



Review Article

Cross-Linguistic and Cross-Disciplinary Investigation of Lexical Bundles in Academic Writing

Kashiha, H.* and Chan, S. H.

Department of English, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

ABSTRACT

The present paper reviews the use of lexical bundles in academic writing from two different viewpoints namely linguistic and discipline, directed at how academic writers belonging to different disciplines or linguistic backgrounds construct their discourses through lexical bundles. As cohesive devices, lexical bundles are an indispensable part of the text and play a crucial role in shaping propositions, evolving the text, guiding readers through the flow of information and gaining the writer's proffered meaning. By using lexical bundles, academic writers are able to attain naturalness in their writings and create a more reader-friendly approach to the unfolding text. Bearing the significance of lexical bundles in mind, this review paper aims to examine the effect of disciplinary variation and linguistic differences on the use of lexical bundles in academic writing. Most researchers believe that the frequency as well as the use of lexical bundles is different across disciplines and from one language to another language. Therefore, through a review of previous studies, there is a systematic investigation of evidence to support the above claims. Possible limitations of previous studies are discussed and some implications for further research are presented.

Keywords: Academic writing, linguistic variations, disciplinary variations, lexical bundles

ARTICLE INFO

Article history:

Received: 6 June 2013

Accepted: 12 September 2013

E-mail addresses:

hadi.kashiha@yahoo.com (Kashiha, H.),

shchan@upm.edu.my (Chan, S. H.)

* Corresponding author

INTRODUCTION

Language learners have always been interested in learning and using multi-word expressions to show that they have a good command of the target language. These multi-word fixed expressions are one of key elements of fluent linguistic production

and play an important role in the process of language learning (Hyland, 2008). They also attract many academic writers who use the language for specific or academic purposes. With the arrival of computer technology, the calculation and analysis of word-combinations has become much easier (Jablonkai, 2010). *The Longman Grammar of Spoken and Written English* (Biber, Conrad & Cortes, 1999) was the first corpus framework for studying and calculating the frequency of longer word combinations. The next step was to introduce lexical bundles as corpus investigation; they are, also known as clusters and chunks (Hyland, 2008) as a new term to differentiate these recurrent word chains from other types of multi-word expressions (e.g. idioms). Biber and Conrad (1999) define lexical bundles as “multi-word expressions which occur frequently and with accidental sequences of three or more words (e.g. *in the case of the, do you want me to*)” (p. 183). Bundles are frequently occurring series of unexpected word forms that do not have any special sequence pattern and appear to come together by chance. Nonetheless, they play an important role in understanding the meaning of specific contexts and contribute significantly to the coherence of the text (Hyland, 2008). In addition, they are manifested differently according to the nature of the text. Phrases like *what I want to say* or *as far as I know* would refer to group discussions and conversations while *as it was noted before* or *in accordance with* would belong to an academic corpus.

Recent studies (Biber *et al.*, 1999; Wray & Perkins, 2000; Biber, 2006; Cortes, 2002, 2004; Hyland, 2008) have witnessed the significance of lexical bundles as a major source of coherence in academic texts of different discourse communities. A majority of the studies conducted on the notion of lexical bundles have focused on structural and functional analysis of these recurring expressions (Hyland, 2008; Cortes, 2004; Strunkyt & Jurkūnait, 2008; Adel & Erman, 2012; Bal, 2010; Chen & Baker, 2010). Structural categorization of bundles was first introduced by Biber and his colleagues (1999). They found lexical bundles structurally complex, usually incomplete and not fixed and they classified them into 8 broad structural categories: *noun phrase + of, other noun phrases, Prepositional phrase + of, other prepositional phrases, passive + prep phrase fragment, anticipatory it + verb/adj, be + noun/adjectival phrase* and *others* (Biber *et al.*, 1999). Accompanying the identification of the forms, there was also the need to have other frameworks for analyzing the functions or meanings of lexical bundles in a text. Biber *et al.* (2004) identified three major discourse functions for lexical bundles: *stance bundles, discourse organizer bundles* and *referential bundles*. They define stance bundles as the “overt expression of an author’s or speaker’s attitudes, feelings, judgments, or commitment concerning the message”. They argue that discourse organizer bundles try to “indicate the general overview of the sentence”, while, referential bundles

“signalize and point out some important features of a unit to be important in a way” (pp. 386-388). Hyland (2008) proposed another functional taxonomy as a more refined version of Biber’s classification. Hyland’s classification identified three major discourse functions for lexical bundles and introduced sub-categories which specifically reflect the concerns of research writing. The main functions are: *research-oriented bundles*, *text-oriented bundles* and *participant-oriented bundles*. Research-oriented bundles “help writers to structure their activities and experiences of the real world”. Text-oriented bundles have to do with the cohesion of the text and its meaning in the form of logical arguments and “participant-oriented bundles focus on the writer or reader of the text” (p. 14).

A number of differences can be found studying the two frameworks regarding their realizations to functional categories. For example, Biber *et.al.* (2004) assign stance as a super-ordinate category while Hyland (2008) has grouped it into a category in which such bundles refer to the connection they can make to either the writer or reader. Another difference is, while Biber *et. al.* (2004) categorize *as a result of* and *on the basis of* as referential expressions stating the representation of the reality, Hyland (2008) takes them in the category of text organizers that establishes logical relations in the discourse. It seems that each framework has its own approach and reasons for the explanation and classification. The advantage of Biber’s taxonomy is that it can be used on a wide corpus and

can cover studies of lexical bundles in both spoken and written registers. On the other hand, Hyland’s classification is more focused on research writing and cannot specifically be applicable for studies on spoken discourse, and this is seen as a limitation. As Hyland (2008, p.13) himself argues, Biber *et al.* (2004) had introduced a “much broad corpus of both spoken and written registers including casual conversation, textbooks, course packs, service encounters, institutional texts, and so on,...” while he had concentrated more on the idea of lexical bundles in relation to research focused genres.

In the last 20 years, many researchers have also branched into the issue of disciplinary variation while analyzing lexical bundles (Cortes, 2004; Hyland, 2008; Strunkyt & Jurkūnait, 2008). Some other researchers have considered linguistic variation as a factor that influences the use of lexical bundles (Chen & Baker, 2010; Rafiee, Tavakoli & Amirian, 2011; Adel & Erman, 2012; Karabacak & Qin, 2013). Yet, in the existing studies, little research has been done on the roles of linguistic and disciplinary variations in the use of lexical bundles in academic writing. Therefore, the purpose of this review is to shed more light on the notion of lexical bundles in academic writing focusing on two different perspectives, linguistic and discipline. The aim is to gather evidence that can explain how native and nonnative writers from different disciplinary and cultural backgrounds are influenced in the use of lexical bundles in order to

generally contribute to the coherence in a text, create a realistic academic voice and attain naturalness in writing. This is particularly with reference for the authors in different fields and contexts to come up with disciplinary and linguistic specific bundles. Secondly, it is also concerned with the method that enables the investigation. Based on the non-quantitative and non-empirical nature of review papers, the methodology section will be discussed first and then followed by substantial review of related studies.

METHODOLOGY OF REVIEW

In meeting the aim of the study, there were some steps taken in finding and synthesizing the reviewed articles related to this study. First, different online websites were searched using key words such as lexical bundles, disciplinary variation, cultural background, academic writing, functional and structural classification of bundles. Then, reading through reference sections of the findings introduced some new related articles. The electronic databases used for the literature search for the related studies were Google Scholar and *Scencedirect*. The criterion behind choosing the two electronic databases is mainly their popularity among the scholars in searching for empirical published articles. *Scencedirect* is considered as one of the largest online databases including collections of journals with published academic articles. Google Scholar is also one of the mostly visited web search engines which encompass a variety of scholarly literature across many publishing

topics and disciplines. On the whole, the exploration included 35 published articles on lexical bundles that spanned the last 20 years. Out of them, six most related articles were selected to be reviewed. The reason for not selecting the remaining articles was because they were not straightforwardly or directly relevant to the focus of this study. Some of them were only disciplinary, focusing on one single discipline rather than being cross-disciplinary. A few others investigated the use of lexical bundles in only one language, rather than being cross-linguistic. However, they were studied and used as background literature. Articles in the cross-disciplinary section examined the use of bundles across a variety of fields including, biology, history, electrical engineering, applied linguistics, business studies, physics and a few others. In the same line, articles in the cross-linguistic section investigated the manifestation of bundles in different cultural contexts such as Swedish, Persian, Chinese, Turkish and English.

This paper employs the review method proposed by Creswell (1994), in that, the major purpose of a review is first to summarize the current state of knowledge on the basis of what has been reviewed so far and second to shed light on the important parts that have not been focused on in previous studies. The method of the review in this study started by dividing the selected articles into two sections: cross-disciplinary and cross-linguistic. Each section was assigned three most relevant articles to be reviewed. Each article was then reviewed on

the basis of its topic, corpus, model, method, and findings. Then next step was to draw attention to the positive and negative points and possible shortcomings in the survey. Finally, the last stage discussed the strengths, flaws, and limitations of the reviewed studies and put forwards some implications for future research. The description of the corpus is shown in Table 1.

REVIEW OF RELATED STUDIES ON LEXICAL BUNDLES

Review of related studies was done on the basis of two main categories: 1) Cross-disciplinary studies, and 2) Cross-linguistic studies. The most related studies in each category are reviewed and explained in detail.

Cross-disciplinary Studies

As mentioned, having control over lexical bundles or other multi-word expressions is one of the most important aspects in linguistic production that leads to fluency. These recurring expressions are also referred to as extended collocations (Cortes, 2004) that can help shape the meaning in context and build coherence in a text (Hyland, 2008). Therefore, they attract many English for Academic Purposes (EAP) scholars but only scant attention had been paid to their difference in use across disciplines.

Hyland (2008) addressed this issue in his study. He analyzed the frequency, structure and function of 4-word lexical bundles in research articles, doctoral dissertations and master's theses, across four disciplines namely electrical engineering and biology from the applied and pure sciences, and business studies and applied linguistics from the social sciences, forming a 3.5 million word corpus. The corpus of research articles included 120 published papers with 30 in the most important journals of each of the four disciplines. The PhD and master's corpora contained 20 texts in each discipline written by Cantonese L1 speakers at Hong Kong universities. The purpose was to calculate the frequency of lexical bundles to find out the disciplinary variation. The results revealed that electrical engineering articles have the greatest number of bundles, with 213 different bundles. Biology articles on the other hand, have the smallest range of lexical bundles, 131 different bundles. He also found that the other three disciplines did not seem to use the bundles found in the engineering texts leading to the conclusion that engineering writers appear to depend more on pre-fabricated structures compared to writers in other fields. To provide a firm reason for the phenomenon is difficult, but "speculatively it could be a consequence of the relatively abstract and graphical nature of technical communication" (Hyland,

TABLE 1
Description of the corpus

Perspectives	Cross-disciplinary	Cross-linguistic	Total
No. of reviewed articles	3	3	6

2008, p. 11). Both Biology and Electrical engineering appear to rely more on statistics and visual representations to support their arguments. Hyland (2008) concluded that writers of different fields make use of a variety of ways to “develop their arguments, establish their credibility and persuade their readers, with less than half of the top 50 bundles in each list occurring in any other list” (p. 19).

Addressing the question, as to what extent fixed expressions are exclusive to particular registers and also as to the difference between disciplinary professional and beginner writing in relation to the use of frequent word combinations, Cortes (2004) conducted a study on the use of the most frequent four-word lexical bundles (called *target bundles*) in the writing of university students of two disciplines, history and biology. Through a comparison of the published writings from history and biology journals, the results showed some disciplinary differences in the use of epistemic–impersonal/probable–possible stance markers in biology articles. To Cortes (2004), these bundles are used in a variety of ways to show “the effect of an affirmation, or to make an affirmation or argument more tentative” (p. 411) in the corpus of published biology writings. History writers, on the other hand, did not use these bundles frequently. The findings proposed that in order for the history writers to communicate the same functions, they prefer to use much ‘simpler’ vocabularies which are probably shorter than those of the lexical bundles investigated.

The frequency of occurrence, and structural and functional analysis of lexical bundles are affected by the disciplinary variation. To affirm this contention, Strunkyt and Jurkūnait (2008) investigated the use of lexical bundles in research articles in two disciplines: humanities, represented by research articles in linguistics and educology, and natural sciences, represented by research articles in physics and astronomy. The research analyzed and compared structural and functional types of lexical bundles in 40 research articles (20 from each discipline). The findings revealed that writers in the humanities used a larger number of lexical bundles compared to those in natural sciences. They also employed more structural types of bundles which indicated that the language of humanities is more mixed and varied than the language of the research articles in natural sciences. As for the functional analysis, research articles in natural sciences resorted to more text organizing bundles in order to set up a communication with the reader. Stance and referential bundles were more frequent in humanities than in natural sciences. However, the referential bundles included a higher level of transactional information, such as exemplification, relationship between topics, evaluations, qualifications in both humanities and natural sciences.

Cross-linguistic studies

In general, investigation and research in the area of second language acquisition has always been integrated by a comparison

of different or similar aspects of language use among native and non-native users. Comparing native and non-native learners in terms of using combinations of words that carry out specific meaning in a text is also of a great interest in the area of academic writing. Adel and Erman (2012) compared L1 speakers of Swedish with British native-speakers in terms of using lexical bundles in advanced academic writing of undergraduate university students of linguistics. The investigation involved both quantitative and qualitative analysis of four-word lexical bundles in the Swedish local corpus; Stockholm University Student English Corpus (SUSEC) which included 325 essays and over one million words. The results revealed that native speakers in general, used more lexical bundles than the non-natives, with 130 different bundles as compared with 60. In terms of types of bundles, they also found some similarities and differences between native and non-native learners. Both native and non-native speakers used bundles such as *the result from the* and *can be used to*, while bundles like, *as a result of*, *at the beginning of*, *the aim of this* and *to look at the* were only found in the writing of native speakers.

Teaching both L1 and L2 apprentice writers to use lexical bundles in their writing could to a great extent play a crucial role in competent English academic writing (Karabacak & Qin, 2013). With this concern and on the basis of their hypothesis, Karabacak and Qin (2013) conducted a comparative study and looked

at cross-cultural variations concerning the use of lexical bundles in writings of Turkish (EFL), Chinese (EFL) and American (native speakers of English) university students. To this end, 29,532 articles were collected from the *New York Times* and *SF Gate* newspapers as a corpus in order to extract target bundles out of one million words. Then the students' papers were analyzed to see the extent they had used such bundles in comparison. Results showed that there were some considerable differences between three groups of writers concerning the frequency of the types of bundles. American papers had the largest number of 5-word lexical bundles, while the number used is the lowest for the Chinese writers. Regarding target and reference bundles, the result showed that American and Turkish students employed these bundles quite similarly. Again, Chinese students used the lowest number of target bundles. They concluded that natural acquisition of some lexical bundles through simple exposure may be difficult even for advanced English learners. Therefore, they need to be taught explicitly to hasten their acquisition process.

Analyzing the frequency of lexical bundles in journalistic writing would be a great help to both native and non-native academic authors and at the same time, guide readers to follow the information in the text and thus get a better understanding of the point of the writer. With regards to this idea, Rafiee, Tavakoli and Amirian (2011) analyzed lexical bundles to investigate the frequency and distribution of structural

types of lexical bundles between English newspapers published in native and non-native contexts. To this end, a 3 million word corpus of four English newspapers, two published in Iran (the Iran Daily and Tehran Times) and the two others published in England (The Times and Independent) from 1/1/2009 to 15/1/2010 were used. Their first finding was in line with the findings of previous studies conducted by Biber *et. al.* (1999) and a few others, in that most of the bundles in their corpus were phrasal rather than clausal bundles. In fact, Biber *et. al.* (1999) made an important note that the register of academic writing is distinguished from other modes such as conversation, in that it includes more phrasal rather than clausal bundles.

The investigation showed marked similarities and differences between two groups of journalists in their use of lexical bundles. Overall, Iranian journalists used more lexical bundles compared with native speaker journalists. To Rafiee, Tavakoli and Amirian (2011), the reason for such a higher tendency to use lexical bundles among non-native writers “could be due to the fact that they have already been exposed to such word sequences several times in their prior readings of various kinds of English literature” and thus the ability to use these lexical bundles is well inculcated and influenced by this EFL experience (p.13). There was a remarkable inclination to use the structural types of lexical bundles among both Iranian and English journalists. The analysis based on structures of bundles showed that both Iranian and English

journalists used the same structures in constructing lexical bundles, probably showing the high degree of competence needed in journalistic writing.

SUMMARY OF FINDINGS

After reviewing the above related studies in regards to their methodology, total data, obtained results, explanation of ideas, discussion of findings and, of course, recommendation and implication, the researchers found that each of the above studies has its own strong and weak points. Both the research gap and the purpose of the study were fulfilled and described in all the reviewed studies. From the overlaps, frequency and corpus-based investigation of lexical bundles, many researchers established that such word combinations are very text-dependent in academic writing. On the other hand, the nature of lexical bundle use in oral discourse has yet to be firmly established, and this leads to growing interest among researchers to study lexical bundles in spoken academic discourse. The supporters of studies on spoken discourse argue that the distribution of lexical bundles could be dependent on the genre and the mode. As Biber and Barbieri (2007) pointed out, “the extent to which a speaker or writer relies on lexical bundles is strongly influenced by their communicative purposes, in addition to general spoken/written differences. The explanation for the infrequent use of lexical bundles in the academic written registers (textbooks and academic prose) apparently lies in the restricted communicative goals of

those registers—focused on informational communication—rather than the written mode *per se*”. (p. 273)

Concerning the discussion of findings, among the studies, Rafiee, Tavakoli and Amirian (2011) and Karabacak and Qin’s (2013) works suffered from the fact that they did not provide explicit details on the examples of their corpus while discussing their findings. In contrast, however, Hyland (2008), Cortes (2004), and Adel and Erman (2012) supported their results by giving ample examples and explanations exploited from their analyzed data and discussed them in a lucid and comprehensive way. From the critique, it is also noted that Biber *et. al.*’s (1999, 2004) framework stands out as the dominant one in analyzing the structural types of lexical bundles. Almost all the studies have applied the framework proposed by Biber *et. al.* to analyze the frequency and distribution of structural types of lexical bundles but concerning functional analysis, Hyland (2008) used his own taxonomy which was extended from Biber’s (Biber 2006, Biber *et. al.*, 2004). Studies conducted by Cortes (2004) and Strunkyt and Jurkūnait (2008) also made use of functional categorization designed by Biber *et. al.* (1999, 2004). Studies done by Rafiee, Tavakoli and Amirian (2011) and Adel and Erman (2012) compared and contrasted their results with those of previous researchers but other studies did not compare nor contrast their findings with earlier studies. Points of comparison and contrast would have helped in establishing greater credence to any findings.

In terms of the amount of data used in the related studies, findings showed that the corpora used in most of the studies such as those by Hyland (2008), Cortes (2004), Rafiee, Tavakoli and Amirian (2011), and Adel and Erman (2012), were sufficiently large and could help in the generalization and representation of the results. Biber (2006) believes that a corpus must be large enough to adequately represent the occurrence of the features being studied. The study by Strunkyt and Jurkūnait (2008) was restricted in the number of disciplines studied leading to the use of a small corpus, thus affecting the reliability of the claims made. Another study by Karabacak and Qin (2013) also lacked in number of adequate texts (with only two American newspapers) as a reference corpus needed for representing written English in general and the results in specific. For the genre in question, therefore, a larger corpus would be needed to discover lexical bundles within each sub-corpus which could serve as a point of departure for further research.

In comparing the findings to point out similarities or differences, some studies such as those by Adel and Erman (2012) and Rafiee, Tavakoli and Amirian (2011) were quite lucid, comparing their findings to those of previous studies or were able to evoke similarities and differences between them, but such comparison could not be seen in those by Karabacak and Qin (2013) and Strunkyt and Jurkūnait (2008) whose findings lacked comparison and related elaboration. Cortes (2004) compared the writing of university students of two

disciplines, history and biology, with those of published writing from history and biology journals in terms of using lexical bundles but she discussed the disciplinary differences in two separate sections. Instead, she could have compared the use of target bundles in history and biology in one specific section since she was conducting a cross-disciplinary study.

Regarding recommendation and implication, Cortes (2004) and Hyland (2008) did not provide any recommendations for further research but identified some implications for their studies. Hyland (2008) addressed EAP practitioners and course designers to increase the learning of these multi-word units in their EAP courses through greater exposure and

use of activities like item matching and identification, and productive tasks such as consciousness raising which require learners to produce the items in their extended writing. Adel and Erman (2012), on the other hand, directed researchers towards several issues in design and methodology but did not talk about the implications of their findings. Karabacak and Qin (2013) did not take into account these two factors. They neither direct researchers to make further studies nor provide any implications. A section on recommendation and implication would be a great help to novice writers and could direct them to future research. The description of the strengths and weaknesses of the reviewed studies is illustrated in Table 2 and Table 3.

TABLE 2
Description of the reviewed cross-disciplinary studies

Study	Author(s)/ Year of publication	Strengths	Weaknesses
1. As can be seen: Lexical bundles and disciplinary variation	Hyland (2008)	1. Using ample examples from the data to support the results. 2. Data was sufficiently large.	1. No recommendation for further research
2. Lexical bundles in published and student disciplinary writing: examples from history and biology	Cortes (2004)	1. Using ample examples from the data to support the results. 2. The data was sufficiently large. 3. Compared the findings to those of previous studies	1. No recommendation for further research
3. Written Academic Discourse: Lexical bundles in humanities and natural sciences	Strunkyt and Jurkūnait (2008)	-----	1. Suffering from inadequate corpus size. 2. Findings lacked comparison and related elaboration. 3. No implication and recommendation for further research.

CONCLUSION

This paper aimed to look at the effects of disciplinary variation and linguistic differences on the use of lexical bundles in academic writing. In general, all the above mentioned cross-disciplinary and cross-linguistic studies have been conducted using different languages, context, frameworks, methodologies, analyses and points of view. Many researchers agree that lexical bundles are one of the most important aspects of fluent linguistic production and thus play an important role in the process of language learning. Lexical bundles also help many

academic writers to use the language for specific or academic purposes. Scholars also attest to the fact that the manifestation of lexical bundles is restricted and conditioned by the discipline's public goals, norms, and conventions and writer's native language writing culture. It is suggested that more work needs to be done on the notion of lexical bundles cross-disciplinarily and cross-linguistically, through other genres of academic written discourse. This includes a variety of published written materials such as textbooks, theses and dissertations in different disciplines. Of some significance

TABLE 3
Description of the reviewed cross-linguistic studies

Study	Author(s)/ Year of publication	Strengths	Weaknesses
1. Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach	Adel and Erman (2012)	1. Using ample examples from the data to support the results. 2. Compared the findings to those of previous studies 3. Data was sufficiently large	1. The study did not provide any implication.
2. Comparison of lexical bundles used by Turkish, Chinese, and American university students	Karabacak and Qin (2013)	1. Significant findings regarding the use of lexical bundles in the three languages.	1. No implication and recommendation for further research. 2. The study did not provide explicit details from the examples to support the findings. 3. It lacked in the number of texts used
3. Structural analysis of lexical bundles across two types of English newspapers edited by native and non-native speakers	Rafiee, Tavakoli and Amirian (2011)	1. Compared the findings to those of previous studies 2. Data was sufficiently large	1. The study did not provide explicit details from the examples to support the findings.

is the notion that lexical bundles should be extended in investigation with reference to its use in different genres and modes.

As an extension, there is the possibility of giving more details to building a framework that deals with spoken data thus widening the scope on the investigation of lexical bundles for such different text purposes. Findings reveal that having a good command of lexical bundles can highly ensure the degree to which members of different disciplinary communities acquire the discipline-specific knowledge as these multi-word expressions contribute to the coherence in a text, create a realistic academic voice and attain naturalness in the language. Findings of this study are beneficial to experts and academic writers from different disciplinary or cultural backgrounds in using formulaic expressions in their specific disciplines or contexts. In the same line, course designers can also realize the significance of multi-word sequences to include them in their disciplinary or linguistic specific syllabus.

REFERENCES

- Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81–92
- Bal, B. (2010). Analysis of four-word lexical bundles in published research articles written by Turkish scholars. Unpublished MA Thesis. Georgia State University.
- Biber, D. (2006). *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam: Benjamin.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263–286.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of Corpora: Studies in Honour of Stig Johansson* (pp. 181–190). Amsterdam: Rodopi.
- Biber, D., Conrad, S., & Cortes, V. (2004). *If you look at...: Lexical bundles in university teaching and textbooks*. *Applied Linguistic*, 25(3), 371–405.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Chen, Y., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language learning and technology*, 14(2), 30–49.
- Cortes, V. (2002). Lexical bundles in published and student academic writing in history and biology. *Unpublished Ph. D 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.
- Creswell, J. W. (1994). *Research design*. Thousand Oaks, CA: Sage publications.
- Haswell, R. (1991). *Gaining Ground in College Writing: Tales of Development and Interpretation*. Dallas: Southern Methodist University Press
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, 4–21
- Jablonkai, R. (2010). English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes*, 29 (4), 253–267
- Karabacak, E., & Qin, B. (2013). Comparison of lexical bundles used by Turkish, Chinese, and

- American university students. *Procedia. Social and Behavioral Sciences*. 70, 622-628
- Rafiee, M., Tavakoli, M., & Amirian, Z.(2011). Structural analysis of lexical bundles across two types of English newspapers edited by native and non-native speakers. *Modern Journal of Applied Linguistics*. 3(2), 218-236.
- Strunkyt, G., & Jurkūnait, E.(2008). *Written academic discourse: Lexical bundles in humanities and natural sciences*. (Unpublished master's thesis). Vilnius University.
- Wray, A., & Perkins, M. (2000). The functions of formulaic language. *Language and Communication*, 20, 1–28

APPENDIX

Cross-disciplinary Studies

Study	Sample	Purpose	Type of data	Main findings
Hyland (2008)	120 published papers- 20 dissertations	To analyze the frequency, form and function of four-word lexical bundles to find out the disciplinary variations.	Academic research articles, PhD and Master dissertation written by authors in four disciplines.	Disciplinary variations in the use and discourse functions of lexical bundles among the disciplines.
Cortes (2004)	277 university papers	Aimed to investigate the difference between disciplinary professional and beginner writing in relation to the use of frequent word combinations.	University papers written by university students belong to two different disciplines.	Differences in the use of epistemic- impersonal/probable-possible stance markers in the two disciplines. Unlike biology, these bundles were not frequently found in history.
Strunkyt and Jurkūnait (2008)	40 research articles	Attempted to investigate the frequency of occurrence, and structural and functional analysis of lexical bundles in two disciplines.	Experimental research articles written by authors in each fields	Cross-disciplinary differences regarding the use of bundles in the two disciplines. Natural sciences resorted to more text organizing bundles. Stance and referential bundles were more frequent in humanities than in natural sciences.

Cross-linguistic Studies

Study	Sample	Purpose	Type of data	Main findings
Adel and Erman (2012)	325 essays	Compared the use of lexical bundles in advanced academic writing of L1 speakers of Swedish and British native-speakers undergraduate university students to scrutinize the possible cultural variations.	Advanced academic papers in the form of essays	Some similarities and differences were found across two languages. Native speakers in general, used more lexical bundles than the non-natives. Both writers used bundles such as the result from the and can be used to, while bundles like, as a result of, the aim of this were only frequent in the writing of native speakers.
Karabacak and Qin (2013)	51 academic papers	Explored the cross-cultural differences between academic papers written by Turkish, Chinese and American university students regarding lexical bundle use.	University students' papers to be compared with 29,532 articles collected from the New York Times and SF Gate newspapers	They found that American papers had the largest number of 5-word lexical bundles, while the number used is the lowest among the Chinese writers. American and Turkish students employed target and reference bundles quite similarly.
Rafiee, Tavakoli and Amirian (2011)	4 English newspapers	Aimed to supply a cross-cultural investigation on the use and structure of lexical bundles in English newspapers published in native and non-native contexts.	2 English newspapers published in Iran and another 2 published in England.	They found that both Iranian and English journalists use the same structural groups in constructing lexical bundles, probably showing the high degree of competence needed in journalistic writing. Most of the bundles in their corpus were phrasal rather than clausal.

