

Prevalence of Prefabricated Structures in Academic Discourse: A Corpus-Based Study

Muhammad Yousaf¹ & Wasima Shehzad¹

¹ Air University, Islamabad

Correspondence: Muhammad Yousaf, Air University, Islamabad. E-mail: usaf_pak@yahoo.com

Received: May 12, 2018 Accepted: July 16, 2018 Online Published: August 3, 2018

doi:10.5539/ijel.v8n5p297 URL: <https://doi.org/10.5539/ijel.v8n5p297>

Abstract

Multiword structures that appear in a text more than expected frequency are called lexical bundles. These prefabricated structures vary in length but the most common lexical bundles are four-word lexical bundles which have been explored by many scholars worldwide. The current study aimed to explore five-word lexical bundles, which have lesser been researched. For this purpose a corpus of about 4.7 million words was compiled which consists of PhD dissertations written in Pakistani context. Moreover, the dissertations were selected from three different disciplines to make the study cross disciplinary. The corpus was analyzed according to the taxonomy of lexical bundles given by Biber et al. (1999). The analysis shows that lexical bundles are predominant feature of PhD dissertations in Pakistani context. Moreover, frequency of lexical bundles varies from discipline to discipline, and the structural variation of lexical bundles is also found across disciplines. Dominant structures across disciplines are not fixed as Prepositional Phrase Fragments is the dominant category in the corpus of English Studies and corpus of Social Sciences, whereas, Verb Phrase Fragments is the dominant category in the corpus of Bio Sciences.

Keywords: lexical bundles, polyword structures, multiword structures, prefabricated chunks

1. Introduction

English is a very important language for academics as it is understood by a vast majority of people in the world. Moreover, the new knowledge that is produced is either in English at first hand or is translated in English to reach a wider audience. This has benefited scholars across the world as they address wide audience without getting restricted by geographical or, in specific, linguistic boundaries. In addition, they can also have an easy access to the knowledge produced anywhere in the world without getting in the trouble of having it translated themselves. This makes English as lingua franca of the academics (Hyland, 2009a). Along with all the advantages of English in the world of academia there is a disadvantage in form discourse of native and non-native speakers or writers. This discourse gives privilege to those who have native like proficiency in English language. In this scenario, many of the scholars and students are left out because of having nonnative expression. To overcome this difficulty, scholars have been striving hard for last few decades to understand the reasons of lacking native like proficiency in English and finding out ways to help learners to get native like expression.

Although, there has been a lot of research in the above mentioned area but the advancement in the field of computers has brought revolution in the field as the researchers have been enabled to compare a large bulk of data from native and non-native context. Along with many other revelations, it has been explored that the language structures are not novel every time instead these are prefabricated in nature. After years of research the researchers and scholars have come to conclusion that these prefabricated structures enhance accuracy and fluency of the speaker of a language. The native speakers/writers of English use these structure frequently and they are also well aware of how and when to use them, which is lacking in non-native expressions. Moreover, researches (such as Hyland, 2008; Chen & Baker, 2010) show that use of prefabricated structures varies not only in native and non-native expression but it also has variation from discipline to discipline, and from one level of education to the other. There has been research exploring use and variation of prefabricated structures across the world but no such great efforts in Pakistani context are found. Hence, this study is an attempt to explore the prefabricated structures called lexical bundles in Pakistani academic setting.

The study has been limited to PhD dissertations from three different disciplines of academic discourse, which are: English Studies, Social Sciences and Bio Sciences. The research questions set for the current study are as follows:

- 1) How far does academic discourse rely on five word lexical bundles?
- 2) What are structural similarities/differences of lexical bundles across disciplines?
- 3) Which of the structural categories are dominant in different disciplines?

2. Literature Review

Over the period of last few decades there has been increasing interest in the co-occurring words. These words have been explored and discussed from different perspectives. Among the first to draw attention towards such combinations was Firth (1957). These combinations were observed to occur more with than expected frequency in the given text, and Firth (1957) termed them as collocations. Despite of naming them he did not give a proper definition to these word combinations. The interest in the word combination units kept on increasing by the time Halliday (1964) et al. defined collocation as a group of words having a tendency to co-occur with each other. Similarly, some other scholars (such as Yorio, 1980; Cowie, 1988; Moon, 1992, Cortes 2004) suggested these expressions as conventionalized language forms, speech formulas, readymade expressions, multi-word units and fixed expressions. Similar to the linguistic terms used for these expression, these multiword units have been studied from different aspects and perspectives but researchers and scholars have not given a unanimous definition for categorizing similar or distinguishing types of word combinations.

The focus on the research increased with the research in this field by Biber et al. (1999) who explored a large amount of data with the help of computer softwares that was not possible before the advancement in the field of information technology. The computer tools used by them helped explore word patterns of varying lengths in a large bulk of data. As described above, there have been various names for the combinations of such words prior to Biber et al. (1999:990), such as prefabricated structures, fixed expressions, formulaic structures, lexical phrases etc., but they called them lexical bundles. According to Biber et al. (1999), regardless of their idiomaticity and their lexical status, lexical bundles are recurrent word combinations that co-occur in natural language use. Moreover, these are frequency-driven word combinations that comprise of three or more than three words functioning as building blocks for any text (Biber et al., 2004). Similarly, Nattinger and DeCarrico (1992) explored the multiword expressions and called them prefabricated language. According to them it could be easily retrieved as this exists ready made in the mind. This helps the speakers to get speak fluently owing to the fact of readiness (Yorio, 1980). Such expressions save the speaker from the efforts to make new structure every time they engage in the speech activity. Similarly, it also facilitates the listener to interpret the message quickly. Hence, communication becomes fluent, spontaneous and easy for the communication counterparts. Moreover, the speakers find prefabricated chunks of language helpful as they save them from the time required to find appropriate choices of words and word forms as Cortes (2002) also states that the lexical bundles are helpful in fluent speaking and writing and also accelerate language acquisition. Along with this, lexical bundles help establish contact in the communication as these perform phatic function (Drazdauskien, 1981). Moreover, a study by Yu & Kim (2017) states that learners of different proficiencies show significant improvement in their error control if they are taught with the help of lexical bundles. The most frequent error by learners in the study was omission of articles which are within bundles which can be avoided if are taught with the help of lexical bundles. The results also highlight that bundles, article-including expressions, can be helpful to teach article uses in context.

It is important to note down the difference between different types of multi word units. A point in case is idioms which have fixed structure and meanings. Whereas, lexical bundles have their frequency count that makes them different from the other fixed structures. Apart from the frequency the other difference is the method of their extraction (Cortes, 2002; Biber, 1999).

To be identified as lexical bundle a multiword unit needs to occur at least ten times in a million words and five texts of the same genre. Hence, the word structures that meet this frequency cut off are known as lexical bundles (Biber et al., 1999).

After the identification of the bundles the next stage is for the categorization of lexical bundles as per their structures and functions. These can be categorized in polyword units, sentence building expressions, phrasal constraints and institutionalized expressions (Nattinger & DeCarrico, 1992). Among these types, polyword units function similar to individual words, on the other hand, sentence builders are short phrases that become a basis for a sentence. The third type of the categorization is phrasal constraint structures which are phrases of varied

length which most of the time are sentence beginners, and the last type of is institutionalized structures that are basically short sentences which are fixed. Among these polyword units and phrasal constraints can be both canonical and non-canonical, and are continuous. But in respect of their variability, polyword units show no variability, whereas, phrasal constraints allow variation. On the other hand, institutionalized expressions are canonical and invariable, whereas, sentence builders can be canonical or non-canonical and continuous or discontinuous. Moreover, sentence builders show a lot of variation too. Although this categorization is an important contribution for the explanation and investigation of lexical bundles, but still it is not a clear scheme of categorization where sometimes a bundle may belong to multiple categories or it may not fall in any of these categories.

In this regard the other problem in the placement of prefabricate structures in different categories is because of the exclusion of discourse function from the features of these structures as discourse is decisive in the production of language variety in terms of its form and function (Bhatia, 1993). For example, research in the field of lexical bundles show that the bundles explored in science text contain assertive sentences, similarly, Hyland (1998) shows the descriptions of the scientific procedures and results, moreover, these also have the features of academic discourse which have been highlighted by other scholars (such as Nattinger & Decarrico, 1992). Majority of the lexical bundles found in academic discourse are fragments of noun phrase (Cortes, 2004).

The functional aspect of lexical bundles in academic discourse shows that these work as discourse devices to connect different part of discourse and help in meaning making process. Biber (1999) explored conversation and academic prose for lexical bundles, whereas, Conrad and Biber (2005) researched students' writing in the disciplines of history and biology for lexical bundles. They proposed that lexical bundles function as building blocks of academic discourse. A qualitative analysis was conducted by Wang & Ying (2017) for functional variations of *I don't know if*. The study shows that genre and discipline are both very important factors in understanding academic ELF communication and idiomaticity which cannot be ignored; moreover, study of lexical bundles provides useful understanding of genre and disciplinary variation.

The above discussion shows that a lot of research has been carried out worldwide but perhaps no attention has been paid to explore lexical bundles in Pakistani academic discourse. Therefore, the current study attempts to investigate lexical bundles in Pakistani academic discourse. The study has been limited to PhD dissertations from three different disciplines of academic discourse, which are: English studies, Social Sciences and Bio Sciences, and it focuses only on five-word lexical bundles.

3. Research Methodology

3.1 Corpus

For the purpose of this study a corpus of PhD dissertation has been compiled which consists of 4.7 million words. For the collection of data three disciplines i.e. English Studies, Social Science and Bio-Sciences were selected. Furthermore, thirty dissertations from each discipline (ten from three subjects each) were included in the corpus. The dissertations selected for this study were the latest ten dissertations in each subject area. The details of the corpus data are as follows:

Table 1. Corpus for the study

English Studies	Linguistics	Literature	ELT
Social Sciences	Political Science	Education	Psychology
Bio Sciences	Bio Technology	Botanical Science	Zoological Science

3.2 Tools for Corpus Analysis

For the analysis of the corpus AntConc 3.4.4w (Windows) 2014 software was used. This corpus software helped in identification of the lexical bundles as it can show frequency as well as range of a structure in a number of texts which is the necessary of the identification and analysis of lexical bundles.

3.3 Theoretical Framework

The lexical bundles which were extracted from the corpus are categorized according to Biber et al.'s (1999) taxonomy. The lexical bundles which did not fall in any of the existing categories were placed in new categories that made the analysis in depth. The details of the categorization have been provided in the section of results and discussion.

4. Results and Discussion

This section has further been divided in four parts where three parts consist of the result and discussion of lexical bundles in three different disciplines, whereas, the final part deals with the comparison of lexical bundles across disciplines.

For the identification of five-word lexical bundles, the length of N-Gram was fixed at 5, the range on 5, and frequency on 10 per million. This setting gave a list of lexical bundles which were then categorized according to Biber et al.'s (1999) taxonomy according to their structures. Following is the detailed discussion:

4.1 Five Words Lexical Bundles in English Studies

Table 2 shows that the three subjects included in the discipline of English Studies all have lexical bundles. The number of lexical bundles is not very high, which is unlike the case with 4 word lexical bundles (Yousaf, 2018), but still all the categories have representation of lexical bundles. The table shows that the largest category of lexical bundles in linguistics is *Prepositional Phrase Fragments* which consists of 17 lexical bundles, that makes 50% of the total lexical bundles in the corpus of linguistics. The same category is the dominant category in the corpus of Literature where it has 05 lexical bundles which are 45.5% of the total lexical bundles in the corpus of literature. Similarly, the corpus of ELT too has *Prepositional Phrase Fragments* as the largest category which consists of 09 lexical bundles, which make 31% of the total five-word lexical bundles in the corpus of ELT in this study.

The second largest category of lexical bundles in the corpus of linguistics is *Verb Phrase Fragments*. It has 05 lexical bundles which is 14.70% of the five-word lexical bundles in the corpus of linguistics. On the other hand, *Other Expressions* is the dominant category in the corpus of literature; this category has 03 five-word lexical bundles which is 27.27% of the total bundles in the corpus. Moreover, making 27.59% of the total bundles in this corpus, *Noun Phrase Fragments* is the largest category in the Corpus of ELT which has 08 five-word lexical bundles. This shows that unlike the most dominant category, the second highest number of lexical bundles do not belong to the same category which confirms the disciplinary variation.

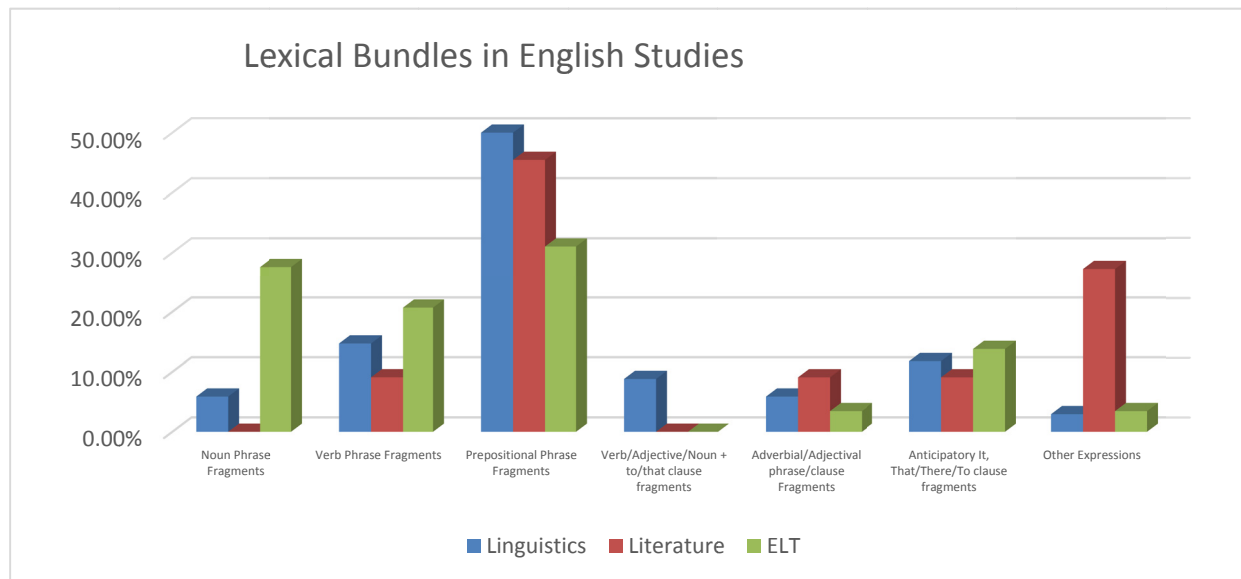
The third largest category of five-word lexical bundles in the corpus of Linguistics is *Anticipatory It, That/There/To Clause Fragments* consisting of 04 bundles it makes 11.76%% of the total lexical bundles in the corpus. Whereas, the corpus of Literature, in this study, has 01 bundle each in *Verb Phrase Fragments, Adverbial/Adjectival Phrase/Clause Fragments* and *Anticipatory It, That/There/To Clause Fragments* which is 09.09% in each category. This number is very small as compared to the four-word lexical bundles in the same corpus. On the other hand, the third largest category in the corpus of ELT, *Verb Phrase Fragments*, has 06 lexical bundles that make 20.69% of the total five-word lexical bundles in the corpus of ELT in this study, which is a comparatively higher percentage.

Lastly, in the table 2 we can see the remaining categories have very low number of five-word lexical bundles or there are no lexical bundles at all. Particularly, *Verb/Adjective/Noun + to/that Clause Fragments* has only 03 lexical bundles in the corpus of Linguistics, and the remaining two corpora have no lexical bundles in this category. Similarly, there are no lexical bundles in *Noun Phrase Fragments* in the corpus of Literature.

Table 2. Five words lexical bundles in English studies

	Noun Phrase Fragments	Verb Phrase Fragments	Prepositional Phrase Fragments	Verb/Adjective/ Noun + to/that clause fragments	Adverbial/ Adjectival phrase/clause Fragments	Anticipatory It, That/There/To clause fragments	Other Expressions
Linguistics	02 (5.88%).	05 (14.70%)	17 (50%)	03 (8.82%)	02 (5.88%)	04 (11.76%)	01 (2.94%)
Literature	00	01 (9.09%)	05 (45.45%)	00	01 (9.09%)	01 (9.09%)	03 (27.27%)
ELT	08 (27.59%)	6 (20.69%)	09 (31.03%)	00	01 (3.45%)	04 (13.79%)	01 (3.45%)
Total	10	12	31	03	04	09	05

The Graph 1 below shows the comparison of five-word lexical bundles in the three corpora of English Studies. The height of bars shows that there is different number of lexical bundles in in different categories across the subjects of English Studies. The graph shows that *Prepositional Phrase Fragment* has highest number of lexical bundles in all three subjects that make this category very significant for the corpus of English Studies.



Graph 1. Five word lexical bundles in English studies

4.2 Five Words Lexical Bundles in Social Sciences

Table 3 shows that the three subjects included in the discipline of Social Sciences all have lexical five-word bundles. The table shows that the largest category of lexical bundles in Political Science is *Prepositional Phrase Fragments* which consists of 19 lexical bundles, which make 46.34% of the total lexical bundles in the corpus of Political Science. On the other hand, *Noun Phrase Fragments* is the dominant category in the corpus of Education which has 06 lexical bundles which are 46.15% of the total lexical bundles in the corpus of Education. Similarly, the corpus of Psychology too has *Noun Phrase Fragments* as the largest category. Consisting of 03 lexical bundles it makes 33.33% of the total five-word lexical bundles in the corpus of Psychology in this study. This shows that unlike English Studies, the highest number of lexical bundles does not belong to the same category in three different subjects of the corpus of Social Sciences.

The second largest category of lexical bundles in the corpus of Political Science is *Noun Phrase Fragments*. It has 13 lexical bundles which is 31.70% of the five-word lexical bundles in the corpus of Political Science. On the other hand, *Prepositional Phrase Fragments* is the dominant category in the corpus of Education. This category has 03 five-word lexical bundles which is 23.08% of the total bundles in the corpus. Similarly, *Prepositional Phrase Fragments* is the largest category in the Corpus of Psychology too where, making 22.22% of the total bundles in this corpus, it has 02 five-word lexical bundles. Moreover, *Adverbial/Adjectival Phrase/Clause Fragments* also has 02 lexical bundles in the corpus of Psychology.

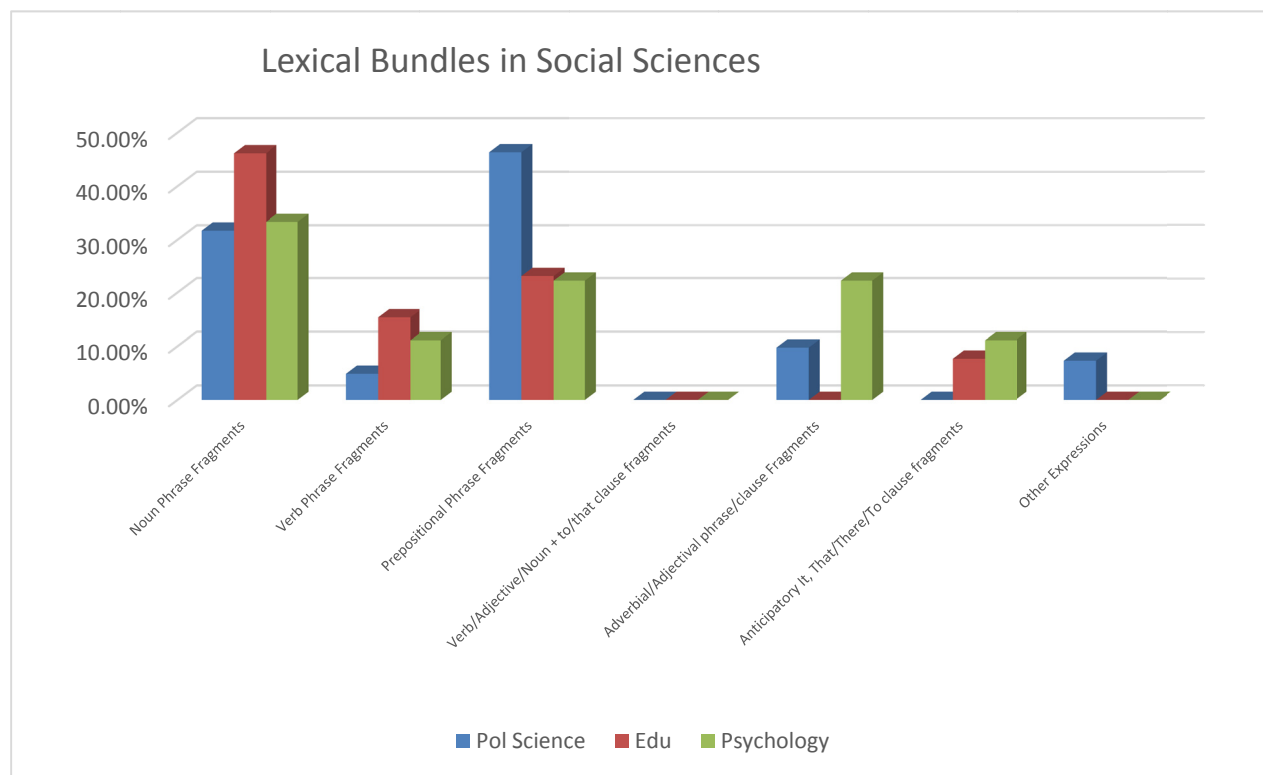
The third largest category of five-word lexical bundles in the corpus of Political Science is *Adverbial/Adjectival Phrase/Clause Fragments*, consisting of 04 bundles it makes 09.76% of the total lexical bundles. Whereas, the corpus of Education has 02 bundles in *Verb Phrase Fragments*, which is 15.38 % of the total bundles in the corpus of Education. On the other hand, *Verb Phrase Fragments* and *Anticipatory It, That/There/To Clause Fragments* both have a lexical bundle each that makes 11.11% of the total five-word lexical bundles in the corpus of Psychology in each category. This shows that the third largest category in the corpus of Psychology has a very low number of lexical bundles.

Lastly, in the table we can see the remaining categories have very low number of five-word lexical bundles or there are no lexical bundles at all. Notably, *Verb/Adjective/Noun + to/that Clause Fragments* has no five-word lexical bundle in any of the three corpora. Similarly, there are no lexical bundles in *Other Expressions* in the corpora of Education and Psychology, and similarly, there are no lexical bundles belonging to *Adverbial/Adjectival Phrase/Clause Fragments* in the corpus of Education in this study. This shows that the structural variety of lexical bundles in Social Sciences is lesser as compared to English Studies.

Table 3. Five words lexical bundles in social sciences

	Noun Phrase Fragments	Verb Phrase Fragments	Prepositional Phrase Fragments	Verb/Adjective/ Noun + to/that clause fragments	Adverbial/ Adjectival phrase/clause Fragments	Anticipatory It, That/There/To clause fragments	Other Expressions
Pol Science	13 (31.70%)	02 (4.88%)	19 (46.34%)	00	04 (09.76%)	01 (2.44%)	03 (07.32%)
Education	06 (46.15%)	02 (15.38%)	03 (23.08%)	00	00	01 (07.69%)	00
Psychology	03 (33.33%)	1 (11.11%)	02 (22.22%)	00	02 (22.22%)	01 (11.11%)	00
Total	22	5	24	00	6	3	3

The Graph 2 below shows the comparison of five-word lexical bundles in the three corpora of Social Sciences. The bars show that there are different number of lexical bundles in different categories across the subjects of Social Sciences. The graph shows that *Noun Phrase Fragments* collectively has highest number of lexical bundles that makes it a very significant category for the corpus of Social Sciences.



Graph 2. Five word lexical bundles in social sciences

4.3 Five-Words Lexical Bundles in Bio Sciences

Table 5 below shows that the three subjects included in the discipline of Bio Sciences all have five-word lexical bundles, but the overall number of lexical bundles is low as compared to English Studies and Social Sciences. The table shows that the largest category of lexical bundles in Bio Tech are *Verb Phrase Fragments* and *Anticipatory It, That/There/To Clause Fragments*, both of which consist of 04 lexical bundles, that makes 26.67% of the total lexical bundles in the corpus of Bio Tech in each category. Similarly, *Verb Phrase Fragments* is the dominant category in the corpus of Botanical Science too where it has 03 lexical bundles which are 42.85% of the total lexical bundles in the corpus. On the other hand, the corpus of Zoological Science has *Noun Phrase Fragments* as the largest category which consists of 02 lexical bundles, that makes 50% of the total five-word lexical bundles in the corpus of Zoological Science in this study.

The second largest category of lexical bundles in the corpus of Bio Tech is *Other Expressions*. It has 03 lexical bundles which is 20% of the five-word lexical bundles in the corpus. On the other hand, *Prepositional Phrase Fragments* and *Noun Phrase Fragments* in the corpus of Botanical Science have 02 lexical bundles each which is 28.57 % of the total bundles in the corpus, separately in both categories. Similarly, *Prepositional Phrase Fragments* and *Verb Phrase Fragments* have a lexical bundle each in the corpus of Zoological Sciences which is 25% of the total bundles in each category.

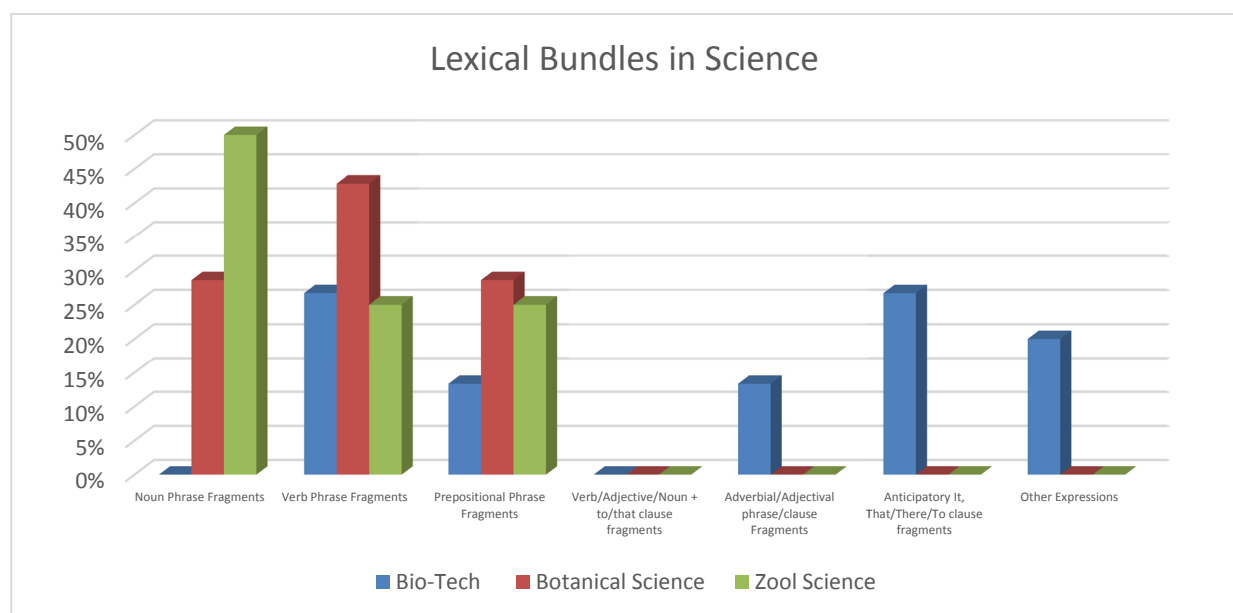
The third largest category of five-word lexical bundles in the corpus of Bio Tech is *Prepositional Phrase Fragments* which consists of 02 lexical bundles, and hence, makes 13.33% of the total lexical bundles. Whereas, the corpora of Botanical Science and Zoological science do not have lexical bundles in any other category.

The table shows that the remaining categories have no lexical bundles at all. Notably, *Verb/Adjective/Noun + to/that Clause Fragments* has no five-word lexical bundle in any of the three corpora. Similarly, there are no lexical bundles in *Other Expressions* and *Adverbial/Adjectival Phrase/Clause Fragments* in the corpora of Botanical Science and Zoological Science. Moreover, there are no lexical bundles present in *Noun Phrase Fragments* in the corpus of Bio Tech in this study.

Table 4. Five words lexical bundles in sciences

	Noun Phrase Fragments	Verb Phrase Fragments	Prepositional Phrase Fragments	Verb/Adjective/Noun + to/that clause fragments	Adverbial/Adjectival phrase/clause Fragments	Anticipatory It, That/There/To clause fragments	Other Expressions
Bio-Tech	00	04 (26.67%)	02 (13.33%)	00	02 (13.33%)	04 (26.67%)	03 (20%)
Botanical Science	02 (28.57%)	03 (42.85%).	02 (28.57%)	00	00	00	00
Zool Science	02 (50%)	01 (25%)	01 (25%)	00	00	00	00
Total	4	8	5	00	2	4	3

The Graph 3 below shows the comparison of five-word lexical bundles in the three corpora of Bio Sciences. The bars show that there is difference in number of lexical bundles in different categories across the subjects of Bio Science. The graph shows that *Verb Phrase Fragments* overall has highest number of lexical bundles that makes this category very significant for the corpus of Bio Sciences. Moreover, *Noun Phrase Fragments* is the dominant category in the corpus of Zoological Science. But overall the corpus of Bio Sciences has very low number of five-word lexical bundles.



Graph 3. Five word lexical bundles in social sciences

4.4 Five Words Lexical Bundles Across the Disciplines

Table 5 shows that all of the three disciplines included in this study have five-word lexical bundles. The table shows that the largest category of lexical bundles in English Studies is *Prepositional Phrase Fragments* which consists of 31 lexical bundles, that makes 41.89% of the total lexical bundles in the corpus. Similarly, *Prepositional Phrase Fragments* is the dominant category in the corpus of Social Sciences too where it has 24 lexical bundles which are 38.09% of the total lexical bundles in the corpus under study. On the other hand, the corpus of Bio Sciences has *Verb Phrase Fragments* as the largest category which consists of 08 lexical bundles, that makes 30.77% of the total five-word lexical bundles in the corpus. This shows that English Studies and Social Sciences both have similarity in the use of majority of lexical bundles whereas Bio Science are different.

The second largest category of lexical bundles in the corpus of English Studies is *Verb Phrase Fragments*. It has 12 lexical bundles which is 16.22% of the five-word lexical bundles in the corpus of English Studies. On the other hand, *Noun Phrase Fragments* is the dominant category in the corpus of Social Sciences. This has 22 five-word lexical bundles which is 34.92% of the total bundles in the corpus. However, *Prepositional Phrase Fragments* is the largest category in the Corpus of Bio Sciences where it has 05 five-word lexical bundles that make 19.23% of the total bundles in this corpus.

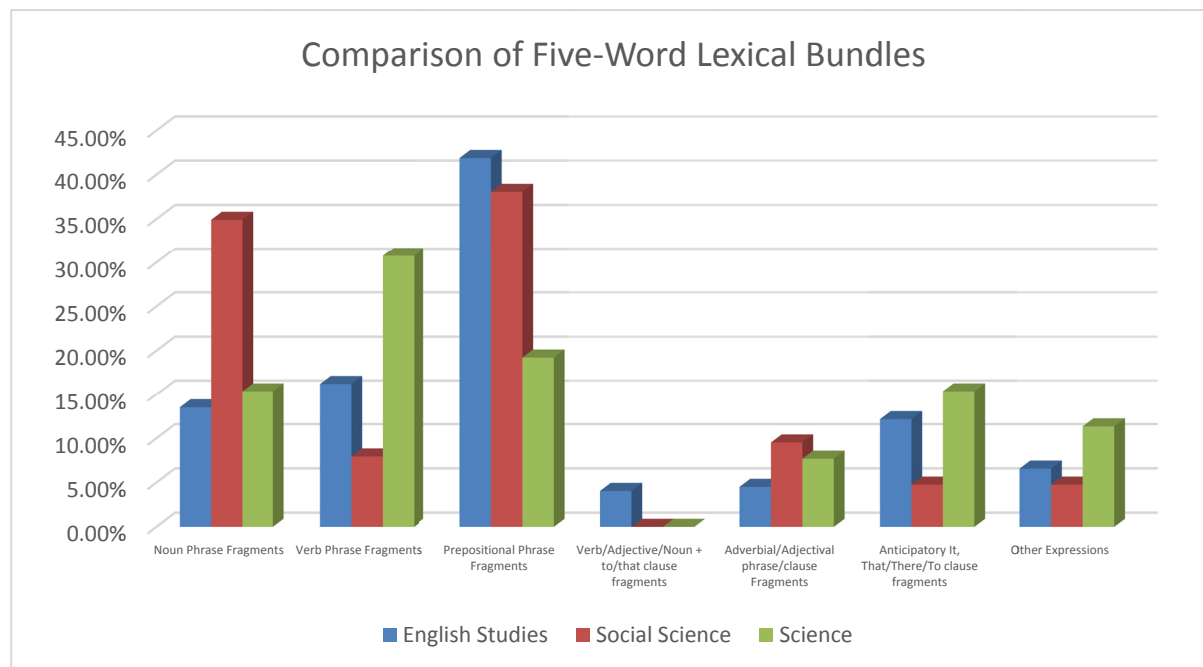
The third largest category of five-word lexical bundles in the corpus of English Studies is *Noun Phrase Fragments* that, consisting of 10 five-word lexical bundles, makes 13.51% of the total lexical bundles in the corpus of English Studies. On the other hand, the corpus of Social Sciences has 06 bundles in *Adverbial/Adjectival Phrase/Clause Fragments* which is 09.52% of the total bundles in the corpus. Whereas, *Noun Phrase Fragments* and *Anticipatory It, That/There/To Clause Fragments* both have 04 lexical bundle each that makes 15.38% of the total five-word lexical bundles in the corpus of bio Sciences in each category, and hence, both stand as the third largest category in the corpus of Bio Sciences.

Lastly, in the table we can see the remaining categories have lower number of five-word lexical bundles or there are no lexical bundles at all. Only *Anticipatory It, That/There/To Clause Fragments* has 09 lexical bundles which have percentage 12.16 in the corpus of English Studies, similarly *Other Expressions* in the corpus of Bio Sciences has 03 lexical bundles which is 11.34% of the total bundles of the Bio Science corpus. Otherwise, all remaining categories have less than 10% of lexical bundles in the respective corpora. As it can be seen, *Verb/Adjective/Noun + to/that Clause Fragments* has only 03 five-word lexical bundles in the corpus of English Studies, whereas, there are no five-word lexical bundles belonging to this category in the corpora of Social Sciences and Bio Sciences.

Table 5. Five word lexical bundles in the corpus

	English Studies Total (74)	Social Science Total (63)	Bio Science Total (26)
Noun Phrase Fragments	10 (13.51%)	22 (34.92%)	4 (15.38%)
Verb Phrase Fragments	12 (16.22%)	5 (7.94%)	8 (30.77%)
Prepositional Phrase Fragments	31 (41.89%)	24 (38.09%)	5 (19.23%)
Verb/Adjective/Noun + to/that clause fragments	3 (4.05%)	00	00
Adverbial/Adjectival phrase/clause Fragments	4 (4.51%)	6 (9.52%)	2 (7.69%)
Anticipatory It, That/There/To clause fragments	9 (12.16%)	3 (4.76%)	4 (15.38%)
Other Expressions	5 (6.57%)	3 (4.76%)	3 (11.34%)

The Graph 4 below shows the comparison of five-word lexical bundles in the three disciplines under study. It can be seen that there is different number of lexical bundles in different categories across the three disciplines. The graph shows that *Prepositional Phrase Fragments* overall has highest number of lexical bundles that makes this category very significant for the corpus. Moreover, *Noun Phrase Fragments* is a significant category in the corpus of Social Science. However, the corpus of Bio Science has highest number of *Verb Phrase Fragments* which is the dominant category in the corpus of Bio sciences.



Graph 4. Five word lexical bundles in the corpus

A table has been presented below which provides some of the examples of five-word lexical bundles found in the corpus. It is important to mention that these examples have been selected randomly.

Table 6. Examples of five word lexical bundles in the corpus

Examples	
the purpose of the study	it is not possible to
an important role in the	same is the case with,
is given in the following	as a result of the
has been used in the	it can be said that
on the basis of the	it has been observed that
on the part of the	and at the same time

5. Conclusion

This study explored five-word lexical bundles in three different academic disciplines (9 subjects). For the purpose of the study a corpus of about 4.7 million words was compiled which consisted of PhD dissertations from the respective subjects of the study from the Pakistani universities. The findings of the study show that all of the subjects (and disciplines) makes use of five-word lexical bundles, which are prefabricated structures having defined frequency per million word and a specific range as discussed above. Moreover, it was found that the use of five-word lexical bundles is not consistent and similar across the disciplines of study; similarly it is not consistent and similar across the subjects of the same discipline. The number of five-word lexical bundles in English Studies is 74, whereas, it is 63 in Social Sciences. On the other hand it is very low in Bio Sciences which is 26 lexical bundles. Moreover, corpus of English Studies uses *Prepositional Phrase Fragments* predominantly, whereas, corpus of Social Sciences also has dominant use of *Prepositional Phrase Fragments*. On the other hand, *Verb Phrase Fragments* is the dominant category used in the corpus of Bio Sciences. Hence, it is concluded that Pakistani academic discourse uses prefabricated structures of language called lexical bundles, and these structures vary across the discipline of study in terms of frequency, structure and use.

References

Bhatia, V. K. (1993). *Analysing Genre: language use in professional settings*. London: Longman.
 Biber, D. et al. (1999). *Longman grammar of spoken and written English*. Edinburgh: Longman.

- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405. <https://doi.org/10.1093/applin/25.3.371>
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, 14(2), 30-49.
- Conrad, S. M., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*.
- Cortes, V. (2002). Lexical Bundles in Freshman Composition. In R. Reppen, S. Fitzmaurice, & D. Biber (Eds.), *Using Corpora to Explore Linguistic Variation*. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.9.09cor>
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423. <https://doi.org/10.1016/j.esp.2003.12.001>
- Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, 3(1), 43-57. <https://doi.org/10.3366/E1749503208000063>
- Cowie, A. P. (1988). Stable and creative aspects of vocabulary. *Vocabulary and Language Teaching*, 139.
- Drazdauskiene, M. L. (1981). On stereotypes in conversation, their meaning and significance. In F. Coulmas, (Ed.), *Conversational Routine*. The Hague-Paris: Mouton. <https://doi.org/10.1515/9783110809145.55>
- Firth, J. (1957). *Papers in linguistics, 1934-1951*. London: Oxford University Press.
- Halliday, M. A. K. et al. (1964). *The linguistic sciences and language teaching*. London. Longman.
- Hyland, K. (1998). *Hedging in scientific research articles*. Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.54>
- Hyland, K. (2008a). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, 4-21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62. <https://doi.org/10.1111/j.1473-4192.2008.00178.x>
- Hyland, K. (2009). *Academic discourse: English in a global context*. London: Continuum.
- Hyland, K. (2009). Corpus informed discourse analysis: The case of academic engagement. *Academic writing: At the interface of corpus and discourse*, 110-128.
- Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford: Clarendon Press.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Wang & Ying. (2017). Lexical bundles in spoken academic ELF. *International Journal of Corpus Linguistics*, 22(2), 187-211(25). John Benjamins Publishing Company.
- Yorio, C. A. (1980). Conventionalized language forms and the development of communicative competence. *Tesol Quarterly*, 433-442. <https://doi.org/10.2307/3586232>
- Yousaf, M. (2018). *A corpus-based analysis of lexical bundles as building blocks of academic discourse*. Thesis submitted to Air University, Islamabad: Pakistan.
- Yu, S., & Kim, Y. J. (2017). Using lexical bundles to teach articles to L2 English learners of different proficiencies. *System*, 69, 79-91. <https://doi.org/10.1016/j.system.2017.08.002>

Copyrights

Copyright for this article is retained by the author, with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).