



Syntactic and Pragmatic analysis of Lexical Bundles in Academic Written Texts of Health Information Management written by the natives: Corpus-Driven Study

Danial Shirzadi¹, Hamid Reza Dowlatabadi^{1*}

¹Department of English Language and Literature, Arak University, Arak, Iran

Abstract

Introduction: Health Information Management, as a well-established field of study which has its own voice, genre, and journals, introduced itself as an independent and significant field of academia which deals with issues like health classification and terminologies, electronic health records (HER), confidentiality, health systems and technology, and health informatics. The current study was set to investigate the most frequent 4-word lexical bundles, the semantic structures of lexical bundles, and the pragmatic functions of lexical bundles in the corpus of Health Information Management academic texts.

Methods: In this study, a content analysis method was used. The health information management corpus included 2,210,466 words from the research articles and course books. Antconc 3.4.4 Software package was used to identify the lexical bundles. Lexical bundles in terms of 3-grams to 5-grams were identified using the software, with a cut-off point frequency of 20 per one million words, as per Biber et al. (2006).

Results: The results showed that the most frequent bundle was “at the time of”. It was also revealed that the most frequent types of phrases in the health information management corpus were noun phrases and prepositional phrases, followed by verb phrases.

Conclusions: The study found that the pragmatic functions of lexical bundles in the Health Information Management Corpus were grouped into three major categories including research-oriented bundles, text-oriented bundles, and participant-oriented bundles.

Keywords: Field, Words, Terminologies, Frequent, Genre

Article History:

Received: 06 February 2023

Accepted: 20 April 2023

Please cite this paper as:

Shirzadi D, Dowlatabadi HR. Syntactic and Pragmatic analysis of Lexical Bundles in Academic Written Texts of Health Information Management written by the natives: Corpus-Driven Study. Health Man & Info Sci. 2023; 10(2): 100-108. doi: 10.30476/JHMI.2023.96525.1152.

*Correspondence to:

Hamid Reza Dowlatabadi,
Department of English Language
and Literature, Arak University,
Arak, Iran
Tel: +98 9188617829
Email: h-dowlatabadi@araku.ac.ir

Introduction

English, as a lingua franca, plays an important role in both everyday social interaction and academic communication (1-4). Accordingly, comprehending and having even native-like proficiency in this global language seems to be urgent if one want to introduce himself/herself as a member of his/her academic field. With the advent of corpus linguistics, there was a sharp interest in finding similarities and regularities in corpora and using them for exploratory and pedagogic aims.

Formulaic language, as one of the key distinctive and valuable features of every discipline and genre, has attracted the researchers' attention for more than half of a century (5-7). It has been scrutinized under different names like 'formulaic sequences' (8), 'formulaic expressions' (9), 'fixed expressions' (10), 'lexical phrases' (11), 'multiword lexical units' (12), 'n-grams' (13), or 'lexical bundles' (14-16). The

pioneering work of (14) led to the coinage of “lexical bundles” that is “recurrent expressions, regardless of their idiomaticity, and regardless of their structural status” (p. 990). Similarly, in (17), this concept was referred to as “extended collocations which appear more frequently than expected by chance, helping to shape meanings and contributing to our sense of coherence in a text” (p. 41). Later, different features of lexical bundles (LBs) have been analyzed and reported by some scholars. LBs have been identified as being frequency-driven, continuous, grammatically incomplete, functionally complete, and meaningfully transparent (18, 19). In (20), the authors delved into a corpus of one million running words which was made up of two sub-parts, academic published articles in the field of telecommunication written by English and Chinese professional authors. Their rationale for carrying out the study was that differences regarding the use of LBs have been investigated between L1

and L2 writers; also, the effects of expertise has been unraveled in some studies; however, no study has considered the differences and similarities of professional writers with different linguistic backgrounds. In (21), the writer investigated the nature of formulaic language for different genres of legal corpora (e.g. academic law, case law, legislation, and documents). The researcher reported cross-generic variations such as the fact that the authors use little formulaic language in the genre of academic law, most of which were specialized terminology to represent abstract concepts. He also found that the most utilized structures for LBs were noun phrases. They contended that the majority of the bundles were to fulfill the function of “intangible framing attributes”.

This exploratory study aims to find the most frequent 3-, 4-, 5-, and 6-word lexical bundles in Health Information Management academic texts, i.e. research articles (RAs) and course books (CBs), and analyze the structures and functions of these bundle sizes. The present study aimed to answer the following research questions: What are the most frequent 4-word lexical bundles in the corpus of Health Information Management academic texts? What are the semantic structures of lexical bundles in the corpus of Health Information Management academic texts? What are the pragmatic functions of lexical bundles in the corpus of Health Information Management academic texts?

Methods

The Corpora and Software

In order to develop a Health Information Management Research Article Corpus (HIMRAC), two content experts (PhD holders and university professors) were consulted and they came up with a long list of 15 journals. Pursuing the investigation for compiling a more representative corpus, content experts reduced the list to 5 most prestigious journals which have been published for more than 10 years and have been hosted by international publishers like Elsevier, Sage, Taylor and Francis, Springer, and Pub Med. They also had an impact factor above 1.00. Representativeness, specificity of the corpus, use of whole documents, and availability in electronic form were also among the selection criteria (22). All articles were published between the period of 2000 to 2017, and they had a balanced length of 2000 to 7000 words. Applying the above-mentioned limitations, we came up with 250 research articles.

In order to develop the Health Information Management Course Books Corpus (HIMCBC), we

consulted with two field experts and came up with 5 important course books that are currently being taught worldwide. To ensure the comprehensibility of our corpus, the researchers surfed some colleges' syllabi (e.g. Midland College: available at www.midland.edu/docs/public_information/paci/hb2504/syllabi/.../HITT1311.pdf). These course books were downloaded and prepared to analyze the same as the HIM research papers corpus.

To achieve a more representative source, the two corpora were integrated and the final draft of health information management corpus (HIMC) was devised. The final corpus contained 2,210,466 running words coming from the above-mentioned research articles and course books.

Having downloaded the research articles and course books in PDF format, the researcher copied all mentioned sections of every research article (abstract, introduction, methods, results and discussion) into a Word file and then converted it to a text file so that it could be read by Antconc 3.4.4 Software package which was used to identify lexical bundles in terms of 3-grams to 5-grams. Following (23), a cut-off point frequency of 20 per one million words was set in the software and each bundle should have been present in at least half of the sub-corpora (i.e. the range element was set to 5) to be considered in the study. Then, each bundle was analyzed in its context, using word smith6 software, to possess the functional and structural categorization of each bundle.

Data Analysis

Having passed the above-mentioned criteria, Lexical bundles were analyzed regarding their structural and functional features. A taxonomy of LBs proposed by (16) was used to categorize the bundles functionally as it has been proposed based on academic written corpora; accordingly, it suited this study better. There are three main functional categories in this model:

- **Research-oriented** bundles help writers to structure their activities and experiences of the real world.
- **Text-oriented** bundles are concerned with the organization of the text and its meaning as a message, or an argument.
- **Participant-oriented** bundles focus on the writer or reader of the text.

Structural category, on the other hand, refers to the grammatical forms lexical bundles may assume. Biber et al.'s taxonomy (24) on structural categories of bundles has been chosen to capture the portions of different grammatical categories that serve formulaic language in HIMC as it is one of the

most comprehensive models of structural types of bundles in academic corpora and it has been widely used.

Each lexical bundle was analyzed in its context to unravel its functional and structural categories. The researchers did the analysis separately and a coefficient-correlation of 0.93 showed an acceptable level of agreement. Items of disagreement were consulted with a third expert and assigned to a category after full agreement.

Results

To answer the first research question, the researchers analyzed the corpus with Range software and the results are summarized in the following table. Table 1 demonstrates the most frequent 4-word lexical bundles in the corpus of HIMC. The results indicated that out of 2,210,466 running words in the corpus a total number of 11,323 tokens of 4-word lexical bundles (290 types) were recorded which accounted for 0.51% of the whole corpus.

Table 1: The most frequent 4-word lexical bundles in HIMC

	Forms	F		Forms	F
1.	at the time of	154	150	in a number of	29
2.	it is important to	149	151	in accordance with the	29
3.	the quality of the	135	152	policies and procedures that	29
4.	can be used to	129	153	the design of the	29
5.	the total number of	128	154	this study is to	29
6.	at the same time	127	155	to meet the needs	29
7.	in the context of	124	156	improve quality of care	28
8.	the use of the	120	157	in the area of	28
9.	as a result of	118	158	is shown in figure	28
10.	as part of the	103	159	is the use of	28
11.	on the other hand	93	160	it is important that	28
12.	perceived ease of use	93	161	should be noted that	28
13.	at the end of	90	162	the basis of the	28
14.	the results of the	87	163	the characteristics of the	28
15.	to ensure that the	86	164	the rest of the	28
16.	in the case of	85	165	to be used in	28
17.	the extent to which	83	166	use of the system	28
18.	for the purpose of	82	167	a small number of	27
19.	at the point of	80	168	could be used to	27
20.	the quality of care	77	169	have the potential to	27
21.	an example of a	75	170	on the number of	27
22.	in addition to the	73	171	the national center for	27
23.	is one of the	73	172	the need for a	27
24.	the purpose of the	73	173	to use the system	27
25.	the use of a	73	174	admitted to the hospital	26
26.	improve the quality of	72	175	an example of the	26
27.	in the form of	71	176	as the number of	26
28.	the end of the	71	177	be used in the	26
29.	that can be used	68	178	for example if the	26
30.	at the university of	67	179	for the purposes of	26
31.	one of the most	67	180	in this case the	26
32.	the international classification of	67	181	is a need to	26
33.	to be able to	65	182	it is likely that	26
34.	as shown in table	64	183	so that they can	26
35.	to improve the quality	64	184	the scope of the	26
36.	the purpose of this	63	185	the validity of the	26
37.	this study was to	63	186	adoption of electronic health	25
38.	in a variety of	62	187	for the use of	25
39.	can be found at	61	188	in a way that	25
40.	is based on the	61	189	it is necessary to	25
41.	the aim of this	61	190	the complexity of the	25
42.	the results of this	61	191	the performance of the	25

	Forms	F		Forms	F
43.	to participate in the	61	192	the use of an	25
44.	a wide range of	60	193	the use of technology	25
45.	in the management of	60	194	to the fact that	25
46.	is an example of	60	195	was found to be	25
47.	the point of care	60	196	with the exception of	25
48.	were more likely to	60	197	a better understanding of	24
49.	on the basis of	58	198	a limited number of	24
50.	the development of a	57	199	a part of the	24
51.	in terms of the	56	200	a summary of the	24
52.	through the use of	56	201	an important role in	24
53.	are more likely to	55	202	as part of a	24
54.	be added to the	55	203	as well as for	24
55.	has the potential to	54	204	beyond the scope of	24
56.	in the use of	54	205	for the management of	24
57.	the majority of the	54	206	how to use the	24
58.	in the process of	52	207	in an effort to	24
59.	the beginning of the	52	208	in order to improve	24
60.	at the beginning of	51	209	in such a way	24
61.	should be able to	51	210	is based on a	24
62.	the needs of the	51	211	of the most important	24
63.	be included in the	50	212	the board of directors	24
64.	in the development of	50	213	the cost of the	24
65.	more likely to be	50	214	the decision making process	24
66.	will need to be	49	215	the difference between the	24
67.	results of this study	48	216	the form of a	24
68.	will be able to	48	217	the importance of the	24
69.	in the number of	47	218	the responsibility of the	24
70.	the development of the	47	219	the success of the	24
71.	the time of the	47	220	when a patient is	24
72.	there is a need	47	221	access to the internet	23
73.	in relation to the	46	222	an analysis of the	23
74.	to the use of	46	223	an important part of	23
75.	can be used for	45	224	an increase in the	23
76.	for each of the	45	225	as one of the	23
77.	in this study the	45	226	in a timely manner	23
78.	the implementation of the	45	227	in line with the	23
79.	the role of the	45	228	in the absence of	23
80.	the size of the	45	229	in the areas of	23
81.	the number of patients	43	230	in the design of	23
82.	a large number of	42	231	in the electronic health	23
83.	a review of the	42	232	in this article we	23
84.	meet the needs of	42	233	in this study we	23
85.	the degree to which	42	234	more than half of	23
86.	with respect to the	42	235	should be included in	23
87.	that need to be	41	236	the data in the	23
88.	the accuracy of the	41	237	the way in which	23
89.	the name of the	41	238	to learn more about	23
90.	the nature of the	41	239	to serve as a	23
91.	a result of the	40	240	a period of time	22
92.	aim of this study	40	241	a response rate of	22
93.	have the right to	40	242	adoption and use of	22
94.	the quality of health	40	243.	allows the user to	22
95.	were included in the	40	244	exchange of health information	22
96.	with the use of	40	245	findings of this study	22

	Forms	F		Forms	F
97.	it is possible to	39	246	in each of the	22
98.	are shown in table	38	247	in the present study	22
99.	to determine whether the	38	248	information about the patient	22
100.	use of the internet	38	249	on the part of	22
101.	all of the above	37	250	participate in the study	22
102.	the institute of medicine	37	251	study was approved by	22
103.	agency for healthcare research	36	252	the information needs of	22
104.	average length of stay	36	253	there are a number	22
105.	purpose of this study	36	254	to the best of	22
106.	the content of the	36	255	to the success of	22
107.	within the context of	36	256	when the patient is	22
108.	can be found in	35	257	a high level of	21
109.	the fact that the	35	258	an integral part of	21
110.	the value of the	35	259	are based on the	21
111.	to the number of	35	260	are likely to be	21
112.	an overview of the	34	261	as shown in figure	21
113.	and the number of	34	262	for a variety of	21
114.	in the current study	34	263	in a manner that	21
115.	to be included in	34	264	in response to the	21
116.	at the bottom of	33	265	in the quality of	21
117.	improving the quality of	33	266	is added to the	21
118.	in the course of	33	267	it was found that	21
119.	the continuum of care	33	268	limitations of the study	21
120.	the impact of the	33	269	more than years of	21
121.	activities of daily living	32	270	one of the major	21
122.	have access to the	32	271	studies have shown that	21
123.	in order to provide	32	272	the average number of	21
124.	in the event of	32	273	the creation of a	21
125.	in the field of	32	274	the integrity of the	21
126.	may be used to	32	275	was based on the	21
127.	on a daily basis	32	276	a broad range of	20
128.	on a regular basis	32	277	a great deal of	20
129.	that the use of	32	278	a higher level of	20
130.	with regard to the	32	279	as well as other	20
131.	can be used as	31	280	have an impact on	20
132.	for example if a	31	281	in order to ensure	20
133.	is responsible for the	31	282	is important to note	20
134.	it is possible that	31	283	is part of the	20
135.	it should be noted	31	284	objective of this study	20
136.	quality of health care	31	285	the concept of a	20
137.	quality of patient care	31	286	the findings of this	20
138.	used in this study	31	287	the introduction of the	20
139.	was approved by the	31	288	the top of the	20
140.	et al found that	30	289	to the patient and	20
141.	for the development of	30	290	used to determine the	20
142.	it is difficult to	30			
143.	must be able to	30			
144.	the majority of respondents	30			
145.	a description of the	29			
146.	and the ability to	29			
147.	are presented in table	29			
148.	be used as a	29			
149.	by the number of	29			

As shown in Table 1, the most frequent bundle was “at the time of” with the frequency of 154 through the whole corpus. On the other hand, the least frequent bundles were “a broad range of, a great deal of, a higher level of, as well as other, have an impact on, to ensure, is important to note, is part of the, is part of the, objective of this study, the concept of a, the findings of this, the introduction of the, the top of the, to the patient and, used to determine the” which occurred at the cut-off point of 20. Being analyzed based on their frequency, LBs are then analyzed based on their structural and functional categories and sub-categories.

Structural Categories and Frequencies of 4-word Lexical Bundles

To answer the second research question, we processed syntactic categories of LBs in the corpus by utilizing Biber et al.’s (2004) model of structural categories of bundles. Matching the bundles extracted from HIMC and those of Biber et al.’s study suggested that this model could be a promising method of classifying bundles in the corpus of Health

Information Management. As illustrated in Figure 1, noun phrases and prepositional phrases (e.g. *the total number of, the extent to which, agency for healthcare research*) encompass the lion share in the corpus with 195 types and 7678 tokens (0.34%) of LBs followed by verb phrase (70 types, 2593 tokens and 0.11% of all corpus) and dependent clause LBs (25 types, 981 tokens and 0.04% of all corpus).

Table 2 displays the structural classification of lexical bundles with its sub-categories used in HIMC.

It has been shown that “verb phrases with passive and non-passive verbs” incorporate the most frequent categories of verb phrase fragment bundles (F=1102 and 1117, respectively). Meanwhile, the most frequent clausal category seems to be “To-clause fragments” with the frequency of 582. Notably, “Noun phrase with of phrase fragments” following by “prepositional phrase expressions” with the frequency of 3962 and 2869 were the most frequent syntactic categories of bundles which incorporate more than 0.17% and 0.12% of the whole corpus, respectively.

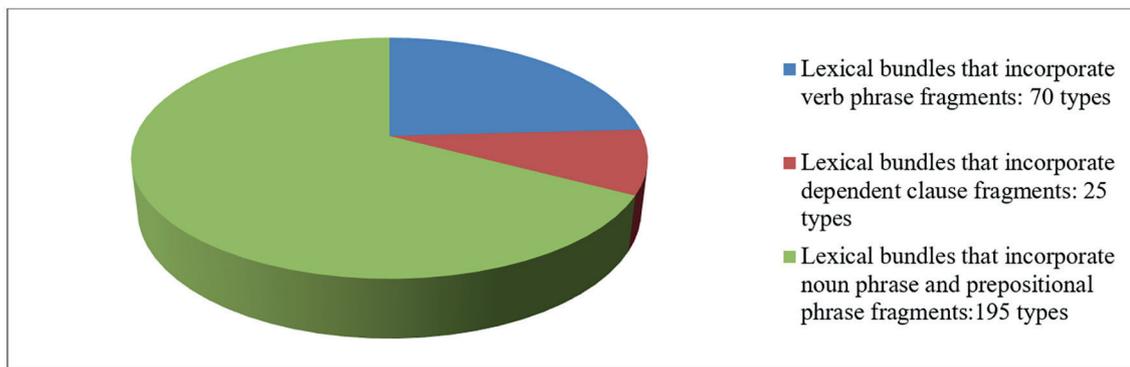


Figure 1: Distribution of structural categories of LBs

Table 2: Structural categories of LBs in HIMC

Structural types	Sub-types	Example Bundles	F.	%
1. Lexical bundles that incorporate verb phrase fragments	1a. 1st/2nd person pronoun+ VP fragments	<i>this study was to</i>	63	0.002
	1b. 3rd person pronoun+ VP fragments	<i>it should be noted</i>	261	0.011
	1c. Discourse markers+ VP fragments	<i>so that they can</i>	50	0.002
	1d. Verb phrase (with non - passive verbs)	<i>has the potential to</i>	1117	0.050
	1e. Verb phrase (with passive verbs)	<i>can be used to</i>	1102	0.049
	1f. Yes. No question fragments			
	1g. WH-question fragments			
2. Lexical bundles that incorporate dependent clause fragments	2a. 1st/ 2nd person pronoun+ dependent clause fragment	<i>it is important to</i>	188	0.008
	2b. WH-clause fragments	<i>how to use the</i>	70	0.003
	2c. If- clause fragments			
	2d. To-clause fragments	<i>to ensure that the</i>	582	0.026
	2e. That-clause fragments	<i>that can be used</i>	141	0.006
3. Lexical bundles that incorporate noun phrase and prepositional phrase fragments	3a. Noun phrase with of phrase fragments	<i>the quality of the</i>	3962	0.179
	3b. Noun phrase with other post-modifier fragments	<i>the extent to which</i>	247	0.011
	3c. Other noun phrase expressions	<i>agency for healthcare research</i>	412	0.018
	3d. Prepositional phrase expressions	<i>at the time of</i>	2869	0.129
	3e. Comparative expressions	<i>more than half of</i>	188	0.008

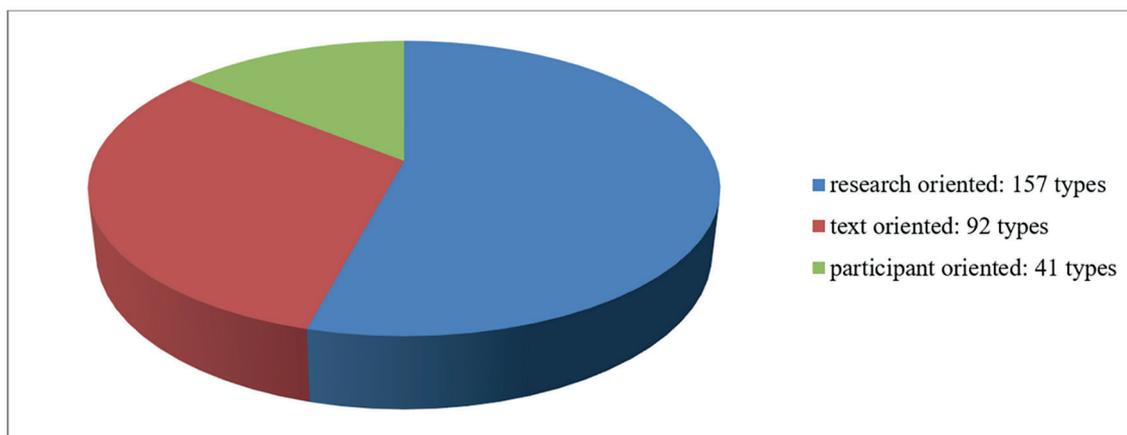


Figure 2: Distribution of functional categories of LBs

Table 3: Structural categories of LBs in HIMC

Functional categories	Sub-categories	Example Bundles	F.	%
Research-oriented	Location	<i>at the bottom of</i>	896	0.040
	Procedure	<i>for the purpose of</i>	1512	0.068
	Quantification	<i>the total number of</i>	1159	0.052
	Description	<i>the quality of care</i>	1501	0.067
	Topic	<i>the international classification of</i>	968	0.043
Text-oriented	Transition signals	<i>on the other hand,</i>	460	0.020
	Resultative signals	<i>as a result of</i>	974	0.044
	Structuring	<i>as shown in table</i>	803	0.036
	Framing	<i>the extent to which</i>	1445	0.065
Participant-oriented	Stance features	<i>it is important to</i>	1370	0.061
	Engagement features	<i>should be noted that</i>	280	0.012

Functional Categories and Frequencies of 4-word Lexical Bundles

Different registers tend to show different functional behaviors regarding the functions that LBs serve (16, 24). Accordingly, the present study investigated the functional categories of LBs in the corpus of Health Information Management (the third research question) using Hyland's 2008a taxonomy, which included three major categorizations: research-oriented bundles, text-oriented bundles, and participant-oriented bundles with various sub-categories for each. The study revealed that research-oriented LBs (e.g. *at the time of, in the context of, as part of a, etc*) with 157 types and 6036 tokens (0.27%) had the highest frequency followed by text-oriented LBs (92 types, 3682 tokens and 0.16%), and participant-oriented LBs (41 types, 1650 tokens and 0.07%). Figure 2 represents the results of functional categorization of the bundles.

Table 3 also illustrates the functional classification of all target bundles identified in Health Information Management articles with their subcategories. These functions are discussed in more details below.

Discussion

The current study aimed to investigate the most frequent 4-word lexical bundles, semantic structures of lexical bundles, and pragmatic functions of lexical bundles in the corpus of Health Information Management academic texts. The findings revealed that the most frequent bundle was "*at the time of*". According to (16), four-word bundles present a clearer range of structures and functions than other types of bundles. He argues that the increase in length from three to four words allows these bundles to be more specific than their shorter bundles; this makes much easier identification of their functions. Further support for the use of four-word bundles belongs to (5). She highlighted that so many four-word bundles contained three-word bundles within themselves. From a pedagogical perspective, then, four-word bundles might seem to be a more efficient target structure than shorter bundles as learners would naturally be exposed to shorter bundles through the learning of longer bundles.

In addition, the results indicated that "*verb phrases with passive and non-passive verbs*" incorporated the most frequent category of *verb*

phrase fragment bundles, “*To-clause fragments*” was the most frequent category of clausal bundles, and “*Noun phrase with of phrase fragments*” followed by “*prepositional phrase expressions*” were the most frequent syntactic categories of bundles. Noun phrases and prepositional phrases had the highest frequency in the corpus followed by verb phrases. The results are consistent with those of previous studies (5, 14, 20). In the same vein, in these studies, phrasal bundles dominated the corpus, both in terms of variety and frequency. The dominance of phrasal bundles can be attributed to the highly informational focus of academic written texts of Health Information Management. As shown in (20), “careful integration of information in academic prose requires utilizing noun phrases and prepositional phrases, leading to a shift from clausal style to phrasal style in academic prose”. However, the finding is contrary to the results of (14) and (24). They concluded that clausal bundles were not varied and frequent.

Furthermore, the results of the current study are in line with another study (21). He found out that noun phrase bundles were the most frequent bundles in the corpus. The findings are also in line with (16). In (16), it is pointed out that social science research articles utilize a large number of bundles beginning with a prepositional phrase due to logical or textual connections among propositional components.

Finally, the findings showed that the pragmatic functions of lexical bundles in the corpus of Health Information Management academic texts consisted of three major categorizations: research-oriented bundles, text-oriented bundles, and participant-oriented, respectively. The findings of previous research suggest that while the most frequent functional type of bundle in academic prose is the referential bundle (23) research carried out on a corpus comprising academic texts from both soft and hard sciences, in social sciences the most prominent function of the lexical bundle is that of discourse organizer (16). According to (16), the reason can be because “the more discursive and evaluative patterns of argument is in the soft knowledge fields, persuasion is more explicitly interpretative and less empiricist” (p. 16).

Moreover, the results of this study are in the same line with those of the research conducted by (17). He concluded that framing devices comprised a high proportion of text-oriented bundles. Based on his findings, writers in some disciplines such as applied linguistics and business studies mostly employ text-oriented bundles.

Conclusion

The results of the current study have some implications. First, pedagogically, it would be fruitful if ESP teachers take syntactic and pragmatic aspects of lexical bundles into consideration in teaching syllabuses as a learning input. They should utilize activities that raise awareness toward lexical bundles and show their structures and functions. Second, the output of the present study would assist Health Information Management researchers in particular to produce more coherent and native-like academic texts. Third, in the information science field, semantics has been at the center of attention in retrieval systems. Therefore, in addition to semantics, syntactic structures would play crucial roles in the amelioration of information systems. This study can be regarded as an introductory step towards more cooperation among health information, management, and linguistics professionals. Accordingly, it would help researchers to make new patterns by using structural and functional analysis of such lexical bundles. Fourth, even though this research explored the four-word lexical bundles, it would be beneficial that future studies identify the lexical bundles in each field separately and compare them with each other.

Future studies are suggested to be carried out on comparing lexical bundles, their functions, and structures in different sections of academic texts. Additionally, the linguistic methods have to resolve word ambiguities and/or generate relevant relationships between words. The development of a sophisticated linguistic retrieval system is a tough task and it needs complex knowledge bases of semantic and syntactic information. Finally, since the applications of this research might concern the place of lexical bundles in the teaching of academic English and may prove that exposure to expert academic writing does not lead to successful acquisition of lexical bundles typical of academic discourse, there is an indispensable need to explore the effective ways for including overt instruction in discipline- and genre-specific bundles in ESP courses.

Authors' Contribution

Both authors contributed to the design and analysis of the study, and read and approved the final manuscript.

Conflict of Interest: None declared.

References

1. Abdollahpour Z, Gholami J. Building blocks of medical abstracts: Frequency, functions and

- structures of lexical bundles. *Asian ESP Journal*. 2018;14(1):82-110.
2. Björkman B. English as an academic lingua franca: An investigation of form and communicative effectiveness: Walter de Gruyter; 2013.
 3. Swales JM. Research genres: Explorations and applications: Cambridge University Press; 2004.
 4. Hyland K. Representing readers in writing: Student and expert practices. *Linguistics and Education*. 2005;16(4):363-77.
 5. Cortes V. The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for academic purposes*. 2013;12(1):33-43.
 6. Dastjerdi HV, Rafiee M. Corpus study of lexical bundles: Journalistic discourse in focus. *Asian ESP Journal*. 2011;7(3):59-78.
 7. Simpson-Vlach R, Ellis NC. An academic formulas list: New methods in phraseology research. *Applied linguistics*. 2010;31(4):487-512.
 8. Wray A. Formulaic sequences in second language teaching: Principle and practice. *Applied linguistics*. 2000;21(4):463-89.
 9. Simpson R. Stylistic features of academic speech: The role of formulaic expressions. *Discourse in the professions: Perspectives from corpus linguistics*. 2004:37-64..
 10. Moon R. Textual aspects of fixed expressions in learners' dictionaries. *Vocabulary and applied linguistics*: Springer; 1992. p. 13-27.
 11. Nattinger J. Some current trends in vocabulary teaching. *Vocabulary and language teaching*: Routledge; 2014. p. 62-82.
 12. Cowie AP. Multiword lexical units and communicative language teaching. *Vocabulary and applied linguistics*: Springer; 1992. p. 1-12.
 13. Cheng W, Greaves C, Sinclair JM, Warren M. Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams. *Applied linguistics*. 2009;30(2):236-52.
 14. Biber D, Conrad S. Lexical bundles in conversation and academic prose. *Out of corpora*: Brill; 1999. p. 181-90.
 15. Cortes V. Lexical bundles in freshman composition. *Using corpora to explore linguistic variation*. 2002;9:131-45.
 16. Hyland K. As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*. 2008;27(1):4-21.
 17. Hyland K. Academic clusters: Text patterning in published and postgraduate writing. *International journal of applied linguistics*. 2008;18(1):41-62.
 18. Biber D, Barbieri F. Lexical bundles in university spoken and written registers. *English for specific purposes*. 2007;26(3):263-86.
 19. Chen YH, Baker P. Lexical bundles in L1 and L2 academic writing.
 20. Pan F, Reppen R, Biber D. Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*. 2016;21:60-71.
 21. Breeze R. Lexical bundles across four legal genres. *International journal of corpus linguistics*. 2013;18(2):229-53.
 22. Sinclair J. *Corpus, concordance, collocation*. (No Title). 1991.
 23. Biber D. University language: A Corpus-based study of spoken and written registers. *TESL-EJ*. 2007;11(2).
 24. Biber D, Conrad S, Cortes V. If you look at...: Lexical bundles in university teaching and textbooks. *Applied linguistics*. 2004;25(3):371-405.