Fixedness of Expressions in Doctoral Research Dissertations: A Corpus Based Analysis

Muhammad Yousaf
Wasima Shehzad

Abstract

Lexical bundles are word structures which occur together more often than expected. The current study aimed to explore lexical bundles in a corpus of Pakistani doctoral dissertations. For this purpose a corpus was compiled which consisted of three broad disciplines of study; English Studies, Social Science and Biological Science. Every discipline consisted of three different subjects with high number of dissertations as has been listed by HEC (Higher Education Commission) Pakistan. The corpus was analyzed using AntConc 3.4.4w (Windows)2014 and for the structural categorization of lexical bundles, Biber et al's (1999) taxonomy was used as theoretical model which was later modified to more deeply investigate the structure of the lexical bundles. The results show that dissertations written in Pakistan rely heavily on prefabricated chunks of language, but the choice and usage of these chunks, or lexical bundles, varies from subject to subject and discipline to discipline. The difference among the structural categories within disciplines (among different subjects of a discipline) is less compared to the difference among the structural categories of different disciplines. The study shows that English Studies and Social Science corpora use more Prepositional Phrase Fragments and Noun Phrase Fragments whereas Science Corpus use more Verb Phrase Fragments and Prepositional Phrase Fragments.

Keywords*:* Subject; Discipline; Structures; Lexical Bundles; Frequency

1. Introduction

It is accepted that English is langua franca and its importance in academics has been ever growing as the knowledge produced in any language is mostly translated into English. This means, English is the langua franca of academics as well (Hyland, 2009). This has been stated by Hoffman (2000) who observes that because of its status English is important in all domains of life, most importantly in academics where we need a high proficiency of English. The importance of English in academics has changed the thinking of researchers, academicians, teachers and students as they have a wider audience belonging to different cultures and languages. This unilingual situation in academics was supposed to bring harmony among the academicians from different linguistic, cultural and regional backgrounds, but it has gone other way round as the distinction of native and non-native has emerged over the period of time.

In this regard there have been multiple studies for how to teach English effectively to non-native users of English in academic contexts (Salazar, 2011). Having said this, the number of research publications in English by non-natives is increasing over time (Hyland, 2006). Studies highlight that non-native scholars, academics, researchers and students need to put in a lot of effort to learn the structures and grammar of English apart from vocabulary and the discipline specified terminology which could make their writing meet international standards of writing, research writing in particular. Learning these dynamics of writing along with specialized vocabulary is not

easy, for this purpose, according to Shmitt (2011) the researchers and scholars dealing with teaching learners of English as a foreign language have been striving hard to create a word bank of vocabulary items of different domains and disciplines so that it helps non-native writers to only learn the specified and required vocabulary which can save time and assist them to gain mastery of expressions relevant to their discipline of study and expertise.

Collecting and learning word banks has helped learners but has not completely overcome the problem as research conducted (such as Genc, 2013) on native and nonnative language use suggest that apart from vocabulary the usage of vocabulary in certain sequences, patterns and structures is different between groups. According to scholars, the process of meaning is widely dependent on the patterns and sequences of the words used in the given textual context as Firth (1957: 195) states that a word is known by the company it keeps. This shows the importance of surrounding words and the patterns of words. These patterns of word are more or less fixed in nature which come natural to native speakers. This has been researched extensively, with the results showing that even with in the language of native speakers there is variety in the use of fixed structures based on the discipline of study and level of study (Hyland, 2008b; Chen and Baker, 2010). However, research has not been carried out in a Pakistani context, hence, it is important to find out the use of fixed structures and their variations across different disciplines of study. For this purpose, the current study has been conducted which deals with the following research questions:

1.1 Research Questions

1. To what extent does research writing rely on Lexical bundles?
2. Which of the structural categories of lexical bundles are predominantly used in research writing?
3. To what extent is there similarity/difference in the use of lexical bundles across disciplines?

1.2. Delimitation

The study has 90 doctoral dissertations written by Pakistani students, examining 4-word lexical bundles. This is explained further in the methods section below.

2. Literature Review

It has been long believed that all structures of language emerge out of a single word and it is the writer who uses the individual words as per their own choice to make a stretch of language. This point of view has been countered with arguments that if it is discretion of a writer to use basic unit of language ‘word’ and make phrases, clauses and sentences/utterances, then there should always be novel and unprecedented expressions every day, whereas, practically language is some sort of fixed entity which has a lot of repetitions. The argument was strengthened with the introduction of corpus linguistics in language research which brought to surface repeated expressions in large language data sources called ‘corpus’ where millions of words could be analyzed with the help of computer. According to Altenberg (1998) 80% of the expressions in language are repeated expressions, which is a very high percentage that can easily challenge the previously dominant view that individual words are used by writers to make novel and innovative structures in language.

These repeated expressions have been named as lexical bundles and are defined by as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status" (Biber, Johansson, Leech, Conrad, and Finegan, 1999; p. 990). The key characteristics of lexical bundles are their frequency and range of occurrence in a number of texts. Along with this fixedness of form and non-idiomatic nature of meaning are the other features which make a group of words as lexical bundle. It is pertinent to mention here that lexical bundles or clusters and chunks (Hyland, 2008b) have multiple structures and serve different functions in a piece of writing. Moreover, studies (such as) have highlighted that the frequency of these fixed expressions varies across different genres, both in written and spoken language. Hyland (2008) shows that lexical bundles are central to the creation of academic discourse as well as these offer means of differentiating written texts according to discipline.

With the further development in computer and technology, the advances in corpus linguistics helped Biber, Johansson, Leech, Conrad, Finegan and Quirk (1999) to explore large amount of data to find out fixed patterns of language called lexical bundles. These fixed patterns have been named and viewed differently by different scholars and researchers as, for example, Altenberg's (1998) formulaic sequences, Wray's (2000) recurrent word combinations and Stubbs' (2007a) n-grams are different terms used for fixed repeated expressions in a language. Moreover, lexical bundles by (Biber & Barbieri, 2007), clusters by Hyland (2008a), lexical clusters (Hyland, 2008b) and recurrent word combinations by Lindquist (2009) are also part of nomenclature. All of these terms refer to the fixed word combinations which could be extracted from large bulk of data with the help of corpus tools. Although there have been many studies previously, as mentioned above, but Biber et al. (1999: 590) was the first one giving a proper definition of the lexical bundles as 'recurrent expressions, regardless of their idiomaticity, and regardless of their structural status'. According to Chen and Baker (2010) these structures are fixed multi-word expressions having a traditional pragmatic and discourse function which is used by the native speakers and recognized by them in a certain context. The traces of such definitions go back to Sinclair's (1990:110) 'idiom principal' that reads 'a language user has available to him or her a large number of semi-pre-constructed phrases that constitute single choices'. These phrases consist of more than one word but the combination of words is considered as one to give meanings. It is to be noted that the pre-fabricated structures of language are meaningful and serve a function of constructing a meaningful discourse.

It has been highlighted by many scholars that many of the fixed patterns of language are meaningful in discourse and discourse construction but these mean a little alone, particularly to a nonnative language learner. Having said this, it is important to mention that different characteristics of formulaic units have been highlighted by scholars some of which are: 1) intuitive recognition, 2) idiomaticity, 3) fixedness, 4) completeness, 5) frequency, 6) length. The last two of these features of the patterns are variable as the frequency and length is not fixed. To term a pattern as lexical bundle Biber et al. (1999) sets a frequency cut off point i.e. 10 occurrences per million words and dispersion in at least 5 texts of the corpus. Similarly, regarding the length of lexical bundle it is argued that a 2-word structure is usually considered as a phrasal verb, and majority of 3-word combination are considered as idioms which have strict fixedness in structure as well as in meanings. But still many scholars consider 3-word structures as lexical bundles. Furthermore, four-word bundles are 'the most researched length for writing studies' (Chen & Baker, 2010, p. 32) because these have 'a wider variety of structures and functions to analyze'

(Cortes, 2004, p. 401). The majority of the studies conducted to explore lexical bundles in corpus set length of bundle on 4 words and one of the reasons stated is that these include a 3-word bundle in their structure and also these are 10 times more frequent than 5-word bundles (Pérez-Llantada, 2014). In addition, Cortes (2004) found out that many 4 & 5 word lexical bundles include 3 word lexical bundles in their structures.

Lexical bundles are not consistent in frequency across genres as Biber & Barbieri (2007) explored the difference of use of lexical bundles between conversation and academic writing. According to their study, in university registers written non-academic registers rely heavily on bundles which shows that lexical bundles vary across academic and nonacademic language. Similarly, Cortes (2004) found that students of history and biology and the authors use lexical bundles differently. The lexical bundles used by authors were very rarely used by the students, and even if they used, the usage was totally different. Moreover, Cortes (2006) found that native graduate students had problems using, discriminating, recognizing, and selecting different lexical bundles, although, they were given instruction about the importance of these word clusters. This serves as an evidence that lexical bundles are to be learnt along with their proper use.

Biber, Conrad and Cortes (2004) found that university tutorials had more lexical bundles as compared to lectures as well as textbooks. In analysis of corpora of research articles, doctoral dissertations and master's theses. Hyland (2008b) found that there were variations in structures and function of lexical bundles across disciplines. Hyland (2008a) observed that in published academic writing and postgraduate writing, postgraduate students used more formulaic expressions than native scholars. Similarly, there is difference between the use of lexical bundles in the writing of published authors, native student writers and non-native student writers in academic writing (Chen and Baker, 2010). Writings of published authors used wide range of lexical bundles as compared to student, whereas some of the students used their own in lexical bundles distinctively.

This has been researched by many researches and the results show that even with in the language of native speakers there is variety in the use of fixed structures based on the discipline of study and level of study. Such research has not been carried out in Pakistani context, hence, it is important to find out the use of lexical bundles and their variations across different disciplines of study. The current study is an attempt to explore the use of lexical bundles by PhD scholars in Pakistani universities, and also it highlights the difference in the use of lexical bundles across academic disciplines.

3. Research Methodology

3.1 Corpora

Three different corpora of three different disciplines were compiled which in total have 4.7 million words. These included PhD dissertations English Studies, Social Sciences and Bio Sciences. The theses were retrieved from thesis repository of HEC (Higher Commission) Pakistan. The repository includes theses from all universities of Pakistan (offering PhD programs) with categorization of disciplines being offered in Pakistani universities in each discipline. Three different subjects were included in each of these disciplines of studies, so strictly speaking it can be said that the corpus consisted on nine sub corpora. The division and detail of the corpora has been given below:

Table 3.1. Sub Corpora

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

The significance to select these subjects is to cover wide range of thesis written in Pakistani universities and compare and contrast the subjects and disciplines.

3.2 Analysis Tool

The corpus was analyzed with the help of AntConc 3.4.4w (Windows) 2014. The software was used to identify lexical bundles and find their frequency in the respective corpus. AntConc has features to adjust the number of words in an N-Gram combination, and it also facilitates seeing the structure in the context.

3.3 Theoretical Model

Lexical bundles found in the corpus were classified into grammatical categories based on Biber et al's (1999) taxonomy. The list of grammatical categories has been presented below. It is important to highlight here that after analysis of the corpora some new categories were added in the existing list which was to have an in-depth view of the bundles used in the larger corpora. The additional categories have been presented in discussion section.

1. NP + of
2. Other NPs
3. Prepositional Phrase + of
4. Other Prepositional Phrase
5. Passive + Prepositional
6. Phrase/That-Complement
7. Anticipatory it+ V/Adj
8. Be + N/Adj Phrase
9. Others

4. Results and Discussion

This section has been divided in four parts, the first three parts deal with the analysis of three disciplines and the final part deals with the comparison of the three disciplines. The section starts with the analysis of four-word bundles in English studies. The detailed discussion is as follows:

To analyze the corpora of English Studies through AntConc, number of words in N-Gram was set on four, frequency and 10 and range on 5, so that four-word lexical bundles could be obtained. The results obtained have been discussed in following section.

4.1 Four-Word Lexical Bundles in English Studies

Table 4.1 shows that the largest grammatical category of lexical bundles in Linguistics is *Prepositional Phrase Fragments* (see table 6 for examples). In total there are 101 lexical bundles which belong to this category that make 32.27% of the total lexical bundles in the corpus of Linguistics. The same category is the largest category in the corpus of Literature too where the number of total lexical bundles is 70 that makes 42.17% of the total lexical bundles in the corpus of literature. Whereas, this category is the not the largest category in the corpus of ELT, it is

second largest in terms of percentage of lexical bundles having 80 bundles that make 30% of the total bundles in the corpus of ELT.

The second largest category in the corpus of linguistics is *Noun Phrase Fragments* that has 82 bundles in total that makes 26.20% of the total bundles in the corpus of linguistics. This category is also second largest category in the corpus of literature where the number of bundles is 34 that makes 20.48% of the total bundles in the corpus of Literature. On the other hand, this category is the largest category in the corpus of ELT where the total no. of lexical bundles is 95 which is 35.71% of the total bundles in the corpus of ELT.

The third category in terms of percentage of lexical bundles in the Corpus of Linguistics is *Verb Phrase Fragments* that has 63 lexical bundles making 20.13% of the total lexical bundles. This category is also significant in the corpus of Literature and ELT as it is 3rd largest in both of the corpora having 15 (9.03%) and 36 (13.90%) lexical bundles respectively. Although, it is third largest category but the variation in percentage shows it is not equally present in three corpora as are the previous categories.

The fourth largest category in the corpus of linguistics is *Anticipatory It, That/There/To clause fragments* that has 34 lexical bundles which make 10.86% of the total lexical bundles in the corpus of Linguistics. The category is 4th largest in the corpora of Literature and ELT as well where it has 15 (9.03%) and 32 (12.03%) lexical bundles respectively. In this category the variation of percentage is not high as we can see in Table 1 and graph 1 that the percentages are 10.86%, 09.03% and 12.03% respectively.

The remaining grammatical categories are not very significant as the percentages of lexical bundles in them are well below 5% in all of the corpora except *Verb/Adjective/Noun + to/that clause fragments* in literature corpus where it has 10 lexical bundles that makes 6.02% of the total lexical bundles in the corpus of Literature, *Adverbial/Adjectival phrase/clause Fragments* in the corpus of linguistics with 16 (5.11%) lexical bundles and *Other Expressions* in the corpus of Literature having 12 lexical bundles which make 7.23% of the total bundles in the corpus of literature in the study. These mentioned categories too have low no. of lexical bundles in the remaining subjects where percentage of lexical bundles is well below 5% of the total bundles in the corpus.

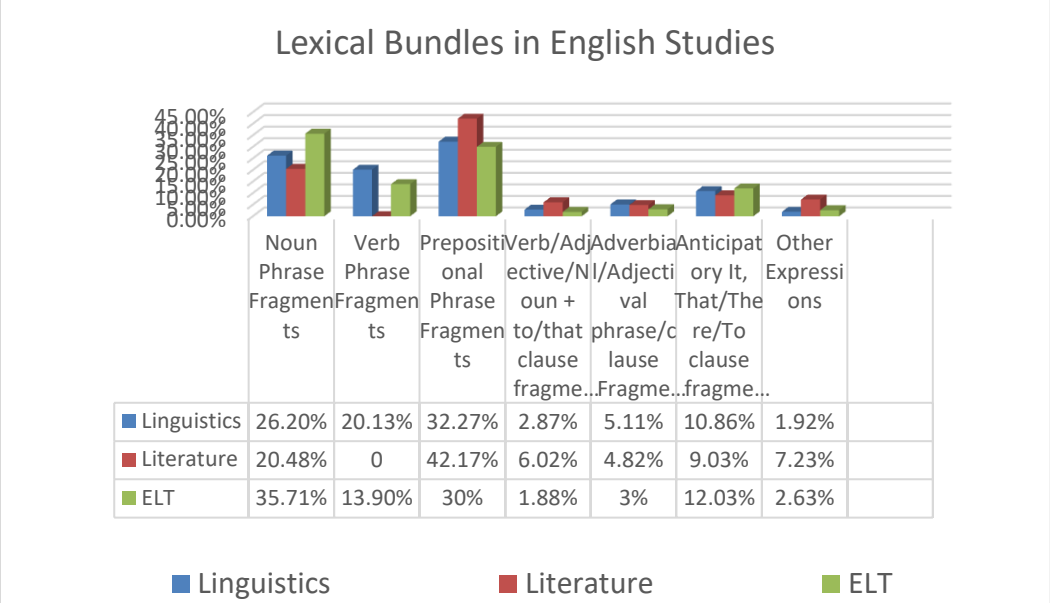
Table 4.1. Four Word Lexical Bundles in English Studies

	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	82 (26.20%)	63 (20.13%)	101 (32.27%)	09 (2.87%)	16 (5.11%)	34 (10.86%)	06 (1.92 %)
Literature	34 (20.48%)	15 (9.03%).	70 (42.17%)	10 (6.02%)	08 (4.82%)	15 (9.03%)	12 (7.23 %)
ELT	95 (35.71%)	36 (13.90%)	80 (30%)	05 (1.88%)	08 (3%)	32 (12.03%)	07 (2.63 %)
Total	211	114	251	24	32	81	25

Graph 4.1 below shows the comparison of grammatical categori

es in three subjects of English Studies. It has been highlighted that there is difference of lexical bundle distribution in different subjects as it is clear from the bars which are of unequal height in almost all the categories. The first category, *Noun Phrase Fragments*, is a dominant category in ELT corpus. Although it is present in Linguistics and Literature corpus too but the percentage shows the difference. The dissertations of literature have least number of lexical bundles in *Noun Phrase Fragments*. The second category, *Verb Phrase Fragments*, also has a wide difference in the percentage in three subjects as there is no *Verb Phrase Fragment* in Literature corpus whereas the percentage of lexical bundles belonging to this category in Linguistics is higher as compared to the category in ELT. The 3rd category, *Prepositional Phrase Fragments*, has high concentration of lexical bundles in the corpus of English Studies. Among the corpora literature has highest percentage of lexical bundles in this category and the other two corpora have almost equal percentages of lexical bundles. The remaining three categories are smaller so the difference of percentage of lexical bundles in the categories is also comparatively lesser.

Graph 4.1. Four Word Lexical Bundles in English Studies



	Noun Phrase Fragments	Verb Phrase Fragments	Prepositional Phrase Fragments	Verb/Adjective/Noun + to/that clause fragme...	Adverbial/Adjectival phrase/clause Fragme...	Anticipatory It, That/There/To clause fragme...	Other Expressions	
Linguistics	26.20%	20.13%	32.27%	2.87%	5.11%	10.86%	1.92%	
Literature	20.48%	0	42.17%	6.02%	4.82%	9.03%	7.23%	
ELT	35.71%	13.90%	30%	1.88%	3%	12.03%	2.63%	

4.2 Four Word Lexical Bundles in Social Sciences

Table 4.2 shows that the largest grammatical category of lexical bundles in the corpus of Political Science is *Noun Phrase Fragments*. In total there are 154 lexical bundles which belong to the category that make 41.85% of the total lexical bundles in the corpus of Political Science. Whereas, this category is not the largest category in the corpora of Education and Psychology. It is second largest in terms of percentage of lexical bundles in the corpus of education having 31 bundles that make 31% of the total bundles in the corpus, and is 3rd largest in the corpus of Psychology where it has 16 bundles that make 16.84% of the total lexical bundles in the corpus of Psychology. This variation of percentage and difference of size of categories shows that the lexical bundles are unevenly present in the three corpora of Social Science.

The second largest category in the corpus of Political science is *Prepositional Phrase Fragments* that has 140 bundles in total that make 38.04% of the total bundles in the corpus of Political Science. This category is largest category in the corpus of Education where the number of bundles is 32 that makes 32% of the total bundles in the corpus of Education. Similarly, the category is the largest category in the corpus of Psychology where the total no. of lexical bundles is 34 which is 35.79% of the total bundles in the corpus of Psychology.

The third category in terms of percentage of lexical bundles in the Corpus of Political Science is *Verb Phrase Fragments* that has 24 lexical bundles making 06.52% of the total lexical bundles in the corpus. This category is also significant in the corpora of Education and Psychology where it is also 3rd largest category having 17 (17%) and 20 (21.05%) lexical bundles respectively. Although, it is third category but the variation of percentage shows it is not equally present in three corpora as are the previous categories.

The remaining grammatical categories are not very significant as the percentages of lexical bundles in them are well below 5% in all of the corpora except *Adverbial/Adjectival phrase/clause Fragments* in the corpora of Political Science and Psychology where it has 20 (5.43%) and 07 (07.37%) lexical bundles respectively, *Anticipatory It, That/There/To clause fragments* in the corpora of Education and Psychology with 10 (10%) and 12 (12.63%) lexical bundles respectively, and *Other Expressions* in the corpus of Education having 06 lexical bundles which make 06% of the total bundles in the Education corpus in the study. These mentioned categories too have low no. of lexical bundles in the remaining subjects where percentage of lexical bundles is well below 5% of the total bundles in the corpus.

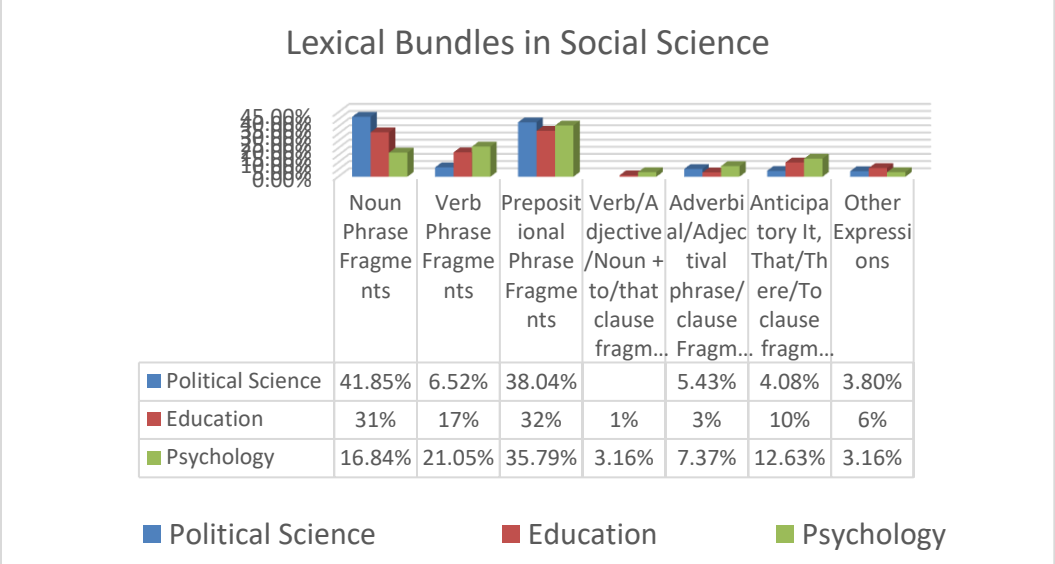
Table 4.2 Four Word Lexical Bundles in Social Science Corpus

	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
Political Science	154 (41.85%)	24 (6.52%).	140 (38.04%)		20 (5.43%)	15 (4.08%)	14 (3.80%)
Education	31 (31%)	17 (17%)	32 (32%)	01 (01%)	03 (03%)	10 (10%)	06 (06%)
Psychology	16 (16.84%)	20 (21.05%)	34 (35.79%)	03 (03.16%)	07 (07.37%)	12 (12.63%)	03 (03.16%)
Total	201	61	206	04	30	37	23

Graph 4.2 below shows the comparison of grammatical categories of lexical bundles in three subjects of Social Sciences. There is difference in percentages of lexical bundle distribution in three subjects as the bars show which are unequal in height in almost all of the categories. The first category, *Noun Phrase Fragments*, is a dominant category in Political Science corpus. Although it is present in Education and Psychology corpora too but there is wide difference in the percentages of bundles belonging to this category in the three subjects. The Corpus of Psychology has least value of lexical bundle percentage in *Noun Phrase Fragments*. The second category, *Verb Phrase Fragments*, also has a wide difference in the percentage in three subjects as percentage of lexical bundles belonging to this category in Psychology is higher as compared to

the other two corpora and it is minimum in Political science. These results are in contrary to the first category. Similar to English Studies Corpus the 3rd category, *Prepositional Phrase Fragments*, has high concentration of lexical bundles. Moreover, the difference in percentage of lexical bundles belonging to this category in different subjects of corpus is also less. The remaining categories are smaller so the difference of percentage of lexical bundles in the categories is also comparatively lesser. The significant finding can be seen in *Anticipatory It, That/There/To clause fragments* where Corpus of Psychology has significant percentage of lexical bundles.

Graph 4.2. Four Word Lexical Bundles in Social Science Corpus



	Noun Phrase Fragments	Verb Phrase Fragments	Prepositional Phrase Fragments	Verb/Adjective /Noun + to/that clause fragm...	Adverbial/Adjectival phrase/ clause Fragm...	Anticipatory It, That/There/To clause fragm...	Other Expressions
Political Science	41.85%	6.52%	38.04%		5.43%	4.08%	3.80%
Education	31%	17%	32%	1%	3%	10%	6%
Psychology	16.84%	21.05%	35.79%	3.16%	7.37%	12.63%	3.16%

4.3 Four Word Lexical Bundles in Bio-Sciences

Table 4 shows that the largest category of lexical bundles in Bio-Tech is *Verb Phrase Fragments*. In total there are 30 lexical bundles which belong to the category that make 39.47% of the total lexical bundles in the corpus. The same category is the largest category in the corpora of Botanical Sciences and Zoological Sciences too where the number of lexical bundles is 23 that makes 30.67% and 18 that makes 36% respectively.

The second largest category in the corpus of Bio Technology is *Prepositional Phrase Fragments* that has 20 bundles in total that make 26.31% of the total bundles in the corpus. This category is also second largest category in the corpora of Botanical Sciences consisting of 16 (21.33%) lexical bundles and Zoological Sciences having 12 (24%) bundles.

The third category in terms of percentage of lexical bundles in Corpus of Bio Technology is *Other Expressions* that has 9 lexical bundles making 11.84% of the total lexical bundles. This category is also significant in the corpus of Zoological Sciences where it is also 3rd largest consisting of 10 (20%) lexical bundles. On the other hand, the 3rd largest category in terms of percentage of lexical

bundles in the corpus of Botanical Sciences is *Noun Phrase Fragments* that has 15 lexical bundles which is 20% of the total lexical bundles in the corpus.

The fourth largest category in the corpus of Bio Technology is *Adverbial/Adjectival phrase/clause Fragments* that has 07 lexical bundles which make 09.21% of the total lexical bundles in the corpus. This category is 4th largest in the corpora of Botanical Sciences as well where it has 07 (9.33%) lexical bundles. Whereas, the 4th largest category in the corpus of Zoological Sciences is *Noun Phrase Fragments* having 5 (10%) lexical bundles.

The remaining grammatical categories are not very significant as the percentages of lexical bundles are low and are almost equal in all corpora as has been shown in table 4 and graph 3. It is to be noted that the category *Verb/Adjective/Noun + to/that clause Fragments* is empty as there is no lexical bundle in any of the corpora in this discipline.

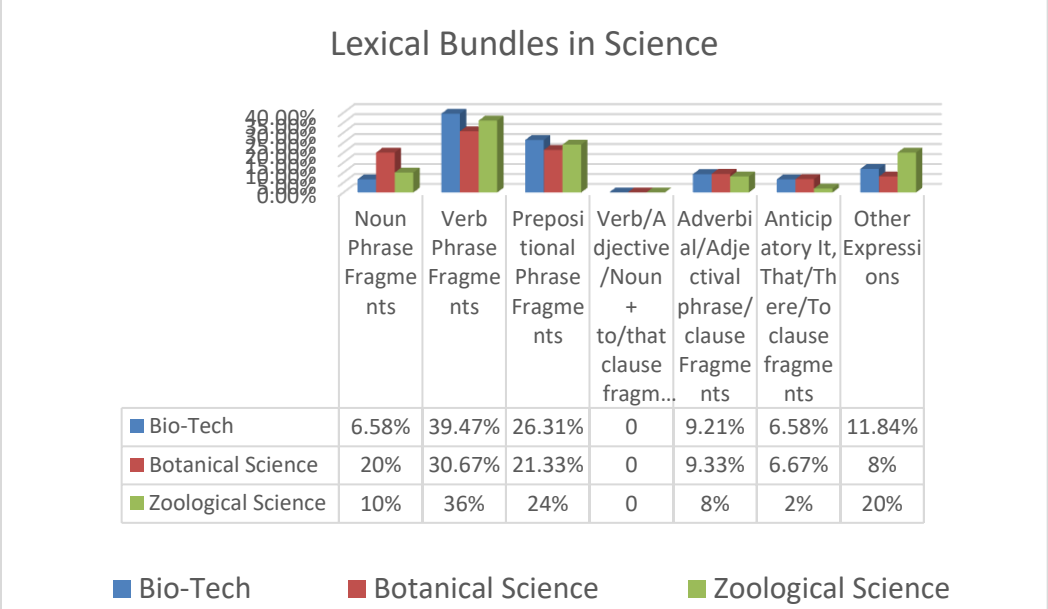
Table 4.3. Four Word Lexical Bundles in Science Corpus

	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	05 (6.58%).	30 (39.47%) .	20 (26.31%)	**00**	07 (9.21%)	05 (6.58%)	09 (11.84%)
Botanical Science	15 (20%)	23 (30.67%)	**16** (21.33%)	**00**	07 (9.33%)	05 (6.67%)	06 (08%)
Zoological Science	05 (10%)	18 (36%).	12 (24%)	**00**	04 (08%)	01 (02%)	10 (20%)
Total	25	71	48	**00**	18	11	25

Graph 4.3 below shows the comparison of categories of lexical bundles in three subjects of Science. The distribution of lexical bundles is totally different as compared to the corpus of English Studies and Corpus of Social Science as the concentration of lexical bundles is in *Verb Phrase Fragments* that is unlike previous disciplines. Moreover, there is difference in percentages of lexical bundle distribution in three subjects as the bars show which are unequal in height in almost all of the categories. The first category, *Noun Phrase Fragments*, has higher percentage in Botanical Science corpus. Although it is present in Bio Technology and Zoological Science corpora too but there is wide difference in the percentages of bundles belonging to this category in three subjects. The Corpus of Bio Tech has least value of lexical bundle percentage in *Noun Phrase Fragments*. The second category, *Verb Phrase Fragments*, is the most concentrated category in the Corpus of Science, but it also has a wide difference in the percentage in three subjects as percentage of lexical bundles belonging to this category in Bio Tech is higher as compared to the other two corpora and it is minimum in Botanical Science Corpus. The 3rd category, *Prepositional Phrase Fragments*, also has significant percentage of lexical bundles. Moreover, the difference in percentage of lexical bundles belonging to this category in different subjects of corpus is lesser as compared to the previous categories. The remaining categories are smaller so the difference of percentage of lexical bundles in the categories is also comparatively

lesser. The significant finding can be seen in *Other Expressions* where Corpus of Science has highest percentage of lexical Bundles as compared to the previous two Corpora (English Studies and Social Sciences), similarly, within the discipline Zoological Science has higher percentage of lexical Bundles as compared to Bio Tech and Botanical Science Corpus.

Graph 4.3. Four Word Lexical Bundles in Science Corpus



4.4 Dominant Categories of Four Word Lexical Bundles

Table 5 shows that the largest category of lexical bundles in English Studies is *Prepositional Phrase Fragments*. In total there are 251 lexical bundles which belong to the category that make 34.01% of the total lexical bundles in the corpus. The same category is the largest category in the corpus of Social Science too where the number of total lexical bundles is 206 that makes 36.59% of the total lexical bundles in the corpus. Whereas, this category is not the largest category in the corpus of Bio Science, it is second largest in terms of percentage of lexical bundles having 48 bundles that make 24.24% of the total bundles in the corpus.

The second largest category in the corpus of English Studies is *Noun Phrase Fragments* that has 211 bundles in total that makes 28.59% of the total bundles in the corpus. This category is also second largest category in the corpus of Social Science where the number of bundles in it is 201 that makes 35.76% of the total bundles. On the other hand, this category is 3rd largest category in the corpus of Science where the total no. of lexical bundles is 25 which is 12.63% of the total bundles in the corpus. This shows the similarity of English Studies corpus and difference of Science Corpus in terms of Lexical bundles.

The third category in Corpus of English Studies is *Verb Phrase Fragments* that has 114 lexical bundles making 15.45% of the total lexical bundles in the corpus. This category is also significant in the corpus of Social Science where too it is 3rd largest category having 61 (10.85%) lexical

bundles. On the other hand, this category is the largest category in the corpus of Science consisting of 71 lexical bundles that make 35.86% of the total bundles in the corpus. This result highlights that the Science corpus has *Verb Phrase Fragments* as a dominant category and other two corpora heavily rely on *Prepositional and Noun Phrase Fragments*.

The fourth largest category in the corpus of English Studies is *Anticipatory It, That/There/To clause Fragments* that has 81 lexical bundles which make 10.97% of the total lexical bundles in the corpus. This category is 4th largest in the corpora of Social Science as well where it has 37 (6.57%) lexical bundles. On the Other hand, this category is the smallest category in the Corpus of Science having 11 lexical bundles that make 5.55% of the total lexical bundles in the corpus.

The remaining categories are not very significant as the percentages of lexical bundles are well below 5% in all of the corpora except *Adverbial/Adjectival phrase/clause Fragments* in Social Science corpus where it has 30 lexical bundles that makes 5.33% of the total lexical bundles in the corpus and in Science corpus consisting of 18 (9.09%) lexical bundles, and *Other Expressions* in the corpus of Science with 25 (12.62%) lexical bundles in the corpus. *Other Expressions* in the corpus of Science is the 4rth largest category whereas it is not a significant category in the corpora of English Studies and Social Sciences.

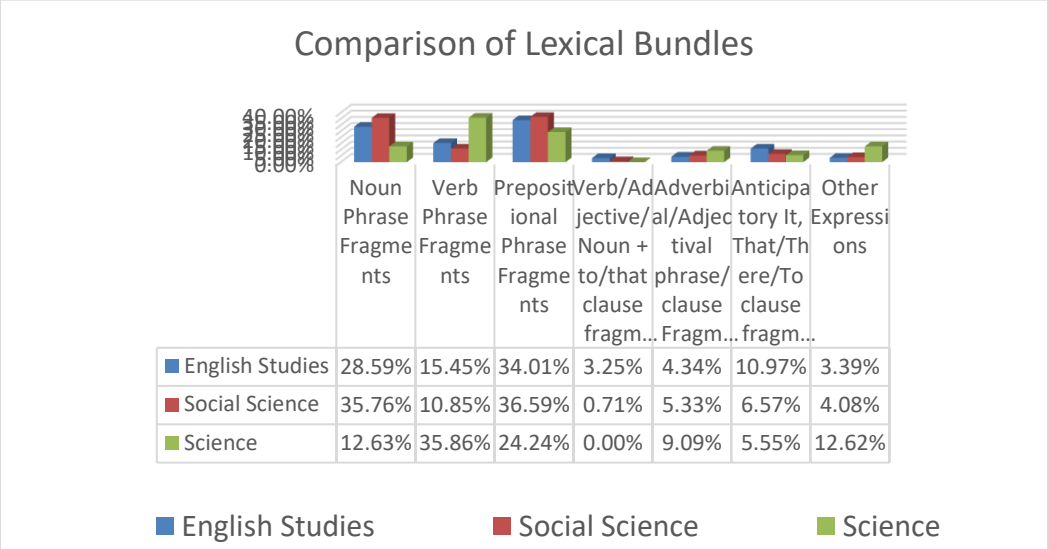
Table 4.4. Comparison of Four Word Lexical Bundles

	English Studies	**Social Science**	**Bio-Science**
Noun Phrase Fragments	211 (28.59%)	201 (35.76%)	25 (12.63%)
Verb Phrase Fragments	114 (15.45%)	61 (10.85%)	71 (35.86%)
Prepositional Phrase Fragments	251 (34.01%)	206 (36.59%)	48 (24.24%)
Verb/Adjective/Noun + to/that clause fragments	24 (3.25%)	04 (0.71%)	00 (0%)
Adverbial/Adjectival phrase/clause Fragments	32 (4.34%)	30 (5.33%)	18 (9.09%)
Anticipatory It, That/There/To clause fragments	81 (10.97%)	37 (6.57%)	11 (5.55%)
Other Expressions	25 (3.39%)	23 (4.08%)	25 (12.62%)

Graph 4.4 below shows the comparison of categories of lexical bundles in three Disciplines of study. As has been shown in previous sections, the concentration of lexical bundles is in first three categories, but the difference of percentage of lexical bundles is wide among the disciplines. The first category, *Noun Phrase Fragments*, has higher percentage in Social Science corpus. Although it is present in English Studies and Bio Science corpora too but there is wide difference in the percentages of bundles belonging to this category in three Disciplines. The Corpus of Science has least value of lexical bundle percentage in *Noun Phrase Fragments* which shows the trend of not using nouns (Noun Phrases) in Science subjects whereas Noun (Noun Phrases) are used more by writer in the discipline of Political Science. The second category, *Verb Phrase Fragments*, is the most concentrated category in the Corpus of Science, and the difference between the percentages of lexical bundles belonging to this category in different disciplines is wide as percentage of lexical bundles belonging to this category in Science is highest as compared to the other two corpora and it is minimum in Political Science Corpus. The 3rd category, *Prepositional Phrase Fragments*, also has highest percentage of lexical bundles in corpora of English Studies and Social Science. Moreover, the difference in percentage of lexical bundles belonging to this category in different disciplines is lesser as compared to the previous categories, but still the difference

between Corpus Social Science and Bio Science is wide. The remaining categories are smaller so the difference of percentage of lexical bundles in the categories is also lesser. The last category, *Other Expressions*, has highest percentage of lexical Bundles in the Science which is more like the 2nd category, *Verb Phrase Fragments*, where Science Corpus has higher percentage of lexical bundles as compared to the other two corpora.

Graph 4.4. Comparison of Four Word Lexical Bundles



4.5 Examples of Four-Word Lexical Bundles

Table 6 below presents some examples of the categories discussed above. It is pertinent to mention here that the examples presented in the table are randomly picked from different subcategories of lexical bundles which are not part of this study and are included in another research project.

Table 4.5. Examples of Four-Word Lexical Bundles

	Category	Examples
1	Noun Phrase Fragments	the frequency of occurrence, the analysis of the, the fact that the, an important role in
2	Verb Phrase Fragments	has been used in, can be found in, can be said that, is used as a
3	Prepositional Phrase Fragments	on the basis of, for the sake of, in the case of, on the other hand, in the following table
4	Verb/Adjective/Noun + to/that clause Fragments	can be used to, does not seem to, the results show that, the table shows that
5	Adverbial/Adjectival phrase/clause Fragments	as a result of, as a foreign language, one of the most, same is the case
6	Anticipatory It, That/There/To clause Fragments	it is important to, it can be said, that most of the, that there is a, there is no need, there is not a
7	Other Expressions	and at the same, and the use of, we come to know, way or the other

5. Conclusion

The current research was an exploratory and descriptive study which explored, identified, categorized and described lexical bundles in 3 broader disciplines (9 subjects) of study where the focus of analysis was on the PhD Dissertations in the respective disciplines. First of all, it was to be established weather all subjects of study do use lexical bundles or they do not. The answer to this query was that all subjects of study heavily rely on the prefabricated structures of language, called lexical bundles. The next point of analysis was to find out how far the subjects within a discipline are similar or different in terms of using lexical bundles. The study proved that every subject, within a discipline, has its own preference of using lexical bundles, every subject has its own choice of lexical structures and the dominant categories of lexical bundles in different subjects are different. Finally, it was to be established that how different disciplines were similar or different in terms of use of lexical bundles. It has been found that there is a difference in use of lexical bundles in different subjects but it is negligible as compared to the difference of use of lexical bundles in different disciplines as there was vast difference in the in the use of lexical bundles among three broader disciplines; the results show Corpus of English Studies has a high variety of lexical bundles and also it dominantly uses *Prepositional Phrase Fragments* and *Noun Phrase Fragments*, whereas, the variety of lexical bundles in Social Science Corpus is lesser and it relies more on *Noun Phrase Fragments* and *Prepositional Phrase Fragments*. On the other hand, Science Corpus heavily relies of *Verb Phrase Fragments* and *Prepositional Phrase Fragments*. Moreover, the use of *Other Expressions* in the corpus Bio Science is higher than that of other two corpora. In addition, it is important to state that *Prepositional Phrase Fragments* have significant use in all disciplines which are part of this study. Hence, it can be concluded that Lexical bundles are essential element in dissertation writing, but still the difference of lexical bundles is strongly influenced by discipline of study and subject of study.

References

Altenberg, B. (1993). Recurrent word combinations in spoken English. In : J. d'Arcy : Proceedings of the Fifth Nordic Association for English Studies Conference. University of Island, Reykjavik 1993.

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

Biber, D. & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In : Hilde Hasselard, Signe Oksefjell : Out of Corpora : Studies in Honor of Stig Johansson. Rodopi, Amsterdam 1999, 181–189.

Biber, D. Johansson, S. Leech, G. Conrad, S. Finegan, E. (1999). The Longman Grammar of Spoken and Written English.Longman, Harlow 1999.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spokenand written English*. Harlow: Longman.

Biber, D. Conrad, S. Cortes, V. (2004). If you look at …Lexical bundles in university lectures and textbooks. Applied Linguistics 25, 2004, 371–405.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers.*English for specific purposes, 26*(3), 263-286. http://dx.doi.org/10.1016/j.esp.2006.08.003

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology, 14*(2), 30–49.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423. http://dx.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. http://dx.doi.org/10.3366/E1749503208000063.

Firth, J. (1957). *Papers in linguistics, 1934-1951*. London: Oxford University Press.

Genc, B. (2013). Differences in English Vocabulary Use: Insights from Spoken Learner Corpus and Native Speaker Corpus. Education & Science/Egitim ve Bilim, 38(167).

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145-160). Oxford: Clarendon Press.

Hoffman, C. (2000). The spread of English and the growth of multilingualism with English in Europe. In J.Cenoz & U.Jessner (Eds.), *English in Europe: The Acquisition of a Third Language (pp.1-21).* Clevedon: Multilingual Matters.

Hyland, K. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing, 13*, 133-151. http://dx.doi.org/10.1016/j.jslw.2004.02.001

Hyland, K. (2006). The 'other' English: Thoughts on EAP and academic writing. *The European English Messenger, 15*(2), 34-38.

Hyland, K. (2008a). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4-21. http://dx.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. http://dx.doi.org/10.1111/j.1473-4192.2008.00178.x

Hyland, K. (2009). *Academic discourse: English in a global context*. London: Continuum.

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics, 32*, 150-169. http://dx.doi.org/10.1017/S0267190512000037.

Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh University Press.

Moon, R. (1997). Vocabulary connections : Multi-word items in English. In : Norbert Schmitt, Michael McCarthy : Vocabulary : Description, Acquisition and Pedagogy. Cambridge University Press, Cambridge 1997, 40–63.

Moon, R. (1998a). Fixed Expressions and Idioms in English : A Corpus-Based Approach. Clarendon, Oxford 1998.

Moon, R. (1998b). Frequencies and forms of phrasal lexemes in English. In : A. Cowie : Phraseology. Oxford University Press, Oxford 1998, 79–100.

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Shared and distinct usage. *Journal of English for Academic Purposes, 14*, 84-94. http://dx.doi.org/10.1016/j.jeap.2014.01.002.

Salazar, D. (2011). Lexical bundles in scientific English: A corpus-based study of native and non-native writing. (Unpublished doctoral dissertation). Universitat de Barcelona, Spain.

Schmitt, N. (2005). Processing advantages of formulaic sequences vs non-formulaic sequences in text passages. In C. Cosme, C. Gouverneur, F. Meunier, & M. Paquot (Eds.),

Proceedings of the Phraseology 2005 conference (pp. 381-382). Universite Catholique de Louvain: Louvain-la-Neuve, 13-15 October 2005.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan. http://dx.doi.org/10.1057/9780230293977.

Sinclair (1991). Corpus, Concordance, Collocation. Oxford University Press, Oxford 1991.

Wray, A., & Perkins, M. (2000): The functions of formulaic language : an integrated model. Language and Communication 20, 2000, 1–28.

Wray, A. (2002): Formulaic language and the lexicon. Cambridge University Press, Cambridge 2002.

Appendix

Intended Learning Efforts

2. I think that I am doing my best to learn English.
7. If an optional assignment in English is assigned by the teacher, I would certainly volunteer to do it.
49. I quite often try to expend a lot of effort in learning English.
63. I would like to spend a lot of time studying English.
82. I would like to concentrate on studying English more than any other subject.
87. If I get some extra time to learn English, I would like to use it.

Ideal L2 Self

3. Whenever I think of my future career, I imagine myself using English.
16. I can imagine myself reading English texts effectively.
37. I can imagine myself writing English e-mails/letters fluently.
71. I can imagine myself speaking English as if I were a native speaker of English.
84. I can imagine myself going abroad and using English effectively for communicating with the locals.
88. I can imagine a situation where I will be speaking English with foreigners.

Ought-To L2 Self

1. I consider learning English important because the people I respect think that I should do it.
9. I have to study English, because, if I do not study it, I think my parents/family will be disappointed with me.
21. Studying English is important to me because other people will give me more respect if I have knowledge of English.
42. Learning English is necessary because people around me expect me to do so.
60. If I study English, my friends and family will considered me as an educated person.
85. If I don't learn English, it will have negative impact on my life.

Attitude Toward Learning English

1. I like English quite a lot.
22. I always look forward to my English classes.
47. I really enjoy my English classes.
67. I would like to have more English lessons at school.
76. Learning English is a wonderful experience.
90. I like the atmosphere of my English classes

Instrumentality—Promotion

4. Studying English can be important to me because I think, someday, it will be useful in having a better future career.
14. Learning English is important to me because I would like to use it if I travel internationally.

26. Studying English is important for me to keep myself updated and informed of the recent events of the world.
41. Studying English can be important to me because I think I will need it to improve my knowledge in the world affairs.
69. English language learning is useful for me to translate and understand the texts written in English.
81. Studying English can be important to me because I think I'll need it to improve my knowledge in Islamic studies.

Instrumentality—Prevention

30. I have to study English; otherwise, I think I will lag behind in my studies.
54. Studying English is important to me because I would feel ashamed if I get bad grades in English.
65. Studying English is important to me because if I don't have knowledge of English, I'll be considered a less educated person.
70. I have to study English; otherwise, I think I cannot be successful in my future career.
73. I have to study English; otherwise, I think I cannot be appreciated as a good learner by the teachers and the fellow students.
83. Studying English is important to me because I do not like to be considered as a poorly educated person.

Fear of Assimilation

23. I think that by learning English ways of living, there is a danger of losing the Pakistani identity.
36. The influence of the culture of English speaking countries, have affected the social structure of the Pakistani society.
46. Due to the influence of the English language, I think there is a danger that the Pakistani people may forget the importance of their own culture.
51. I think the morals of the Pakistani people are becoming worse due to the influence of the English-speaking countries.
79. I think the cultural and artistic values of English have badly affected the Pakistani values.
86. I think, the Pakistani languages are becoming corrupt due to the influence of the English language.

Integrativeness

5. I think, I should improve my English to understand more about the English people.
10. I like to live a life as the English-speaking people do.
17. I would like to speak and write English as the native speakers do.
27. I find learning English really interesting.
33. I would you like to become similar to the people who speak English.
50. I think learning English is important for me in order to learn more about the culture and art of its speakers

Attitudes Toward L2 Community

8. I like the people who live in English-speaking countries.
13. I like meeting people from English-speaking countries.
24. I would like to travel to English-speaking countries.
44. I would like to know more about people from English-speaking countries.
59. I think that English-speaking countries have an important role to play in the world.
77. I think that English-speaking countries are advanced and developed nations.

Cultural Interest

20. I would like to see the cultural gatherings and festivities taking place in English-speaking countries.
43. I want to know the culture and art of English-speaking countries
61. I like the dresses and foods of the English-speaking countries
72. I like to read English literary texts written by the people of English speaking countries.
74. I like English magazines, newspapers, or books.

Milieu

12. My family puts a lot of pressure on me to study English.
15. People around me think that I must learn English to become an educated person.
35. My parents encourage me to take every opportunity to learn English.
45. My peer group encourages me to learn English.
55. I get support from my friends in learning English.
57. I am often told by my teachers that English is important for my future.