

Examining Novices' Selection of Lexical Bundles: The Case of EFL Postgraduate Students in Applied Linguistics

Hassan Jalali

Assistant professor, Isfahan University of Technology, Iran

Abstract

Previous research shows that part of mastering a given genre is correct, adequate, and appropriate use of a set of word combinations called chunks, clusters, and/or lexical bundles as these word combinations serve important discourse functions and are for the most part peculiar to and characteristic of a disciplinary field. While some few previous studies have demonstrated that even native speaker educated students may find it difficult to use these word sequences properly and sufficiently in their written academic production as compared with experts' choices, almost no work has been done to examine the extent to which EFL students are able to discriminate between different bundles and choose them appropriately. Using a reading measure of selecting word sequences which was administered to a group of EFL postgraduate students in applied linguistics, this study showed that in many cases, students' choices of these word combinations did not match those of experts. This finding suggests that such students who might still be novice members of their field were likely to encounter serious difficulties later in adhering to preferred accepted writing practices of the discipline especially when they would want to get their voices heard in a high-stake genre like research article. The paper closes with some pedagogical implications as well as suggestions for further research.

Key words: applied linguistics, lexical bundles, postgraduate students, selection reading task

INTRODUCTION

The past four decades have witnessed a growing interest in the study of formulaic expressions in both theoretical and practical terms in different languages (Conklin and Schmitt, 2008) as well as a more robust pedagogical focus on them especially because the research has shown that the acquisition and use of these expressions among both natives and non-natives could be a difficult learning task (Cortes, 2011; 2006). Although it is yet far from clear what counts as formulaic and what does not (Wray & Perkins, 2000), different categories of word combinations like idioms, proverbs, collocations, clichés, and sentence frames (see Wray & Perkins, 2000; Wray, 2000; & Cortes, 2002 for an extensive list of formulaic expressions) are all put under the umbrella term of

formulaic language to contrast it with the more productive and analytic language (Conklin & Schmitt, 2008).

Introduced relatively recently, lexical bundles were coined by Biber, Johansson, Leech, Conrad, and Finegan (1999) for the first time to refer to a group of word combinations (made of three, four, five, or six words) which are identified empirically and just on the basis of their frequency. These word clusters have for the most part a non-idiomatic meaning, and are structurally incomplete and act as building blocks of coherent discourse in different registers. However, it is not yet evident whether these word combinations really have a formulaic nature (Biber & Barbieri, 2007; Cortes, 2006), and there is no one single all agreed-upon cut-off frequency for the identification of a word cluster as a bundle. In most cases, a word combination like *at the same time* has to occur at least twenty times in a corpus of one million words and five different texts to count as a bundle (Hyland, 2008a, 2008b). These word combinations are used in different registers such as fiction, conversation, journalism, and academic writing. As examples of typically occurring bundles in academic writing, one can name *on the other hand*, *in the case of*, *as well as the*, *it should be noted*, to name only a few (see Biber et al, 1999, & Biber, 2006 for an extensive list of bundles).

Corpus-based studies of lexical bundles across different disciplines, registers, and genres have been the main focus of mainstream research on these groups of word combinations for more than a decade. It has been shown that lexical bundles are also for the most part discipline-specific, i.e., each discipline draws on a unique set of bundles in the development of its characteristic discourse (Cortes, 2004; Hyland, 2008a, 2008b). The more considerable presence of lexical bundles in certain registers like classroom teaching and management registers (Biber et al, 2004; Biber, 2006; Biber & Barbieri, 2007), and their multi-functionality (Biber et al, 1999; Biber & Conrad, 1999; Biber & Gray, 2013; Cortes, 2004; Biber et al, 2004; Hyland, 2008a, 2008b) have also been shown in the prior literature. Along with their high frequency (Biber et al, 1999; Wray & Perkins, 2000; Wray, 2000), such recurrent word combinations have been noted for the important functions that they can serve in spoken as well as written discourse such as acting as processing shortcuts (Wray & Perkins, 2000; Conklin & Schmitt, 2007), expressing identity with a group (Wray, 2000), and reflecting maturity and expertise in writing (Cortes, 2004, 2006).

The use of these recurrent word combinations has been found to be a good indicator of proficient and competent language use in different registers including academic writing (Cortes, 2006). For example, Cortes (2002, 2004) studied the use of these word clusters in native speaker college students' (undergraduate, graduate lower level, and graduate upper level) essays in two fields of history and biology. It was shown that generally students' use of bundles was quite rare, infrequent, and in many cases, different from those of published academic research in the functions that they served in students' written production. In fact, a wide discrepancy was found between academic writers

and novices (e.g. postgraduate students) in that the latter often failed to use lexical bundles correctly and adequately.

More attention to the use and acquisition of these word clusters can find more plausibility when one finds that simple and even frequent long exposure to lexical bundles may not guarantee students' correct and sufficient use of these clusters, whether they be native or non-native or undergraduate or graduate (Cortes, 2006). This failure has been ascribed to lack of any systematic instructional focus on these word combinations although there have been some major attempts to introduce formulaic expressions into L2 curricula (e.g., Willis, 2003; Nattinger & Decarrico, 1992; Lewis, 1997, 2000). In spite of all these, some studies (e.g. Cortes, 2006) have shown meager progress in students' productive use of these word combinations even after some formal instruction. So, it is yet far from clear to what extent students are able to choose the right lexical bundle in a given text.

THIS STUDY

As the age of research on bundles is not that long, especially the research of a more acquisitional and pedagogical nature is not sufficient, this study aims to examine the extent to which postgraduate EFL students in applied linguistics are able to choose among lexical bundles. Accordingly, the following questions were posed to serve as a guide in this study:

1. To what extent are EFL postgraduate students in applied linguistics as novice members of this disciplinary field able to choose the right lexical bundles in the right context?
2. Are there any differences between different lexical bundles in the degree to which they are chosen correctly by the postgraduate students?

METHOD

Participants

13 second-year college students at master's level in applied linguistics at a state university and an EFL context took part in this study. These students had all entered master's programs through a strict entrance exam-nationwide and had a high level of language proficiency and also relatively good knowledge of discipline. They were studying in the third semester and preparing themselves for writing their thesis research proposals.

Getting the consent of one of the instructors for taking a whole session of one of the courses (around 90 minutes), the researcher himself first explained the purpose of the study to these postgraduate students and assured them that the results would be kept confidential. Although all necessary instructions had already been given in the selection

measure itself, the researcher preferred to reiterate them orally to guard against any misunderstanding at the very beginning before students started to do the task.

Material

A reading measure was used in this study in order to investigate students' ability to choose the appropriate bundle in a text. This reading measure consisted of thirty short texts. Each of these texts was incomplete with a lexical bundle missing. Such tests have been widely used in the language teaching field for the teaching of collocations (Cortes, 2006). Many of these texts were no more than a paragraph. The lexical bundles of interest were taken from a wide variety of research articles from different journals of applied linguistics. These articles were selected from a relatively large corpus of research articles (Author et al., 2008; Author, 2009) comprising around two hundred articles from seven different journals in the diverse field of applied linguistics.

Data Analysis

The selection of the most frequent lexical bundles in applied linguistics was based on the study of Hyland (2008a). Accordingly, fifty word combinations frequently employed and used by published writers in the field were chosen. By using Antconc 3.2.1.w concordancer (Anthony, 2007), articles were identified which contained the thirty most frequent bundles in the top fifty list of word clusters and then the contexts within which these bundles had been used were cut from the articles. Due to the high frequency of each of these word clusters in research articles, some decision had to be made about the articles within which the target bundles of interest in this study had been used. The text fragments used in the reading task were finally taken from 23 research articles. An attempt was made to choose the texts from those journals students often refer to in their courses. By studying and examining the use of every bundle in its context of use, a decision was also made about the amount of textual context that needed to be included in the reading test to incorporate enough contextual clues to help the reader in the selection of bundles. Because of this consideration, all text fragments used in the study were not exactly of the same length, with some being shorter and some even longer than a paragraph.

RESULTS AND DISCUSSION

Overall, the results of the study showed that in many cases, students' selection of the bundle did not match that of published writers. As table1 shows, it was found that in just 208 cases (out of the whole 390), students' selection of the bundles corresponded to the original choices of published writers, indicating that only in 53.33% of choices made by the participants, they were able to choose the right bundle. Almost in half of the cases, participants were either unable to choose any bundle or make the right choice. Furthermore, looking at the number of correct uses for each bundle, one could find that there were differences between bundles of interest and the contexts within which they had been used in terms of the difficulty they posed for the students. For

example, in texts 1, 12, 13, 18, 19, 25, 28, 29, and 30, the number of correct uses (out of the whole possible 13 in this study) was 10 or more than that, while in the case of some texts like 6, 7, 20, 21, and 26, the overall accurate use was three or less than that. The number of correct selections in other texts like 2, 4, 5, etc. varied from 4 to 9.

Table 1. Postgraduate students' correct use of lexical bundles

Texts	Correct bundle	The number of correct answers
Text1	At the end of	10
Text2	Can be seen as	8
Text3	In relation to the	1
Text4	The nature of the	5
Text5	In the case of	5
Text6	In the form of	2
Text7	In the present study	2
Text8	In the sense that	7
Text9	Is one of the	9
Text10	On the other hand	9
Text11	The end of the	5
Text12	One of the most	10
Text13	The fact that the	10
Text14	The relationship between	9
Text15	The ways in which	4
Text16	The role of the	9
Text17	The result of the	6
Text18	To the fact that	13
Text19	On the one hand	11
Text20	At the beginning of	3
Text21	As a result of	3
Text22	As well as the	7
Text23	In the process of	7
Text24	At the same time	4
Text25	Can be found in	10
Text26	In terms of the	1
Text27	In the context of	7
Text28	It is important to	10
Text29	It should be noted	11
Text30	On the basis of	10
Total	30	208

Students were also found to have different degrees of difficulty with different bundles and texts. Some bundles like *to the fact that*, *at the end of*, *one of the most* posed little difficulty while some others like *in relation to the*, *in terms of the*, and *in the form of* were more demanding. Interestingly, there were two bundles (i.e., *in relation to the*, *in terms of the*) that were used correctly only by a single participant. It seems that the selection of such bundles has been quite difficult. This could be partly due to the more abstract and cognitively complex nature of these word combinations. There were also some other bundles that were chosen very infrequently by the participants in the study

(e.g., *in the form of, in the present study, at the beginning of, as a result of*). Participants may have already been exposed to such bundles several times, but they are not perceptually salient (Cortes, 2006), they may have easily gone unnoticed. On the other hand, there were some bundles that seemed to pose little if any difficulty for the participants in their selection as used in the respective texts (e.g., *at the end of, one of the most, to the fact that*).

As said in the introduction, different classifications have been developed for the functional description of lexical bundles in different registers including academic writing (e.g., Biber et al, 1999; Biber et al, 2004). In this study, in order to further explore whether there were differences between lexical bundles in the extent to which they were used by the participants in the study, they were classified functionally using the functional taxonomy proposed by Hyland (2008a, 2008b). According to Hyland (2008a, 2008b), lexical bundles serve three major functions that correspond to Halliday's (1994) tripartite functional typology of language: Research-oriented (ideational), text-oriented (textual), and participant-oriented (interpersonal). Research-oriented bundles serve a more ideational and information-focused function encoding activities, experiences, time, procedures, etc. in the world. Text-oriented bundles connect different parts of texts and announce different stages in the evolving discourse. Finally, participant-oriented bundles play a more interpersonal role expressing writer's different attitudes and feelings toward the subject matter and the readers.

Table 2 shows functional classification of lexical bundles tested in this study based on the taxonomy developed by Hyland (2008a, 2008b). To make comparisons between different functional types easier, the frequencies corresponding to lexical bundles have been given in parentheses.

As can be seen, there were some relatively slight differences between the functional types of lexical bundles in the extent to which they were used correctly by the participants in the study. While some bundles showing quantification (number, quantity, amount, etc.) and the relation between the writer and readers (participant-oriented bundles) were used in a correct way relatively frequently, some word combinations acting as location markers and framing signals seemed to be difficult for participant to choose correctly in the respective texts. However, these differences cannot be emphasized as the number of participants in the study was relatively small and judging on the differences just on the basis of frequencies may not be very reliable. It can be argued that regardless of the function they serve in the text, different lexical bundles can pose different degrees of difficult for the students. What is noteworthy and needs to be explained is that in many cases, participants' selection of lexical bundles did not match that of published writers. While the effect of reading proficiency, task and topic familiarity, and specific background knowledge should be acknowledged, the results of this study showed that generally learning to understand, recognize, discriminate, and use lexical bundles could be challenging but, at the same time, inevitable task even for postgraduate students. The results of the study were in line

with those of Cortes (2006) who found that even native speaker undergraduate students had problems using lexical bundles even after a period of instruction by which they were made aware of the importance of these word clusters. As this study found that students had difficulty in discriminating, recognizing, and selecting different lexical bundles, it is more likely that they could even have more problems in sufficient and accurate use of these word clusters in their written production.

Table 2. Functional Classification of Lexical Bundles Tested in this Study
(Based on Hyland, 2008a, 2008b)

Major functions	Sub-categories	Lexical bundles
Research-oriented: help writers to structure their activities and experiences of the real world includes:	Location – indicating time/place	at the beginning of (3), at the same time (4), in the present study (2), at the end of (10), the end of the(5)
	Procedure	
	Quantification	is one of the (9), one of the most (10)
	Description	the nature of the (5), in the process of (7), in the context of (7) , the result of the (6), the role of the (9)
	Topic – related to the field of research	
Text-oriented – concerned with the organization of the text and its meaning as a message or argument includes:	Transition signals – establishing additive or contrastive links between elements	<i>on the other hand(9), on the one hand(11), as well as the (7)</i>
	Resultative signals – mark inferential or causative relations between elements	<i>as a result of (3), on the basis of (10)</i>
	Structuring signals – text-reflexive markers which organize stretches of discourse or direct reader elsewhere in text	
	Framing signals – situate arguments by specifying limiting conditions	<i>in relation to the(1), in the case of(5), in the form of(2), in the sense that(7), the relationship between the (9), the ways in which (4), in terms of the (1)</i>
Participant-oriented – these are focused on the writer or reader of the text	Stance features – convey the writer’s attitudes and evaluations	<i>the fact that the (10), to the fact that(13), it is important to (10)</i>
	Engagement features – address readers directly	<i>can be found in (10), can be seen as(8), it should be noted(11)</i>

Probably also, lexical bundles are retrieved and stored whole from memory through holistic rather than analytical processes (Conklin & Schmitt, 2008), and therefore, postgraduate students may have difficulty not only in understanding but also in producing lexical bundles in this study and many others used quite pervasively in academic writing. While 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 & Schmitt, 2008), it can also be postulated 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. It seems that postgraduate students, unlike published writers, need more exposure and practice in the use of these building blocks of discourse. Furthermore, automatic acquisition of lexical bundles should not be taken for granted as this study showed that there are lexical bundles in applied linguistics published writing on which students may not draw quite frequently. 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 (Cortes, 2006).

Given that these word clusters are typically used by established published members in any discipline and that these experts commonly use and rely on them for achieving a wide variety of different meanings and functions, there seems to be a need for EAP practitioners, writing instructors, supervisors, as well as students, both undergraduate and graduate, to be more aware of the pervasive and differential presence of these word combinations in different written genres of the academic register. As one of the main pedagogical implications of this study, one can highlight the importance of a more focused, longer, systematic, and explicit pedagogical treatment of bundles. Introducing and including lexical bundles as part of the syllabus of a given writing course, especially for students who are at intermediate or advanced level of language proficiency, could be one of the indispensable things that EAP (English for academic purposes) researchers and practitioners should address in their courses. Future research should also address more rigorously instruction on lexical bundles can help students use these word combination correctly and appropriately.

CONCLUSION

The main purpose of this study was finding the extent to which a group of postgraduate EFL students as novices to their disciplinary field were able to select frequently-used lexical bundles correctly. For this purpose, students were asked to perform on a selection reading task in which they had to complete a set of textual fragments taken from some journal articles by choosing the bundles that had been removed. The results showed that in almost half of the cases, students were unable to choose the right bundle. It may be assumed that these word combinations are so common and straightforward that they can be acquired easily (Biber & Conrad, 1999). The results of this study, however, showed that acquisition and appropriate use of these expressions cannot occur automatically and students may not be able to learn their correct use in the relevant registers at least within a short time. Given that Lexical bundles are very

recurrently used in published academic writing and they are for the most part discipline-bound (Cortes, Jones, & Stoller, 2002), students should be helped to find out that each discipline has different purposes or ways of seeing the world associated with distinct communicative conventions (Cortes, 2006).

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 and 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, which they may usually draw on in their writing, 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, along with some other previous studies like Cortes (2004), and Hyland (2008a, 2008b), was able to show that different academic writers in different disciplines and genres do draw 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, 2000), 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.

REFERENCES

- Anthony, L. (2007). Antconc 3.2.1: A free text analysis software. Available on line at <http://www.antlab.sci.waseda.ac.jp>.
- Biber, D, & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263-286.
- Biber, D, Conrad, S & Cortes, V. (2004). If you look at ...: lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371-405.
- Biber, D, Johansson, S, Leech, G, Conrad S, & Finegan, E. (1999). *Longman grammar of spoken and written English* (pp. 990). Harlow: Pearson.
- Biber, D. & Conrad, S., (1999). Lexical bundles in conversation and academic prose. In Hasselgard, H. & Oksefjell, S., (Eds.), *Out of corpora: Studies in honor of Stig Johansson*, (pp. 181-189). Amsterdam: Rodopi.

- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: Benjamin.
- Biber, D., & Gray, B. (2013). Discourse characteristics of writing and speaking task types on the TOEFL iBT test: A lexico-grammatical analysis (TOEFL iBT Research Report No. 19). Princeton, NJ: Educational Testing Service.
- 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/Main:Peter Lang.
- 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, 31-64.
- 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. (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(4), 391-406.
- Cortes, V. (2011). Genre analysis in the academic writing class: with or without corpora? *Cuadernos de Filologia. Estudios Linguistics*, XVI, 41–64.
- Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for academic purposes*, 12(1), 33-43.
- Cortes, V., Jones, J., & Stoller, F. (2002). *Lexical bundles in ESP reading and writing*. Paper presented at the TESOL Conference, Salt Lake City, UT.
- Halliday, M.A.K. (1994). *Functions of language* (2nd ed.). London: Arnold.
- Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Hyland, K. (2008b). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18 (1), 41-62.
- Lewis, M (2000). 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. (1997). *Implementing the lexical approach*. Hove: Language Teaching Publications.

- Nattinger, J. & DeCarrico, J.(1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Pang, W. (2010). Lexical Bundles and the Construction of an Academic Voice: A Pedagogical Perspective. *Asian EFL Journal*, 47, 1-13.
- Schmidt, R.W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, pp. 129-158.
- Staples,S., Egbert,J., Biber, D & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for academic purposes*, 12 (3), 214-225.
- Willis, D. (2003). Rules, patterns, and words: Grammar and lexis in English language teaching. Cambridge: Cambridge University Press.
- 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), 1-28.