The Journal of Applied Linguistics and Discourse Analysis
Vol. 2, Issue 2, Oct. - Mar., 2014 pp. 67-83

A Comparative Study of Lexical Bundles in Soft Science Articles Written by Native and Iranian Authors

Faramarz Pourmusa
MA in TEFL, Department of English Language and Literature,
Azarbaijan Shahid Madani University, Tabriz, Iran
E-mail: calmnesseverf@yahoo.com

Abstract

Lexical bundles have been focused on by linguists because they pave the way for learners to utilize a series of co-occurring expressions in their productions instead of single words. The present paper sheds light on the lexical bundles utilized in Soft Science articles written by Native and Iranian Authors (NA and IA) with the elementary purpose of analyzing the structural and functional similarities and differences. The secondary purpose was to present a list of explored lexical bundles employed in these articles. In the light of structural classification, *Noun phrase + of-phrase fragments and other prepositional phrases* were the most widespread and *Other passive fragments and Verb phrases with personal pronoun we* were the least employed structures in the articles of NA and IA. Considering the functional classification, the most commonly employed function by both NA and IA was *procedure*. Native authors employed *citation* and Iranians utilized *generalization* with *the least frequency*. The author recommends course developers to incorporate a list of the most common lexical bundles beside the existing lists of single words to enrich the students' knowledge of vocabulary.

Keywords: Lexical Bundles, Soft Science Articles, Corpus, Frequency, Formulaic Expressions

ARTICLE INFO

Article history:

Received: Thursday, April 20, 2017
Accepted: Monday, July 24, 2017
Published: Thursday, September 28, 2017
Available Online: Tuesday, October 10, 2017
DOI: http://dx.doi.org/10.22049/jalda.2017.13651

E-ISSN:
ISSN: 2383-2460

Introduction

During the past two centuries, English managed to prove itself as the most dominant language prevailing the world of research and science. In fact, English has been serving as a *lingua franca* in various fields of modern life (Hoffman, 2000). Academe, more than any other field, has been benefiting from the ever-increasing popularity of English.

Since many scholars have come to an agreement about the distinguishing features of academic discourse, there have been many attempts to clarify those features. Different aspects of academic writing have been investigated such as verb classes (Hunston, 1995), the organization of discourse (Ferguson, 2001), academic registers (Flowerdew, 2002; Hewings, 2001), and expressions of stance (Charles, 2003; Crompton, 1997; Grabe & Kaplan, 1997; Holmes, 1986; Hyland, 1994, 1996a, b; Meyer,1997; Myers, 1989, 1990; Salager-Meyer, 1994; Silver, 2003; and Varttala, 2003).

Academic vocabulary is another interesting quality of academic writing which has been focalized through several studies (Coxhead, 2000; Nation, 1990, 2001). It is believed that the academic prose employs a specific set of vocabularies with their own norms (Nation, 2001). Based on the premise that community norms affect the type of expressions utilized, several reference tools have been designed in order for the non-native and novice writers to fill any gaps between their written products' level and that of their native counterparts. One of those references is the SciE-Lex Electronic combinatory dictionary (Verdaguer, Poch, Laso & Gimenez, 2008). It has been designed to serve as a database of biomedical English for the applicants and members of the Spanish medical community.

HSC in its initial version relied on co-occurrence of words following a frequency-based approach which zoomed on the number of times a series of words were used together. In fact, an inventory of the words' co-selection or *collocation* (Manning & Schutze, 1999; Sinclair, 1991; Stubbs, 2002) in scientific writing provided a worthy contribution to empowering the Spanish writers' potential lexical literacy. During later editions of HSC, its authors found out that another analysis had to be done in addition to collocations. They decided to explore and analyze the continuous sequences of repeatedly co-occurring words. This was a great achievement since it led to the next worthy reference *Longman Grammar of Spoken and Written English* (LGSWE) (Biber et. al, 1999, chap. 13). Chapter 13 of that dictionary was based on a study of a multi-million corpora collected from conversation and academic prose. Lexical bundles consisting of as many as six words were identified on a frequency-based approach. Biber, Conrad, and Cortes (2003) later improved their first attempt by performing an analytical investigation of the discourse functions of lexical bundles. A year later these authors worked on the use of lexical bundles in university teaching and textbooks (Biber, Conrad, & Cortes, 2004). To fill the gap in functional properties of lexical bundles, Hyland (2008a) developed a functional taxonomy and classification for written research genres. Simpson-Vlach and Ellis (2010) developed a valid inventory of the most common multiword expressions. All these studies became a springboard for the second edition of SciE-Lex project, which provided additional data on the function, composition, and textual distribution of the three-to-five-word bundles (Verdaguer, Comelles, Laso, Gimenez, & Salazar, 2009).

The present paper defines corpus and studies based on corpus as well as defining the relationship between formulaic language and corpora. Corpus is a Latin word, meaning *body.* Once pertained to linguistics, it holds the meaning of '*body of texts'.* John Sinclair (2005), who can rightfully be assigned as one of the most prominent figures in the field of corpus studies, defines the term corpus as "a collection of pieces of language texts in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (p.16). One of the first modern corpus-based studies worthy of mention was conducted by Francis and Kucera at Brown University in 1961. Known as the Brown Corpus, it contained a one-million-word corpus on the basis of randomly selected materials written in American English in 1961 in a diverse set of genres.

The history of formulaic patterns in applied linguistics dates back to Jespersen (1924) and Firth (1951), who popularized the term "collocation". Formulaic sequences are considered to be multi-word combinations that are stored and retrieved holistically from the mental lexicon upon speech when in need. They minimize encoding work for the speaker and decoding work for the addressee, with the ultimate function of allowing for the construction of fluent spoken discourse (Erman, 2007; Wood, 2006). It has also been found that the proper use of formulaic sequences is critical for the acquisition of native-like language competence (Dufon, 1995; House, 1996).

Biber et al. (1999) were the first to bring into existence the notion of lexical bundles in their corpus-based study of English grammar book the Longman Grammar of Spoken and Written English (LGSWE). Biber and colleagues in a chapter of this book defined lexical bundles as "bundles of words that show a statistical tendency to co-occur" (p.989) and as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status" (p.990). Salazar (2009) gives a similar definition of lexical bundles as "frequently occurring lexical sequences automatically extracted from a given corpus using a computer program" (p. 13).

The main objective of the present paper is to identify the three to six word lexical bundles in soft science articles written by native and Iranian authors. Hence, the paper attempts to address the following questions:

1. What are the most frequently occurring lexical bundles in native and Iranian soft science articles?
2. What are the structural and functional characteristics of the target lexical bundles? How can they be classified according to these features?
3. What are the differences between the native and Iranian articles in terms of the frequency, structure and functions of the target bundles?

Methodology

The corpus analyzed here consists of a collection of 200 published articles by native authors (NA) (whose native language is English) and Iranian non-native authors (IA). Articles were selected from among published ones in seven main disciplines, but because of the insufficiency of valid articles written by Iranian authors in the field of History it was decided to bring together History and Culture as one discipline. All articles were published between the years 2013-15. Equal proportions of articles were allocated for both native and

Iranian authors in order to avoid norming and for the findings to be more reliable. As Table 1 illustrates, the whole number of the words in the corpus comes up to 1,666,884. Considering the size of the corpus, the criteria Biber (2006) set was used. He believes that a corpus must be "large enough to adequately represent the occurrence of the features being studied" (p. 89).

The compiled corpus was analyzed using AntConc 3.4.4, a freeware concordance program developed by Laurence Anthony at the Center for English Language Education in Science and Engineering (CELESE), Waseda University (Japan). This software has features completely appropriate for studying lexical bundles, including the word and frequency generators as well as the N-gram and cluster analysis. The N-gram function of this software was used to identify the existing lexical bundles in each corpus along with their frequency. Since the focus was on three to six word lexical bundles, the "N" in N-gram was assigned three to six in different stages to identify the targeted lexical bundles. This way, the extraction process was done carefully because each time the focus was just on one type of bundles. For meeting the frequency criterion, the minimum frequency function of the N-gram was set to 15 in order to automatically ignore the expressions with a lower frequency rate. The range function of N-gram could eliminate those expressions occurring in less than 5 texts. This was done to avoid any potential unreliability originating from idiosyncratic preferences.

Table 1: Disciplines, number of articles, number of words

Disciplines	No. of words (NA)	No. of words (IA)	No. of words (NA+ IA)	No. of articles (NA+ IA)
Psychology	82598	70175	152773	32
Applied Linguistics	140360	166961	307321	40
Anthropology	102521	187261	289782	30
Sociology	294992	213891	408883	40
History & Culture	127478	92418	219896	30
Politics	86446	101783	188229	28
	834395	832489	1666884	200

Since AntConc 3.4.4 is based on raw text files, all compiled data were converted into txt. file types using another freeware software called AntFileConverter 1.2.0.

Lexical bundles have been classified structurally several times, among which Biber et al. (1999) has been mostly relied on in other studies (Cortes, 2002, 2004; Hyland, 2008a, 2008b). For this paper, an adaptation was performed through adding five new categories: other noun phrases, other adjectival phrases, verb phrases with personal pronoun *we*, other passive fragment, and other verbal fragments. Table 2 depicts the adapted structural classifications adapted from Biber et al. (1999). The adaptations were made in order to narrow down structural divisions, with the ultimate purpose of assigning each lexical bundle a more precise classification.

Classifying lexical bundles in terms of their discoursal and pragmatic functions was the next stage. To meet the determined aims, Hyland's (2008a) functional classification was utilized, but, like the structural classification, the functional taxonomies of Hyland were

changed a little to handle the purpose of the paper more precisely. Hyland's classification had three broad classifications which were kept unchanged. The adaptations occurred in subcategories, which are mentioned here.

In research-oriented broad category four subcategories of *location, procedure, quantification,* and *description* were preserved, but the *topic* subcategory was eliminated because the author did not want to include discipline-based lexical bundles among the findings. The reason refers to the original purpose, which is providing the novice and/or non-native authors with a list of the most common lexical bundles utilized in soft science articles without any resort to any specific discipline-oriented expressions. Instead of the *topic* subcategory, *grouping* was added which is used in grouping, categorization, classification, and ordering of research elements.

The second broad category, the text-oriented functions, was also modified. The *contrastive* function was substituted with *additive* and *comparative*. *Inferential* and *causative* functions substituted the *resultative* function. *Structuring* and *framing* were retained, and three new subcategories were added: *citation,* assigned to bundles which are employed for citing the sources and supporting data; *generalization,* for expressing generally accepted facts and principles; and *objectives,* utilized for stating the authors' aims.

The participant-oriented category underwent just one change: the *acknowledgment* subcategory for bundles employed to express thanks or appreciation to institutions or people who have contributed to the study.

Table 2: Structural patterns of lexical bundles in soft science articles
(adapted from Biber et al., 1999, pp.1015-1024)

Noun phrase with *of*-phrase fragment	***the existence of, a variety of***
Noun phrase with other post-modifier fragment	*the difference in, no effect on*
Other noun phrases	*the present study*
Prepositional phrase + *of*	*as a consequence of*
Other prepositional phrases	*with respect to*
Passive + prepositional phrase	*are shown in*
Other passive fragment	*has been reported*
Anticipatory *it* + verb or adjectival phrase	*it is likely that*
Copula *be* + adjective phrase	*is consistent with*
(Verb phrase or noun phrase) + *that*-clause fragment	*this suggests that*
(Verb or adjective) + *to*-clause fragment	*to account for*
Adverbial clause fragment	*as described before*
Verb phrase with personal pronoun *we*	*we were unable to*
Other verbal fragments	*does not require*
Other adjectival phrase	*similar to that*
Other expressions	*as well as*

Tabe 3: Functional taxonomy of target bundles (adapted from Hyland, 2008a, pp.13-14)

A. Research-oriented bundles help writers to structure their activities and experiences of the real world	
Location: Indicating time / place: Indicate place, extremity, and direction	*at the end of*
Procedure: Indicate events, actions, and procedures	*the use of*
Quantification: Indicate measures, quantities, and proportions	*a wide range of*
Description: Indicate quality, degree, and existence	*the structure of*
Grouping: Indicate groups, categories, parts, and orders	*a group of*
B. Text-oriented bundles concerned with the organization of the text and its meaning as a message or argument	
Additive: Establishing additive or contrastive links between elements	*on the other hand*
Comparative: Compare and contrast different elements	*in contrast to*
Inferential: Signal inferences and conclusions drawn from data	*we conclude that*
Causative: Mark cause and effect relations between elements	*as a result of*
Structuring: Text-reflexive markers that organize stretches of discourse or direct the reader elsewhere in the text	*as described previously*
Framing: Situate arguments by specifying limiting conditions	*with respect to*
Citation: Cite sources and supporting data	*as reported previously*
Generalization: Signal generally accepted facts or statements	*is thought to be*
Objective: Introduce the writer's aim	*in order to*
C. Participant-oriented bundles focused on the writer or reader of the text	
Stance: Convey the writer's attitudes and evaluations	*are likely to be*
Engagement: Address readers directly	*it should be noted that*
Acknowledgment: Recognize people or institutions that have participated in or contributed to the study	*kindly provided by*

To purify the final list of the identified lexical bundles, some exclusion criteria were set. The lexical bundles which had the following features were not included in the final inventory: fragments of other bundles, bundles ending in articles, topic-specific bundles, bundles whose components were exclusively function words, bundles consisting of cardinal and/or ordinal numbers, random section titles, and bundles that express time, temperature, volume, and length. The exclusion factors mentioned above decreased the overall number of originally detected lexical bundles to a large extent. That is, the original number of the explored bundles was 1914 and it decreased to 1041 after applying the exclusion criteria. In nearly all previous studies there is almost no exclusion criteria as strict as the one adapted in this paper. Although after applying the exclusion criteria the whole sum of lexical bundles decreased, the remaining ones are purified and reliable bundles empty of any useless information. The main focus was on the quality and functional analysis of the lexical bundles identified, not the number.

Results and Discussions

The whole sum of the explored lexical bundles was 1041. The final list mainly was composed of three-word expressions, which account for 84.64% of the whole strings in the articles of IA and 91.31% in the articles of NA. The average of three-word strings came up to 86.96%. As previously stated by Salazar (2009), three-word expressions were found ten times more than the four-word strings. The results of the present study showed that the three-word lexical bundles were nearly eight times more than the four-word ones.

Table 4: Top 50 lexical bundles in order of frequency

Rk.	Fq.	Rg.	Lexical bundles in NA	Rk	Fq.	Rg	Lexical Bundle in IA
1	284	71	as well as	1	317	79	as well as
2	183	32	international journal of	2	300	68	in order to
3	170	50	the use of	3	259	59	international journal of
4	151	41	in order to	4	209	57	the results of
5	148	32	more likely to	5	206	69	the role of
6	148	16	of psychological studies	6	202	33	of social sciences
7	147	39	the relationship between	7	199	64	in this study
8	140	51	in terms of	8	184	64	in terms of
9	125	49	the role of	9	180	50	the present study
10	124	46	the fact that	10	176	50	the use of
11	110	37	the number of	11	172	20	the journal of
12	92	33	the effect of	12	169	60	on the other
13	92	40	the importance of	13	167	44	the effect of
14	89	44	as a result	14	158	13	of teaching language skills
15	88	42	a number of	15	158	49	the relationship between
16	88	17	the evolution of	16	154	55	in other words
17	86	45	the case of	17	151	15	quality of life
18	82	31	the development of	18	148	60	of this study
19	82	40	the end of	19	140	55	on the other hand
20	79	37	the same time	20	132	55	the fact that
21	78	28	in this study	21	127	40	the development of
22	78	36	the university of	22	122	48	the most important
23	76	38	at the same	23	118	54	as a result
24	76	36	on the other	24	118	8	of foreign affairs
25	75	38	the context of	25	115	52	of the study
26	74	18	to engage in	26	110	43	the number of
27	73	32	the effects of	27	106	9	in the region
28	72	37	the impact of	28	105	44	the process of
29	69	34	at the same time	29	104	49	a number of
30	68	36	the process of	30	90	27	the study of
31	66	34	the nature of	31	88	47	the importance of
32	66	34	the study of	32	87	38	in line with
33	64	32	in the same	33	87	34	in the following
34	63	27	likely to be	34	87	23	of the country
35	62	36	a series of	35	83	31	in this regard
36	61	34	of this article	36	83	36	the end of
37	60	34	in the context of	37	83	18	the history of
38	60	24	the concept of	38	83	29	with respect to
39	57	21	are more likely	39	82	35	it seems that
40	57	21	the politics of	40	81	31	the impact of
41	56	21	annual review of	41	81	12	the ministry of
42	56	32	can be found	42	81	45	this study was
43	54	29	on the other hand	43	79	31	the lack of
44	54	24	the emergence of	44	73	40	the findings of
45	54	20	the issue of	45	69	40	in addition to
46	53	34	a result of	46	69	32	in relation to
47	53	24	ways in which	47	69	32	the analysis of
48	52	26	the idea of	48	68	38	in the present
49	51	33	in the first	49	68	31	the context of
50	50	19	are more likely to	50	67	34	at the same

The lexical bundles utilized in the articles of IA have a higher frequency compared to those of NA. The frequency in the articles of IA ranges from 317 to 67 while its range in

those of NA is from 284 to 50. This means that Iranian authors utilized the lexical bundles with a higher frequency than their native counterparts. In addition, Iranian authors had a higher minimum and maximum frequency rate than native authors. Second, 23 out of 50 top lexical bundles are present in both lists. This means that 46% of the lexical bundles with the highest frequency have been utilized by both native and Iranian authors with a frequency rate higher than 50.

Table 5 represents the number and percentage of three to six word lexical bundles in the soft science articles of NA and IA.

Table 5: The number and percentage of three-to-six-word lexical bundles from the articles of NA and IA after the application of exclusion criteria

	No. in NA	Percent in NA	No. in IA	Percent in IA	No. in IA + NA	Percent in IA + NA
Three-w	368	91.31	540	84.64	908	86.96
Four-w	33	8.18	90	14.10	123	11.81
Five-w	2	0.49	7	1.09	9	0.86
Six-w	-	-	1	0.15	1	0.096

According to the information presented in Table 5, three-word bundles make up more than 91% and 84% of the whole identified bundles in the articles of NA and IA, respectively. The findings are in accordance with those of Salazar (2009). She found out that three-word lexical bundles were used in a frequency ten times more than that of four-word ones. Native authors employed three-word bundles nearly 11 times more than four-word ones, and Iranian authors did so 8 times more than four-word lexical bundles.

Structural Characteristics of Target Bundles

Figures 1 and 2 show the percentage and distribution of the classified lexical bundles identified in the articles of NA and IA. The vertical axis in each figure was set to a number close to the highest percent in order for the design to be legible and for the numbers to be accessible more easily.

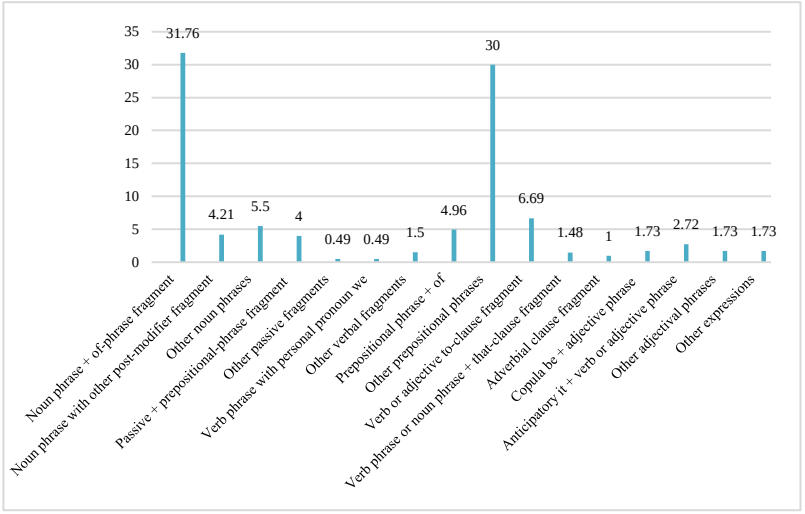

Figure 1. The distribution of the structural types of lexical bundles in the articles of NA

As the figure clearly shows, the category *noun phrase + of-phrase fragment* is the most frequently used structure in the articles of NA. *Other prepositional phrases* structure is ranked as the second common structure chosen and utilized by NA. Noun phrase and prepositional groups make up nearly 76% of the whole identified lexical bundles. Most expressions made by these two structures are mainly three-word bundles which are shorter than other multiword strings. *Other passive fragments* and *Verb phrase with personal pronoun we* are the lowest preferred structures. These two form only 1% of the employed lexical bundles.

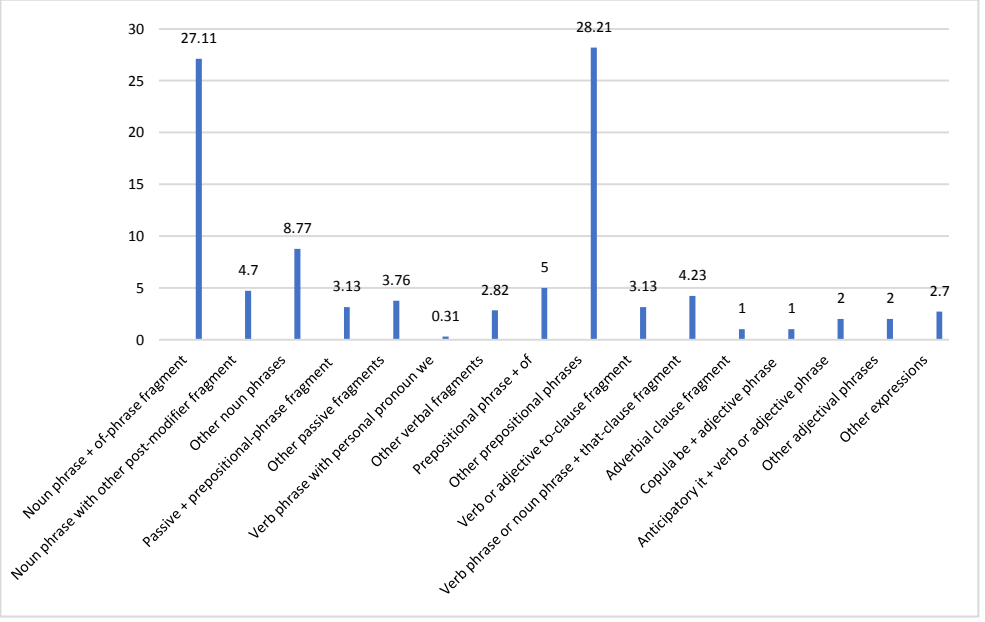

Figure 2. The distribution of the structural types of lexical bundles in the articles of IA

Like native authors, Iranian authors preferred *other prepositional phrases* and *noun phrase + of-phrase fragment* over other structures. *Verb phrase with personal pronoun we* was again the least favorite structure employed in the writings of Iranian researchers.

Noun Structures

Table 6: Noun structures in the articles of NA

Noun phrase + *of*-phrase fragment	international journal of – the use of – the role of – the number of – the effect of – the importance of – a number of – the evolution of – the case of – the development of – the end of – the university of – the context of – the effects of – the impact of – the process of – the nature of – the study of – a series of – the politics of – annual review of – the emergence of – the issue of – a result of – the idea of – the risk of – a variety of – the history of –the influence of – the basis of – the majority of – the presence of – the age of – the creation of –the question of – the rest of –the time of – a theory of – the rise of – a sense of – a set of – the form of – the order of – the result of – the department of – the level of – the formation of – the purpose of – the possibility of – the results of – the likelihood of – the analysis of – a study of – the future of – the making of – a history of – an example of – the construction of – the division of – a form of – the language of – the notion of – the significance of – a measure of – the value of – a function of – a lack of – the course of – the extension of – the intersection of – the meaning of – the work of – a range of – the field of – the power of – the state of – a review of – an analysis of - the size of – the structure of – a host of – a member of – the amount of – the beginning of – the growth of – the implications of – the absence of – the distribution of – the origins of – the dynamics of – the establishment of – the government of – the legitimacy of – the part of – the problem of – the range of – a kind of – a wide range of – the quality of – this type of – a group of – a total of – our understanding of – the experience of – the terms of – a matter of – the findings of – the name of – the rate of – the lives of – the rules of – the subject of – a discussion of – the center of – the degree of – the discovery of – the magnitude of – the production of – the practice of – the scope of – a consequence of – an extension of – national survey of – the ability of – the focus of –the heart of – the theory of – in the case of
Noun phrase with other post-modifier fragment	the relationship between – ways in which – the ways in - the ability to – the ways in which – the need to – the extent to which – an increase in – the need for – a decline in – the right to – the difference between – way in which – an effort to – gender differences in – the case for – their ability to
Other noun phrases	the same time – the middle class – this spatial issue – the same way – division of labor – the twentieth century – point of view – version of this article – the present study – the one hand – evolution of human – each of these – a case study – purpose of this – the first time – the past years – twenty first century – an important role – science and education – the current study – a central role – a great deal

Table 7: Verb structures in the articles of IA

Passive + prepositional-phrase fragment	is related to – is based on – was used to – were asked to – be considered as – were used to – are related to – are presented in – considered to be – is defined as – is considered as – is shown in – are based on – be regarded as – is associated with – can be attributed to – are expected to – based on this – be divided into – be used in
Other passive fragments	can be seen – should be noted – can be used – should be considered – based on their – was based on – be said that – be seen in – be concluded that – can be seen in – can be concluded – can be said - presented in table – referred to as – used in this – taken into account – carried out in - shown in table – the results indicated – the results obtained – participants were asked – are presented in table – has been shown – used in this study
Verb phrase with personal pronoun *we*	we found that –we know that
Other verbal fragments	this study was – the participants were – this study is – the results showed – this study was to – had a significant – present study was – studies have been – test was used – the present study was – this research is – the students were – this article is – is the results of – the results revealed – has its own – present study is – the results are

Table 8: Prepositional structures in the articles of NA

Prepositional phrase + of	in terms of –in the context of – in the case of – as a result of – at the end of – as part of – in the form of – at the time of – by means of – in front of – on the part of – by virtue of – in the process of - in favor of – in light of – in the study of – about percent of – in favor of – in the field of – to the study of
Other prepositional phrases	in order to – of psychological studies – of this article – as a result – in this study – at the same – on the other – at the same time – in the same – of this study – on the other hand – in the first – in the case – in relation to – in this article – with respect to – in other words – of the world – in contrast to – in the past – at the time – of the same – at the end – in the early – in this case – on the basis - in the world – of the social – of the state – of this study – in addition to – in line with – of the participants – of the first – of their own – with regard to – in the middle – over the past – of association between – of the human – as a whole – in response to – in the social – between the two – in this way – of the child – of the mind – of the nation – under the terms – in the region – in the study – of the brain – of the relationship – to what extent – in some cases – in the late – in the present – of the time – of the total – in recent years – of the past – of a nation – of the relationship between – as a political – in the field –in this content – of research in – of the new – to the fact that – for this article – in this paper – of political science – to the same – in doing so – in the last – in the mid – in the sense – on the one hand – for the study – in the following – in the same way – in this regard – in ways that – of the family – of the study – of this paper – in the appendix – in the new – in the next – in the text – in the years – of a new – of the middle – of the war – to the extent – as a means – for more than – in a way – in the age – in the most – in this respect – of the national – of the population – of the second – of the self - to the study – as a consequence – for this article is – in our study – in the human – in the model – of the group – at least in – for a new – in the future –in the labor – of the period – of the research – of the term – with each other – in the case

Table 9: Other structures in the articles of NA

Verb or adjective to-clause fragment	more likely to – to engage in – likely to be - are more likely to – is important to – to account for – likely to engage in – need to be – more likely to be – tend to be – are likely to – appear to be – likely to have – to deal with – seems to be – is likely to – to be more – allows us to – are able to – is necessary to – to note that – to say that – is difficult to – to do so – to refer to – appears to be – was able to
Verb phrase or noun phrase + *that*-clause fragment	the fact that – the idea that –the possibility that – is clear that – the assumption that - this suggests that
Adverbial clause fragment	as opposed to – when it comes –does not necessarily – when it comes to
Copula *be* + adjective phrase	are more likely – is associated with – is consistent with – is an important – is an open – are less likely – is due to
Anticipatory *it* + verb or adjective phrase	it is possible – it is important to – it is important –it is possible that – it is clear – it is difficult – it is possible to – it comes to – it is necessary – it is clear that - it is difficult to
Other adjectival phrases	less likely to – the most important –distributed under the terms – the same as – not due to – micro and macro – more or less
Other expressions	as well as – this article is –is part of – so as to – there is little – this article we – this essay will

Distribution of Target Bundle Functions

The identified bundles were classified in terms of their functions according to a modified version of Hyland's (2008a) functional taxonomy. In order for the results to be more reliable, those lexical bundles which were listed in structural classifications but did not belong to any subcategory of the functional classification were deleted from the final list of functional classification. Figure 3 shows the distribution of target bundles in the writings of NA and IA.

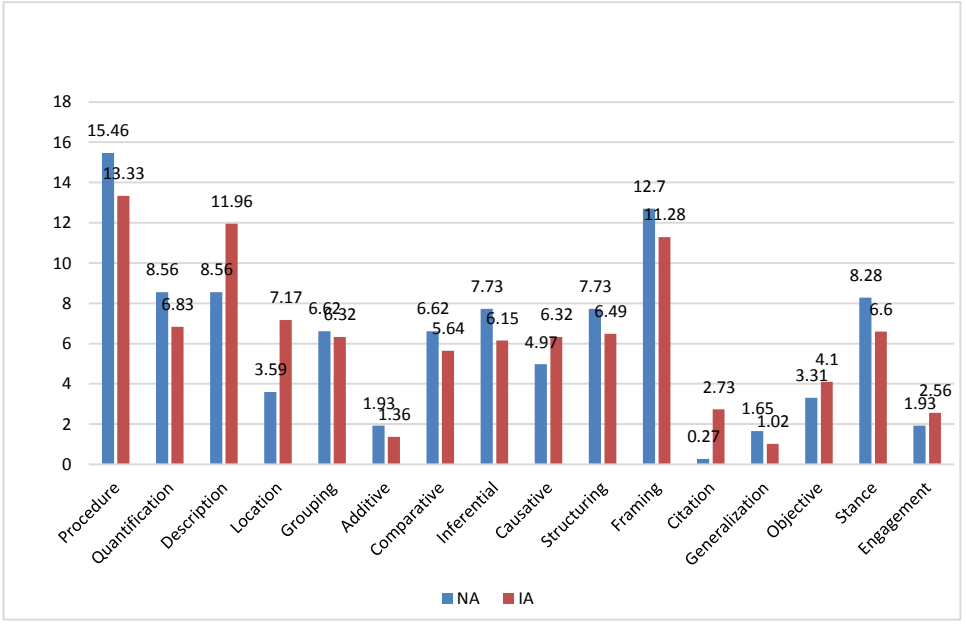

Figure 3. Distributions of lexical bundles' functions

Research-oriented bundles

The functions classified under research-oriented (Table 12) come up to five subcategories with the focal point of the processes and procedures involved in manipulating the research and study. As the name denotes, research-oriented bundles characterize the functions which are observable in academic writing. Different subcategories are properties of the bundles in relation to the research itself. For example, quantification clearly stands for the amounts, numbers, volumes, and measures apparent in the research papers.

Table 10. Research–oriented bundles in the articles of NA

Procedure	in this study – the use of – the evolution of – the development of – the process of – annual review of – the analysis of – the making of – the division of – the extension of – the intersection of – division of labor – of research in – the beginning of – the formation of – tend to be – the work of – a review of – the implications of – in doing so – the dynamics of – evolution of human – for the study – in ways that – in the process of – our understanding of – a discussion of – the discovery of – the magnitude of – the practice of – was supported by – an extension of – in the labor – is published by – national survey of – of the research – the assumption that – a study of – can be used – the establishment of – were asked to – the creation of – the emergence of – in this way – the rise of – an analysis of – the origins of – the production of – the construction of – to the study – used in this – the theory of – to deal with – the experience of – a case study – the distribution of
Quantification	the number of – a number of – the majority of – the time of – the age of – of the first – the value of – of the total – of the past – a decline in – the amount of – to be more – in the mid – is part of – to the extent – of the second – the first time – about percent of – at least in – there is little – the extent to which – an increase in – in recent years – the size of – as part of – the degree of – to what extent – over the past – the rate of – a great deal – a total of
Description	the importance of – in line with – a lack of – in the following – the rules of – is related to – the level of – the basis of – the quality of – the presence of – the form of – in the sense – the nature of – a sense of – in the form of – the issue of – this special issue – the ability to – the ability of – the idea of – the idea that – was able to – are able to – their ability to – the power of – the absence of – a form of – the focus of –the subject of – the problem of – the structure of
Location	in the region – in front of – in the following – the center of – the heart of – the end of – in the world – of the world – at the end – of the state –at the end of – in the middle – the university of
Grouping	a series of – the middle class – as part of – each of these – of the group – with each other – of the participants – the department of – an example of – between the two – a variety of – a set of – a group of – a kind of –a range of – the range of – a member of – the scope of – the rest of – a wide range of – the part of – on the part of – as a whole – different types of

Research-oriented bundles constitute nearly 46% of the total bundles. This is in line with the most outstanding feature of academic writing: giving a precise account of the subject being studied. In fact, the abundance of research-oriented bundles confirms the scientific nature of academic writing.

Table 11: Participant-oriented lexical bundles in the articles of IA

Stance	it seems that – be noted that – we found that – it is important – should be noted – to the fact – is important to – it is possible – should be noted that – it should be noted that – it is important to – should be considered – is possible to – be attributed to – it is possible to – it is obvious that – to note that – be said that – considered to be – important role in – can lead to – seems to be – has led to – is considered as – be regarded as – can be concluded - is associated with – an important role in – can be attributed to – important to note that – it can be concluded – be mentioned that – can be concluded that – can be said – is important to note that – it is important to note – are expected to – it is necessary – be argued that
Engagement	be noted that – it is important – should be noted – is important to – should be noted that – it should be noted that – it is important to – should be considered – to note that – the possibility of – important to note that – be mentioned that – is important to note that – it is important to note – as can be seen
Acknowledgments	

Conclusion and Implications

A) Points about the structural classification of the identified lexical bundles:

1. *Noun phrase + of-phrase fragments* and *other prepositional phrases* were the most widespread structures of the identified lexical bundles in the articles of NA and IA.The above mentioned structures comprised almost 62% of the bundles in the articles of NA and 56% in those of IA. According to Biber et al. (1999) and Hyland (2008a) noun and prepositional phrases in academic writings shift the focus in the text from the writer to the action being done and the kind of relationship which exists between different elements of the text.

2. Noun structures dominated others in classification taxonomy in both articles of NA and IA. Three subcategories (noun phrase+ *of*-phrase fragments, noun phrases + other post modifier fragment, and other noun phrases) made up almost the same amount of the final lexical bundles in both native and Iranian authors' articles. Almost 41% of the whole bundles consisted of noun phrase structures.

3. *Other passive fragments* and *Verb phrases with personal pronoun we* were the least employed structures by both NA and IA. These two structures comprised the minimum percent in both types of articles. They barely come up to 1% of the total identified lexical bundles.

4. Other structures had a normal distribution in both NA and IA articles. The remaining types of structures did not have any outstanding dispersion in any kinds of articles. This means that NA and IA employed these bundles when necessary and they cannot be a distinctive feature of academic writing.

5. Iranian writers employ more verb structures than their native counterparts. The percentage of verb structures for native writers was 6.5 and for Iranian authors it came up to 10.

B) Points about the functional classification of the identified lexical bundles**:**

1. The most commonly employed function by both NA and IA was procedure. It comprised 15.5% and 13.5% of the whole functions in the articles of NA and IA, respectively.

2. Native authors employed *citation* with the least frequency (0.27%) and Iranians utilized *generalization* (1.02%) less than other functions in their writings.

3. In the articles of IA, research-oriented and text-oriented bundles were employed in an almost equal proportion (45%). But, NA preferred text-oriented (47%) to research-oriented bundles (43%).

4. Participant-oriented bundles were the least employed in the articles of NA (10%) and IA (9%). This may be due to the characteristics of academic writing compared with other types of writing such as fictional or advertising. In academic settings, authors try to decrease personal tone in their writings through minimizing the direct addressing of their readers as well as the least use of personal pronouns *I* and *we*.

5. Native and Iranian authors utilized functions in similar ways. If a specific bundle was used in more than one function by NA, IA also employed it in similar functions. The

bundle *is important to*, for instance, was employed to serve the functions of *stance* and *engagement* by both NA and IA.

Having a ready-made inventory of the most common lexical bundles can not only help students who intend to write for international magazines and journals, but also saves the time and energy to a large extent since writers do not need to think of sequencing several words one after another or have doubts about the accuracy of using certain words together.

Curriculum developers may utilize the lists of lexical bundles as a complementary inventory to be added to the list of new words or expressions English books usually contain. Alternatively, a list of commonly used expressions can replace the traditional lists of new words at the end of each course. That way, students do not memorize only long lists of single words without gaining any deep insight on the way these words combine with other strings. Moreover, lists of lexical bundles give the chance for EFL students to be able to learn longer stretches of words, thereby enhancing their writing skill level. Students can also benefit from these lists through getting familiar with different functions each lexical bundle serves in sentences and conversations. As a result, their speaking skill may be positively influenced through using multiword strings for their intended functional interactions.

References

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: Benjamin.

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

Biber, D., Conrad, S., & Cortes, V., (2003). Lexical bundles in speech and writing: An initial taxonomy. In A. Wilson, P. Rayson, & T. McEnery (Eds.). *Corpuslinguistics by the lune: A festschrift for Geoffrey Leech*. 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.

Charles, M. (2003). 'This mystery. . .': A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes, 2(*4), 313-326.

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*(4), 397-423.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Crompton, P. (1997). Hedging in academic writing: Some theoretical problems. *English for Specific Purposes, 16(*2), 271-287.

Dufon, M. (1995). The acquisition of gambits by classroom foreign language learners of Indonesian. In M. Alves (Ed.), *Papers from the 3rd annual meeting of the Southeast*

Asian Linguistic Society (pp. 27-42). Tempe: Arizona State University, Program for Southeast Asian Studies.

Erman, B. (2007). Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics*, *12*(1), 25-53.

Ferguson, G. (2001). If you pop over there: A corpus-based study of conditionals in medical discourse. *English for Specific Purposes, 20*(1), 61-82.

Firth, J. R. (1951). Modes of meaning. *Essays and Studies of The English Association, n.s,. 4*(123-149).

Flowerdew, J. (Ed.). (2002). *Academic Discourse*. New York, NY: Longman.

Fries, C. & Traver, A. (1940). *English word lists: A study of their adaptability and instruction*. Washington, DC: American Council of Education.

Grabe, W., & Kaplan, R. B. (1997). On the writing of science and the science of writing: Hedging in science text and elsewhere. In R. Markkanen & H. Schroder (Eds.), *Hedging and discourse: Approaches to the analysis of a pragmatic phenomenon in academic texts* (pp. 151-167). Berlin: Walter de Gruyter & Co.

Hewings, M. (Ed.). (2001). *Academic writing in context: Implications and applications.* Birmingham: The University of Birmingham Press.

Hoffman, C. (2000). The spread of English and the growth of multilingualism with English in Europe. In C. Jasone & J. Ulrike (Eds.). *English in Europe: The acquisition of a third language*, (pp.1-21). Clevedon: Multilingual Matters.

Holmes, J. (1986). Doubt and certainty in ESL textbooks. *Applied Linguistics, 9*(1), 21-43.

House, J. (1996). Developing pragmatic fluency in English as a foreign language. *Studies in Second Language Acquisition, 18*(2), *225-252*.

Hunston, S. (1995). A corpus study of some English verbs of attribution. *Functions of Language, 2*(2), 133-158.

Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for Specific Purposes, 13*(1), 239-256.

Hyland, K. (1996a). Talking to the academy: Forms of hedging in science research articles. *Written Communication, 13*(1), 251-281.

Hyland, K. (1996b). Writing without conviction? Hedging in science research articles. *Applied Linguistics, 17*(1), 433-454.

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41-62.

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.

Jesperson, O. (1924). *The philosophy of grammar*. London: Allen & Unwin.

Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge MA: The MIT Press.

Meyer, P. G. (1997). Hedging strategies in written academic discourse: Strengthening the argument by weakening the claim. In R. Markkanen, & H. Schroder (Eds.), *Hedging and discourse: Approaches to the analysis of a pragmatic phenomenon in academic texts* (pp. 21-41). Berlin: Walter de Gruyter, & Co.

Myers, G. (1989). The pragmatics of politeness in scientific articles. *Applied Linguistics, 10*(1), 1-35.

Myers, G. (1990). *Writing biology: Texts in the social construction of scientific knowledge*. Madison, WI: University of Wisconsin Press.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York, NY: Newbury House.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes, 13*(1), 149-170.

Salazar, D. (2009). *Lexical bundles in scientific English: A corpus-based study of native and non-native writing*. Universsitat de Barcelona.

Silver, M. (2003). The stance of stance: A critical look at ways stance is expressed and modeled in academic discourse. *Journal of English for Academic Purposes, 2* (1), 359-374.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (2005). Corpus and text: Basic principles. In M. Wynne (Ed.). *Developing linguistic corpora: A guide to good practice* (pp. 1-16). Oxford: Oxbow Books.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512.

Stubbs, Michael. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics, 7*(2), 215-244.

Varttala, T. (2003). Hedging in scientific research articles: A cross-disciplinary study. In G. Cortese, & P. Riley (Eds.). *Domain-specific English: Textual practices across Communities and Classrooms* (pp. 141-174). New York, NY: Peter Lang.

Verdaguer, I., Poch, A., Laso, N. J., & Giménez, E. (2008). Proceedings of the XIII EURALEX International Congress. In E.B. Gallen, & J. D. C. Ward (Eds.), *A lexical database of collocations in scientific English for Spanish scientists* (pp. 145-173).

Verdaguer, I., Comelles, E., Laso, N. J., Gimenez, E., & Salazar, D. (2009). eLexicography in the 21st century: New challenges, new applications: Proceedings of eLex 2009, Louvain-la-Neuve. In S. Granger & M. Paquot (Eds.), *SciE-lex: An electronic lexical database for the Spanish medical community* (pp. 325- 334).

Wood, D. (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *Canadian Modern Language Review, 63*(1), 13-33.