

## A Corpus-driven Study of *it* Lexical Bundles in Applied Linguistics Postgraduate Genres

Hassan Jalali

Assistant professor, English Language Center, Isfahan University of Technology, Isfahan-84156-83111, Iran

### Abstract

Lexical bundles are recurrent word combinations that commonly occur in different registers. They have been the subject of much research as they serve important functions as building blocks of coherent discourse. While much of previous research has been mainly concerned with variations in the use of these word sequences across different registers and a number of disciplines, very few studies have focused on their use within postgraduate genres. This study zoomed on possible generic variations in the use of *it* bundles as a particular group of these word combinations with important interpersonal roles in academic register. More specifically, this study addressed range, frequency and function of these word clusters in EFL postgraduate genres by examining applied linguistics master these and doctoral dissertations. The results obtained indicated that *it* bundles were generally used infrequently in both postgraduate genres. The study also showed that while there were some overlaps between the two genres, doctoral students seemed to rely more on *it* bundles in the development of their texts. Functional analysis of lexical bundles showed that *it* lexical bundles served a wide variety of functions. The findings call for a more increased pedagogical focus on different multi-word sequences like *it* lexical bundles. They also stress a more genre-focused EAP (English for academic purposes) especially in advanced writing courses.

**Keywords:** corpus linguistics, genre, applied linguistics, postgraduate writing, *it* lexical bundles

### INTRODUCTION

The study of formulaic patterns in general and specific groups of word combinations in particular has a history of more than five decades (Cortes, 2002). Among diverse categories of formulaic sequences, lexical bundles, also known as clusters and chunks (Hyland, 2008a, 2008b), were first introduced and defined by Biber, Johansson, Leech, Conrad, and Finegan (1999) in their extensive treatment of English grammar as a new group of word combinations with important functions. They defined lexical bundles as "recurrent expressions, regardless of their idiomaticity, and regardless of their

structural status" (p. 990). More importantly, they referred to frequency as the most salient and defining characteristic of bundles; in order for a word combination (e.g. *on the other hand, at the same time, it is necessary to*, etc.) to count as a bundle, it must occur at least twenty 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. Lexical bundles are identified empirically just on the basis of frequency and breadth of use (Cortes, 2002, 2004). Fixedness in form (e.g., *on the basis of* not *\*on a basis of*) and non-idiomatic meaning (e.g., the meaning of a four-word bundle like *in the presence of* is almost easily retrievable from the meaning of its individual parts) are other properties of bundles. Among other registers, lexical bundles have been found to be an important part of academic discourse (Biber et al, 1999).

Lexical bundles have been classified structurally (Biber et al, 1999; Biber, Conrad, and Cortes, 2004; Biber, 2006; Jalali, 2009; Jalali, Eslami Rasekh and Tavangar Rizi, 2008; Staples, Egbert, Biber and McClair, 2013) as well as functionally (Cortes, 2013, 2011, 2006, 2002; Biber, Conrad, and Cortes, 2003; Biber and Barbieri, 2007; Jalali, 2009; Jalali et al., 2008; 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 (Biber and Barbieri, 2007). Interestingly, there is also usually a correlation between the structural type of bundles and the function they serve in the discourse (Biber et al, 2004); for example, *it* bundles followed by *is*, an adjective and an infinitival *to* or the complementizer *that* (e.g. *it is necessary to, it is clear that*), the subject of the present study, are usually used to act as metadiscourse elements (Hyland, 2000, 2008a, 2008b) or expressions of stance (Biber, 2006). Biber et al (1999) show that *it* clauses or bundles followed by either *to* (as in *it is important to note that this relationship may always be true*) or *that* (as in *it is clear that this policy is unlikely to lead to fruitful results*) are common in academic writing and their relatively frequent presence has been substantiated in research articles (Hewings and Hewings, 2002).

According to Hewings and Hewings (2002), clauses starting with an anticipatory *it* have four metadiscoursal or interpersonal roles: hedges (showing speaker or writer's tentativeness and uncertainty about the following proposition), attitude markers (expressing writer's attitude toward the content), emphatics (stressing writer's certainty about the force, and credibility of the propositional meaning), and attribution (convincing the reader through a general or specific reference). It is, however, noteworthy that very few studies have focused on the use of *it* bundles within some key relatively over-looked genres of the academy (see Hewings and Hewings, 2002; Hyland, 2008a). Especially important is the scarcity of studies that would address specific phraseological practices in academic students genres like master theses and doctoral dissertations (this study makes a distinction between theses and dissertations with the former being written by students at the master's level and latter by those in the doctoral level) especially with an aim to describe and explain possible differences

and/or similarities between these two groups in their use of these word combinations in their respective high-stakes genres.

## THIS STUDY

The purpose of this study is to compare the use of *it* lexical bundles as one structural class of bundles with important metadiscursive functions in EFL master theses and doctoral dissertations in one single disciplinary area of applied linguistics through the use of two corpora of academic writing. This structural group of lexical bundles was investigated in this study for two reasons. First, there is some evidence to suggest that for many non-natives, this structure can pose serious degrees of difficulty mostly because of the absence of such an *it* structure in some languages (Jacobs, 1995, Hewings and Hewings, 2002). Second, recognizing the importance of this structure as reflective of metadiscursive elements or stance expressions, the study seeks to identify the range of interpersonal meanings conveyed by such word clusters. It has been previously highlighted that *it* clauses or bundles could be usually good means by which writers can express their opinions, evaluate the subject matter, and engage with their audience (Hewings and Hewings, 2002). Meanwhile, exploring possible variations in the use of such word combinations across the two postgraduate genres could be a good contribution to a better understanding of phraseological preferences and practices in these two relatively similar discourse communities. This study, therefore, addresses the following questions:

- What are the most frequent four-word *it* lexical bundles in two postgraduate genres of applied linguistics?
- To what extent is there evidence to support similarity or contrast in the range, frequency and function of *it* lexical bundles across the two postgraduate genres?

## METHOD

### Corpora

Two corpora were used in this study. The first corpus included master theses written in the discipline of applied linguistics by EFL postgraduate students, and the second one consisted of doctoral dissertations in the same disciplinary area. Both corpora had been originally prepared by Jalali (2009) for his study on variations in the use of lexical bundles within applied linguistics (see also Jalali et al., 2008). Table 1 displays the numbers of texts and the number of words in these two corpora.

**Table 1.** Corpora word count

Corpora	Number of texts	Number of words
Master theses	22	441033
Doctoral dissertations	12	476922

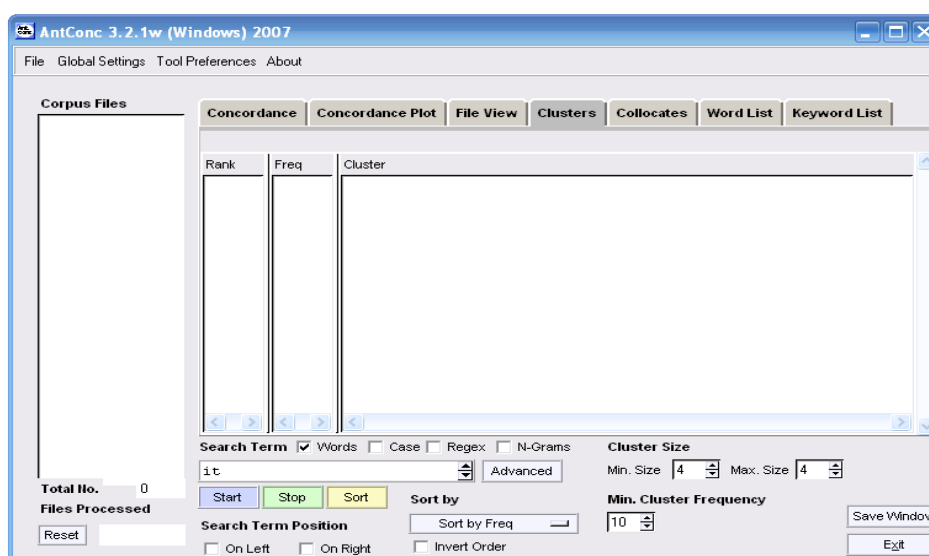
## Data analysis tools

### *Computer programs*

Two computer programs were used in this study: Antconc3.2.1w (Anthony, 2007) and Wordsmith (Scott, 2008). The former was used for the identification of lexical bundles and concordancing while the latter was only used to find the number of texts within which each bundle had been used. These two programs are described more below.

Antconc3.2.1.w is a free concordance program designed and developed by Anthony (2007) (see Figure1). This study used it to identify "it" lexical bundles and find their frequency. It is an extremely practical tool that is free on the internet. It has useful tools such as concordance, concordance plot, file view, N-grams (part of clusters), collocates, word lists, and keywords list that are used to analyze the texts. For example, the "word list" tool takes the words in the corpus and places them in a ranked order based on the most frequent words, or "file view" allows the researcher to see the word or concordance in its context in order to re-contextualize how it is used. The concordancer also makes it possible to see each of the clusters in their actual textual context within which they had originally been used.

But among all of these tools, there is a tool by which it is possible to identify word combinations, clusters, or lexical bundles of different lengths and frequencies in small or large corpora. All lexical bundles in corpora of different sizes with their actual frequencies are found and displayed by inserting a set of commonly key words with which the bundles collocate such as prepositions (e.g., at, of, on, etc), modals (e.g., can, should, could, may, etc), etc, and deciding on the minimum optimal frequency (e.g. ten, twenty or forty in a corpus of one million words) and the number of words in clusters (e.g., three, four, five, or six).



**Figure1.** Tools of Antconc3.2.1. Software

WordSmith tools5 (Scott, 2007) was another computer program applied for the identification of lexical bundles (see figure 2). This program is similar to Antconc3.2.1.w, but as Antconc3.2.1.w could not count and display the number of different texts, wordsmith tool5 was employed in order to count the texts. So when all candidate lexical bundles were identified by the first computer program, each of them was again searched on Wordsmith tools5 to find the number of texts with which they had been used.

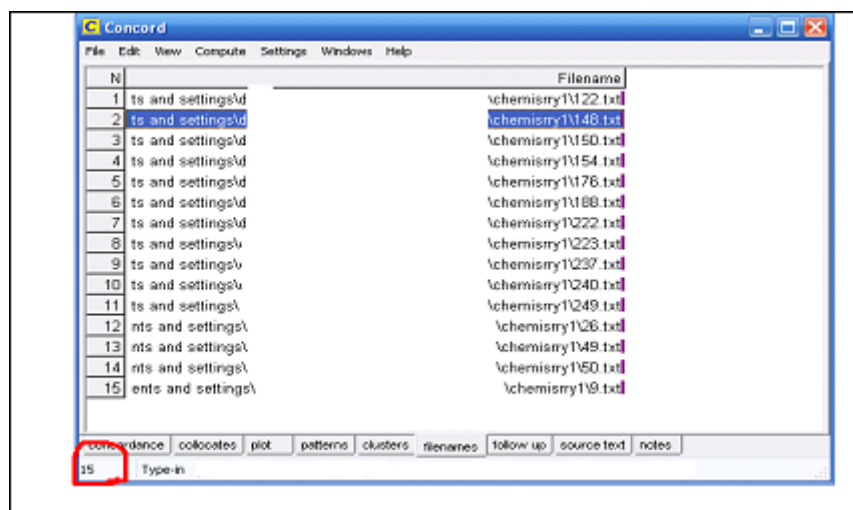


Figure 2. Wordsmiths tools

### ***Functional analysis of bundles***

The focus of this study was on 4-word bundles because as Hyland (2008a) proposed, "they are far more common than 5-word strings and offer a clearer range of structures and functions than 3-word bundles" (p.8). Bundles are essentially extended collocations defined by their frequency of occurrence and breadth of use, but the actual frequency cut offs are somewhat arbitrary. As the size of the two corpora was relatively, this study did not adopt any pre-determined cut-off frequency as this would exclude some potentially important bundles from consideration. Therefore, lexical bundles were explored in a more exploratory manner. So all *it* bundles that had been used in at least 10% of texts, i.e. the word combinations had to appear in three or more texts, were regarded as lexical bundles.

The data were analyzed in three steps. First, all *it* lexical bundles of interest were identified in the two corpora along with their actual frequencies and the number of texts in which they had been used. Second, by using the functional typology of *it*-clauses developed by Hewings and Hewings (2002) (see table 2) and the tools of AntConc 3.2.1 concordancer (Anthony, 2007) and WordSmith tool5 (Scott, 2008) for the analysis of lexical bundles, the researcher analyzed all lexical bundles identified in their contexts and decided on the most predominant functions to which they had been put. In the third stage, the results from each of the two corpora were compared to determine the extent

to which the two EFL postgraduate genres of applied linguistics were similar and/or different from each other in terms of range, frequency and function of *it* bundles.

It should be explained that while there are already some functional classifications of lexical bundles (e.g. Biber and Conrad, 1999; Cortes, 2002; Biber et al, 2004; Hyland, 2008a, 2008b), Hewings and Hewings' functional taxonomy of *it*-clauses (2002) was used in this study since it specifically deals with the interpersonal functions of this structural group. However, as the developers of this model confirm themselves, it should be noted that no functional classification of language can be totally objective and watertight; therefore, some subjectivity in the functional analysis could be inevitable. There are also no clear-cut divisions between all categories especially between the second, attitude markers, and the third one, emphatics.

**Table2.** Interpersonal Functions of *it* clauses (Hewings and Hewings, 2002: 372)

Interpersonal functions <i>It</i> -clauses of	subcategories	Example realization
1. hedges	1a likelihood/possibility/ certainty; importance/value/necessity etc. 1b what a writer thinks/assumes to be//will be/ was the case	It is likely, it seems improbable, it would certainly appear, it could be argued, it was felt
2. attitude markers	2a the writer feels that something is worthy of note 2b the writers evaluation	It is of interest to note; it is worth pointing out; it is noteworthy; it is important
3. emphatics	3a the writer indicates that a conclusion/deduction should be reached; that a proposition is true 3b the writer strongly draws the reader's attention to a point 3c the writer expresses a strong conviction of what is possible/ important/necessary, etc.	It follows; it is evident; it is apparent It is important to stress; it should be noted; it must be recognized It is clear; it is impossible; it is safe to assume
4. attribution	4a specific attribution ( with a reference to the literature) 4b general attribution ( no referencing)	It has been proposed (+ reference) It is estimated (+ no reference)

## RESULTS

As table 3 shows, overall, there were only six different bundles with this particular structure in the two corpora with *it is important to* being the top most frequent bundle in the two postgraduate genres. Surprisingly, this shows that this structural group of bundles is not very frequent in comparison to other structural groups of bundles in postgraduate genres (Biber et al, 1999; Hyland, 2008a). Although the size of the two corpora was not very large, it seems that postgraduate students make a relatively infrequent use of these bundles. In fact, given the low frequency of some *it* bundles in students' texts (*i.e. it is possible that, it is difficult to, and it is clear that* in doctoral

dissertations and *it is necessary to*, *it is clear that*, and *it is possible to* in master theses), it is dubious whether they would really count as bundles if they were explored in corpora as large as one million words. There were also some differences in the extent to which the two groups of writers used certain functional groups of bundles. These are discussed in more details and with some examples below.

**Table 3.** Frequency of *it* bundles in the two postgraduate genres

Lexical bundles	Doctoral dissertations No# No of texts	Master theses No# No of texts
it is important to	21# 6	15# 10
it is possible that	3 #3	8 #7
it is difficult to	6 #5	12# 8
it is necessary to	11# 7	7# 6
it is clear that	5 #4	6#5
it is possible to	16# 7	3 #3
Total	62	51

## Hedges

The study found that *it is possible that* and *it is possible to* were the only two bundles used mostly to show writers' tentative stance toward the following propositions. While postgraduate students used both of these bundles to withhold their complete commitment from the arguments and express rather hypothetical statements, they did not seem to employ such bundles quite commonly in their discourses. The scarce use of *it is possible that*, especially in the corpus of doctoral texts, could be attributed to the relatively low number of such texts in comparison to master theses used in this study. However, as far as the function of this bundle is concerned, it seems that postgraduate students used this bundle in the right way to serve its interpersonal function as the following examples can show:

(1) This finding may indicate that as a result of greater exposure to L2 input, *it is possible that* a negative correlation exists between length of stay in the target community and the degree of pragmatic transfer. (Corpus of doctoral dissertations)

(2) *It is possible that* an L2 learner tends to access the topic knowledge in the L1 in which the topic knowledge was processed and acquired if the information has never been reprocessed in an L2. (Corpus of master theses)

With regard to *it is possible to*, there was a considerable difference between students at the master's level and doctoral students in the frequency with which they used this bundle. The study showed that doctoral students drew on this bundle quite recurrently, in comparison to master's students, to mitigate the force of their claims, findings, and interpretations. The following examples from the corpus of doctoral texts can show how they used this *it* bundle:

(3) These examples point to the fact that *it is possible to* transfer a range of literal expressions from a concrete semantic domain (e.g., money) and use them metaphorically to describe abstract experiences of another semantic domain (e.g., time).

(4) This finding again probably backs up the claim that *it is possible to* beef up L2 learners' conceptual fluency and metaphorical competence.

(5) Therefore, *it is possible to* consider UG as the cognitive module that constrains syntactic constructs during acquisition but itself remains invariant during this process.

The relatively infrequent use of this bundle in master theses could be partly accounted for by referring to generic expectations. The most important purpose that such students follow in their theses is to first show their familiarity with disciplinary knowledge, research and practices, and second to report on the results of their studies (Hyland, 1996, 1999, 2008b). They may assume that drawing on their own interpretations and inferences about the study may not be so much part of their job at this level, so they try to adhere mostly to the study itself and minimize their own presence in the text. Furthermore, students at this level may rely on some other simpler expressions (Cortes, 2004, 2006) to show their tentativeness and lack of certainty. Students' preference for such expressions may be simply because they are less difficult to use (Jones and Haywood, 2004). While research article writers may also draw on adverbs like *maybe*, *probably*, and modals like *may* and *might* to convey a state of tentativeness and uncertainty in their discourses, students at the master's level were found to be more dependent on these apparently simpler ways of expressing tentativeness.

### Attitude markers

There were two *it* bundles that were put in the category of attitude markers based on the analysis of their functions in their contexts of use: *it is important to*, and *it is difficult to*. Of course, it must be noted here that the former had also a perceptibly emphatic tone as well and therefore, could overlap with the third category, emphatics (see Hewings and Hewings, 2002). Although *it is important to* was found to be the most frequent bundle in both corpora, it seemed that it was employed more frequently in doctoral dissertations than master theses. In doctoral texts, *it is important to* was found to be often part of a longer bundle: *it is important to note* and to a less extent *point out*. It seemed that through conjoining this bundle with a verb like 'note', doctoral students could explicitly direct readers' attention to an important point and engage them in the evolving discourse. In the case of *point out*, the main purpose seemed to be stressing a point that is very important for readers' understanding of the whole study or there is a fear on the part of writers that otherwise, something may be missed or mistaken on the part of the readers. Other verbs collocating more with this bundle in students' texts were *know*, *see*, *make*, *keep*, *recall*, *realize*, *specify*, *consider*, *understand*, *emphasize*, and *inquire*. The following two examples showcase the use of this bundle by master's and doctoral students, respectively:



(6) *It is important to* note that their classification was not only theory-based but also has been fairly accepted by both teachers and researchers in the field.

(7) It is important to note that the Full-Access Hypothesis does not deny the existence of differences between L1 and L2 acquisition, nor is it incompatible with the existence of linguistic development through time. Within this framework, however, the source of these differences is not a lack of access to UG in L2 acquisition.

Interestingly, *it is difficult to* was used more frequently by students at the master's level. The scarce use of this bundle in doctoral texts was surprising given that this bundle had been used more by less proficient and expedient students at the master's level. This could be because of the smaller number of doctoral texts (12) in comparison with master texts (22). *It is difficult to*, which was usually used in the final parts of texts, usually described the difficulty in doing an action, or reaching a conclusion as can be seen in these examples:

(8) With these results, *it is difficult to* assess the effect of writing tasks on the L1 use involved (Corpus of doctoral dissertations).

(9) On the other hand, in the case of some ungrammatical sentences in the collected data, *it is difficult to* single out what specific principle or parameter is exclusively violated because, in any sample of a language, there might obviously be more than one single principle or parameter involved. (Corpus of master theses)

## Emphatics

There were two *it* bundles with a mostly emphatic function in the two corpora based on the analysis of bundles in their contexts of use: *it is clear that*, and *it is necessary to*. Drawing on the classification developed by Hewings and Hewings (2002), both of these two bundles can be put in the subcategory 3c: "the writer expresses a strong conviction of what is possible/ important/necessary, etc." (p.372). By using *it is clear that*, writers try to project the following proposition in the subordinate clause as an undisputed and almost certain argument. The use of this bundle, therefore, can help writers to overtly express their position regarding the factual status of the following proposition and commit themselves more to the accuracy of the ensuing argument. Perhaps, this can account for students' relatively rare use of these bundles as they may not feel confident enough to frankly voice their own personal judgment about the truth of a given proposition. Postgraduate students, either at the master's or doctoral level, did not like to run the risk of using the strong, authoritative, and somehow imposing language that *it is clear that* and *it is necessary to* implies. Probably, the most important difference between these two emphatic bundles is that while the latter mostly works to serve as an indicator of the factual and nontentative status of a given proposition, the former is usually used to invite or urge the writer, readers, future potential researchers, and consumers of research (e.g. teachers and other practitioners) to a future action or a way of thinking.

Although, like *it is clear that*, the frequency of *it is necessary to* in both postgraduate genres was relatively low, it seemed that doctoral students drew on this bundle more than students at the master's level. The more infrequent use of this bundle in master texts could be again attributed to these developing writers' incipient growing disciplinary identity and confidence. The use of highly persuasive *it is necessary to* implies the voice of a disciplinary knowledgeable expert who in one way or another wants to make the readers come to a particular kind of thinking or do a possible future action. While doctoral students seem to have developed this confidence at least partly, students at the master's level are not so much at ease with this overtly expressive bundle. The following final examples can show doctoral students' use of this bundle:

(10) Although one may not consider text analysis as an instrument of data collection but rather as a method, *it is necessary to* mention that text analysis is used in this study to see the quality of business correspondence in terms of culture load.

(11) It has to be shown that the evidence for parameter resetting in SLA is convincing, otherwise *it is necessary to* apply complementary perspectives on SLA to reach a better understanding of this issue.

## DISCUSSION

Postgraduate students' relatively infrequent use of *it* bundles, both in range and frequency, in their writing could lend support to the findings of the previous research (e.g., Cortes, 2004, 2006) that had shown generally students, whether native, non-native, graduate or undergraduate, as still novices to any particular disciplinary community, tend to rely less on some groups of bundles in the development of their discourses as they are associated with the more overt expressions of stance (Jalali, 2009; Jalali et al., 2008). In fact, the analysis of both doctoral dissertations and master theses used in this study showed that the number of different lexical bundles used by Iranian EFL students was relatively low. It seemed that students both at the master's and doctoral levels tended to use *it* lexical bundles infrequently. While part of this gap could be attributed to generic differences, writers' purposes and readers' expectations, it can also suggest that EFL postgraduate students, because of their lack of enough confidence or expertise, rely less on *it* lexical bundles. Like Cortes (2006) and Jones and Haywood (2004), this study, therefore, reflects the fact that good acquisition of lexical bundles seems to be a long-term goal as far as their production in developing writers is concerned.

Students' relatively infrequent use of *it* bundles could also be surprising as they had already been exposed to such word-sequences several times in their prior readings of applied linguistics published literature. There is almost no doubt that postgraduate students have repeatedly observed different lexical bundles, including *it* bundles, in different research articles they may have studied for doing and writing their own research. Furthermore, given that *it* lexical bundles are very pervasive in university written language (Biber et al., 1999; Biber and Barbieri, 2007) and they may have a

formulaic status (Wray, 2000, Wray and Perkins, 2000), it may be expected 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 writing expertise.

This infrequent use of *it* bundles in both corpora could be partly attributed to the absence of this structure in many languages (Hewings and Hewings, 2002) including Persian, and also the association of this particular structural group of bundles with the relatively explicit expression of writers' stance, something that postgraduate students may not feel comfortable with. In addition, in line with findings of some previous research (e.g., Biber et al, 1999; Cortes, 2002, 2004; Biber et al, 2004; Hyland, 2008a, Jalali et. al., 2008), it may be postulated that as *it* lexical bundles, among other word combinations, serve a wide range of interpersonal functions in the academic register, postgraduate students may be apprehended by directly bringing themselves in to the texts by using such lexical bundles. Furthermore, the potential influence of other factors like students' L1, writing conventions in the first language, the topics and areas of research, students' prior reading experience, supervisors' preferences and students' degree of familiarity and their expertise in the use of clusters cannot easily be underestimated and need to be investigated further.

Another important reason for the infrequent use of *it* bundles could be that they are usually encouraged to avoid overt personal presence in the texts by many style guides and in some cases, their instructors. They may not be helped to realize that academic writing, like many other registers, cannot be absolutely objective and depersonalized (Biber, 2006; Hyland, 2004). Probably, many postgraduate students whose works were examined here had no problem at least understanding *it* bundles given that they may have been exposed to such clusters quite often in their prior readings, but they were simply trying to avoid some of them on the basis of a mistaken assumption that the use of such word sequences (e.g. *it is clear that, it is necessary to*) may signal unsubstantiated claims, strong language or even impoliteness.

Students should understand that gaining acceptance and recognition in the community of expert published members is one of the main responsibilities of a member of academy in almost any given field of study (Swales, 1990). One of the factors that could affect the success or failure of novice postgraduate students in getting their work published may lie in the degree to which they adhere to those word sequences as part of disciplinary conventions, which, if not peculiar and exclusively favored in a given discipline, are typically used by established academics (Cortes, 2004). Understanding to use *it* bundles is, in fact, part of this familiarity with the writing practices.

This shows that lexical bundles should be at the forefront of explicit instruction at the initial stages of language learning given that their acquisition could be a time-consuming process. Although there are already some models on how to introduce students to different word combinations (e.g. Nattinger and Decarrico, 1992; Lewis,

1997; Willis, 2003), the findings of this study can call for a more increased pedagogical focus on different multi-word sequences like lexical bundles. 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. Exposing students to good samples of published writing in their disciplines, especially those usually introduced to students to take as models in their own writing and making them notice the form, frequency and function of such bundles, may help them come to a better understanding of these word clusters and their often necessary functional contribution in the academic discourse. These implications are discussed more in the next section.

## CONCLUSION

Lexical bundles are not idiomatic in meaning and hence they 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). It seems, therefore, necessary for EAP practitioners in general and academic writing instructors in particular to leave a good space in their instruction for a more pedagogically focused treatment of anticipatory *it* bundles, which are for the most part a characteristic of academic writing (Biber et al, 1999). The use of noticing (Cortes, 2004, 2006), conscious raising tasks (Lewis, 2000a, 2000b), clusters lists, and concordances (Hyland, 2008a) could be some of the means by which students could come to a possibly better understanding and more frequent and appropriate use of these word combinations. These implications may also hold true for native-speaker developing writers as the infrequent and rare use of target bundles in their production has been almost well attested in some previous research (e.g. Cortes, 2002, 2004, 2006).

In spite of two decades of research on lexical bundles, much still remains to be explored about this group of word combinations which can contribute to an almost overlooked dimension of genre analysis (Hyland, 2008b). Identifying lexical bundles in other disciplines, registers, and genres, examining the formulaic status of these multi-word sequences (Biber and Barbieri, 2007) and probing the effect of a pedagogical treatment on their acquisition could be areas worth exploring in future research.

## 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., 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.
- 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? *Cuaderns de Filologia. Estudis 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.
- 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.
- Hyland, K. (1996). Writing without conviction? Hedging in scientific research articles. *Applied Linguistics*, 17, 433-454.
- 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. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, 13, 131-151.
- 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.
- Jacobs, R.A. (1995). *English syntax: A grammar for English language professionals*. New York: Oxford University Press.
- Jalali, H. (2009). *Lexical bundles in applied linguistics: Variations within a single discipline*. Unpublished doctoral dissertation: Isfahan University, Isfahan.

- Jalali, H., Eslami Rasekh, A., Tavangar Rizi, M (2008). Lexical bundles and intradisciplinary variation: The case of applied linguistics. *Iranian Journal of Language Studies*, 2(4), pp. 447-484.
- Jones, M and Haywood, S. (2004). Facilitating the acquisition of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 269-292). John Benjamins: Philadelphia.
- Lewis, M (2000a). Learning in the lexical approach. In M. Lewis (Ed.), *Teaching collocation. Further developments in the lexical approach* (pp. 155-185). Hove: Language Teaching Publications.
- Lewis, M (2000b). Materials and resources for teaching collocation. In M. Lewis (Ed.), *Teaching collocation. Further developments in the lexical approach*. 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.
- Scott, M. (2008). *Wordsmith Tools 5*. Oxford: Oxford University Press.
- Staples, S., Egbert, J., Biber, D & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- 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.