Coding Stance through *it* bundles: The case of applied linguistics

Hassan Jalali, English Language Center, Isfahan University of Technology, Isfahan, Iran
h.jalali@cc.iut.ac.ir

Abstract

  The study of linguistic devices variously referred to as stance expressions is one of the best means by which the relationship between the writer, the reader, and propositional meaning could be examined. This paper looks at a particular structural group of lexical bundles encoding stance expressions. These are bundles starting with an anticipatory *it* followed by *is*, a predicative adjective and finally ending with infinitival to or complementizer *that* (e.g. *it is important to, it is possible that*). The use of these bundles is compared in three corpora of research articles, doctoral dissertations, and master theses in the discipline of applied linguistics to explore possible generic variations and identify possible differences between published students writing. Using Hewings and Hewings's functional typology of interpersonal roles of *it* clauses (2002), this group of bundles is found to have three stance expressions of hedging, marking attitude, and stressing emphasis. The major difference is discovered to be between students' genres and research articles, with the former drawing less in their expression of interpersonal meanings. The differences are accounted for by referring to generic expectations, and students' growing disciplinary identity. The findings of the study have some implications for academic writing.

Keywords: anticipatory *it* bundles*,* stance expressions, academic writing, research articles

Introduction

  Lexical bundles were first introduced and defined by Biber, Johansson, Leech, Conrad, and Finegan (1999) in their well-known rendering 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 refer 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 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. Lexical bundles are identified empirically 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 form the meaning of its individual parts) are among other properties of bundles. Among other registers, lexical bundles have been found to be an important part of academic discourse (Biber et al., 1999; Hyland, 2008a, 2008b).

  Lexical bundles have been classified structurally (Biber et al., 1999; Biber, Conrad, and Cortes, 2004; Biber, 2006a) as well as functionally (Cortes, 2001, 2002; Biber, Conrad, and Cortes, 2003; Biber et al., 2004; Biber and Barbieri, 2007; Hyland, 2008a, 2008b). The most widely-used structural typology of bundles is that of Biber et al. (1999), where these multi-word sequences were arranged into eight categories (see table 1). 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 (Cortes, 2002, 2004; Hyland, 2008a, 2008b), registers, such as conversation, fiction, news, academic prose, classroom teaching and non-conversational speech) (Biber et al., 1999; Biber and Conrad, 1999; Biber et al., 2004, Biber and Barbieri, 2007), genres (Hyland, 2008b), and different degrees of writing expertise (Cortes, 2002, 2004; Levy, 2003). Overall,

these studies indicate that lexical bundles are strong discipline, genre, and register discriminators (Biber et al., 1999). This means that apart from some overlaps, each discipline, genre, or register draws on its own specific set of bundles to organize its discourse, express stance, and refer to different parts of the evolving text or elements outside the text. The findings also indicate that many lexical bundles favored by experts in any given disciplinary area may not be used by students with varying degrees of language proficiency and disciplinary expertise (see for example Cortes, 2004, 2006). As building blocks of coherent discourse, these word clusters can serve such a wide range of discursive functions as organization of discourse, expression of stance, and reference to textual or external entities (Biber and Barbieri, 2007). Interestingly, there is sometimes a correlation between structural type of certain bundles and the function they serve in the discourse (Biber et al., 2004); for example, anticipatory *it* bundles are usually used to act as expressions of stance (Biber, 2006a).

The clause-initial anticipatory *it* is often part of a multi-word fixed word combination or bundle which can act like a frame within which the following propositional meaning could be embedded (Biber and Barbieri, 2007). A straightforward and possibly accessible way for writers to express their stance toward the reader, and the content could be the use of those bundles beginning with anticipatory *it*, copula *is*, a predicative adjective (e.g. *interesting*, *necessary*, *clear*, etc.) followed by a subordinate clause usually introduced with complemetizers *to* or *that* (*e.g. it is interesting to*, *it is possible that*). Such bundles seem to convey a range of epistemic, evaluative, and attitudinal meanings. Most extraposed complement clauses beginning with anticipatory *it* can also reflect the speaker or writer's assessment (Hewings and Hewings, 2002).

There are different reasons for the use of this type of *it* clausal bundles (Hewings and Hewings, 2002). Grammatically, there is a marked tendency in English to put the longer subjects at the end of the clause (Quirk, Greenbaum, Leech, and Svartvik, 1985). Thematically, elements with higher information usually occur at sentence final position (Hasselgard, Johansson, Lysvag, 1998). From a metadiscursive point of view, the use of *it* bundles enables the writer to distance herself or himself away from the propositional content and thereby project an objective and impartial persona (Quirk et al., 1985; Hyland, 2004). One the other hand, by embarking on such a structure, the writer is able to evaluate the ensuing proposition (Hunston and Sinclair, 2000), and finally to depersonalize the opinions (Hewings and Hewings, 2002) as can be seen in the following examples taken from the corpus of research articles used in this study:

(1) First, *it is important to* help student-teachers to look at teaching through multiple but complementary lenses.

(2) *It is necessary to* bear in mind the mix of regional and national languages and accents that the participants represent as well as those that they are familiar with.

Biber et al. (1999) show that *it* clauses followed by either *to* (as in *it is important to note that this relationship may always be true*) or complementizer *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 a range of academic genres (Hewings and Hewings, 2002). According to Hewings and Hewings (2002), clauses starting with an anticipatory *it* have four interpersonal roles (see table 2): 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). As will be discussed later, this same model is also used in

classifying the interpersonal or metadiscursive functions of anticipatory *it* bundles in the present study.

Table 1, Most common patterns of 4-word bundles in academic writing (Biber et al.., 1999, pp. 997–1025)

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

The purpose of the present study is to compare the use the mentioned group of bundles in three key written academic genres of one single disciplinary area. The structure of these bundles is made of anticipatory *it*, *is*, a predicative adjective (*e.g. necessary*) and one of two complementizer, *to* (as in *it is necessary to*) or *that* (as in *it is clear that*). This structural group of lexical bundles is here investigated for two reasons. First, there is some evidence to suggest that for many non-native speakers of English, this structure can pose serious degrees of difficulty mostly because of the absence of anticipatory *it* structure in some languages (Jacobs, 1995; Hewings and Hewings, 2002). Second, recognizing the importance of this structure as a representation of stance expressions, can help to identify the range of interpersonal meanings conveyed by such word clusters as they are usually good means by which writers can express their opinions, evaluate the subject matter, and engage with readers (Hewings and Hewings, 2002—See table 2 below).

More specifically, this study focuses on the use of anticipatory *it* lexical bundles in three corpora of research articles, master theses, and doctoral dissertations, all in the single discipline of applied linguistics, to find the extent to which published academics in this area are different among postgraduate students in the variety of used bundles and the functions to which they have been put.

Table 2, Interpersonal functions of it clauses (Hewings and Hewings, 2002: 372)

Interpersonal functions	Subcategories	Example realizations
1. Hedges	1a likelihood/possibility/certainty; Importance/value/necessity etc. 1b what a writer thinks/assumes, to be\will	it is likely, it seems probable, it would certainly appear it could be argued, it seems reasonable

	be\was the case	to assume, it was felt
2. Attitude markers	2a the writer thinks that something is worthy of note 2b the writer's evaluation	it is of interest to note; it is worth pointing out; it is noteworthy it is important; it was interesting; it is surprising
3. Emphatics	3a the writer indicates that a Conclusion/deduction should be Reached; that a proposition is true Reader's attention to a point	it follows; it is evident; it is apparent it is important to stress; it should be noted; it must be recognized; it is Essential to understand
4 attribution	4a specific attribution (with a Reference to the literature) 4b general attribution(no referencing)	it has it is estimated (+no reference) been proposed (+reference)

This study is similar to Hewings and Hewings (2002) in that it examines the use of anticipatory *it* structures in student published writing. However, the study has a narrower focus of examining a group of anticipatory *it* bundles as a set of frequently-occurring word combinations which are structurally incomplete by themselves (Biber et al., 1999). More specifically, the study is concerned with those multi-word sequences whose elements are the first part of a larger clause. It is also distinctive in that it works with three relatively large corpora especially in the case of research articles (see table 3 and 4).

Two rather important points must be mentioned here. First, this study is only looking at those anticipatory *it* bundles made of anticipatory *it*, *is*, a predicative adjective, and complementizer *to* or *that*, and therefore, other potential anticipatory *it* bundles like *it should be noted*, *it seems that the*, and *it can be argued* were not included. Second, it should be noted that 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).

Method
Corpora
In order to explore possible differences between students' genres and research articles, three corpora were employed in this study. The first and second corpora represented master theses and doctoral dissertations in applied linguistics written by EFL postgraduates (see table 3), while the third corpus comprised more than two hundred articles from seven different journals in applied linguistics (see table 4), based on the advice of expert informants, previous corpus-based studies in this field (*e.g.* Ruiying and Allison, 2003, 2004), and access to electronic files of journal articles. As can be seen, the first and second corpora were not as large as one million words while the corpus of research articles was larger than a million

words. Given such an apparently large difference in the size of the corpora and the possible unreliability of employing a normalization procedure (Biber and Barbieri, 2007), a more qualitative approach, similar to that used by Cortes (2004), was followed in this study, as described below.

Procedures

First, the corpus of research articles was explored to identify anticipatory *it* bundles in published writings in applied linguistics. As in this study the more conservative frequency cut-off of twenty in one million was adopted, to find lexical bundles in research articles corpus, anticipatory *it* bundles 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. The other two corpora were searched to see the extent to which they were used by postgraduate students. As stated before, because of the relatively small size of students' corpora, which were less than half a million words, and hence the unreliability of employing a normalization procedure (Biber and Barbieri, 2007; Cortes, 2002), lexical bundles in students' production were not explored separately. It should be noted that this is one of the distinctive aspects of this study.

Table 3, 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 4, 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

Computer 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 as well as their actual contexts of use: Antconc3.2.1w (Anthony, 2007), and Wordsmith5 (Scott, 2008). Antconc3.2.1w is usually used for identification of N-grams (word combinations clustering together), but it can also be used for the identification of such word combinations as lexical bundles. Another computer program was also used, i.e. Wordsmith5, to find the number of

texts within which each bundle had been used. Both of these two computer programs are described more below.

Developed by Anthony (2007), AntConc 3.2.1 has a number of useful features and tools (e.g. concordances, concordance plot, file view, N-Grams (part of clusters), collocates, word List, keyword list). This free text analysis computer program has a tool by which it can identify word combinations of different lengths and frequencies in small or large corpora. Just by giving it a set of commonly 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 word sequences in corpora of different sizes with their actual frequencies. The concordancer also makes it possible to see each of the clusters in actual textual context within which it had originally been used. Probably, 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 with which a given bundle had been used. So, when lexical bundles were identified by the first computer program, each of them was again searched on Wordsmith tools5 to find the number of texts and only those that had appeared in five different texts could enter the analysis.

It is to be also noted here that in this study only four-word anticipatory *it* bundles having the aforementioned structural constituents were analyzed since, generally, bundles of this length are more frequent than five-word clusters and serve more varied functions than three-word combinations, which are for the most part too frequent to be managed in a study of this kind (Hyland, 2008a, 2008b).

Data analysis

The data were analyzed in three steps. First, all anticipatory *it* lexical bundles were explored in the corpus of the published writings in applied linguistics. Then, the two students' corpora were explored to find which of the bundles identified in the previous corpus were used. Second, by using the functional typology of *it*-clauses developed by Hewings and Hewings (2002) (see section 1.1 and table 2) and the AntConc 3.2.1 concordancer, an attempt was made to probe the context in which bundles had been used, and in this way the most predominant functions to which they had been put were spotted and classified. In the third stage, the results were compared to determine the extent to which research articles might be different from EFL students' postgraduate writings in applied linguistics with regard to frequency and function of this group of anticipatory *it* bundles.

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's functional taxonomy of *it*-clauses (2002) was used in this study since it is specifically dealing with interpersonal functions of this structural group. But as the developers of this model confirm themselves, no functional classification of language can be totally objective and watertight; therefore, as well as a good degree of subjectivity in functional grouping, there are no clear-cut divisions between all categories (see table 2).

Results and discussion

As table 5 below shows, overall, there were only six different bundles with this particular structure in the three corpora with *it is important to* as the top most frequent bundle. Not surprisingly, this shows that this structural group of bundles is not very frequent and

almost seems to be the least used in comparison to other structural groups of bundles (Biber et al., 1999; Hyland, 2008a). Although all six anticipatory *it* target bundles identified in the corpus of research articles were also found in students' corpora, published academics made a noticeably heavier use of these bundles than student writers. In fact, given the low frequency of some target 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 three groups of writers used certain functional groups of bundles. For example, there was a wide discrepancy between research article writers and postgraduate students in the considerable reliance of the former on such bundles as *it is important to*, *it is possible that*, and *it is clear that* than the latter. These functional differences are discussed more in the next three parts.

Table 5, *Frequency of anticipatory it bundles in the three genres*

Lexical bundles	Articles No# No of texts	Doctoral dissertations No# No of texts	Master theses No# No of texts
it is important to	88# 58	21# 6	15# 10
it is possible that	38# 23	3 #3	8 #7
it is difficult to	36 #31	6 #5	12# 8
it is necessary to	36 #29	11# 7	7# 6
it is clear that	33# 26	5 #4	6#5
it is possible to	25 #22	16# 7	3 #3
Total	254	62	51

As can be seen from table 5, *it is possible that* and *it is possible to* were the only two bundles used mostly to show writers' tentative stance towards the following propositions. While published academics used both of these bundles relatively frequently to withhold their complete commitment from the arguments and rather hypothetical statements, postgraduate students did not often employ such bundles in their written discourse. 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. The following two examples show this kind of interpretative and rather inferential use in the corpus of research articles:

(3) *It is possible that* the learners in the control group did not report noticing of certain forms because their focus was not oriented towards them even if they noticed them.

(4) Therefore, *it is possible that* the Japanese participants transferred their L1 sociolinguistic norm to their role play interactions in L2 English.

But, as far as the function of this bundle is concerned, it seems that postgraduate students use this bundle in a similar way to that of the published authors to show this stance, as can be seen in these examples:

(5) 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)

(6) 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)

But with regard to *it is possible to,* there did not seem to be a considerable difference between research article writers and doctoral students except that the frequency with which the latter used this bundle was higher than the former. Doctoral students, who drew on this bundle, even more recurrently than published academics, used it commonly to mitigate the force of their claims, findings, and interpretations. The following two examples show the typical use of this bundle in published writings:

(7)This shows that *it is possible to* combine activities which involve relatively simple language with complex thinking (and that activities involving complex thinking do not necessarily require complex language).

(8) There are three ways to address this problem. First, *it is possible to* compare learners in terms of the amount of time spent on foreign language learning (e.g., Sellers, 2000). Second, *it is possible to* situate learners on standardized tests (e.g., CELT, TOEFL). Third, it is advisable to have the same learners take all performance measures used in previous studies, compare relative difficulty of the measures, and estimate the learners' proficiency levels across the studies.

The following 3 examples from the corpus of doctoral texts can show how similar they were to the published writers in the way they used this bundle:

(9) 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).

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

(11) 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 show their familiarity with disciplinary knowledge, research, and practices, and report on the results of their studies (Hyland, 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).

Attitude markers
There were two anticipatory *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, overlapped with the third category, emphatics (see

Hewings and Hewings, 2002). Although *it is important to* was found to be the most frequent bundle in all the three corpora, it seemed that it was employed much more frequently in research articles than doctoral dissertations and master theses (see table 5). In research articles, *it is important to* was found to be often part of a longer bundle, i.e. *it is important to note* and to a less extent, *point out*. It seemed that through conjoining this bundle with a verb like 'note', research article writers explicitly directed readers' attention to an important point and engaged them as can be seen in the following examples:

(12) *It is important to* note here that the design of an elicited imitation test can largely determine to what extent it is either a measure of a learner's internal language system or a measure of his/her ability to imitate given stimuli verbatim.

(13) However, *it is important to* note that, nevertheless, no empirical evidence of any kind is offered in support of this interpretation – no interviews or any other means of verification were used. Instead, it is solely Holliday's own views that form the basis of the analysis.

Similarly, in the case of *point out*, the main purpose seems 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:
(14) *It is important to point out* that the computer does not leave the final choice with   the student.

There were some other verbs collocating with *it is important to* which almost served the same functions as *note* and *point out* (*e.g. distinguish, stress, highlight, notice, take into account, realize, remember, recognize, emphasize* etc.).

In doctoral dissertations and master theses, the frequency with which *it is important to* occurred with *note* was also higher than that of other verbs. Other verbs collocating more with this bundle in students' texts were those like *know, see, make, keep, recall, realize, specify, consider, understand, emphasize,* and *inquire*. The following two examples show the use of this bundle by master's and doctoral students, respectively:

(15) *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.

(16) *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 research article writers and master students. 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 the examples below taken from the three corpora:

(17) From the interview *it is difficult to* establish whether Philip repeats the notion of connecting potential buyers and sellers because he has not understood the meaning of the textbook extracts or whether, as the second excerpt from his interview suggests, he simply had not bothered to edit for repetition because the essay was written in haste. (corpus of research articles).

(18) With these results, _it is difficult to_ assess the effect of writing tasks on the L1 use involved (Corpus of doctoral dissertations).

(19) 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 again two anticipatory _it_ bundles with a mostly emphatic stance in all the three corpora based on the analysis of bundles in their contexts of use (see table 5): _it is clear that_, and _it is necessary to_. Drawing on 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 them more to the accuracy of the ensuing argument. Perhaps, this can account for students' relatively rare use of this bundle 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, do not like to run the risk of using the strong, authoritative, and somehow imposing language which _it is clear that_ implies. The following examples show typical uses by research article writers:

(20) _It is clear that_ academic literacy specialists across institutional and regional contexts are increasingly being called upon to design a variety of developmental programmes in order to assist research students—of any language background—in becoming more effective research writers in English.

(21) _It is clear that_ there have been considerable changes in the nature of tests from early forms to those that are in use today.

The frequency of _it is necessary to_ was almost the same as _it is clear that_ in the corpus of research articles. Probably, the most important difference between these two emphatic bundles was that while the latter mostly worked to serve as an indicator of the factual and non-tentative status of a given proposition, the former was 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 as can be seen in the following examples from the corpus of research articles:

(22) In fact, to develop familiarity with another culture, to improve one's real inter-cultural skills, _it is necessary to_ live within that culture for a good period of time, to be what Byram (1997: 1) terms a 'sojourner' rather than a tourist.

(23) It is difficult to draw conclusions from the NNES responses regarding accents, With reference to the variety of accents exemplified as 'easy' and 'difficult' to understand, _it is necessary to_ bear in mind the mix of regional and national languages and accents that the participants represent as well as those that they are familiar with.

Although, as in _it is clear that_, the frequency of this bundle in postgraduate genres was relatively low, it seemed that doctoral students drew on this bundle more than students at the

master's level. The rare use of this bundle in master's texts could be attributed to these developing writers' incipient growing disciplinary identity and confidence. The use of highly persuasive *it is necessary to* imply the voice of a disciplinary knowledgeable writer who seeks 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:

(24) 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.

(25) 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.

Implications and concluding remarks

Studies done in the last two decades demonstrate quite well that academic language is not as impersonal and objective as it has long been assumed (Hewings, and Hewings, 2002; Swales, 1990; Hyland, 1996, 1999, 2000; Crismore and Farnsworth, 1990; Biber, 2006a, 2006b). Written communication is not just mere transfer of propositional, ideational, and referential meaning. Equally important is the expression of affective, attitudinal, evaluative, and interpersonal stance meanings. In line with these studies, the present investigation showed that just a subset of anticipatory *it* bundles was also performing such a relatively wide range of interpersonal functions as hedging, marking the writer's attitude, and emphasizing the writer's conviction, or certainty about the truth of a given proposition or a state of affairs. The study also showed that postgraduate students' use of such bundles encoding stance was comparatively infrequent.

Unlike the study of *it* clauses by Hewings and Hewings (2002), this study showed that at least in so far as this group of anticipatory *it* bundles was concerned, not only was students' use of them relatively scarce and rare in comparison to that of the published experts, but also it relied on those bundles with a mostly emphatic and hedging functions, much less frequently than research article writers. One the other hand, the relatively more frequent employment of these bundles by doctoral students could suggest the possibility of progress in the use of these multi-word sequences. However, it seemed that both groups of students needed to receive instruction about the use of such bundles, especially because all of them were found to serve important interpersonal and evaluative functions in academic discourse.

Gaining acceptance and recognition in the community of expert published members is one of the main aspirations of an academic member 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). The present study showed that in so far as anticipatory *it* bundles are concerned, there is a wide gap between EFL postgraduate writing and that of the published experts, which may not be easy to bridge. While part of this gap could be attributed to generic differences, writers' purposes, and readers' expectations, it can also suggest that EFL postgraduate students, at both master's and doctoral level, because of their lack of enough confidence or expertise, rely less on anticipatory *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. Perhaps lexical bundles should be at the forefront of explicit instruction at the initial stages of language learning even if 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 academic discourse.

While students are usually encouraged to avoid overt personal presence in the texts by many style guides and in some cases their instructors, they should be helped to realize that academic writing like many other registers cannot be absolutely objective and depersonalized (Biber, 2006b; Hyland, 2004). Probably, many postgraduate students whose works were examined here had no problem at least understanding anticipatory *it* bundles given that they might 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. It seems, therefore, necessary for EAP practitioners to invest on their instruction for a more pedagogically focused treatment of anticipatory *it* bundles (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 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, and 2006).

Finally, despite one decade of research on lexical bundles, as Hyland (Hyland, 2008b) puts it, much still remains to be explored about this group of word combinations which can contribute to an almost overlooked dimension of genre analysis. In fact, identifying lexical bundles in other disciplines, registers, and genres, examining the formulaic status of these multi-word sequences (Biber and Barbieri, 2007), and probing more 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*: free text analysis software. Available on line: 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, and 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 Hasselgard, H. and Oksefjell, S., (Eds.), *Out of corpora: Studies in honor of Stig Johansson*, (pp. 181–189). Amsterdam: Rodopi.

Biber, D., Conrad, S. and 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/Main: Peter Lang.

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

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

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

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.

Crismore, A., and Farnsworth, R. (1990). Metadiscourse in popular and professional science discourse. In W. Nash (Ed.), *The writing scholar: Studies in academic discourse*. Newbury Park, CA: Sage.

Hasselgard, H., Johansson, S., and Lysvag, P. (1998). *English grammar*. Oslo: Universitetsforlaget.

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

Hunston, S., and Sinclair, J. (2000). A local grammar of evaluation. In S. Hunston, and G. Thompson (Eds.), *Evaluation in text* (pp. 74–101). Oxford: Oxford University Press.

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: Metadiscousre 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.

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

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

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

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

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

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

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). A comprehensive grammar of the English language. London: Longman.

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

Ruiying, Y, and Allison, D. (2004). Research articles in applied linguistics: structures from a functional perspective. *English for Specific Purposes*, 23 (3), 264-279.

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

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.