A Corpus Based Study Language Fixity in Journalistic Discourse: A Corpus-Based Study of Pakistani Editorials and Blogs

Muhammad Yousaf
Lecturer
National University of Modern Languages, Islamabad

Azhar Habib
Lecturer
National University of Modern Languages, Islamabad

Ijaz Ali Khan
Lecturer
Govternment Degree College Lahore, Sawabi

Farzana Masroor
Assistant Professor
Air University, Islamabad

Abstract

Lexical bundles are the fixed patterns of language that appear in texts more than expected which according to Biber and Conrad (1999) are "word forms that often co-occur in longer sequences" (p. 989). The length of these structures varies but the most common lexical bundles are three words strings followed by the four words strings. The aim of the current study is to explore 3-word lexical bundles in Newspaper editorials and news blogs written by Pakistani writers to verify the fixity of language, which means language relies on the prefabricated structures, in the journalistic discourse. For this purpose a corpus of 25641 words was compiled which consisted of 20 newspaper editorials and 20 news blogs. The corpus was explored in the light of Biber and Conrad's (1999) structural taxonomy of lexical bundles. The analysis shows that the news editorials have higher use and variety of lexical bundles compared to news blogs which highlights the fluency and higher linguistic competence of the editorial writers. The dominant category of the lexical bundles was noun phrase fragments in the newspaper editorials.

Keywords: fixed phrases, language of editorials, phrase fragments

1. Introduction

Lexical bundles are the fixed patterns of language that appear in texts more than expected which according to Biber and Conrad (1999) are "word forms that often co-occur in longer sequences" (p. 989). The patterns are given different names by different scholars which in this study have been used interchangeably (Altenberg, 1998; Banerjee & Pedersen, 2003; Biber & Conrad, 1999; Scott, 1997; Strzalkowski, 1998; Stubbs, 2002). Studies (such as Hyland, 2004; Yousaf, 2019; Yousaf & Shehzad, 2018) have shown that lexical bundles vary across different disciplines of study across academic discourse. Similarly, the length of these structures varies in different discourses but the most common lexical bundles are three words strings followed by the

four words strings which highlights that length of a lexical bundle in inversely proportional to its frequency.

Majority of the studies conducted in the area under discussion were in the field of academic writing (Biber & Conrad, 1999; Biber, Conrad, & Cortes, 2004; and Biber & Barbieri, 2007) which focus on the written or spoken language in the academic discourse. Chen and Baker (2010) found difference of lexical bundles in the writings of novice and expert writers, whereas Cortes (2004), Hyland (2008), Yousaf and Shehzad (2018, 2019) and Yousaf (2019) explored fixed phrases across different disciplines of study. All of these studies propounded large difference in terms of use of fixed phrases across different texts. Similarly, Yousaf, Khan, and Habib (2019) explored difference of fixed phrases in the blogs by male and female bloggers which also showed difference of the use of lexical chunks, and the preferred usage of different fixed structures by male and female writers.

As stated above the majority of the previous studies aimed at academic discourse therefore the presence of lexical bundles in the journalist discourse and its prevalence in different types of the texts of journalist discourse is not much known. Hence, the present research attempts to explore the journalistic texts with reference to fixed phrases. For this purpose the presence of fixed expressions in the news editorials and blogs has been explored to address the research questions given below. The study employs Biber and Conrad's (1999) framework for the structural analysis of the fixed expressions.

The aim of the current study is to explore 3-word lexical bundles in Newspaper editorials and news blogs written by Pakistani writers to verify the fixity of language in the journalistic discourse. For this purpose a corpus of 25641 words was compiled which consisted of 20 newspaper editorials and 20 news blogs. The corpus was explored in the light of Biber and Conrad's (1999) structural taxonomy of lexical bundles. The analysis shows that the news editorials have high use and variety of lexical structures compared to news blogs which highlights the fluency and higher linguistic competence of the editorial writers.

2. Literature Review

Firth (1957) drew attention towards fixed pattern in language use where he observed some of the expression used with a frequency more than expected and called them collocations. Although he discussed collocations but he did not provide proper definition for them. A proper definition of collocations was given by Halliday, MacIntosh, and Strevens (1964) who stated that collocations are the word units which have a higher tendency to occur together in language use. The previous few decades have focused on the research in the area of fixed expression in language which have been explored and studied from different perspectives. Scholars have suggested different names such as fixed expressions, lexical bundles, language patterns, formulaic patterns, prefabricated structures, etc. (see Cortes, 2004; Cowie, 1988; Moon, 1992; Yorio, 1980; Yousaf & Shehzad, 2018; Yousaf & Shehzad, 2019). Although there has been difference of nomenclature for the expression, but they have given invariable definition of these bundles.

Advancement in corpus linguistics took research in the area of fixed expressions to an altogether different level since computer assisted corpus compilation and analysis enabled the

researchers to explore large sized, real life language data which was not possible manually. Biber and Conrad (1999) studied a large corpus and stated that lexical bundles are frequency driven three or more than three words combinations which co-occur in language used in real life. Moreover, regardless of their idiomaticity and lexical status they serve as a building block of language. According to Yorio (1980) because of their readiness such fixed expression increase fluency of the language users. Fixed expressions allow them to save the efforts and time required for making new grammatically correct structures. These expressions are ready made in the mind of the speaker and the listener and can easily be retrieved (Nattinger & DeCarrico, 1992). Apart from the readiness of the structures, the language users are equipped with the larger linguistic repertoire. According to Cortés (2002), prefabricated chunks help in building fluency and these quicken the process of language learning too. According to Chen and Baker (2010), Cortes (2004), Haswell (1991), and Hyland (2008) the frequent use of lexical bundles signifies linguistic competence of the language users and the language users who lack these expressions are novice users. Moreover, the frequently used lexical bundles provide a natural language in use (Ellis, Simpson-Vlach, & Maynard, 2008).

The criterion of considering an expression as lexical bundles varies in different studies but the three elements which are considered to be essential for a structure to be lexical bundle are i) normalized frequency ii) dispersion cross number of texts of the same genre iii) number of words in a structure (see Biber & Conrad, 1999; Cortes, 2004; Hyland, 2008; Yousaf & Shehzad 2018; Yousaf, 2019; Yousaf & Shehzad, 2019; Yousaf et al., 2019). The same yardstick was considered as the basis of the lexical bundles in the current study which has been explained in the research methodology section.

Majority of the studies conducted in the area under discussion were in the field of academic writing (Biber & Conrad, 1999; Biber et al., 2004; Biber & Barbieri, 2007) which focus on the written or spoken language in the academic discourse. Chen and Baker (2010) found difference of lexical bundles in the writings of novice and expert writes, whereas Cortes (2004), Hyland (2008), Yousaf and Shehzad (2018, 2019) and Yousaf (2019) explored fixed phrases across different disciplines of study. All of these studies propounded huge difference in terms of use of fixed phrases across different texts. Similarly, Yousaf et al. (2019) explored difference of fixed phrases in the blogs by male and female bloggers which also showed difference of the use of lexical chunks, and the preferred usage of different fixed structures by male and female writes.

Attempts to explore the journalistic texts with reference to fixed phrase have been rare (see Yousaf et al., 2019) thus the current research aims to study the presence of fixed expressions in the news editorials and blogs to address the research questions. The study employs Biber and Conrad's (1999) framework for the structural analysis of the fixed expressions. The current study aims to answer the following research questions:

1. How far do the Pakistani newspaper editorials and blogs rely on fixed lexical structures?

2. What are the dominant lexical patterns used in the text?

3. How far are the editorials and blogs similar or different in terms of fixed phraseology?

3. Research Methodology

The following section discusses the theoretical framework and corpus tools used in this study to analyze the data.

3.1. Corpus

For the purpose of the current study a corpus of News editorials and news blogs was complied. The editorials and the blogs were collected from the online sources and converted into text file. It is important to mention here that the editorials and blogs were written by Pakistanis writers and they dealt with the broader social issues of Pakistan. This criterion ensured that the writers of the blogs were nonnative writers of English and their writings were without professional jargon of a specific field. The total number of words in the corpus and the details of the data are as follows:

Table 1: Corpora

	Number of word Tokens	Number of Word Types	Token-Type Ratio
Editorials	8396	2304	3.64
Blogs	17245	4574	3.77

3.2. Corpus Tool

For the analysis of the corpus AntConc 3.4.4w (Windows) 2014 was used as a corpus analysis tool. The software has built-in features to explore N-grams, frequency of N-grams, range of N-grams and concordance of N-grams. For this purpose the range of N-gram was set on 2 texts, which is 10% of the total number of texts, and the length of N-gram was specified on 3 words. The N-grams which fulfilled the criterion of lexical bundle for the current study was identified by the software which was then further analyzed for structural pattern according to the framework given below.

3.3. Theoretical Framework

After the lexical bundles were extracted from the corpus these were placed in the structural categories provided by Biber and Conrad (1999). For the structures which did not fall in any of the categories, some new categories were created. The taxonomy suggested by Biber and Conrad (1999) has been shown below and the details of the categorizations have been given in the result and discussion section.

1. Noun Phrase Fragments
2. Verb Phrase Fragments
3. Noun + Verb Phrase Fragments
4. Prepositional Phrase Fragments
5. Verb/Adjective/Noun + to/that Phrase Fragments
6. Adverbial/Adjectival Phrase Fragments
7. Anticipatory It, That/There/To Fragments
8. Other Expressions

4. Results and Discussion

Table 2 shows the number of 3-word tokens and types in the corpora of Pakistani newspaper editorials and news blogs. The corpus of editorial has 519 3-word token and 205 types which means the token-type ratio in the corpus is 02.53. On the other hand, the number of 3-word token and type in the corpus of news blogs is 394 and 116 respectively which make the token-type ration 03.40. The token-type ratio shows that the repetition of a lexical chunk in news blogs is higher as compared to newspaper editorials. Moreover, the number of tokens per thousand words highlights the significant fact that editorials have a number of bundle tokens in a thousand words than the blogs which is 61.82 against 22.85. The normalized frequency of bundles per thousand words means that the newspaper editorial writers use them three times more than blog writers. It shows their competence and reliance on the prefabricated structures.

Table 2: 3-word Lexical Bundles

	3-Word Fragment Tokens	Tokens per 1000 words	3- Word Fragment Types	3-Word Token-Type Ratio
Editorials	519	61.82	205	02.53
Blogs	394	22.85	116	03.40

4.1. Noun Phrase Fragments

These structures start with a noun phrase fragment which may or may not be followed by a prepositional phrase. Table 3 shows that there are 82 bundles that belong to this category in the corpus of Pakistani news editorials which make 40% of the total lexical bundles in the corpus of newspaper editorials whereas there are 27 3-word lexical bundles in the corpus of Pakistani blogs which make 23.27% of the total bundles. The number of lexical bundles as well as the percentage of lexical bundles in this category show that this is the dominant category for the news editorials. The studies of Yousaf (2019), Yousaf and Shehezad (2018), and Hyland (2008) show that in academic discourse majority of the expert writers in the discipline of humanities use noun phrase fragments. This means that the editorial writers are more close to the expert writers in the discipline of humanities as compared to the blog writers. Moreover, it is important to mention here that the predominate use of Noun Phrase Fragments show that the Editorial writers rely more on argument building than presenting facts and figures (see Yousaf et al., 2019) as compared to the blog writers.

Table 3: Noun Phrase Fragments

	Editorials	Blogs
1. Bundles	82	27
2. Percentage	40%	23.27%

4.2. Verb Phrase Fragments

These structures start with a verb phrase fragment. Table 4 shows that there are 23 bundles that belong to this category in the corpus of Pakistani news editorials which make 11.22% of the total lexical bundles in the corpus of newspaper editorials whereas there are 17 3-

word lexical bundles in the corpus of Pakistani blogs which makes 14.66% of the total bundles. The number of lexical bundles as well as the percentage of lexical bundles in this category show that this is not a dominant category for the news editorials as well as for the blogs. The previous studies of Yousaf (2019), Yousaf and Shehzad (2018), and Hyland (2008) show that in academic discourse majority of the expert writers in the discipline of science use verb phrase fragments. The use of verb phrase fragments gives prominence to actions than that of in writing. This means, although the difference is not great but still, the blog writers are more close to the expert writers in the discipline of sciences as compared to the editorial writers. This means that the blog writers are slightly more prone to present facts and figures to make their argument strong as compared to the Editorial writers. The studies of Yousaf (2019), Yousaf and Shehzad (2018), and Hyland (2008) show the predominant use of Verb Phrase Fragments in science subjects where writers and scholars rely heavily on facts and figures to strengthen their argument.

Table 4: Verb Phrase Fragments

	Editorials	Blogs
1. Bundles	23	17
2. Percentage	11.22%	14.66%

4.3. Noun + Verb Phrase Fragments

The lexical structures in this category start with a noun or pronoun and end with a verb phrase fragment which make them have a subject and a predicate. Table 5 shows that this category has only 10 bundles in the corpus of newspaper editorials that is 04.88% of the total bundles in the corpus. On the other hand, the number of lexical bundles in the corpus of Pakistani blogs is higher, i.e. 15 which makes 12.93% of the total bundles in the corpus of blogs. Although this category is not dominant in either of the corpora but still the lexical items in the bundles make it interesting to be discussed.

Table 5: List of Noun + Verb Phrase Fragments

List A	List B
1. the government is	1. he further explained
2. the government needs	2. i am not
3. government needs to	3. it was a
4. present government has	4. we used to
5. previous government have	5. you can not
6. the government has	6. you have a
7. the situation is	7. i am not
8. he said that	8. we have to
	9. i am the
	10. i have mentioned

The lexical items in the editorial consist of noun + verb predominantly. However, in news blogs the structures are predominantly pronoun + verb. Lists a and b given above indicate that the use of personal pronouns make the blog writers show their presence in the texts, whereas the editorial writers are absent from the text. The results of this category show that the blog writers are more subjective compared to the editorial writers and they display more assertive stance.

Table 6: Noun + Verb Phrase Fragments

	Editorials	Blogs
1. Bundles	10	15
2. Percentage	4.88%	12.93%

4.4. Prepositional Phrase Fragments

Table 6 shows that there are 44 lexical bundles in the corpus of editorials and 20 lexical bundles in the corpus of blogs. The percentage of this category in editorials and blogs is 21.46 and 17.24 respectively. The results show that the editorial writers rely more on prepositional phrase fragments. It means the difference in the use of this category between blog and editorial writers is not great. The overall percentage of this category is low which is in contrast to the results of research in academic discourse (see Hyland, 2008; Kwary, 2017; Yousaf, 2019; Yousaf & Shehzad, 2018). The list of the lexical bundles generated for this category shows that the editorial writers use a variety of prepositional structures as compared to the blog writers which shows their higher language competence and fluency.

Table 7: Prepositional Phrase Fragments

	Editorials	Blogs
1. Bundles	44	20
2. Percentage	21.46%	17.24%

4.5. Verb/Adjective/Noun + to/that Phrase Fragments

Table 7 below shows that editorial and blog writes do not use these structures in their writings. Yousaf (2019) found out that these structures are used by the scholars in the discipline of English studies which means that the writers with advanced competence in the language have higher probability to use complex structures.

Table 8: Verb/Adjective/Noun+to/that Phrase Fragments

	Editorials	Blogs
1. Bundles	xxxx	xxxx
2. Percentage	00%	00%

4.6. Adverbial/Adjectival Phrase Fragments

The lexical structures in this category are adverbial or adjectival phrase fragments. Table 8 shows that this category has only 07 bundles in the corpus of newspaper editorials that is 03.41% of the total bundles in the corpus. The number of lexical bundles in the corpus of Pakistani blogs is higher, i.e. 18 which makes 15.52% of the total bundles in the corpus of blogs. Although this category is not dominant in either of the corpora but still the higher percentage in the corpus of Pakistani blogs makes it significant. Higher number of adjectives makes writing closer to creative and literary writing and lack of adjectives signifies scientific nature of the writing. Hence, it can be stated that the blog writers are more literary and use more boosters to

emphasize their personal stance, as described in Noun/Pronoun+Verb category, whereas the editorial writers are more scientific and avoid their subjective evaluation of the matter under discussion.

Table 9: Adverbial/Adjectival Phrase Fragments

	Editorials	**Blogs**
1. Bundles	07	18
2. Percentage	3.41%	15.52%

4.7. Anticipatory It, That/There/To Fragments

The structures in this category start with *it, that, there, to,* and may have varied endings (the variation of endings has been shown in lists c and d below).

Table 10: Variation of Endings in Anticipatory It, That/There/To Fragments

List C	**List D**
1. there is a	1. there is a
2. it is a	2. there is no
3. that the government	3. that women are
4. it has been	4. that you are
5. it is important	5. that you can
6. it to the	6. there was no
7. that in the	7. to be taken
8. that the public	8. it was the
9. that there is	9. that i am
10. there is an	10. there is no
11. this is why	11. it was a
12. to take the	12. that she has
13. to work towards	13. that we have
	14. there is nothing
	15. there was a
	16. to live in
	17. to enjoy the
	18. to make a

Table 9 below shows that there are 13 bundles that belong to this category in the corpus of Pakistani news editorials which make 06.34% of the total lexical bundles in the corpus of newspaper editorials whereas there are 18 3-word lexical bundles in the corpus of Pakistani blogs which make 15.52% of the total bundles. The number of lexical bundles as well as the percentage of lexical bundles in this category shows that the category is not dominant for both the corpora; however, its percentage is too low in the corpus of editorials. The studies of Hyland (2008), Yousaf (2019), and Yousaf and Shehezad (2018) show that in academic discourse majority of the expert writers in the discipline of science use *anticipatory it fragments*. This means that the blog writers are more close to the expert writers in the discipline of science as compared to the editorial writers. The predominant use of Noun Phrase Fragments shows that the

editorial writers rely more on argument building than presenting facts and figures (see Yousaf et al., 2019) as compared to the blog writers.

Although these patterns are used to passivize the structure to show absence of the writer from the text and make the text unbiased (Hyland, 2008; Yousaf, 2019) but here the editorial writers are not using them which is against the norm set by the previous categories where the editorial writers are most of the times absent from the writing and try to build an unbiased argument. But at the same time the lower percentage of the structures justifies the fact that the editorial writing is based on opinions and views instead on reporting of facts and referring to evidence. This means bloggers rely more on description and reporting and they want to keep the source in the foreground.

Table 11: Anticipatory It, That/There/To Fragments

	Editorials	**Blogs**
1. Bundles	13	18
2. Percentage	6.34%	15.52%

4.8. Other Expressions

This category has the structural patterns which could not be placed in any of the above mentioned categories; hence the lexical bundles encapsulated in this category have varied structures. The table below shows that the number of lexical bundles in both the corpora is very low as there are 11 lexical bundles in the corpus of editorials and only 6 lexical bundles in the corpus of blogs. The percentage of the bundles in editorials and blogs is 5.36% and 5.17% respectively which is very low in both the corpora. Majority of the lexical bundles belonging to this category in the academic written discourse is citations and references (see Yousaf 2019); however, some of the structures may stand as simple sentence. One such example is: *I was thinking*.

Table 12: Variations of the Lexical Items

List E	List F
1. and Asian development	1. and they were
2. and the government	2. and tried to
3. and after the	3. i was thinking
4. and lack of	4. and all other
5. and multilateral donor	5. and his family
6. as well as	6. but there was
7. but the government	
8. than the ones	
9. if the government	
10. to percent of	
11. I insaaf pti	

The other dominant lexical items are conjunctions. The variations of the lexical items can be seen in lists e and f above.

Table 13: Other Expressions

	Editorials	Blogs
1. Bundles	11	06
2. Percentage	5.36%	5.17%

5. Comparison of Structural Categories

Table 11 highlights the difference in the different structural categories used by Pakistani editorial writers and bloggers. The highest percentage of lexical bundles is the *Noun Phrase Fragments* in the corpus. This category is not dominant in the corpus of editorials but it dominates both the corpora and it is the highest category in the corpus of blogs which is 23.27%. The second highest category in the corpus is of *Prepositional Phrase Fragments*. Similar to Noun Phrase Fragments, this category is also prevalent in both the corpora, although in the corpus of editorials the difference of percentage with respect to highest category is very high. The use of Prepositional Phrase Fragments is similar to the use of this category in academic discourse where it is frequent category in humanities, social science and science disciplines (see Hyland, 2008; Yousaf, 2019). The remaining categories have significantly low percentages in the corpus of editorials, whereas the corpus of blogs has equal distribution of lexical bundles across all categories except Verb/Adjective/Noun+to/that Phrase Fragments and Other Expressions. The dominance of lexical bundles in Noun Phrase Fragments and Prepositional Phrase Fragments show the emphasis of the writers on the argument building and the lower percentage of *Verb Phrase Fragments* shows the lack of reference to the evidence, and description of facts and figures. All of these features are higher in the editorials which make them better composed and effective writings in the journalistic discourse.

Table 14: Comparison of Structural Categories

Structural Categories	Percentage	
	Editorials	**Blogs**
1. Noun Phrase Fragments	40 %	23.27 %
2. Verb Phrase Fragments	11.22 %	14.66 %
3. Noun + Verb Phrase Fragments	04.88 %	12.93 %
4. Prepositional Phrase Fragments	21.46 %	17.24 %
5. Verb/Adjective/Noun + to/that Phrase Fragments	00%	00%
6. Adverbial/Adjectival Phrase Fragments	03.41 %	15.52 %
7. Anticipatory It, That/There/To Fragments	06.34 %	15.52 %
8. Other Expressions	05.36 %	05.17 %

6. Conclusion

The paper aimed to explore lexical chunks in Pakistani newspaper editorials and Pakistani news blogs. The study focused to compare the language of editorials and blogs with reference to fixed fragments for the purpose of which three questions were set in the beginning. The first question dealt with the use of different structural patterns of fixed phrases. The corpus of the study was analyzed with the help of framework provided by Biber and Conrad (1999) for the structural analysis of lexical bundles. The results of the study confirmed that all structural categories were present in the corpus Verb/Adjective/Noun+to/that Phrase Fragments. The

second question focused on the similarity and difference in the use of lexical bundles in the editorials and blogs by the Pakistani writers. The results highlighted that both the newspaper editorial writers and blog writers rely on fixed phrase for writing editorials and blogs. The third question explored the dominant categories of lexical bundles used by the writers in editorials and blogs. The results and discussion section highlighted that both the editorials and blogs have the use of *Noun Phrase Fragments* predominantly, but the percentage of this category in editorials is higher in the corpus of editorials than that of the corpus of the blogs. The second dominant category used by the editorial and blog writers was *Prepositional Phrase Fragments* which is actually in line with the norms of academic writing. The remaining categories have very low percentage in the corpus of editorials, but the percentages in the corpus of blogs are moderate and lexical bundles are almost equally distributed which is the other major difference in the use of lexical bundles in editorials and blogs. Therefore, it can be concluded that both editorial and blog writers of Pakistani origin rely on fixed phrases but their use of fixed phrases and distributions of the fixed phrases across structural categories is significantly different which in return highlights the difference of linguistics competence and fluency of the writers. The study suggests that phraseology of the newspapers editorials should be explored to find out the functions of lexical bundles too.

References

Altenberg, B. (1998). Connectors and sentence openings in English and Swedish. In S. Johansson & S. Oksefjell (Eds.), *Corpora and cross-linguistic research: Theory, method and case studies* (pp. 115-144). Leiden, Netherlands: Brill Publishers.

Banerjee, S., & Pedersen, T. (2003). The design, implementation, and use of the ngram statistics package. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 370-381). Berlin, Germany: Springer.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263–286.doi:10.1016/j.esp.2006.08.003

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgård & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 181–189). Amsterdam, Netherlands: Rodopi.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371–405. doi:10.1093/applin/25.3.371

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

Cortés, K. H. (2002). Youth and the study of foreign language: An investigation of attitudes. *Foreign Language Annals, 35*(3), 320-332. doi: 10.1111/j.1944-9720.2002.tb01857.x

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397–423. doi: 10.1016/j.esp.2003.12.001

Cowie, A. P. (1988). Stable and creative aspects of language use. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 126–139). London, England: Longman.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 42*(3), 375-396.

Firth, J. R. (1957). Ethnographic analysis and language with reference to Malinowski's views. In R. Firth (Ed.), *Man and culture – An evaluation of the work of Bronislaw Malinowski* (pp. 93-118). London, England: Routledge & Kegan Paul.

Haswell, R. H. (1991). *Gaining ground in college writing: Tales of development and interpretation.* Dallas, TX: Southern Methodist University Press.

Halliday, M. A. K. (2014). Language as social semiotic. In R. Wodak, J. Angermuller, & D. Maingueneau (Eds.), *The discourse studies reader: Main currents in theory and analysis* (pp. 263-272). Amsterdam, Netherlands: John Benjamins.

Halliday, M. A., MacIntosh, A., & Strevens, P. (1964). *The linguistic sciences and language teaching.* London, England: Longman.

Hyland, K. (2004). *Genre and second language writing.* Ann Arbor, MI: University of Michigan Press.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4–21. doi: 10.1016/j.esp.2007.06.001

Kwary, D. A., Ratri, D., & Artha, A. F. (2017). Lexical bundles in journal articles across academic disciplines. *Indonesian Journal of Applied Linguistics*, *7*(1), 132-140. doi:10.17509/ijal.v7i1.6866

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development, 16*(4), 495-500.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching.* New York, NY: Oxford University Press.

Scott, M. (1997). PC analysis of key words—And key key words. *System, 25*(2), 233–245. doi: 10.1016/S0346-251X(97)00011-0

Strzalkowski, T. S. G. (1998). Wise GB Pérez-Carballo J. Tapanainen P. jarvinen T. Voutilainen A. and Karlgren J. Natural Language Information Retrieval: TREC-7 Report. In Proceedings of TREC-7 Conference.

Strzalkowski, T., Stein, G., Wise, G. B., Perez-Carballo, J., Tapanainen, P., Jarvinen, T., Voutilainen, A. & Karlgren, J. (1998). Natural Language Information Retrieval: TREC-7 Report. In proceedings of the *Seventh Text REtrieval Conference (TREC 7).* Gaithersburg, Maryland: NIST Special Publications.

Stubbs, M. (2002). Some basic sociolinguistic concepts . In J. K. Dowdy & L. Delpit (Eds.), *The skin that we speak: Thoughts on language and culture in the classroom* (pp. 63-85). New York, NY: The New Press.

Yorio, C. A. (1980). Conventionalized language forms and the development of communicative competence. *TESOL Quarterly, 14*(4), 433-442. doi: 10.2307/3586232

Yousaf, M., & Shehzad, W. (2018). Prevalence of prefabricated structures in academic discourse: A corpus-based study. *International Journal of English Linguistics, 8*(5), 297–306.

Yousaf, M. (2019). *A corpus-based study of lexical bundles as building blocks of academic discourse* (Unpublished doctoral dissertation). Air University, Islamabad, Pakistan.

Yousaf, M. & Shehzad, W. (2019). Fixedness of expressions in doctoral research dissertations: A corpus based analysis. *Kashmir Journal of Language Research, 21*(2), 27-46.

Yousaf, M., Khan, I. A., & Habib, A. (2019). Study of gendered language in blogs: A corpus-based analysis of fixed phrases. *Corporum: Journal of Corpus Linguistics, 2*(1), 37-49.