

Phraseological Expressions: Gender-Based Corpus Analysis of EFL/ESL Academic Research Articles

Ismat Jabeen

Department of English Language and Literature, Prince Sattam Bin Abdulaziz University, Sharia Abdullah bin Amir, 16278, Alkharj, Saudi Arabia, KSA

Nuha Alsmari

Department of English Language and Literature, Prince Sattam Bin Abdulaziz University, Sharia Abdullah bin Amir, 16278, Alkharj, Saudi Arabia, KSA

Abstract—Phraseological or multi-word-pattern corpus-driven analysis of language in use has offered significant insights in recent years into how linguistic discourse can vary. This variation has been researched across genres, registers, disciplines, and native or non-native differences. However, very few studies have presented the gender-based analysis of academic research discourse within the EFL/ESL perspective. The current study explored the use of lexical bundles practiced by male and female researchers working in the EFL/ESL academic context within KSA. Corpora comprising almost 300,000 words including 68 research articles, 36 by female and 32 by male researchers were collected and run through Lancsbox 6.0 software package. The analysis was based on the frequency and structural patterns across the selected data. For the critical analysis of structural patterns, the structural taxonomy framework offered by Gezegin-Bal (2109) adapted from Biber et al. (1999) was employed. As established by the findings of the study, prepositional and noun phrases remained overwhelmingly more frequent and common in both corpora. There were no significant gender-based differences in the use of lexical bundles found which reflects that both male and female researchers practiced similar expressions in their use of the English language.

Index Terms—lexical bundles, corpus analysis, gender-based comparative analysis, EFL/ESL research

I. INTRODUCTION

Language, a uniquely human characteristic, is not simply a means of communication rather it is an embodiment of one's shared knowledge, belief, culture, and identity (Kramsch, 2014). Resultantly, when one uses spoken or written discourse, they not only express their own expressions but rather that discourse reflects the nature of the language itself as well as the cultural and prescriptive meanings and beliefs associated with it. In addition, linguistically, every language user's individualistic variety of language also depends on the factors such as their background, discipline, genre, and context. One apt example of this can be the distinction between the practice of formal and informal language such as the language of academia and social media respectively. So, how can such linguistic varieties be explored and studied systematically?

One widely explored and investigated phenomenon in the last few decades is named the phraseological or formulaic use of language as pointed out by Wray (2006) "when we speak, we select particular turns of phrase that we perceive to be associated with certain values, styles and groups" (p. 593). The phraseology or formulaic language use has been studied under diverse names such as "chunks, phraseological sequences, formulaic language, lexical bundles, collocations, multi-word items, recurrent sequences, n-grams, lexical phrases" (Ang & Tan, 2018, p. 82). The cluster expressions or more commonly referred to as lexical bundles are considered an essential component of natural discourse serving as the "lexico-grammatical base" (Dastjerdi & Rafiee, 2011, p. 61) of a language owing to their repetitive characteristic. The term 'lexical bundles'- the focal point of this study, was first introduced and defined by Biber et al. (1999) as "recurrent words that occur together (e.g. the fact that, I don't think so, etc.) regardless of their idiomaticity (i.e. the quality of being idiomatic), and regardless of their structural status" (p. 990). Three fundamental features set lexical bundles apart from other multiword items or sequences i.e., they recurrently occur in various registers; are not idiomatic in nature, and generally are not complete-unit expressions (specifically in academic discourse) (Dastjerdi & Rafiee, 2011).

Phraseological constructions and context-relevant use of them in a given language indicate users' level of linguistic competence in and mastery over that language (Dastjerdi & Rafiee, 2011). When it comes to academic discourse, competence in phraseological sequences appears to be inevitable to access and become a part of the "academic community" (Coxhead, 2008, p. 151). Academic research is an extensively relied upon medium by a massive research community for dispersing theoretical knowledge, new insights, and hypotheses. However, the aim of receiving and

imparting knowledge and information effectively can only be achieved when the writers are proficient in “using standard words, phrases, and a formulaic language” both at the lexical and syntax levels (Budiyanto & Suhardijanto, 2020, p. 292). Furthermore, while writing scholarly journal articles in itself a daunting task, there it also needs to adhere to discipline-specific rhetorical and linguistic patterns and variations (Hyland, 2016).

Linguists and researchers probe how the writers use language individually as well as collectively as a part of a larger research community. The possibility of accessing, compiling, assessing, and evaluating large corpora for such purposes opened a new world of possibilities and opportunities in the field of corpus analysis (Adel, 2010a; Biber, 2010). Computational tools facilitate the researchers in systematically analyzing the real-time and context-based use of linguistic structures such as phraseological expressions within a target language. For example, from classroom discourse to newspaper content and political speeches, numerous researchers have endeavored to analyze statistically and qualitatively how language is used and manipulated in different day-to-day contexts. A significant amount of corpus research focuses on how a second or foreign language is used specifically within academic contexts by learners, teachers, common speakers, researchers, etc.

English is considered a lingua franca for reaching the multilingual research community and has been opted for by researchers across disciplines and countries for academic purposes (Montgomery, 2013). It is highlighted in some of the previous research studies (Ellis et al., 2008; Güngör & Uysal, 2016; Salazar, 2014) that the variation in the use of lexical bundles and recurrent structures not only varies from discipline to discipline, rather from native and non-native perspectives too. It has been established that non-native speakers make use of different lexical bundles than native speakers such as “overuse of clausal or verb-phrase based lexical bundles” by non-native writers as compared to “noun and prepositional phrase-based lexical bundles” opted by their native counterparts (Güngör & Uysal, 2016, p. 176).

Taking into account the established consistency of the use of lexical bundles across specific fields and disciplines, the current study aimed at studying this phenomenon further by analyzing the use of phraseological expressions in the research articles produced by male and female researchers published in the field of EFL within the Saudi. First, the study intended to explore the nature and frequency of lexical bundles used by both male and female researchers. Then it aimed to examine the frequently used structural patterns of lexical bundles by male and female researchers. Finally, the research focused to investigate whether or not there were any gender-based similarities and/or differences present in the use of lexical bundles by both genders. In light of these objectives, the following questions were formulated and answered:

Q1. How do the lexical bundles used by both male and female researchers differ in nature and frequency?

Q2. What are the frequently use structural patterns of lexical bundles by male and female researchers?

Q3. What type of gender-based similarities and/or differences do emerge (if any) in the use of lexical bundles?

Keeping in view the prevalent veracity of genre, register, and discipline-based use of specific recurrent expressions and clusters in English, it became a worthwhile endeavor to explore and understand how the male and female academic researchers working in the field of EFL within KSA employ English in their research. And, for the linguistic community specifically, the study may contribute to advancing the research held in the area of corpus-based linguistic analysis focusing on the emerging phraseological patterns and sequences within the selected data. Moreover, the study addressed the gap existing in the current research conducted in KSA in the fields of the corpus as well as gender-based use of word clusters. Future researchers, teachers, and students interested in the arenas of the corpus, and gender-based linguistic analysis may benefit from the interpretation and the findings of this study.

II. LITERATURE REVIEW

With the advancement in computer technologies, researchers focused on the exploration of formulaic language practice by incorporating large corpora to establish how language has been used in real contexts (Adel, 2010a; Biber, 2010; Gu, 2006). This advancement opened remarkable vistas for investigating and probing language practices in a variety of ways and perspectives. For instance, early studies focused on analyzing how a particular language is used by native and non-native speakers and writers (Meunier, 2002; Muller, 2005; Schmid, 1990). In this aspect, the majority of the researchers have comparatively analyzed native English and other language speakers’ and writers’ use of English in both spoken and written discourse. For example, in spoken discourse, the comparative studies of discourse markers (Buisse, 2010; Huang, 2011; Muller, 2005), phonological and intonational properties (Gut, 2009; Verdugo, 2005) lexicalised language (De Cock, 1998; Foster, 2013) are only a few examples to quote.

Along with spoken discourse, research in the use of written discourse also gained momentum. Numerous research studies focused on probing the primary and advanced level functions of a target language through a large corpus within the academic sphere e.g. use of sentence connectors and conjunctions (Granger & Tyson, 1996; Kuswoyo et al., 2020) recurrent word combinations (Adel & Erman, 2012; Güngör & Uysal, 2016) novice and expert differences in the of lexical bundles (Zhang & Zhang, 2021) or simply academic writing as a genre in itself (Evison, 2013; Gray, 2011).

Later, along with academic discourse, the other variants of genre and discipline-based analysis of language also started to emerge, such as legal discourse (Breeze, 2011; Upton, 2004; Phillips & Egbert, 2017), medical discourse (Ferguson, 2001; Zinn & McDonald, 2015), marketing discourse (Kheovichai, 2014; Labrador et al., 2014; Simon & Dejica-Cartis, 2015), political discourse (Adel, 2010b; Taylor, 2022; Willis, 2017), feminist discourse (Jaworska &

Krishnamurthy, 2012; Santaemilia, 2017), media discourse (Callies & Levin, 2019; O'Halloran, 2010) to name a few endeavors in corpus linguistic analysis.

Since Lakoff's (1975) claims that women use more polite language containing more hedges, tag questions, intensifiers, and empty adjectives, resultantly exhibiting less power than men, numerous studies have sought to explore the gendered language phenomenon in length offering valuable insights. For instance, Drass (1986) established that men speak more than women whereas this claim was refuted by another study later conducted by Brizendine (1994) finding that women produce more speech than men do. Schmid (2003) demonstrated through an interesting research that "women and men actually live in different cultures" (p. 12) as the female discourse is more involving whereas male discourse is more detached. An important contribution was Rühlemann's (2010) study which focused on the probing of the comparatively less researched area of conversational grammar, establishing that female discourse is particularly more grammar oriented than their counterparts. Similarly, studies have highlighted gender-based differences emerging in the use of English by the natives. Such as Fuchs (2017) investigated the sociolinguistics of intensifier usage in two corpora of spoken BrE from the early 1990s and early 2010s, concluding that men use overall fewer intensifiers than women remaining consistent irrespective of the time period which supports earlier claims of Lakoff.

Within the Saudi context, a number of studies have employed the corpus approach to analyze English language use in varied contexts. For example, various studies have explored English language practice within academic settings (Alattar, 2014; Almujaivel, 2018; Jawhar, 2012). Corpus-based research has also been applied in media discourse analysis, such as in the studies by Alotaibi (2021) and Altoaimy (2018). However, there is not even a single research study that highlights and investigates how male and female researchers use English in their published research articles with regard to phraseological or formulaic language expressions practiced within the academic context to see whether or not there emerge any gender-based differences. This research addressed this gap by analyzing the corpora based on research published by both male and female researchers.

III. RESEARCH METHODOLOGY

This study followed a corpus-driven approach to let the linguistic choices of the male and female writers emerge from the data and the analysis itself instead of adopting a corpus-based approach which is for "the validity of linguistic forms and structures derived from linguistic theory" (Biber, 2015, p. 1).

A. *Corpora of the Study*

The research corpora for the current study comprised 68 articles containing approximately 300,000 words. Out of 68 articles selected, 36 articles (words 150085) were produced by females, whereas 32 articles (words 150126) were written by male researchers. The articles were published in recognized, scientific journals within the EFL/ESL/ELT fields. The selection of corpora was based on the following criteria. Firstly, the articles falling within the Saudi academic context were selected. Secondly, articles had to be related to EFL/ESL fields. Also, all the articles were published within the last decade. Lastly, the number of articles selected from both male and female researchers depended on generating almost an equal number of words so that corpora are balanced.

B. *The Corpus Tool & Framework Applied*

In the next stage, the selected corpus was processed and analyzed using commonly adopted software named Lancsbox 6.0, developed by Lancaster University for the analysis of language corpora (Brezina, 2015; McEnery & Wattam, 2015). Considering the arbitrariness of the frequency threshold, a 10 times frequency threshold was deemed suitable for the current research as it offered a significant variety of lexical bundles employed in the selected corpora. The 10 times frequency criteria refer to the number of times lexical bundles occurred within the selected corpora of 300,000 words, as for a small size corpora like this study, "a higher frequency threshold" (Hsu & Hsieh, 2018, p. 529) was regarded more effective. Similarly, the phraseological structures based on four words length were extracted. To critically evaluate these bundles, the structural taxonomy framework offered by Gezegin-Bal (2019) slightly adapted from Biber et al. (1999) was applied. The functional analysis of the lexical bundles was not included for the purpose of focusing more on the frequency differences and the structural patterns of the lexical bundles found. The subsequent table presents the framework adopted for the analysis of the collected corpus.

TABLE 1
STRUCTURAL TAXONOMY OF LEXICAL BUNDLES

S. No	Category	Sub-category	Types	Examples
1.	Phrasal	NP-based	- (connector +) NP with of- phrase fragment - NP with other post-modifier fragment	the end of the the extent to which
2.		PP-based	- PP with embedded of-phrase fragment - Other Prepositional Phrase (fragment)	as a result of at the same time
3.		VP-based	- Anticipatory it + VP/adjective P + comp. cl - Passive verb +PPf - Copula be + noun phrase/adjective phrase - Pronoun/NP + be	it is important to is based on the is one of the this is not the
4.	Clausal	-	- (verb/adjective +) to-clause fragment - (VP +) that-clause fragment - Adverbial clause fragment	is likely to be should be noted that as shown in the figure
5.	Other expressions	-	-	as well as the

Gezegin-Bal (2019).

IV. DATA ANALYSIS AND DISCUSSION

A. Data Analysis Procedures

Data were processed and categories were established through the following procedures. The collected data was run through Lancsbox 6.0. The lexical bundles (60 in number) were identified in both corpora through Ngrams, employing the frequency-based approach. Once the lexical bundles were retrieved, they were manually reviewed to address any discrepancies or exclude irrelevant items such as subject-specific expressions like English as a foreign/second language. Lastly, the identified bundles were comparatively analyzed first, as per their structural patterns.

Structural Analysis of the Lexical Bundles

In Table 2, the lexical bundles identified in both corpora are presented.

Lexical Bundles Identified

A cursory glance over the data presented in Appendix A demonstrates with regard to frequency, males used the lexical bundles more often than their counterparts with a frequency ratio of 41: 33 which is notable. Similarly, prepositional and verb phrases were employed more frequently by the male researchers than the females. The females showed a tendency of using a higher number of noun phrases (though not very significantly), and clausal phrases than the males. The next sections offer the analysis of identified bundles under the structural taxonomy framework.

B. Phrasal Structures

Prepositional Phrase Lexical Bundles

The majority of the phraseological expressions identified fall in the prepositional and noun phrases conforming to the findings of the previous studies (Chen & Baker, 2010; Gezegin-Bal, 2019; Salazar, 2014). A prepositional phrase (PP) contains a preposition followed by a noun, pronoun, or noun phrase such as on time, in the room, called complements. The complements are an inseparable part of a PP and a modifier can be an optional element (Downing & Lock, 2006).

TABLE 2
PREPOSITIONAL PHRASE LEXICAL BUNDLES

Rank	Male Corpus		Female Corpus	
	Frequency	Expressions	Frequency	Expressions
1	41	on the other hand	33	in the Saudi context
2	24	as it is evident	32	on the other hand
3	24	of Saudi efl learners	25	in the field of
4	23	in the Saudi context	22	at the same time
5	23	of the target language	22	outside of the classroom
6	21	in the target language	22	of the current study
7	20	of fl reading anxiety	21	of cognitive and metacognitive
8	18	in Saudi Arabia the	21	in the learning process
9	17	to find out the	20	in the target language
10	16	in the process of	18	of the target language
11	15	to the target language	16	to find out the
12	15	by Saudi efl learners	15	in the preparatory year
13	15	in the field of	15	as a result of
14	14	in the process of	14	of the study in
15	14	at the graduate level	14	at the end of
16	13	at the end of	14	to participate in the
17	12	of a third language	13	in the use of
18	12	between teachers and student	12	in this study the
19	12	of the targeted students	12	in the current study
20	11	in the following excerpt	12	to the fact that
21	11	in the form of	12	on the use of
22	10	with the findings of	11	of the study in
23	10	for the purpose of	10	at the end of
24	10	to the fact that	10	in the process of
25	10	in the following excerpts	10	in the efl classroom
26	10	at the same time	10	in order to answer
27	10	in addition to the	10	as an English learning
28	-	-	10	to answer the research
29	-	-	10	of the effectiveness of

Source: own data.

The majority of the phraseological expressions in the concerned corpora produced by either male or female researchers belong to the PPs category. As presented in the table, there is a significant variety of lexical bundles under the PP category, employed by both genders, however, the frequency of the usage varies between male and female corpus. The male writers tended to use PPs noticeably more often as compared to the female writers. On the other hand, female researchers have used slightly more variety of PPs than their counterparts. The internal structure of most of these phrases appears to be PPs with embedded of-phrase fragments in both the corpora. Also, a significant number of PPs contain NPs construction such as in the process of, for the purpose of, in the field of, etc. Furthermore, some of the instances in the table above, contain adjectives as complements i.e. in the following excerpt, at the same time, in the target language, in the current study, etc.

Noun Phrase Lexical Bundles

A noun phrase is a phrase that contains a noun or pronoun as its head or performs the same grammatical function as a noun. The following table represents the instances of noun phrases identified in both male and female writers' corpora.

TABLE 3
NOUN PHRASE LEXICAL BUNDLES

Rank	Male Corpus		Female Corpus	
	Frequency	Expressions	Frequency	Expressions
1	21	use of Arabic in	23	the results of the
2	18	the findings of the	19	female Saudi teachers in
3	18	the use of the	18	both teachers and students
4	17	the end of the	18	the results showed that
5	17	the ministry of education	17	the findings of this
6	16	the result of the	16	the end of the
7	16	kinds of corrective feedback	15	a positive attitude towards
8	15	the analysis of the	14	the findings of the
9	15	the use of Arabic	14	duration of studying English
10	13	use of L1 in	14	teachers' perceptions of the
11	12	knowledge of a third	14	findings of this study
12	12	results of the study	14	the effectiveness of the
13	12	both teachers and students	13	the majority of the
14	11	promotion of learner autonomy	13	the use of L1
15	11	foreign language reading anxiety	13	a result of the
16	11	the promotion of learner	12	the results of this
17	11	proficiency in English reading	12	the purpose of this
18	10	an online peer review	12	the extent to which
19	10	teachers in Saudi Arabia	11	the use of social
20	10	scores of the targeted	11	the process of learning
21	-	-	11	affective factors involved in
22	-	-	11	half of the teachers
23	-	-	10	the target language and
24	-	-	10	results of the study

Source: own data.

As evident in the table above, interestingly female researchers' corpus exhibits higher numbers and frequency of noun phrases (hence NP) as compared to the male corpus. As far as the nature of the NP is concerned, one visible similarity that emerges in both corpora is an abundant use of prepositional phrases as post-modifiers. Also, noun phrases with pre and post-modifiers are perceptible in both corpora. Moreover, both male and female writers rely heavily on NP with of-phrase fragment structures, similar to the lexical bundles in PPs. In addition, irrespective of gender, in both corpora, the majority of the pre-modifiers fall into the identifier category (a, an, the), with the to be the most commonly used. Occasionally, a determiner is followed by an adjective in the NP, for example, an online peer review (male corpus, rank 18), a positive attitude towards (female corpus, rank 15). However, the difference emerges in the number of noun phrase bundles with either pre or post-modifiers. The male corpus contains a similar percentage of pre and post-modifier noun phrases, whereas, in the female corpus, the percentage of pre-modifier phrases remains overwhelming, 78% to be exact. Moreover, the noun phrases identified in the female corpus offer more variety of lexical bundles such as containing an adjective as a pre-modifier as in 'affective factors involved in' (rank 21) or a quantifier 'half of the teachers' (rank 22).

TABLE 4
VERB PHRASE LEXICAL BUNDLES

Rank	Male Corpus		Female Corpus	
	Frequency	Expressions	Frequency	Expressions
1	24	it is evident in	20	is one of the
2	22	is one of the	13	it is important to
3	21	is evident in	10	there is a need
4	13	it was found that	10	the most frequently used
5	11	be attributed to	10	involved in learning efl

Source: own data.

Verb Phrase Lexical Bundles

As compared to NPs and PPs, the occurrences of VP were minimum with similar frequency in both corpora. A few VPs followed copula be + noun phrase/adjective phrase i.e. is one of the, is evident in. Some others had anticipatory it + VP/adjective P + comp. cl structures such as it is evident in, it was found that, it is important to. At the frequency level, the males' corpus indicated slightly more VPs than the females.

Clausal & Miscellaneous Lexical Bundles

TABLE 5
CLAUSAL & MISCELLANEOUS LEXICAL BUNDLES

Rank	Male Corpus		Female Corpus	
	Frequency	Expressions	Frequency	Expressions
1	-	-	10	that there is a (clausal)
2	16	as well as	22	one of the most
3	14	one of the most	16	as well as
4	11	that most of the	12	than half of the

Source: own data.

The corpora under investigation demonstrated the use of only one clausal fragment i.e. that there is a in the female corpus. Moreover, the same number of instances of other expressions not falling into any major categories were also noticed in both corpora. However, these instances were not very significant.

V. CONCLUSION

This study aimed at conducting a comparative analysis of the phraseological practices of male and female researchers published in the EFL/ESL paradigm within the Saudi context. The corpora of the study comprised 68 articles of 300,000 words. The analysis offers some interesting insights into the linguistic practices of both genders. Firstly, the most significant categories of lexical bundles that emerged were prepositional and noun phrases that conform to the majority of the previous research findings (Chen & Baker, 2010; Gezegin-Bal, 2019; Salazar, 2014). Also, the structural patterns including of-phrase fragments and noun-phrase constructions in prepositional phrases, and the use of prepositional phrases as post-modifiers, and phrases with pre and post-modifiers in noun phrases remained significant. With respect to gender differences, males' use of PPs was prominently more frequent than their counterparts, who in turn offered more variety of PPs than quantity. As for NPs, female writers exhibited higher numbers and frequency of noun phrases than male writers with an overwhelming tendency of using adjectives as pre-modifiers which fortifies some of the previous studies which highlight that the female language is more intensified and involving than the males (Rühlemann, 2010; Schmid, 2003; Xia, 2013). There were no substantial differences identified based on gender-oriented variances in the practice verb and clausal phrases. It is important to mention that this study was limited in its scope and focused only on gender-based differences in the frequency and structural patterns of the identified lexical bundles excluding the functional analysis. Further studies with larger corpora may yield varied results and findings and can add to the discussion developed in this study.

APPENDIX

Rank	Male Corpus		Female Corpus	
	Frequency	Expressions	Frequency	Expressions
1	41	on the other hand	33	in the Saudi context
2	24	as it is evident	32	on the other hand
3	24	it is evident in	25	in the field of
4	24	of Saudi efl learners	23	the results of the
5	23	in the Saudi context	22	one of the most
6	23	of the target language	22	at the same time
7	22	is one of the	22	outside of the classroom
8	21	is evident in the	21	of the current study
9	21	in the target language	21	of cognitive and metacognitive
10	21	use of Arabic in	21	in the learning process
11	21	evident in the following	20	is one of the
12	20	of fl reading anxiety	20	in the target language
13	18	the findings of the	19	female Saudi teachers in
14	18	the use of L1	18	both teachers and students
15	18	in Saudi Arabia the	18	the results showed that
16	18	statistically significant difference between	18	of the target language
17	17	the end of the	17	the findings of this
18	17	the ministry of education	16	the end of the
19	17	to find out the	16	to find out the
20	16	as well as the	16	as well as the
21	16	in the process of	15	a positive attitude towards
22	16	the result of the	15	in the preparatory year
23	16	in English language classes	15	as a result of
24	15	kinds of corrective feedback	14	of the study in
25	15	to the target language	14	the findings of the
26	15	the analysis of the	14	duration of studying English
27	15	by Saudi efl learners	14	at the end of
28	15	the use of Arabic	14	teachers' perceptions of the
29	15	in the field of	14	to participate in the
30	15	that the use of	14	findings of this study
31	14	in the process of	14	the effectiveness of the
32	14	at the graduate level	13	the majority of the
33	14	one of the most	13	the use of L1
34	13	at the end of	13	in the use of
35	13	use of L1 in	13	a result of the
36	13	it was found that	13	it is important to
37	12	knowledge of a third	12	than half of the
38	12	results of the study	12	the results of this
39	12	of a third language	12	in this study the
40	12	both teachers and students	12	the purpose of this
41	12	promotion of learner autonomy	12	in the current study
42	12	between teacher and students	12	to the fact that
43	12	foreign language reading anxiety	12	on the use of
44	12	of the targeted students	12	the extent to which
45	11	in the following excerpt	11	half of the teachers
46	11	be attributed to	11	the process of learning
47	11	the promotion of learner	11	the use of social
48	11	that most of the	11	affective factors involved in
49	11	in the form of	11	in the process of
50	11	proficiency in English reading	10	the target language and
51	10	with the findings of	10	that there is a
52	10	an online peer review	10	there is a need
53	10	for the purpose of	10	in the efl classroom
54	10	to the fact that	10	results of the study
55	10	in the following excerpts	10	the most frequently used
56	10	at the same time	10	in order to answer
57	10	teachers in Saudi Arabia	10	involved in learning efl
58	10	scores of the targeted	10	as an English learning
59	10	in addition to the	10	to answer the research
60	-	-	10	of the effectiveness of

ACKNOWLEDGEMENTS

This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444).

REFERENCES

- [1] Ädel, A. (2010). Using corpora to teach academic writing: Challenges for the direct approach. *Corpus-based approaches to English language teaching*, 39-55.
- [2] Ädel, A. (2010). How to use corpus linguistics in the study of political discourse. In *The Routledge handbook of corpus linguistics* (pp. 591-604). Routledge.
- [3] Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for specific purposes*, 31(2), 81-92.
- [4] Alattar, A. (2014). *Content-Based Instruction and Corpus Linguistics Curriculum for Early Advanced EFL Saudi Students* (Publication No. 91) [Master Thesis, University of San Francisco]. USF Scholarship: A Digital Repository@ Gleeson Library | Geschke Center.
- [5] Almujaawel, S. (2018). Analysing culture and intercultural in Saudi EFL textbooks: A corpus linguistic approach. *English Language Teaching*, 11(2), 31-43.
- [6] Alotaibi, N. (2021). *Social media and feminist activism in Saudi Arabia: a corpus-aided critical discourse analysis of the End Male Guardianship System campaign* (Doctoral dissertation, University of Sheffield).
- [7] Altoaimy, L. (2018). Driving change on Twitter: A corpus-assisted discourse analysis of the Twitter debates on the Saudi ban on women driving. *Social Sciences*, 7(5).
- [8] Ang, L. H., & Tan, K. H. (2018). Specificity in English for Academic Purposes (EAP): A Corpus analysis of lexical bundles in academic writing. *3L, Language, Linguistics, Literature*, 24(2), 82-94.
- [9] Biber, D. (2010). Corpus-based and corpus-driven analyses of language variation and use. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of linguistic analysis* (pp. 159–191). Oxford: Oxford University Press.
- [10] Brizendine, L. (2006). *The female brain*. Broadway Books.
- [11] Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 159–191). Oxford: Oxford University Press.
- [12] Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2). London: longman.
- [13] Breeze, R. (2011). Disciplinary values in legal discourse: a corpus study. *Ibérica, Revista de la Asociación Europea de Lenguas para Fines Específicos*, (21), 93-115.
- [14] Brezina, V., McEnery, T., & Wattam, S. (2015). Collections in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139-173.
- [15] Budiwiyanto, A., & Suhardijanto, T. (2020). Indonesian lexical bundles in research articles: Frequency, structure, and function. *Indonesian Journal of Applied Linguistics*, 10(2), 292-303.
- [16] Buysse, L. (2010). Discourse marker so in native and non-native spoken English (unpublished doctoral dissertation). *Gent: Universiteit Gent*.
- [17] Callies, M., & Levin, M. (Eds.). (2019). *Corpus Approaches to the Language of Sports: Texts, Media, Modalities*. Bloomsbury Publishing.
- [18] Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language learning & technology*, 14(2), 30-49.
- [19] Coxhead, A. (2008). Phraseology and English for academic purposes. In *Phraseology in language learning and teaching* (pp. 149–161). John Benjamins Publishing Company.
- [20] De Cock, S. (1998). A Recurrent Word Combination Approach to the Study of Formulae in the Speech of Native and Non-Native Speakers of English. *International Journal of Corpus Linguistics*, 3(1), 59-80.
- [21] Dastjerdi, H. V., & Rafiee, M. (2011). Corpus study of lexical bundles: Journalistic discourse in focus. *Asian ESP Journal*, 7(3), 59-78.
- [22] Downing, A. Locke, P. (2006). *English Grammar: A University Course*. Second Edition. Routledge, Taylor & Francis Group, London and New York.
- [23] Drass, K. A. "The effect of gender identity on conversation". *Social Psychology Quarterly* (1986): 294-301.
- [24] Ellis, N. C., Simpson-Vlach, R. I. T. A., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375-396.
- [25] Evison, J. (2013). A corpus linguistic analysis of turn-openings in spoken academic discourse: Understanding discursive specialisation. *English Profile Journal*, 3, E4. doi:10.1017/S2041536212000049
- [26] Ferguson, G. (2001). If you pop over there: a corpus-based study of conditionals in medical discourse. *English for Specific Purposes*, 20(1), 61-82.
- [27] Fuchs, R. (2017). Do women (still) use more intensifiers than men? Recent change in the sociolinguistics of intensifiers in British English. *International Journal of Corpus Linguistics*, 22(3), 345-374.
- [28] Gezegin-Bal, B. (2019). Lexical bundles in published research articles: A corpus-based study. *Journal of Language and Linguistic Studies*, 15(2), 520-534.
- [29] Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15(1), 17-27.
- [30] Gray, B. E. (2011). *Exploring academic writing through corpus linguistics: When discipline tells only part of the story* (Doctoral dissertation, Northern Arizona University).
- [31] Gu, Y. (2006). *Multimodal text analysis: A corpus linguistic approach to situated discourse*, 26(2), 127-167. <https://doi.org/10.1515/TEXT.2006.007>
- [32] Güngör, F., & Uysal, H. H. (2016). A Comparative Analysis of Lexical Bundles Used by Native and Non-native Scholars. *English Language Teaching*, 9(6), 176-188.
- [33] Hsu, C. C., & Hsieh, S. K. (2018). Identifying lexical bundles in Chinese: Methodological issues and an exploratory data analysis. *Language and Linguistics*, 19(4), 525-548.

- [34] Huang, L. F. (2011). *Discourse markers in spoken English: A corpus study of native speakers and Chinese non-native speakers* (Doctoral dissertation, University of Birmingham).
- [35] Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, 31, 58-69.
- [36] Jawhar, S. (2012). *Conceptualising CLIL in a Saudi context: A corpus linguistic and conversation analytic perspective* (Doctoral dissertation, Newcastle University).
- [37] Jaworska, S., & Krishnamurthy, R. (2012). On the F word: A corpus-based analysis of the media representation of feminism in British and German press discourse, 1990–2009. *Discourse & Society*, 23(4), 401-431.
- [38] Kheovichai, B. (2014). *Marketization in the language of UK university recruitment: A critical discourse analysis and corpus comparison of university and finance industry job advertisements* (Doctoral dissertation, University of Birmingham).
- [39] Kramsch, C. (2014). Language and culture. *AILA Review*, 27(1), 30-55.
- [40] Kuswoyo, H., Sujatna, E. T. S., Indrayani, L. M., & Rido, A. (2020). Cohesive conjunctions and so as discourse strategies in English native and non-native engineering lecturers: A corpus-based study. *International Journal of Advanced Science and Technology*, 29(7), 2322-2335.
- [41] Labrador, B., Ram n, N., Alaiz-Moren, H., & Sanjurjo-Gonz lez, H. (2014). Rhetorical structure and persuasive language in the subgenre of online advertisements. *English for specific purposes*, 34, 38-47.
- [42] Lakoff, R. (1975). Language and woman's place. *Language and society*, 2, 45-79.
- [43] Meunier, F. (2002). The pedagogical value of native and learner corpora in EFL grammar teaching. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp.119-142). Amsterdam: Benjamins.
- [44] Montgomery, S. L. (2013). *Does science need a global language?: English and the future of research*. University of Chicago Press.
- [45] Muller, S. (2005). *Discourse markers in native and non-native English discourse* (Vol. 138). John Benjamins Publishing.
- [46] O'Halloran, K. (2010). How to use corpus linguistics in the study of media discourse. In *The Routledge handbook of corpus linguistics* (pp. 563-577). Routledge.
- [47] Phillips, J. C., & Egbert, J. (2017). Advancing law and corpus linguistics: Importing principles and practices from survey and content-analysis methodologies to improve corpus design and analysis. *BYU L. Rev.*, 1589.
- [48] Ruhlemann, C. (2010). Conversational grammar-feminine grammar? A sociopragmatic corpus study. *Journal of English Linguistics*, 38(1), 56–87.
- [49] Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching* (Vol. 65). John Benjamins Publishing Company.
- [50] Santaemilia, J. (2017). A Corpus-Based Analysis of Terminology in Gender and Translation Research: The Case of Feminist Translation. In *Feminist Translation Studies* (pp. 15-28). Routledge.
- [51] Schmid, J. (1990). Corpus linguistics and non-native varieties of English. *World Englishes*, 9(3), 255-268.
- [52] Schmid, H. J. (2003). *Do women and men really live in different cultures? Evidence from the BNC*. na.
- [53] Simon, S., & Dejica-Cartis, D. (2015). Speech acts in written advertisements: Identification, classification and analysis. *Procedia-Social and Behavioral Sciences*, 192, 234-239.
- [54] Taylor, C. (2022). Corpus linguistics in the study of political discourse: recent directions. *The Routledge Handbook of Corpus Linguistics*, 602-614.
- [55] Upton, T. A. (Ed.). (2004). *Discourse in the professions: Perspectives from corpus linguistics* (Vol. 24). John Benjamins Publishing.
- [56] Verdugo, D. R. (2005). The nature and patterning of native and non-native intonation in the expression of certainty and uncertainty: Pragmatic effects. *Journal of Pragmatics*, 37(12), 2086-2115.
- [57] Willis, R. (2017). Taming the climate? Corpus analysis of politicians' speech on climate change. *Environmental Politics*, 26(2), 212-231.
- [58] Wray, A. (2006). Formulaic language. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (pp. 590–597). Oxford: Elsevier.
- [59] Xia, X. (2013). Gender differences in using language. *Theory and Practice in Language Studies*, 3(8), 1485-1489.
- [60] Zhang, S., Yu, H., & Zhang, L. J. (2021). Understanding the sustainable Growth of EFL students' writing skills: differences between novice and expert writers in their use of lexical bundles in academic writing. *Sustainability*, 13(10), 5553.
- [61] Zinn, J. O., & McDonald, D. (2015). Changing discourses of risk and health risk: A corpus analysis of the usage of Risk Language in the New York Times. In *Medicine, risk, discourse and power* (pp. 207-240). Routledge.

Ismat Jabeen was born in Pakistan. She earned a Doctor in English at the National University of Modern Languages, Islamabad, Pakistan, in 2015.

She is an Assistant Professor in the English Language & Literature Department at Prince Sattam bin Abdulaziz University, KSA. Formerly, she served at the University of Central Punjab (UCP), Lahore, Pakistan as Director Program (MPhil/MS English Linguistics, MPhil/MS English Literature & BS English) and Assistant Professor. She also served at Air University, Islamabad, and the University of Lahore (UOL) as Assistant Professor.

Dr. Jabeen has considerably published in the fields of corpus linguistics, sociolinguistics, and applied linguistics. She has been a part of various academic bodies and committees in distinguishing universities in Pakistan such as Member Board of Studies (UCP, UOL, Riphah International University), Member Graduate Research Committee (UCP), Member Academic Council (UCP), Member Self-Assessment Review Committee (UCP), Member MPhil & PhD English Admission Test Committee (UCP), Member Curriculum Review Committee (UOL). At Prince Sattam bin Abdulaziz University, KSA, Dr. Jabeen is: Member Quality Assurance Committee (BA, MA), Member Graduate Affairs Committee, Member Alumni Affairs Committee, Member Community Services Committee.

Nuha Alsmari serves as an Associate Professor of Applied Linguistics at Prince Sattam bin Abdelaziz University, KSA. She earned her PhD in Linguistics in 2017 from Imam Muhammed Bin Saud University, KSA. She has published in various Web of Science and Scopus indexed journals. Her research interests encompass pragmatic instruction, discourse analysis, teaching English as a foreign language (TEFL), and media-based language learning. ORCID <https://orcid.org/0000-0002-9121-0868>