A Comparative Qualitative Study of Lexical Bundles in Three Academic Genres of Applied
Linguistics by 1.Hassan Jalali & 2. Somayeh Ghayoomi

A Comparative Qualitative Study of Lexical Bundles in Three Academic Genres of Applied Linguistics

Hassan Jalali
Somayeh Ghayoomi

English Language Center, Isfahan University of Technology, Isfahan, Iran
Department of English, Faculty of Humanities, Najafabad Azad University, Isfahan, Iran

Hassan Jalali is an assistant professor at Isfahan University of Technology in Iran. He got his Ph.D from Isfahan University in 2009. His main areas of interest are corpus linguistics, second language writing, and genre analysis.

Somayeh Ghayoomi is an M.A. student at Najafabad University in Iran. Currently, she is doing her thesis research on a group of lexical bundles in two disciplines of chemistry and applied linguistics.

Corresponding author's e-mail: Hasan_jalali1@yahoo.com

Content
Introduction
2. The study
2.1. Corpora
2.2. Text analysis programs
2.3. Data analysis methods
3. Findings
3.1. Lexical bundles in research articles in applied linguistics
3.2. Functional description of lexical bundles in research articles
3.3. Target lexical bundles in postgraduate genres of applied linguistics
4. Discussion and conclusion
References

Abstract

Lexical bundles, frequent word combinations that commonly occur in different registers, have attracted researchers in corpus linguistics in the last decade. While most previous studies of bundles have been mainly concerned with variations in the use of these word sequences across different registers and a number of disciplines, no major comparison has yet been made between some key written academic genres of one single disciplinary area. This qualitative study explored the extent to which target bundles in the discipline of applied linguistics, as identified in research articles, are used in master theses and doctoral dissertations as two postgraduate genres of the same discipline.

A Comparative Qualitative Study of Lexical Bundles in Three Academic Genres of Applied Linguistics by 1.Hassan Jalali & 2. Somayeh Ghayoomi

Although the relatively small size of students' corpora makes it difficult to make strong claims, the study, surprisingly enough, found that almost all target bundles are also used by postgraduate students relatively frequently and there is little if any difference between the three groups of writers as far as their overall use of bundles is concerned. However, there are some differences between the three groups in the extent to which they draw on some specific bundles. The paper ends with some directions for future research on lexical bundles.

Keywords: Lexical bundles; Corpus linguistics; Research articles; Doctoral dissertations; Master theses; Applied linguistics

1. Introduction

Lexical bundles, as a particular and relatively recent category of word combinations with a possibly formulaic status (see Biber and Barbieri, 2007), were coined and defined by Biber, Johansson, Leech, Conrad, and Finegan (1999). They defined lexical bundles as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status" (p. 990). More importantly, they refer to frequency as the most salient and defining characteristic of bundles; in order for a word combination (e.g. *on the other hand*.) 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 in form and non-idiomatic meaning are among other properties of bundles. 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; Biber et al, 2004; Hyland, 2008a, 2008b).

   There have been a number of studies that have developed functional classification of these word combinations (e.g., Cortes 2001, 2002, 2004; Biber, Conrad, and Cortes, 2003; Biber et al, 2004; Biber and Barbieri, 2007; Hyland, 2008a, 2008b). Such studies have shown that 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.

A Comparative Qualitative Study of Lexical Bundles in Three Academic Genres of Applied Linguistics by 1.Hassan Jalali & 2. Somayeh Ghayoomi

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. 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. More extensively, Biber and Barbieri (2007) investigated the use of bundles in a wider range of university registers. Among the results such studies, one can report the differential pervasiveness of bundles in different university registers and the heavier reliance of written non-academic registers (e.g. course syllabi) on bundles.

It seems that few studies have focused on the study of bundles within one single disciplinary area (see Hyland, 2008b) especially with an aim to describe and explain possible differences and/or similarities between published experts and developing postgraduate students in the use of these word combinations in their respective high-stakes genres: research articles, master theses, and doctoral dissertations. To address possible intradisciplinary variations in the use of bundles, this more qualitative corpus-based 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.

The paper proceeds with the description of corpora, text analysis programs, and data analysis methods used in the study. In the findings 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. The paper concludes with some suggestions for future research on lexical bundles.

2. The study

This study, therefore, seeks to address the following questions:

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

2. How can lexical bundles identified in research articles be described structurally and functionally?

A Comparative Qualitative Study of Lexical Bundles in Three Academic Genres of Applied Linguistics by 1.Hassan Jalali & 2. Somayeh Ghayoomi

3. To what extent target lexical bundles, as identified in research articles, are used in postgraduate genres of applied linguistics?

2.1. Corpora

Three corpora were used in this study. The first two were doctoral dissertations and master theses in applied linguistics that all had been written by Iranian (L1 Persian) postgraduate students of applied linguistics at one of the Iranian state universities.

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 form 2006-2007 time period.

2.2. Text analysis programs

Two computer programs were used in this study to explore lexical bundles, their frequencies, the number of texts in which they had been used, and their actual contexts of use: Antconc3.2.1w (Anthony, 2007), and Wordsmith tools5 (Scott, 2008). The former was used for 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. In this study like some other previous studies of lexical bundles (e.g. Cortes, 2002), only four-word combinations or bundles were investigated.

2.3. Data analysis methods

We employed a more qualitative approach, similar to that used by Cortes (2004). According to this approach, first the corpus of research articles was explored to identify candidate bundles in published writing in applied linguistics. 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 searched to see the extent to which they are used by postgraduate students.

All lexical bundles identified in research articles were classified structurally using the widely-used structural taxonomy of bundles developed by Biber et al (1999). As for functions, Hyland's functional taxonomy of bundles in academic writing (2008a, 2008b) was used as an initial framework for classification of bundles.

3. Findings

A Comparative Qualitative Study of Lexical Bundles in Three Academic Genres of Applied Linguistics by 1.Hassan Jalali & 2. Somayeh Ghayoomi

3.1. Lexical bundles in research articles in applied linguistics

Overall, the results of this study showed that in the corpus of research articles in applied linguistics, there were 121 lexical bundles with *on the other hand* as one of the most frequent bundles. It is important to reflect on this finding in the light of the relatively large number of texts used in this corpus. Table 1 shows the variety and overall frequency of bundles in research articles.

Table 1
 Variety and Overall Use of Bundles in Research Articles

Genres	Research articles
Number of bundles	121
Actual frequency	5697

Structurally, as can be seen from table 2, it seems that more than eighty percent of all bundles in research articles are phrasal rather than clausal lending support to findings of some previous studies like Biber et al (1999) 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. Interestingly, just less than ten percent of all bundles in research articles are clausal. Among clausal bundles in research articles, those beginning with anticipatory *it* are preferred more than the other two groups (passive+ prepositional fragments, be+ noun\adjectival phrase).

Table 2

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

A Comparative Qualitative Study of Lexical Bundles in Three Academic Genres of Applied Linguistics by 1.Hassan Jalali & 2. Somayeh Ghayoomi

Be +noun\adjectival phrase	2	1.22
Others	10	7.81
Total	121	100

3.2. Functional description of lexical bundles in research articles

Table 3 shows the varieties and overall frequencies 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 4 also shows the functional classification of all target bundles identified in published writing.

Table 3

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

The results obtained show that around half of all bundles in research articles are mostly employed to encode experiences, activities, and events in the real world (research-oriented bundles). These are clusters that focus more on the external relations in the world describing especially in the case of academic texts time and place relations, size and magnitude, the study itself and research procedures.

On the other hand, more than forty percent of bundles in research articles serve 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 as well. Table 3 also shows that participant-oriented bundles, which are used to express different stance

A Comparative Qualitative Study of Lexical Bundles in Three Academic Genres of Applied Linguistics by 1.Hassan Jalali & 2. Somayeh Ghayoomi

meanings and encode engagement features (Hyland, 2008a, 2008b), are the least used in comparison to the two previous categories(Biber, 2006a).

3.3. Target lexical bundles in postgraduate genres of applied linguistics

As can be seen, almost all lexical bundles identified in the corpus of research articles are also used in both of the two postgraduate genres. However, on the face of it, the overall use of target bundles in the two postgraduate bundles seems to be quite phenomenal, a surprising finding which runs 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 novice members of disciplinary communities.

However, there are still a good number of target bundles on which postgraduate students do not draw quite frequently (e.g. *the extent to which*, *the end of the*, *in the*

Table 4
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

A Comparative Qualitative Study of Lexical Bundles in Three Academic Genres of Applied Linguistics by 1.Hassan Jalali & 2. Somayeh Ghayoomi

		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 4 (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
	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
Participant-oriented bundles	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,

A Comparative Qualitative Study of Lexical Bundles in Three Academic Genres of Applied Linguistics by 1.Hassan Jalali & 2. Somayeh Ghayoomi

		interesting to note that

Table 5

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

context of, *the use of the, at the end of,* and *it is important to*).

On the contrary, there are some fewer bundles that seem to be used more by both or one group of postgraduate students than research articles writers given that students' corpora were much smaller than the corpus of published writing (e.g., *as a foreign language* and *English as a foreign*). Actually, one of these bundles (i.e. *the results of the*) was the top most frequent one in the corpus of master theses and its rate of use was as much as that of the most frequent bundle in the corpus of research articles (*on the other hand*). At the same time, there are also some differences between the two postgraduate genres in the extent to which they draw on some target bundles. For example, while doctoral students use some clusters like *as a foreign language* much more than students at the master's level, the latter draw on some bundles like *in the present study, in the process of,* and *to the use of* more than the former.

4. Discussion and conclusion

 Students' relatively frequent use of target bundles could be due to the fact that they have already been exposed to such word sequences several times in their prior readings of applied linguistics published literature. There is no doubt that postgraduate students have repeatedly observed different lexical bundles in different research articles they may have studied for doing and writing their own research; furthermore, given that lexical bundles are very pervasive in different registers especially university language (Biber at al, 1999; Biber and Barbieri, 2007) and may have a formulaic status (Wray, 2000, Wary and Perkins, 2000), the acquisition of such word combinations may not confront students with a very difficult task especially at this level given their relatively high level of language proficiency and disciplinary expertise. Probably also, lexical

332

MJAL 2:4 June 2010   **ISSN 0974-8741**
A Comparative Qualitative Study of Lexical Bundles in Three Academic Genres of Applied Linguistics by 1.Hassan Jalali & 2. Somayeh Ghayoomi

bundles are retrieved and stored whole from memory through holistic rather than analytical processes (Conklin and Schmitt, 2008), and therefore, postgraduate students may have little if any 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 easier to use (Conklin and Schmitt, 2008). However, automatic acquisition of lexical bundles should not be taken for granted as this study also showed that there are some target lexical bundles on which students do not draw quite frequently (for examples those encoding time, place, size, and overt expressions of stance). These word sequences are not idiomatic in meaning and therefore, may be easy to understand, but they do not seem to be marked and perceptually salient. Consequently, there may still be a need to leave a particular place in any L2 syllabus or EAP (English for academic purposes) course for an increased pedagogical focus on lexical bundles especially those that students need to understand and use in their future target genres (Hyland, 2008b).

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

Biber, D. and 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, & 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*

A Comparative Qualitative Study of Lexical Bundles in Three Academic Genres of Applied Linguistics by 1.Hassan Jalali & 2. Somayeh Ghayoomi

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.

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.

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.

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

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

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