Iranian Journal of Applied Language Studies
Vol (10), No 1, 2018, pp. 94-124
http://ijals.usb.ac.ir

A Corpus-Based Study of Lexical Bundles Discussion Section of Medical Research Articles

Zahra Sadat Jalali
(Corresponding author), Ph.D Candidate, Alzahra University, Iran
z.s.jalali25@gmail.com

Mohammadraouf Moini
Assistant Professor, University of Kashan, Iran
raoufmoini@yahoo.com

Abstract

There has been increasing interest in utilizing corpora in linguistic research and pedagogy in recent years. Rhetorical organization of different sections of research articles may appear similar in various disciplines, but close examination may show subtle differences nonetheless. One of the features that have been at the center of attention especially in recent years is the idiomaticity of a discourse which can be examined from the perspective of multi-word units captured by the automatic retrieval of lexical bundles. This study takes a corpus-based approach for the identification of lexical bundles. A corpus of 801,894 words from 790 articles was collected. In order to fulfill the purposes of the present study, ABBYY FineReader 10 professional edition, Total Assistant, Antconc 3.2.3, and WordSmith Tools 5 were used to identify lexical bundles. Then these bundles were classified structurally and functionally based on the presented taxonomies in the literature. The results of the current study indicated that the writers of medical research articles mostly rely on text-oriented bundles in the discussion section of research articles to establish academic discourse.

Keywords: Lexical Bundles, Academic Discourse, Discussion, Research Articles, Corpus

Received: June 2016; Accepted: August 2017

1. Introduction

In recent years much more attention has been paid to corpus linguistics. Recent studies carried out in corpus linguistics have pointed to EAP (English for Academic Purposes) - specific phraseology and multi-word combinations. Most of these studies have used corpus data in order to analyze formulaic expressions used in different registers. For instance, some of them have focused on spoken vs written registers or some have had a look at academic vs. non-academic registers. For example, in series of studies carried out on lexical bundles, Biber and colleagues (Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2003, 2004; Biber, Johanson, Leech, Conrad & Finegan, 1999) found that conversation and academic prose present distinctive distribution patterns of lexical bundles and that most of the bundles are clausal in conversation and phrasal in written academic discourse. Other studies have focused on expert and non-expert writing. For example, Cortes (2004) compared lexical bundles in published and student academic writing in history and biology. The findings indicated that lexical bundles were frequently used in published writing but that students' use of these expressions was rare and that many lexical bundles are discipline bound. Specialized academic corpora may concentrate on just one genre, or a wide variety of genres or even they may concentrate on one discipline or many. However, most of them tend to be made up of professionally written texts. The importance of genre knowledge which can help language learners to master academic, professional and educational discourse has been widely acknowledged for decades. Rapid development of science and technology has forced researchers to participate actively in the international academic discourse community. Research articles (hereafter referred to as RAs) as the central knowledge production are a much-studied genre and have received extensive attention in genre analysis.

Chang and Kuo (2011) discuss that some of the studies carried out in EAP provide fruitful results with respect to the genre of RAs at two levels of macrostructure and linguistic features. The former level takes the genre-analysis approach to the rhetorical functions and organizational patterns of RA sections. A large number of studies have been carried out regarding the moves and steps included in different sections. For example, the study that was carried on the introduction section of RAs (Swales, 1990), another one on the result section (Brett, 1994; in sociology RAs) and the other one on the discussion section of research articles in three disciplines of sociology, political science, history RAs (Holmes, 1997) as well as economics, business and financial articles (Lindeberg, 1994). The latter level focuses on the analysis of linguistic features such as hedging, reporting verbs, modals, personal pronouns and meta-discourse (Butler, 1990; Hyland, 1998; Kuo, 1999; Tarone, Dwyer, Gillette, & Icke, 1998; Thompson & Ye, 1991). Combinations of words which fulfills specific functions and that are called upon automatically by native speakers have come to be known under the term of formulaic language (Schmitt & Carter, 2004). These are particular expressions being specific to an academic community. So, individuals who have not been members of such communities may not hit on the most appropriate expressions, and their production may, therefore, seem not quite right to insiders (Kjellmer, 1990). Different terms have been used to refer to word sequences and linguists do not concur with the defining characteristics which differentiate one type of word combinations from another. Reviewing the related literature showed that there are different terms that are used to refer to multi-word combinations. These terms are *recurrent word combinations* (Altenberg, 1998; De Cock, 1998), *phrasicon* (De Cock, Granger, Leech, & McEnery, 1998), *clusters* (Hyland, 2008a; Schmitt, Grandage & Adolphs, 2004),

n-grams (Stubbs, 2007a, 2007b) and *lexical bundles* (Biber & Barbieri, 2007; Cortes, 2002).

Series of studies were conducted on lexical bundles. For instance, Hyland (2008a) identified lexical bundles in various genres of research articles, master thesis, and doctoral dissertations in four disciplines of electrical engineering, business studies, applied linguistics, and biology. In another study by Alipour, Jalilifar and Zarea (2013) lexical bundles were identified and compared in the genre of research articles in three disciplines of physics, computer engineering, and applied linguistics. The results of this study showed significant differences between structures and functions of bundles in the mentioned disciplines and that the writers of these disciplines rely on different norms in order to communicate appropriately with the members of their own communities.

Hong and Hua (2018) identified lexical bundles in the corpus of journal articles in the field of international business management (IBM) and found that lexical bundles are discipline specific. Other studies have focused on the similarities and differences of lexical bundles across different genres within one discipline (Cortes, 2004; Hyland, 2008a; Jalali, 2009; Breeze, 2013). For instance, Breeze (2013) investigated lexical bundles in four legal genres: academic law, case law, legislation and documents. In this study, major differences were found between types of bundles and their functions in various corpora. In another study, Łukasz Grabowski (2015) investigated lexical bundles across samples of patient information leaflets, summaries of product characteristics, clinical trial protocols and chapters from academic textbooks on pharmacology and found salient links between situational, linguistic and functional features of the four pharmaceutical registers under scrutiny and showed that patterns of language use differ considerably due to topic- and function-related differences between the text types, despite their dealing with a

similar theme, namely with medicines or medicinal products. Some studies have investigated articles and different sections of articles across various disciplines (Biber & Finegan, 1994; Martinez, 2003; Valipoor, 2010; Parvizi, 2011). For instance, Cortes (2013) studied lexical bundles in the introduction section of research articles in different disciplines, and the most frequent structural correlates and functions were found.

For the last few decades, some of the studies have adopted a genre-based approach in order to investigate the discussion sections of RAs in disciplinary areas such as social sciences (Lewin, 2001), Biomedicine (Dubois, 1997), chemical engineering (Peng, 1987) and dentistry (Basterkmen, 2012). Dudley-Evans (1997) focused on the discussion section of RAs since he believed that students have the greatest difficulty with this section in the academic discourse and that this section has received less attention than the introduction section. Related literature and studies which were carried out on the discussion sections of RAs in different disciplines showed that few studies have focused on the lexical bundles in the discussion section of RAs. Since EAP research has indicated that phraseology or formulaic expressions in academic written discourse are often problematic for non-native or novice writers (Cowie, 1998; Gledhill, 2000; Jalali, 2014; Güngör & Uysal, 2016; Safarzadeh, Monfared & Sarfeju, 2013; Pan, Reppen & Biber, 2016; Bychkovska & Lee, 2017; Shin, Cortes & Yoo, 2018), the current study presents such lexical phrases that can be linked to communicative purposes of a section through the analysis of lexical bundles in the discussion section of medical research articles. Thus, this study poses the following three specific research questions:

1. What are the most frequent lexical bundles used in the discussion section of Medical RAs?

2. What are the different forms of lexical bundles used in discussion section of MRAs?
3. What roles do lexical bundles play in the discussion section of MRAs?

2. Method

To collect the required RAs for establishing the corpus (Corpus of Discussion section of Medical Research Articles, referred to as CODMRA hereafter in this study), the Science Direct Online (SDO) was used. All the written medical research articles gathered in the corpus were downloaded from this authentic database, i.e., SDO which is the world's largest electronic collection of science, technology and medicine. Over 1800 journals related to 24 disciplines ranging from natural sciences to social sciences are included in SDO, and consequently, it can be considered as the most representative and authoritative database.

In the discipline of Medicine and Dentistry of SDO, there are 33 subject areas. Following Wang and Ge (2008), in this study, almost all areas of Medical Science were included. All journals in the 33 subject areas published during 2009-2011 were used for the establishment of the corpus. In each year, two issues of each volume were randomly selected. Totally, about 24 articles in each 33 subject areas were selected while each article on average included about 3000 words. As a conclusion, 790articles were obtained in order to produce the corpus of 801,894 words of the discussion sections (i.e., all sections of articles were deleted except the discussion ones).

The frequency of 20 times per million words with the requirement that this rate of occurrence is realized in at least five different texts was considered as criteria in the present study although the cut-off frequency was determined according to the purpose of the study. However, some studies have used a cut-off frequency of 10 per million words (Biber, Johansson, Leech, Conrad &

Finegan, 1999). Biber (2006), Cortes (2004), Hyland (2008a, 2008b), Jalali (2009), Valipoor (2010), Parvizi (2011) considered a cut-off frequency of 20 times per million words, while Biber and Barbieri (2007) raised the idea of 40 occurrences per million words in the study of spoken and written university language. Identification of 4-word lexical bundles was at the center of attention in the CODMRA because 4-word lexical bundles were far more common than 5-word strings and offered a clearer range of structures and functions than 3-word bundles (Hyland, 2008a). Furthermore, many 4-word strings hold 3-word bundles in their structure (Cortes, 2004). After data was collected, different computer software products were used in order to identify the lexical bundles. The following section introduces these computer programs.

2.1. Computer Programs and Software

The software products utilized in the current study were: ABBYY FineReader 10 professional edition, Total Assistant, Antconc 3.2.3 (Anthony, 2007) and WordSmith Tools 5 (Scott, 2008). ABBYY FineReader is an intelligent document processing software which is used to convert scanned documents, PDF files and documents, and image files into an editable format. Regarding the purpose of the current study, all collected RAs which were in a PDF format were subjected to ABBYY FineReader 10 to be converted to plain text (txt.) format files. Using ABBYY FineReader enabled us to produce plain texts which could be uploaded to Antconc.

Next, the Total Assistant software was used for counting the words and characters of the CODMRA. Then, txt. format files were fed into Antconc software, and a concordance tool of this software was used and the cluster size, a 4-word string (for min and max size), was set. Then different keywords or search terms such as articles, “to be” verb, modals, prepositions, demonstrative

adjectives were typed in a specific part of this software. Also, a cut-off frequency of 20 in one million words was considered as a criterion, and consequently the minimum cluster frequency of 16 for a corpus of 801,894 words was given to Antconc software. Provided with these features, Antconc displayed clusters of words that surrounded a search term and ordered them alphabetically or by frequency. Like Antconc, another software program such as WordSmith tools 5 (Scott, 2008) was used to extract and identify lexical bundles in different texts. The difference between these two programs is that the WordSmith has the additional advantage of showing the number of texts in which lexical bundles happen.

The next stage was the structural and functional classification of lexical bundles. The former was based on Biber et al.'s (1999) structural taxonomy, and for the latter Hyland's (2008a) functional taxonomy was used. This taxonomy is based on academic registers. Since the focus of this study was on academic register and a specific section of research articles, this kind of taxonomy seemed to be more useful in the functional classification of bundles. The Table1 presents these major functions and their sub-categories of Hyland's taxonomy (2008a).

Table 1. *Functional Taxonomy of Lexical Bundles by Hyland (2008)*

Major functions	Sub-categories	Examples
Research-oriented: Help writers to structure their activities and experiences of the real world.	**Location**- indicating time and place	*In the present study, at the end of*
	Procedure-indicating methodology or purpose of research	*The purpose of this, was used as a*
	Quantification- describing the amount or number	*Is one of the, one of the most*
	Description- detailing qualities or properties of material	*in the control group, the size of the*
	Topic- related to the field of research	*in the United States*
Text-oriented: These clusters are concerned with the organization of the text and the meaning of its elements as a message or argument.	**Transition signals** – establishing additive or contrastive links between elements	*on the other hand, as well as the*
	Resultative signals – mark inferential or causative relations between elements	*the results of the, been shown to be*
	Structuring signals – text-reflexive markers which organize stretches of discourse or direct reader elsewhere in text	*as shown in fig, are shown in table*
	Framing signals – situate arguments by specifying limiting conditions	*on the other and, in the presence of*
Participant-oriented: These are focused on the writer or reader of the text.	**Stance features** – convey the writer's attitudes and evaluations	*were more likely to, it is possible that*
	Engagement features- address readers directly	*it should be noted, is important to note*

3. Results

Based on the size of the corpus, the frequency of 16 was set for the identification of bundles. Results of the software showed 94 bundles in the discussion section of MRAs which are presented in Table 2 with their overall frequency (FRQ) in the corpus. Finally, the bundles of this section were classified structurally and functionally using the taxonomies mentioned previously.

Table 2.*Lexical Bundles in Discussion Section*

	Lexical bundles	FRQ	No of texts		Lexical bundles	FRQ	No of texts
1	this is the first	69	66	**48**	in agreement with the	21	20
2	In our study, the	63	58	**49**	is in line with	21	18
3	the present study we	57	50	**50**	may be explained by	21	19
4	In this study, the	54	46	**51**	TNF a and IL	21	9
5	In our study, we	48	36	**52**	can be explained by	20	19
6	in the pathogenesis of	42	31	**53**	explained by the fact	20	18
7	studies are needed to	42	39	**54**	research is needed to	20	19
8	It is known that	41	37	**55**	the presence of a	20	18
9	knowledge, this is the	40	40	**56**	the small sample size	20	20
10	we found that the	38	33	**57**	the validity of the	20	16
11	It is likely that	37	36	**58**	we have shown that	20	19
12	These findings suggest that	37	34	**59**	with the findings of	20	17
13	It should be noted	36	32	**60**	are consistent with the	19	17
14	To our knowledge, this	36	35	**61**	be noted that the	19	17
15	is possible that the	35	31	**62**	be one of the	19	16
16	be related to the	34	28	**63**	consistent with previous studies	19	17
17	The results of our	33	32	**64**	due to the fact	19	19
18	results of the present	32	26	**65**	is the first to	19	18
19	should be noted that	31	28	**66**	this study is the	19	19
20	Our results suggest	30	26	**67**	we were unable to	19	17
21	This is consistent with	29	26	**68**	are similar to those	18	17
22	to the fact that	29	29	**69**	et al. found that	18	17
23	et al. reported that	28	21	**70**	in this study is	18	18
24	Further studies are	28	25	**71**	is in agreement with	18	15
25	study, we found that	28	21	**72**	patients in our study	18	17
26	important to note that	27	24	**73**	results suggest that the	18	17
27	in this study was	27	26	**74**	the presence of the	18	15
28	results of our study	27	26	**75**	to be due to	18	18
29	is important to note	26	22	**76**	a decrease in the	17	12
30	it is difficult to	26	25	**77**	be more likely to	17	15
31	the first study to	26	26	**78**	be used as a	17	16
32	Our results showed	24	23	**79**	been reported in the	17	16
33	patients with heart failure	24	7	**80**	finding is consistent with	17	15

	Lexical bundles	FRQ	No of texts		Lexical bundles	FRQ	No of texts
34	study is the first	24	23	**81**	present study showed	17	15
35	that the presence of	24	22	**82**	that the use of	17	15
36	be attributed to the	23	20	**83**	the duration of the	17	17
37	could be explained by	23	21	**84**	to be involved in	17	15
38	is also possible that	23	22	**85**	were not able to	17	15
39	It is also possible	23	22	**86**	which is consistent with	17	16
40	we were able to	23	23	**87**	and an increase in	16	14
41	findings of this study	22	19	**88**	are in line with	16	14
42	is in accordance with	22	19	**89**	are likely to be	16	16
43	reported in the	22	19	**90**	be responsible for the	16	16
44	results are consistent with	22	22	**91**	findings are consistent with	16	14
45	this study is that	22	20	**92**	is due to the	16	16
46	at least in part	21	18	**93**	may contribute to the	16	16
47	have shown that the	21	21	**94**	to an increase in	16	15

Table 2 presents 94 different lexical bundles in the CODMRA. When the overall frequency of bundles is divided by the number of words in the corpus, the outcome shows the percentage of bundles in the corpus (Jalali, 2009). As a consequence, it can be said that just 0.3% of the whole corpus of discussion section is formed by lexical bundles. As it is shown in Table 2, the most frequent lexical bundle is *this is the first* with the frequency of 69 in the corpus which is four times more than the cut-off frequency of 16. This bundle is used as a kind of study-based lexical bundle. The table also shows that *in our study the, the present study we, in this study the, In our study, we* are the most frequent lexical bundles in the CODMRA. On the other hand, there are 8 bundles with the least frequency of 16 which have occurred in different number of texts. In fact, the number of texts in which these lexical bundles have happened is very few, as presented in Table 2. These bundles with the frequency of 16 have occurred at least in 14 texts which is five times less than the number of texts in which most

frequent lexical bundles have occurred. Although these eight bundles with the same frequency show different distribution order, it can be claimed that the fewer the number of texts in which the bundle has occurred, the more frequent the bundle.

3.1. Structural Classification of Bundles in Discussion Section

Table 3 shows the structural classification of 94 identified lexical bundles in the corpus of 801,894 words in discussion section of MRAs. In this table, the number, the overall frequency and the percentage of bundles being categorized and encompassed under each main category and its subcategories are presented.

Table 3.*Structural Classification of Bundles in CODMRA*

Category	No of bundles	Overall FRQ	Percentage%
Noun phrase+of	10	230	9.40

Category	No of bundles	Overall FRQ	Percentage%
Noun phrase+oher post-modifier fragment	4	75	3.05
Noun phrase+prepositional phrase fragment	4	75	
Prepositional phrase+of	2	62	2.52
Other prepositional phrase	6	223	9.09
Anticipatory it+verb/adjective phrase	5	163	6.64
Anticipatory it+adjective phrase	3	86	
Anticipatory it+verb phrase	2	77	
Copula be+noun/adjective phrase	9	191	7.8
Copula be+noun phrase	3	54	
Copula be+adjective phrase	6	137	
Verb phrase+that-clause fragment	14	351	14.30
Verb phrase+that-clause	3	71	
Noun+verbphrase+that-clause	11	280	
Verb/adjective+to-clause fragment	6	125	5.09
Predicative adjective+to-clause	2	33	
Passive verb phrase+to-clause	2	57	
To-clause	2	35	
Passive verb+prepositional phrase fragment	7	140	5.70
Pronoun/noun phrase+be+….	3	117	4.76
This+be+…	3	117	4.76
Other expressions	26	659	26.86
Total	94	2453	100

Note: Those categories that are in bold are main categories, and the others are subcategories

Based on what is presented in Table 3, it can be claimed that clausal bundles are more frequent than phrasal bundles as they form 49.05% of the whole corpus of the discussion section. Phrasal bundles in comparison with clausal bundles make up 24.06% of the whole bundles. Among those which can be considered as phrasal bundles, noun phrase+of bundles indicate the highest rank of about 9.40%, although noun phrases (with and without 'of')constituted about 12.45%

of phrasal bundles generally. Another group which can be subsumed under phrasal bundles is prepositional phrases which account for about 12% of the bundles.

Based on the results, clausal bundles are used twice the rate of phrasal ones by the medical experts and make up about 49.05% of the whole corpus. Among those categories which can be subsumed under clausal bundles, verb-phrase+that-clause fragments are the most frequently used lexical bundles which make up about 14.30% of the whole bundles. This kind of bundles incorporate that-clause and are of two main types: 1) those bundles that include verb-phrase and 2) those which are comprised of only that-clause. In this study no bundle could be subsumed under that-clause category, but there were three bundles which could be encompassed under the category of the verb phrase+that-clause, e.g., *have shown that the, be noted that the* and *should be noted that* and eleven bundles under the category of noun+verbphrase+that-clause, e.g., *we found that the,* and *these findings suggest that.*

The next most frequent type among clausal bundles is related to the category of copula be+noun/adjective phrase which constitutes 7.8% of the whole bundles. Based on whether the subject predicative is a noun phrase or an adjective phrase, this category can be divided into two subcategories, as shown in Table 3. All in all, 9 bundles could be included in this category with three bundles in the subcategory of copula be+noun phrase such as *be one of the, is the first to, be responsible for the* and 6 bundles in the other subcategory of copula be+adjectivephrase; e.g., *is possible that the, is important to note, is also possible that.*

In addition to the above-mentioned categories, other expressions comprise a large part of the discussion corpus and make up about 27% of the whole corpus.

These bundles cannot be classified based on structural taxonomy as Biber et al. (1999) mentioned.

3.2. Functional Analysis of Bundles in Discussion Section

Following structural analysis of bundles, lastly the turn was given to the functional analysis of bundles in the discussion section. As it was said earlier, for functional classification of bundles the Hyland's taxonomy (2008a) was used. However, in order to fulfill the purpose of the present study, some new categories to be encompassed in the main categories were obtained that are boldfaced in Table 4. In sum, functional classification of 94 bundles identified in the 801,894 word corpus of the discussion is presented in the following table accompanied by their overall frequency and percentage.

Table 4. *Overall Functional Analysis of Bundles in Discussion Section of MRAs*

categories	Subcategories	No of bundles	Overall FRQ	percentage	Examples
Research-oriented bundles	Location	1	17	0.72	*The duration of the*
	Quantification	5	88	3.73	*The small sample size*
	Procedure	2	34	1.44	*That the use of*
	Study-focusing	15	540	22.91	*The present study we*
	Evaluation	1	20	0.84	*The validity of the*
	Discipline–bound	2	63	2.7	*In the pathogenesis of*
Text-oriented bundles	Resultative Signals	15	375	15.91	*We found that the*
	Framing signals	4	86	3.64	*Patients with heart failure*
	Literature-reference	6	126	5.34	*Reported in the literature*
	Confirmation	13	255	10.81	*This is consistent with*
	Suggestion	3	90	3.81	*Studies are needed to*
	Relation	3	73	3.1	*Be related to the*
Participant-oriented bundles	Attitude markers	1	26	1.10	*It is difficult to*
	Epistemic-certain	8	186	7.9	*It is known that*
	Epistemic-uncertain	10	236	10.01	*It is likely that*
	Engagement features	5	142	6.2	*It should be noted*
	Total		**2357**	**100**	

Note: Those categories that are in bold are those found in the present study

As it is shown in Table 4, the highest concentration of discussion corpus is on text-oriented bundles since they make up about 42.61% of the whole bundles

in this corpus. The discussion corpus is concerned with the organization of the text through which the writers of academic texts can convey a message or set an argument. The current study shows that only six types of functional categories could be subsumed under text-oriented bundles. Among text-oriented bundles, resultative signals show the highest overall frequency of 375 and 15.91%. One of the new categories is related to literature-reference bundles used to refer to some previous studies conducted in a specific field of study. MRAs writers have used these bundles to confirm the results of their own study. These bundles were considered as text-oriented bundles since they convey a message and writers set an argument by using them. These bundles form about 5.34% of the whole bundles in the discussion corpus. It was found that the highest frequency of 28 was related to *et al. reported that* while the lowest frequency of 17 was presented by *been reported in the*.

The second new category is related to confirmation bundles which refer to those clusters that are mostly used by writers in the discussion section of MRAs to confirm and approve the validity, reliability or correctness of their results. The findings indicated that just 13 bundles were included in this functional category. It was indicated that *this is consistent with* and two other bundles, i.e., *are in line with* and *findings are consistent with* representing the most and the least frequently used bundles, respectively.

The third new sub-category is suggestion bundles. As the name of this type of bundles implies, the bundles included in this functional category are mostly used by writers of RAs especially in the discussion section of articles to refer to the suggestions for further researches. As it was found, there were three bundles named suggestion bundles with an overall frequency of 90 and 3.81% such as *Studies are needed to, Further studies are needed, Research is needed to*.

The last sub-category is relation bundles**.** These bundles are used to show the relation between two or more elements in a study. Totally, there were three bundles that could be included in the relation functional category with an overall frequency of 73.These bundles are: *Be related to the*, *Be attributed to the*, *May contribute to the*. This group of bundles represents the least frequently used category in text-oriented bundles.

According to Table 4, research-oriented bundles constitute 32.34% of bundles in the discussion corpus. Not only do study-focusing bundles show the highest overall frequency of 540 and percentage of 22.91% among the sub-categories of research-oriented bundles but also they represent the highest percentage in the whole discussion corpus. It should be noted that although the most frequently used sub-category is included in research-oriented bundles category, the least frequently used sub-category such as location is encompassed in this major category as well. Some of the sub-categories of research-oriented bundles in Hyland's taxonomy were used, and some new sub-categories were also formed for the purposes of the present study. Study-focusing bundles as new sub-category developed by Jalali (2009) refer explicitly to the study being carried out and reported by writers of MRAs among which *this is the first* was considered as the most frequent bundle with the highest frequency of 69.

Evaluation bundles that are used instead of description bundles for the purpose of the present study refer to some of the evaluations that are made by the researcher during the study through which he/she can come to a conclusion. This bundle with an overall frequency of 20 comprises 0.84% of the whole bundles and is considered as the second least frequent bundle in the discussion corpus. The last new sub-category is related to discipline-bound bundles that are distinctive common word combinations in a specific field (medicine in this study) (Jalali, 2009). These bundles can signal the writers' knowledge of a specific field

of study. Totally, there were two bundles that could be classified as discipline-bound bundles in the discussion corpus. They form about 3% of the whole bundles in this corpus. Examples are: *In the pathogenesis of, TNF and IL.*

Based on what is presented in Table 4, the lowest proportion is devoted to participant-oriented bundles as they comprise 25.03% of the whole bundles among which the highest percentage is recommended by epistemic-uncertain bundles with an overall frequency of 236. The categories subsumed under the participant-oriented bundles are based on categories presented by Cortes (2002). Bundles serving as attitude-markers are used to show writers' overt stance toward a subject or what she/he is talking about. Totally, only one bundle was classified as attitude markers in the whole discussion corpus. This group of bundles represents an overall frequency of 26 with 1.10%.

According to Biber and Barbieri (2007), epistemic-certain bundles act as a frame and project the propositions as unhedged and undisputed arguments. In other words, they are about the certainty of the writer toward what he/she is talking about. It was indicated that totally there were eight bundles which could be classified as epistemic-certain bundles with an overall frequency of 186 and 7.9% of the whole bundles in discussion corpus. Some of the examples are *it is known that* and *were not able to.*

Unlike epistemic-certain bundles, epistemic-uncertain bundles qualify their propositions by expressing a tentative and less certain stance toward them (Hyland, 2004). It can be said that most writers use these bundles to distance themselves away from making a mistake and being accused by others (Jalali, 2009). It was shown that the highest frequency of 37 and the lowest frequency of 16 are related to *it is likely that* and *are likely to be*, respectively.

Engagement bundles subsumed under the category of participant-oriented bundles are used to address the reader directly. It can be said that most of the

writers use this type of clusters to engage the reader in the text and draw her/his attention to a specific point. It can be inferred from the results that all of the bundles belong to two 6-word and 5-word clusters such as *it should be noted that the* and *is important to note that* as it is obvious that *it should be noted, should be noted that* and *be noted that the* belong to the former cluster while *important to note that* and *is important to note* belong to the latter one although they represent various frequencies.

4. Discussion

Gaining control of new language or register calls for experts following users' preferences for certain sequences of words called lexical bundles. It can be said that learning the most frequent lexical bundles can contribute to gaining communicative competence since these bundles can be considered as a paramount component of fluent or coherent linguistic production and a key factor in successful language learning.

According to the results of the present study, there were 94 bundles identified in the discussion section of MRAs. In a study by Jalali (2009), he found that three corpora of research articles, master thesis, doctoral dissertation in applied linguistics included 121, 255 and 141 bundles, respectively. In another study, Valipoor (2010) found that there were 223 bundles in the corpus of 4,000,000 words of chemical research articles (CRAC) and 83 bundles of them were devoted to discussion section. The result of another study by Parvizi (2011) indicated that there were just 24 bundles in the corpus of 2,000,000 words in the field of education and just 15 bundles were identified in discussion section. The structural analysis of bundles showed that clausal bundles took leading priority over phrasal bundles since they formed 49.05% of the whole corpus in comparison with phrasal bundles which made up 24.06% of the whole bundles.

The results of the present study are in contrast with the studies by Jalali (2009), Valipoor (2010) and Parvizi (2011) who found that phrasal bundles formed about 75%, 55% and 84% of their corpora, respectively. In these studies, prepositional phrase + of bundles showed the highest frequency in phrasal bundles. Regarding the functional classification of bundles, it was found that the highest concentration of discussion section of MRAs was on text-oriented bundles which showed 42.61% of use. Since text-oriented bundles were prominent in the discussion sections of MRAs, it can be claimed that most writers of academic texts such as RAs use these types of bundles in order to elaborate arguments discursively over a greater span of texts. One of the outstanding functional sub-categories of text-oriented bundles in the discussion corpus was resultative signals which attract readers' attention toward writers' understanding and interpretations of research processes and outcomes. Most of these resultative signals were shown through verb-phrase+ that-clause fragment structure which substantiates Hyland (2008)'s idea about the connection between functions and structures of the bundles. The results of this study are in agreement with the findings of the study carried out by Hyland (2008), who found that most of the writers of doctoral dissertations and research articles in disciplines of applied linguistics and business studies relied heavily on text-oriented bundles. In order to reach a comprehensive analysis of bundles in the current study, a comparison was made in terms of functional categorization of bundles in discussion sections of RAs among recent previous studies (Valipoor, 2010; Parvizi, 2011) and current study which are presented in Table 5.

Table 5. *Comparison of Functional Categories among Different Disciplines*

Authors	**Research-oriented**		**Text-oriented**		**Participant-oriented**	
Valipoor (2010) CRAC	Result	Discussion	Result	Discussion	Result	discussion
	9.41%		44.7%		45.9%	

Parvizi (2011) education	Discussion	Discussion	Discussion
	53.33%	33.33%	13.33%
Jalali (2012) medicine	Discussion	Discussion	Discussion
	32.34%	42.61%	25.03%

As it is shown in Table 5, there were some differences among the previous two studies and the present study regarding the type of sections selected for the identification and functional analysis of bundles. For instance, Valipoor has lumped result and discussion sections together under one single section. Regarding research-oriented bundles, it is shown that this type of bundles takes a leading position in education corpus in comparison with the other two corpora (Valippor & Jalali). It should also be pointed out that all three corpora of CRAC, education and CODMRA showed that they contained a high proportion of text-oriented bundles in the discussion sections of RAs with 44.7%, 33.33% and 42.61% of use, respectively.

As for participant-oriented bundles, it can be said that all the three corpora of CRAC, education and CODMRA showed a high percentage of use of this type of bundle in the discussion section and CRAC represented a higher percentage of 45.9% in comparison with other two corpora. On the other hand, regarding the functional sub-classification of bundles, it was found that study-focusing, confirmation and epistemic-uncertain bundles took a leading position in the discussion sections of MRAs. It is noticeable that most academic discourse writers discuss their own studies and other studies which are in agreement or in contrast with them even though they discuss in an almost uncertain way as they want to distance themselves from making mistakes and being accused by others. It can be said that most of the writers of academic prose use this type of bundles

to show their reluctance to direct commitment to a statement. In other words, writers of medical research articles used mostly uncertain-epistemic bundles to protect themselves from likely false interpretations. However, medicine is a kind of empirical science, and it is expected that writers talk about the results of their studies with more certainty. Thus, the use of uncertain-epistemic bundles can show either the uncertainty of the researchers of the results of their studies or the lack of their knowledge of bundles. So, it is likely that the writers of medical articles are not completely aware of the correct uses or functions of lexical bundles or they maybe do not know how/where to use them. The following examples are from the corpus:

As apoE-deficient mice develop severe atherosclerosis with age associated with a relatively small increase in MP concentration, *it is possible that* the lack of an increase in plasma MP concentrations in apoE2-KI mice under Western diet is due to the relatively mild atherosclerosis development.

Thus, the effects on intimal cells of a given quantity of LDL- derived hydrolysis products *are likely to be* enhanced in advanced atherosclerotic lesions displaying acidic intimal fluid.

Young Chinese adults were found to *be more likely to* consume alcoholic beverages with their friends and parents than with other people (Lu et al., 1997), and thus their friends' and parents' beliefs about and behaviors towards driving after the consumption of alcohol may greatly shape their own beliefs and attitudes toward this behavior.

One interesting point is that writers of MRAs have heavily relied on those studies which have been in agreement with their own studies and have not pointed to the results of the studies which have been in contrast with theirs! In the corpus of CRAC, Valipoor (2010) found structuring signals, framing bundles and stance features as outstanding functions of discussion section. Parvizi (2011)

found topic bundles, framing signals and stance features as the most frequent key bundles in the discussion section. A comparison was also made among these three corpora regarding the number of common bundles and their functions, and just 11 bundles were found as common between the corpus of medicine and chemistry but not with corpus of education. Functional analysis of these bundles indicated that there were three bundles which performed the same functions in the two corpora and the other eight bundles, despite their commonality, played different roles in the corpora of chemistry and medicine. In all, the results of the current study and its comparison with other studies carried out on lexical bundles in different fields confirm the idea that was proposed by Hyland (2008a): " different academic discourses rely on different repertoires of lexical clusters" (p. 46). So, it was indicated that medicine has particular lexical bundles with specific structures and functions which are different from those identified in chemistry, education, applied linguistics, history, biology, etc.

There is no doubt that like other texts used in academic contexts, research articles contain lexical bundles which are pervasive in the academic discourse. So, students encounter these clusters and failure to understand their textual meaning or function leads to failure in their production and comprehension. It should also be noted that students may know the meaning of individual words in each lexical bundles, but the problem arises when they encounter these bundles with different functions even in the same text as it was shown in some examples from the corpus previously. Consequently, it would be useful to help students to get familiar with these expressions and their discourse functions and be encouraged to use them in academic discourses. Therefore, some pedagogical techniques should be applied in order to encourage students to learn how to use lexical bundles as part of their writing repertoires. Cortes (2006) discussed that the exposure to lexical bundles should be long enough for the students. Pang

(2010) introduced text analysis, disciplinary ethnographies, concept or semantic maps, writing sentences and comparing registers as techniques which can improve students' awareness of lexical bundles. Pedagogically, it would be useful if teachers of EAP or EMP (English for Medical Purposes) courses include lexical bundles in teaching syllabi as a learning input. They should encourage activities which raise awareness toward lexical bundles and show their structures and functions.

Although this study has investigated the 4-word lexical bundles in all 33 fields in the discussion section of MRAs, it would be useful that future studies identify the lexical bundles in each field separately and compare them with each other. Further studies are needed to be done in order to find out how these lexical bundles and their functions could be introduced to the learners in a way that they do not have any difficulties in their comprehension and production.

References

Alipour, M., Jalilifar, A., & Zarea, M. (2013). A Corpus Study of Lexical Bundles across Different Disciplines. *The Iranian EFL Journal*, 9(6), 11-35.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 101–122). Oxford: Oxford University Press.

Anthony, L. (2007). Antconc 3.2.3: *A free text analysis software*. Available on line at http://www.antlab.sci.waseda.ac.jp/.

Basturkmen, H. (2012). A genre-based investigation of discussion sections of research articles in Dentistry and disciplinary variation. *Journal of English for Academic Purposes*, 11, 34–144.

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, 263–286.*

Biber, D., & Finegan, E. (1994). Intra-textual variation within medical research articles. In N. Oostdiijk, & P. deHaan (Eds.), *Corpus-based research into language* (pp. 201–221). Amsterdam: Rodopi.

Biber, D., Conrad, S. & Cortes, V. (2003). Lexical bundles in speech and writing: an initial taxonomy. In A. Wilson, P. Rayson & T. McEnery. *Corpus Linguistics by the Lune.* (pp. 71-92). Frankfurt: Peter Lang.

Biber, D., Conrad, S. & Cortes, V. (2004). If you look at lexical bundles in university teaching and textbooks. *Applied Linguistics 25 (3),* 371– 405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English.* London: Longman.

Breeze, R. (2013). Lexical bundles across four legal genres. *International Journal of Corpus Linguistics, 18*(2), 229-253.

Brett, P. (1994). A genre analysis of the Results section of sociology articles. *English for Specific Purposes, 13* (1), *47–59.*

Butler, C. S. (1990). Qualifications in science: Modal meanings in scientific texts. In W. Nash (Ed.), *The writing scholar: Studies in academic discourse* (pp. 137–170). Newbury Park, CA: Sage.

Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes, 30*, 38-52.

Chang, C. F. & Kuo, C. H. (2011). A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes, 30, 222–234.*

Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds.) *Using corpora to explore linguistic variation* (pp. 131–145). Amsterdam: John Benjamins Publishing Company.

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 from a writing intensive history class. *Linguistics and Education, 17,* 391-406.

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, 33–43.

Cowie, A. P. (1998). Introduction. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications*. Oxford: Oxford University Press.

De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics, 3* (1)*, 59–80.*

De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on computer* (pp. 67–79). London: Longman.

Dubois, B. L. (1997). *The biomedical discussion section in context*. Greenwich, Connecticut: Ablex Publishing.

Dudley-Evans, T. (1997). Genre: how far can we should we go?. *World Englishes*, *16* (3), 351–358.

Gledhill, C. (2000). *Collocations in science writing*. Tu¨bingen, Germany: Gunter Narr Verlag Tu¨ bingen.

Grabowski, L. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes, 38*, 23–33.

Güngör, F., & Uysal, H. H. (2016). A Comparative Analysis of Lexical Bundles Used by Native and Non-native Scholars. *English Language Teaching, 9* (6), 176-188.

Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes, 16* (4), 321–337.

Hong, A. L., & Hua, T. K., (2018). Specificity in English for Academic Purposes (EAP): A Corpus Analysis of Lexical Bundles in Academic Writing. *3L: The Southeast Asian Journal of English Language Studies, 24* (2)*,* 82 – 94.

Hyland, K. (1998). *Hedging in scientific research articles*. Amsterdam: John Benjamins.

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. Ann Arbor: University of Michigan Press.

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics. 18* (1), *1-9.*

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes,* 27, 4-21.

Jalali, H. (2009). *Lexical bundles in applied linguistics: Variations within a single discipline* (Unpublished doctoral thesis). University of Isfahan.

Jalali, H. (2014). Examining novices' selection of lexical bundles: The case of EFL postgraduate students in applied linguistics. *Journal of Applied Linguistics and Language Research*, *1*(2), 1-11.

Kjellmer, G. (1990). A mint of phrases. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honor of Jan Svartvik* (pp. 111–127). London: Longman.

Kuo, C. H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes, 18*(2), *121–138*.

Lewin, B. A. (2001). *Expository discourse: A genre-based approach to social science research texts*. London: Continuum.

Lindeberg, A. (1994). Rhetorical conventions in the discussion/conclusion sections of research articles in finance, management and marketing. In M. Brekke, O. Anderson, T. Dahl, & J. Myking (Eds.), *Applications and implications of current LSP research. Proceedings of the 9th European LSP Symposium, Bergen, Norway, August 1993* (pp. 761–779). Bergen, Norway: Fagbokforlaget.

Martinez, I. (2003). Aspects of theme in the method and discussion sections of biology journal articles in English. *Journal of English for Academic Purposes*, *2* (2), *103–123.*

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes, 21*, 60-71.

Pang, W. (2010). Lexical Bundles and the Construction of an Academic Voice: A Pedagogical Perspective. *Asian EFL Journal. Professional Teaching Articles, 47, 1-13.*

Parvizi, N. (2011*). Identification of discipline-specific lexical bundles in education* (Unpublished master's thesis). University of Kashan.

Peng, J. F. (1987). *An investigation of rhetorical and organizational features of the discussion sections of Chemical Engineers' papers* (Master's thesis). University of Birmingham.

Safarzadeh, M. M., Monfared, A., & Sarfeju, M. (2013). Native and non-native use of lexical bundles in discussion section of political science articles. *Iranian Journal of Applied Language Studies, 5 (2),* 138-166.

Schmitt, N., & Carter R. (2004). Formulaic sequences in action: An introduction. In: N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp.1–22). Amsterdam: John Benjamins.

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psychologically valid? In Norbert Schmitt (Ed.). *Formulaic Sequences* (pp. 127-151). Amsterdam and Philadelphia: John Benjamins.

Scott, M. (2008). *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software.

Shin, Y. K., Cortes, V., & Yoo, I. W. (2018). Using lexical bundles as a tool to analyze definite article use in L2 academic writing: An exploratory study. *Journal of Second Language Writing, 39,* 29–41.

Stubbs, M. (2007a). An example of frequent English phraseology: Distribution, structures and functions. In R. Facchinetti (Ed.), *Corpus Linguistics 25 years on* (pp. 89–105). Amsterdam: Radopi.

Stubbs, M. (2007b). Quantitative data on multi-word sequences in English: The case of word 'world'. In M. Hoey, M. Mahlberg, M. Stubbs & W. Teubert (Eds.),

Text, discourse and corpora: Theory and analysis (pp. 163–189). London: Continuum.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Tarone, E., Dwyer, S., Gillette, S., & Icke, V. (1998). On the use of the passive and active voice in astrophysics journal papers: With extensions to other languages and other fields. *English for Specific Purposes, 17(1), 113–132.*

Thompson, G., & Ye, Y. (1991). Evaluation of the reporting verbs used in academic papers. *Applied Linguistics, 12, 365–382.*

Valipoor, L. (2010). *A corpus-based study of words and bundles in chemistry research articles* (Unpublished master's thesis). University of Kashan.

Wang, J., Liang, Sh. & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes, 27, 442–458.*