), were the least used in comparison to the previous two categories. Generally, it seems that such bundles are not very pervasive in published academic writing. As said before, participant-oriented bundles play a more interpersonal role by reflecting different kinds of epistemic, attitudinal, and interactional meanings. These meanings have also already been studied under such varied labels as "metadiscourse elements" (Hyland, 1999; 2000; 2001a; 2001b; 2004; 2005) and "stance expressions" (Biber et al., 1999; Biber, 2006a, 2006b). In this study, while retaining engagement features as one of the main 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 (Biber et al., 2004; Biber, 2006a; Biber & Barbieri, 2007; Cortes, 2002, 2004).These examples can showcase the use of some of such bundles by research article writers:

(19) Finally, _it is important to_ acknowledge that the correlations reported here are not strong enough to reach any definite conclusions. However, they have provided interesting insights and directions for the in-depth interviews which followed.

 (20) By saying that a phenomenon is systematic, on the other hand, _it is necessary to_ observe at least a certain number of cases—ideally, the greater the number of observations, the more reliable the conclusion of systematicity.

(21) *It should be noted* that Pre-textual revisions may also occur when the writer is deliberating changes to be made to part of the text that has already been written. Pre-textual changes may affect only the language or may also affect the conceptual content of the formulation.

(22) It *is important to note* that the use of this instrument results in a 'non-fail' result: the candidate's level of proficiency is assessed according to the instrument; there is no question of a candidate having 'failed' to perform at an officially pre-set level as in many other English language examinations.

(23) It was *interesting to note that* students in advanced levels used social strategies more than any other levels. With increased proficiency came increased confidence, allowing the learners to interact with others by practicing their language knowledge to promote communicative skills.

4.3 The use of target lexical bundles in postgraduate genres of applied linguistics

Table 8 shows the variety and actual overall frequency of target bundles in master theses and doctoral dissertations in applied linguistics. As can be seen, almost all lexical bundles identified in the corpus of research articles were also used in both postgraduate genres (the only exception was the rather special bundle *'the number of word*s', which was not found in the corpus of doctoral dissertations).Notwithstanding this, the overall use of target bundles in the two postgraduate bundles seemed to be quite phenomenal, a surprising finding which ran counter to findings of some previous research (e.g., Cortes, 2002, 2004; Hyland, 2008a, 2008b) that had attested the relatively infrequent use of target bundles by students who could be regarded as novice members of disciplinary communities. Some target bundles used quite frequently by students were *on the other hand*, *the results of the*, *in the use of*, *of the present study*, and *the results of this*. Interestingly, also, while some previous studies (e.g., Hyland, 2008b) found that doctoral students approximated research article writers more in their use of target bundles than students at the master's level, the results did not show any significant difference between these two groups of students as far as their overall use of target bundles was concerned. More importantly, the functions for which bundles were employed by postgraduate students mostly

corresponded to those of published authors in applied linguistics, as further discussed below.

Table 8. Variety and overall use of target bundles in postgraduate genres

Postgraduate genres	Doctoral dissertations	Master theses
Number of target bundles	120	121
Overall frequency	2313	2405

However, the study found a good number of target bundles on which postgraduate students did not draw very often (e.g. *the extent to which, the end of the, in the context of, the use of the, at the end of, it is important to, the ways in which, a wide range of, at the beginning of, can be seen in, as can be seen, at the time of, the part of the, the way in which, a small number of, as part of the, at the university of, as shown in table, it is possible that, the majority of the, in the current study, it is clear that, in the construction of, the language of the, is important to note, the time of the, can be seen as, ways in which the, within the context of, the case of the, as well as to, by the fact that, it is interesting to*). As can be seen, these bundles belonged to different functional categories of bundles (see Table 7) and therefore, it may be difficult to speculate on why they were used infrequently by students.

On the contrary, there were some bundles that seemed to be used more by one or both groups of postgraduate students, as compared with research articles writers. This can be of interest given that students' corpora were much smaller than the corpus of published writing. As examples, we can mention *as a foreign language, native speakers of English, English as a foreign, that there is a, the analysis of the, the results of this, the results of the, in the use of, with respect to the, of the present study*). Actually, one of this bundles (i.e. *the results of the*) was the top most frequent bundle in the corpus of master theses. The heavy use of this bundle could be traced to the heavy adherence of postgraduate students to reporting the original results of their studies.

At the same time, there were also some differences between the two postgraduate genres in the extent to which they drew on some functional categories of target bundles. For example, while doctoral students used some clusters like *as a foreign language, native speakers of English, in terms of*

the, in relation to the, in the sense that, from the perspective of, it is possible to, as well as the, on the part of, on the one hand, and *in each of the* much more than students at the master's level, the latter embarked on some bundles like *in the present study, in the process of, as can be seen, through the use of, in the experimental group, the degree to which, in the area of,* and *to the use of* more than the former.

It seemed that novice postgraduate writers do not draw considerably on those bundles marking time and place (e.g., *at the time of*, *the context of the*, *the beginning of the*, *at the beginning of*), size (e.g., *a small number of*, *the majority of the*, *in a number of*, *for each of the*). It might also be argued that such bundles are more difficult to use for students. Also, those word sequences associated with more overt expressions of stance (e.g., *it is possible that*, *interesting to note that*, *by the fact that*) are not drawn upon very frequently, probably due to the strong association of these bundles with students' incipient disciplinary identity and their lack of enough confidence.

Overall, there were very few differences between published academics and postgraduate students in the way they used target bundles. In fact, both groups of postgraduate students were able to use target bundles for a wide variety of discursive functions as evidenced by the investigation of bundles in their actual contexts of use.

As mentioned in the introduction, only Hyland (2008a, 2008b) has addressed the use of lexical bundles in applied linguistics among other disciplines. As far as applied linguistics, the discipline explored in this study, is concerned, these two studies (i.e., Hyland, 2008a, 2008b) used a small corpus of research articles (no more than 30 texts). In this investigation, a much wider range of research articles (201) was used for the identification of bundles. So, while Hyland (2008b) just finds 71 different bundles in his corpus of research articles, in this study, 121 bundles were found in one single discipline of applied linguistics. It can be argued that corpus size has a considerable effect on the range of bundles identified (Cortes, 2008). Furthermore, it seems plausible to argue that variations in the use of bundles within genres of a single discipline are generally much more than those in the same genre of different disciplines.

Structural classification and comparison of bundles in the three corpora also corroborated quite firmly the findings of previous studies (e.g. Biber et al., 1999; Biber et al., 2004) that unlike conversation, in academic prose, there is the tendency to use phrasal rather than clausal bundles. Different noun phrases and prepositional phrases comprised almost 80% of all bundles. This is an important point that needs to be brought to the attention of students more markedly in the academy.

At the functional level, research articles were found to embark on research-oriented bundles more heavily as compared to the other two major functions of these word combinations. The results, however, were slightly different from those of Hyland's study (2008b) in that they pinpointed that even research article writers were to a great extent dependent on research-oriented bundles. It also seems that postgraduate students, both at the master's and doctoral levels, are generally adept at using lexical bundles for a wide variety of discursive functions just as research article writers can do this with expertise.

Probably more important and noteworthy was postgraduate students' relatively good use of target bundles in their respective genres. Analysis of students' corpora revealed that many target bundles identified in published writing in applied linguistics were also used quite well by both groups of postgraduate students. This could be taken as a surprising finding since some previous research (e.g. Cortes, 2002, 2004, 2006, Jones & Haywood, 2004) shows that target bundles are never or rarely used by developing writers who could be writers with varying levels of language proficiency or differing degrees of disciplinary writing expertise. Some possible explanations for this observation have been discussed below.

Postgraduate students' relatively frequent use of target bundles could be due to the fact that they have already been exposed to such word-sequences repeatedly in their prior readings of applied linguistics published literature. It goes without saying that postgraduate students have recurrently observed different lexical bundles in different research articles they have studied for doing and writing their own research. This can also be further substantiated and understood when one sees that in all theses or dissertations, students devote one single and usually long chapter to the previous research.

Furthermore, given that lexical bundles are very pervasive in different registers, especially university language (Biber et al., 1999; Biber & Barbieri, 2007) and they are likely to have a formulaic status (Wray, 2000, Wary & Perkins, 2000), the acquisition of such word combinations may not confront students with a very difficult task and put a strong burden on them, especially at this level, given their relatively high level of language proficiency and disciplinary expertise.

Also, probably, lexical bundles are retrieved and stored whole from memory through holistic rather than analytical processes, as already discussed in the literature (Conklin & Schmitt, 2008; Gibbs, Bogadanovich, Sykes & Barr, 1997), and therefore, postgraduate students may not have serious difficulty not only in understanding but also in producing lexical bundles. There may be a processing advantage in the use of lexical bundles as some formulaic sequences have been shown to be processed more readily (Conklin & Schmitt, 2008). It can also be argued that lexical bundles can act as handy short-cuts or frames (Biber & Barbieri, 2007) through which writers can scaffold their propositional meanings with a relative ease. However, automatic acquisition of lexical bundles should always not be taken for granted as this study also found out that there were some target lexical bundles on which students did not draw quite often (for example, those encoding time, place, size, and overt expressions of stance). These word sequences are not idiomatic in meaning and hence they may be easy to understand, but they do not seem to be marked and perceptually salient.

5. Conclusion & Pedagogical Implications

This study aimed at finding the way in which target bundles in the discipline of applied linguistics, as identified in research articles, were used by two groups of EFL postgraduate students as novice discourse community members in the same discipline. Surprisingly enough, the study, contrary to some findings of previous research, found that in many cases, postgraduate students were able to use target bundles as published writers did. The study, therefore, revealed little if any difference between the three groups of writers in their actual use of lexical bundles. Notwithstanding this, there

were some discrepancies between the three groups with regard to some structural and functional classes of bundles.

In spite of much corpus-based research on lexical bundles in the last decade or so, there are lots of things not yet explored about this group of word combinations. Such explorations can contribute to genre analysis (Hyland, 2008b). Identifying lexical bundles in other disciplines, registers, and genres, especially through the use of large corpora, examining the formulaic and psycholinguistic status of these multi-word sequences, and further probing the effect of a pedagogical treatment on their acquisition could be areas worth exploring in future research.

This study calls for an increased pedagogical focus on lexical bundles, especially those that students need to understand and use in their future target genres. Therefore, in any L2 syllabus or EAP (English for academic purposes) course, target lexical bundles should be considered (Hyland, 2008b). Developing instructional packages, especially "corpus-enhanced disciplinary writing courses" (Cortes, 2006), through which lexical bundles, their distributions, as well as their functions would be introduced to students could be one of the main frontiers in EFL\ESL writing courses. It is important especially for EAP course designers to be well aware of the pervasive presence of lexical bundles in academic writing and expose students to those clusters that they will likely need to use in their target genres. The use of noticing (Cortes, 2004, 2006) conscious raising tasks, 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, especially within a framework of use. As lexical bundles are very frequent in published academic prose, it is necessary to encourage students to use these expressions. Instructors could introduce some of these lexical bundles (for example, those that have text organizing functions) together with some of those discourse markers or connectors, going beyond single words or two-word combinations to longer phrases, phrases fragments, or clause fragments.

However, while there seem to be potentially enormous benefits in identifying the most frequent forms for teaching, we need to be cautious in making assumptions about the generality of academic bundles. Hoey (2005)

point outs, for instance, that because we all have different textual experiences, we all have a different mental concordance to draw on so that particular patterns are cumulatively loaded with the contexts we participate in. So, just as individual lexical items occur and behave in different ways across disciplines (Hyland & Tse, 2007), we need to be sure we are assisting learners in an appropriate disciplinary set of bundles.

References

Anthony, L. (2007). AntConc 3.2.1w: Freeware corpus analysis toolkit. [On-line]. Available: *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*, 97-106.

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

Biber, D, & Conrad, S., (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard, & S. Oksefjell, (Eds.), *Out of corpora: Studies in honor of Stig Johansson*, (pp.181–189). Amsterdam: Rodopi.

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

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.

Conklin, K, and Schmitt, N. (2008). Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Applied linguistics,* *29*(1*), 72-89.*

Conrad, S. (1996). Academic discourse in two disciplines: Professional writing and student development in biology and history. *Unpublished*

doctoral dissertation, Northern Arizona University. Northern Arizona University, Arizona.

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*, 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.

Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. Corpora, *3*, 43-58.

Gibbs, R, Bogadanovich, J, Sykes. J, & Barr, D. (1997). Metaphor in idiom comprehension. *Journal of Memory and Language*, *37*,141–54. Halliday, M.A.K. (1994). *Functions of language* (2nd ed.). London: Arnold.

Hewings, M., & Hewings, A. (2002). It is interesting to note that… : A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes*, *21*, 367-383.

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

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

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

Hyland, K. (2001a). Brining in the reader: Addressee features in academic articles. *Written communication*, *18*(4), 549-574.

Hyland (2001b). Humble servants of the discipline? Self-mentions in research articles. *English for Specific Purposes*, *20*, 207-226.

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

Hyland, K. (2005) Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, *7*(2), 173–191.

Hyland, K. (2007). Applying a gloss: Exemplifying and reformulating in academic discourse. *Applied linguistics*, *28*(2), 266-285.

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.

Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative that in abstracts. *English for Specific Purposes, 24*(2), 123–139.

Jones, M, & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences* (pp. 269-292). John Benjamins: Philadelphia.

Levy, S.A. (2003). Lexical bundles in professional and student writing. Unpublished doctoral dissertation, University of the Pacific, California.

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

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

Ruiying, Y, & Allison, D. (2004). Research articles in applied linguistics: structures from a functional perspective. *English for Specific Purposes*, *23*, 264-279.

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.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied linguistics*, *21*(4), 463-489.

Wray, A., and Perkins, R. M. (2000). The functions of formulaic language: An integrated model. *Language and Communication*, *20*, 1-28.

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