Formulaic Sequences Used in Academic Writing Register

Abdulrahman Almosa
King Khalid University

The identification of formulaic sequences through a corpus-driven methodology is regarded as a fundamental aspect of achieving smooth linguistic expression and successful communication. The present study examined a prominent category of formulaic language known as lexical bundles. This research endeavors to examine the utilization of lexical bundles as formulaic structures in academic abstracts, introductions, and conclusions. Through the examination of 150 abstracts, introductions, and conclusions, 50 each, sourced from peer-reviewed articles within the domain of Language and Literacy, the quantity of formulaic language, specifically lexical bundles, is quantified and their pragmatic functions are analyzed. The Antconc software was utilized to produce formulaic sets consisting of two-word, three-word, four-word, and five-word lexical bundles, including a few six-word structures. Subsequently, the outcomes are juxtaposed with prior investigations in the respective field. The research indicates that lexical bundles, which are a type of formulaic language, exhibit a slightly lower frequency of usage in introductions and conclusions compared to abstracts.

Keywords: formulaic sequence, lexical bundles, AntConc, academic writing

INTRODUCTION

In recent decades, there have been numerous research paradigms that have explored the concept of formulaic sequences in academic writing, which pertains to the understanding of conventionalized multi-word configurations (Bstgen, 2020; Biber, 2009; Bybee and Beckner, 2015; Demir, 2017). Different studies have categorized formulaic sequences, referred to as "lexical bundles," in academic communication and writing using a frequency-based strategy (Biber and Barbien, 2007; Yoon, 2016; Staple et al., 2013). There is a prevalent argument that every academic genre exhibits a unique collection of lexical bundles that are linked to its characteristic communicative objectives (Biber & Barbieri, 2007, p. 265). The depiction of English academic writing about grammatical compression, syntactic elaboration, and degree of explicitness has been facilitated by the structural and functional characterizations of lexical bundles (Biber, 2009; Al-Hassan and Wood, 2015; Cunningham, 2017). The research tradition has identified the existence of formulaic phrases in different kinds of academic writing.

The investigation of formulaic has been conducted to explore the utilization of lexical bundles by different language user groups, including academic writers, and to suggest instructional strategies for the instruction for the co-occurrence of phrases in creating academic discourse.

The concept of formulaic has been examined through the lenses of the psychology of language and linguistic development as well. There is a claim that possessing acquaintance with academic formulas can enhance one's ability to process language fluently. Additionally, it is suggested that proficiency in utilizing

these formulaic expressions is associated with successful language production (Chen and Baker, 2010). Studies have provided evidence indicating that individuals acquire lexical phrases in their first language as complete units rather than as a series of separate words. Additionally, research suggests that formulaic language is acquired gradually and that proficient language users possess a diverse range of memorized language sequences (Ellis, 2008; Li & Schmitt, 2009). According to Li and Schmitt (2009, p. 86), the nonexistence of "formulaic sequences in language production indicates the deficiency of expertise of a beginner writer in a particular disciplinary group".

The ability to use formulaic phrases in academic writing is never a universally acquired skill (Lei and Liu, 2018; Wray, 2008). Academic writers in both L1 and L2 contexts have the potential to acquire formulaic sequences not exclusively via curricular activities, but also by conducting non-formal supplemental learning. This may include "extensive academic reading and repeated usage of patterns through extensive writing", as noted by Ellis (2008. p.39). According to Granger (2018, p. 38), foreign language learners find it challenging to achieve native-like proficiency in idiomatic expressions. Studies conducted on "small-scale monolingual and multilingual corpora" have indicated that the use of formulaic language is linked to advanced academic writing skills rather than beginner-level proficiency (Cortes 2008, p.40).

LITERATURE REVIEW

Theoretical Framework

This study is grounded in usage-based theories, which posit that formulaic sequences are structures that symbolize the basic components of linguistic representations, characterized by form-meaning or form-function pairs (Slabakova, 2013; Tomasello, 2003; Wulff & Ellis, 2018). Formulaic sequences are linked to specific functional and communication processes that may indicate the standards of a certain linguistic group, as determined by a corpus-driven approach (Elli and Odgen, 2017; Granger and Bestgen, 2018). In this section outlining the theoretical framework, it is essential to explicate the theoretical standing of formulaic sequences within the usage-based assumptions along with corpus linguistics literature.

Theoretical frameworks that rely on language usage as a primary factor typically rely on a limited range of cognitive mechanisms, including but not limited to categorization, analogy, and the process of grouping information into 'chunks' to account for the structure and functionality of the language. This essay focuses on the dimension of formation, learning, and usage of fixed expressions by language users. The significance of the occurrence rate of items within a corpus is evident in numerous usage-based procedures. Entities that exhibit a persistent co-occurrence in language usage and are continually employed for a specific purpose are subject to constraints towards automatization, comparable to that observed in various non-linguistic sensorimotor abilities.

The survival of irregular structures in language can be attributed to their frequency, which enables them to be acquired and utilized independently. Conversely, less frequent items and constructions are more prone to regularization. According to Langacker (1987), the degree of entrenchment of a linguistic unit is directly proportional to its level of cognitive routine or rehearsal in the speaker's mind. The concept of entrenchment pertains to the extent to which a system reinforces its reaction to the stimuli it receives. Conversely, it can be argued that prolonged periods of inactivity can result in a deterioration of cognitive representations. Once a pattern of entrenchment has been established, it can prove challenging to reverse.

According to the usage-based theory, the frequent use of language results in a cognitive process that is characteristic, such as "chunking". According to Chen and Baker (2010, p.72), a chunk refers to "a unit of memory organization" that is created by assembling a group of pre-existing chunks in memory and fusing them into a more extensive unit. According to Ellis (2008), the fusion of frequently co-occurring expressions into larger integrated components is a crucial aspect of cognitive processing as well as developing a language. This phenomenon allows language users to stimulate an increased quantity of information, thereby facilitating the production of fluent language. The co-occurrence of units frequently is attributed to the cognitive process of chunking, which is a consequence of repetitive usage, as stated by Bybee and Beckner (2015). According to Ellis (2003, p.73), "the entrenchment of the form-meaning link" is

impacted by every instance of regularly co-occurring word patterns, resulting in a decrease in the conceptual separateness of the separate segments. According to Simpson-Vlach and Ellis (2010), the form of language is ultimately brought about by the impact of frequency and repetition.

Formulaic Sequences in Academic Writing

Scholars from various fields, including corpus linguistics, psycholinguistics, and education, have demonstrated a sustained interest in the mechanisms by which words are combined (Chen and Baker, 2016; Cunningham, 2017; Garner, 2020; Bestgen, 2020). Biber's (2009) and Bybee and Beckner's (2004) scholarly investigations have examined phraseology as either fixed or semi-fixed units. According to Biber's (2009) proposition, language is wholly reliant on context, and there exists a proclivity for combination. Paquot (2013) identified the concept of "nativelike selection" through the examination of conversational data within a qualitative research framework. This concept refers to the capacity of language users to effectively communicate meanings through the use of phrases that are in addition semantically accurate, yet straightforward and idiomatic.

The study of phraseology has garnered significant attention; however, there exists a lack of agreement regarding the "terminology, descriptive approach, and analytical procedures", as noted by Granger and Paquot (2008, p.71). Various scholars have employed distinct terms to refer to phraseological units. For instance, Nattinger and DeCarrico (1992) used the term "lexical phrases," while Schmitt and Carter (2004) and Wray (2002) referred to them as "formulaic sequences." Additionally, Altenberg (1998) employed the term "recurrent word combinations," and Biber et al. (1999) used "lexical bundles." Furthermore, various descriptive methodologies were employed by researchers. According to Biber et al. (2004), there are variations in studies about phraseological components or combinations of words, focusing on the following dimensions:

i.   The research objectives encompass a comprehensive spectrum of word combinations as opposed to a limited set.
ii. Two criteria commonly used for identifying formulaic sequences are perceptual salience and frequency.
iii. The formal characteristics of formulaic sequences include continuous sequences, discontinuous frames or lexico-grammatical patterns, and the differentiation between two-word collocations and longer sequences.
iv. The quantity of text samples utilized, whether small or large corpora, is a crucial factor to consider.
v.   The inquiry pertains to the existence or non-existence of register comparisons in written and spoken texts, or both.

Approaches to the Study of Formulaic Sequences

According to Granger and Paquot's (2008) summary, previous studies on formulaic have adopted two primary approaches: a conventional approach and a frequency-oriented approach. The conventional methodology for studying formulaic involves placing a series of lexically recognized collocations along a spectrum of rigidity. On one side of the spectrum, there exist pure idioms, while on the opposite end, entirely unrestricted combinations are present. According to Wray (2008, p. 126), solely those idioms that are pure are deemed as the "prototype of the phraseological unit." The conventional theory does not acknowledge the status of combinations as phraseological units in cases where they are solely subject to syntactic and semantic relationships, or when they are entirely "compositional word combinations whose meanings can be anticipated from their constituent parts" (Hyland, 2008, p.33). The traditional approach to language study solely emphasized the examination of firmly established linguistic expressions, such as idiomatic phrases.

The dimension that centers on frequency is concerned with utilizing tools from corpus linguistics to elucidate the occurrence of formulaic language within specific discourses. The utilization of corpus linguistics has facilitated scholars and experts in investigating the formulaic inclination of language usage in an increasingly all-encompassing manner than the conventional method. The conventional method of formulaic was questioned by corpus-based research, based on empirical evidence. The researchers verified

the correlation between lexicon and grammar, as well as the predominance of formulaic sequences and combinations in linguistic expression. The aforementioned results led to the development of the frequency-based methodology for studying formulaic.

A frequency-based stand is a methodological approach that utilizes corpus data to explore formulaic from a bottom-up perspective. The investigation of formulaic through predictive conduct has its foundations in Sinclair's (2004) concept of the "idiom principle". According to Sinclair (2004, p.110), language users possess a vast array "of semi-preconstructed phrases that function as singular choices", despite their apparent segmental analyzability. It can be observed that individuals who utilize language possess the ability to utilize formulaic sequences, and these are interpreted and utilized as complete units. The current strategy towards formulaic encompasses a broader spectrum of lexical combinations and avoids the utilization of predetermined linguistic classifications, as opposed to the conventional method of identifying formulaic sequences. This strategy prioritizes the frequency of utilization and seeks to examine the correlation between a given term and its contextual surroundings. The utilization of frequency-based methodology in the study of formulaic sequences presents novel and diverse possibilities for inquiry into this linguistic phenomenon. This phenomenon results in the proliferation of phrases and approaches to analysis for diverse categories of formulaic sequences that are extracted automatically.

Lexical Bundles as Forms of Formulaic Sequences

The frequency-based approach to the study of formulaic sequences resulted in increased interest in lexical bundles, recently regarded as a form of formulaic sequence instead of a synonym of the term (Candarli, 2021; Garner et al., 2020; Granger and Bestgen, 2018; Yoon, 2016). Lexical bundles have been the subject of significant interest in corpus-based research. The identification process involved two distinct criteria, namely "the frequency of occurrences and the range of texts in which the bundles manifest" (Xia, 2022, p.76). Previous studies have reported a broad spectrum of frequency cut-off thresholds, varying "from 10 to 40 times per million words", as noted by Conrad and Biber (2005, p.29) and Hyland (2008, p.60). The implementation of a frequency threshold was employed to exclusively recognize the most commonly recurring sequences of words as 'lexical bundles'. The range of lexical bundle mining is a crucial factor to consider, as it pertains to the frequency of occurrence of lexical bundles across various texts. The imposition of a range requirement could serve to ensure that the lexical bundles that are identified are not limited to a small number of texts or authors, thereby mitigating the potential impact of idiosyncrasies that are specific to individual texts or authors.

According to previous studies conducted by Biber et al. (2004), Biber and Barbieri (2007), and Hyland (2008), there exists a correlation between the communicative functions conveyed in various registers and the conceptual and operational features of lexical bundles. Biber et al. (2004, p.71) conducted a thorough investigation of lexical bundles in both spoken and written registers. Their findings indicated that the functional as well as structural attributes of lexical bundles in spoken registers, such as "classroom instruction and conversations, differed from those in written registers, such as textbooks".

Special functional and structural attributes in the academic literature are demonstrated by authors who have distinct writing exposure through their use of lexical bundles (Demir, 2017; Lei and Liu, 2018). According to Chen and Baker's (2010, p.44) study, it was observed that learners utilized a greater number of "discourse organizers and bundles" that comprised verb phrase fragments as compared to professional writers. The authors regarded this as an indication of "immature writing". Römer (2009) emphasized the necessity of distinguishing between various levels of proficiency in the context of researching learner corpora. The researchers discovered that intermediate-level language learners primarily utilized verb phrase fragments in their lexical bundles, which aligns with patterns observed in spoken language. Advanced learners, on the other hand, exhibited a preference for utilizing lexical bundles that contained noun phrases, which exhibited more features commonly associated with academic writing (Xia, 2022, p.66).

Research Questions

By the premises of formulaic sequences fully established in the reviewed literature, the following research questions are pursued:

1. What formulaic sequences are predominantly used in the abstracts, introductions, and conclusions of academic writings?
2. What are the contextual implications of the usage of formulaic sequences in abstracts, introductions, and conclusions of academic studies?
3. What is the communication value of the frequency variations in the usage of formulaic sequences in abstracts, introductions, and conclusions of academic research?

The Rationale for Area Selection

The choice of academic writing in the investigation of formulaic sequences in this study is primarily motivated by the growing divergence in methodological values of certain components of academic research. For instance, the density of information required in the abstract, introduction, and conclusions, with limited word counts have set them in certain patterns, wherein word combinations in certain orders are almost inevitable. Also, the choice of lexical bundles, comparable to other forms of formulaic sequences such as the N-grams, is mainly premised on the focus on the use of the consistent lexical unit to pack enormous ideas in limited expressions.

STUDY METHODOLOGY

Study Design

The present investigation employs a combination of qualitative and quantitative methodologies, commonly known as the mixed methods approach. The utilization of a mixed methodology in this research is primarily intended to present a comprehensive analysis of both quantitative and qualitative data. The dataset comprises the frequency distribution of lexical bundles' appearances within the selected abstracts, introductions, and conclusions. The present study is centered on the analysis of qualitative data, with a particular focus on the examination of discourse stylistic implications. This is achieved through the application of the deductive method, which involves the construction of various themes to facilitate the analysis of the relevance of lexical bundles in the abstracts, introductions, and conclusions of the selected studies.

Study Community and Sample

A total of three hundred and fifty abstracts, introductions, and conclusions of different studies were generated from different journals using the Google Scholar feature. This sample includes 150 abstracts, 150 introductions, and 150 conclusions, generated from 350 different studies. The abstracts were generated from 150 different studies published in various journals. This is also maintained in generating the introduction and the conclusions. None of the abstracts, introductions, and conclusions used in this study were from the same study. A total of 301,000 words were generated in the combination, as shown in the table below:

Category	Frequency	Percentage
Abstract	42,900 words from 150 abstracts	14.25%
Introduction	151,800 from 150 introductions	50.44%
Conclusions	106,300 from 150 conclusions	35.31%

Inclusion Criteria

In the selection process, there are certain criteria for inclusion. In the abstract, the content of the abstract must be above 280 words in a single paragraph. The abstract must contain briefs of the study introduction, objectives, methodology, results, conclusions, and recommendations. In the introduction, the content must be above 1000 words and must include a background of the study and a description of the research rationale, dimension, and value proposition. For the conclusion, the content must include an emphasis on what the study has done, the conclusions reached, and possible recommendations or implications of the analysis. The topics searched are mainly language-related, including areas in translation, language development and

learning, language in the classroom, communication studies, and discourse studies. The papers were selected purposefully, using the stated criteria.

Tool for Data Extraction

According to Anthony (2020), the Antconc is a tool that uses corpus linguistics to calculate the frequency and co-occurrence of terms. In this particular research endeavor, the instrument is used. The selection of Antconc as the tool to conduct the inquiry is part of an attempt to narrow its focus. In the field of corpus linguistics, in addition to Antconc, there are various tools available for doing studies of frequency, collocation, concordance, n-grams, and keywords. The validity of the data produced by other research that is comparable to this one was taken into consideration while selecting Antconc as the right tool (Xia, 2022; Yoom, 2016; Bestgen, 2020)

Method of Analysis

On the data, both textual and statistical analyses are carried out. The statistical analysis concentrates on the number of times lexical bundles are used, whilst the textual analysis looks at the different types of lexical bundles that are used in the selected text. Additionally, the data system is generated with the help of Antconc. Analyses of lexical bundle frequency, collocation, concordance, n-grams, and keywords will be performed on the texts using tools such as frequency, percentage, and means.

RESULT AND ANALYSIS

Formulaic in the Abstracts, Introduction, and Conclusions

Appendices A, B, and C indicate that a total of 214 different formulaic sequences were discovered from the 301,000 words culled from 450 abstracts, introductions, and conclusions of different studies. The chart below provides a summary of the distribution across the main categories.

FREQUENCY OF FORMULAIC SEQUENCES
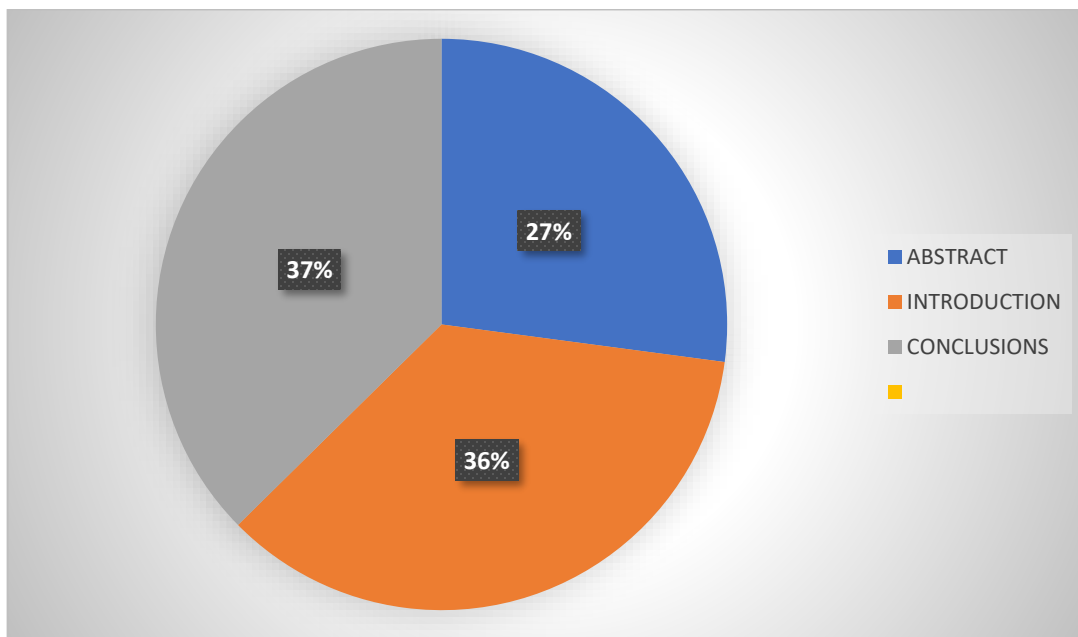
The chart above provides frequency and percentile distribution of the occurrences of formulaic sequences found in the abstracts, introductions, and conclusions of different studies selected for the study. A total of 58 (27%) different formulaic sequences are seen in the 42,000 words culled from 150 abstracts,

as seen in Appendix A. Also, 80 (37%) different formulaic sequences were found in 106,300 words culled from 150 conclusions of different studies in language and literature, as recorded in Appendix B. It can also be seen from Appendix C that a total of 76 (36%) different formulaic sequences were discovered from the 151,800 words culled from 150 introductions of different studies.

Most of the formulaic sequences occurred repeatedly in the corpus. For instance, formulaic sequences occurred 3395 times in the 42,900 abstracts culled from the study. The table below provides insights into the five most frequently occurring formulaic in the abstracts.

TABLE 1
FIVE MOST OCCURRING FORMULAIC IN THE ABSTRACTS

S/N	Formulaic	Frequency	Percentage
1.	*In this research*	106	23.71%
2.	*The result indicated that*	86	19.23
3.	*The findings pointed at*	38	8.50
4.	*The purpose of this*	114	25.50
5.	*The analysis revealed that*	103	23.05
	Total	447	100%

It could be seen that the above five lexical bundles form over 13% of the total 3395 co-occurring instances of formulaic in the selected 150 abstracts of 42,900 words. In the abstracts, the formulaic '*The purpose of this*' dominated the frequency value, occurring 114 times from the 150 articles. This indicates that 'the purpose of' constitutes about 76% of the 150 articles, appearing in 114 abstracts out of the 150.

TABLE 2
FIVE MOST OCCURRING FORMULAIC IN INTRODUCTIONS

S/N	Formulaic	Frequency	Percentage
1.	**The objective/aim of the**	**303**	**34.35%**
2.	**It is essential to**	**42**	**4.77**
3.	**In this research**	**228**	**25.85**
4.	**The purpose of this**	**271**	**30.72**
5.	**Different studies have**	**38**	**4.30**
	Total	**882**	**100%**

To understand the frequency values of the above five formulaic, it is important to emphasize that the discovered 76 formulaic in the 150 introductions of 151,800 words, occurred about 7,860 times in the corpus using Antconc. This implies that some of the formulaic is repeatedly used in the introduction, for instance, the formulaic 'the objective or the aim of this study' occurred a total of 303 times in the 150 articles. It implies that it is used at least twice in the articles. This is followed by 'the purpose of this', and 'in this research', which occurred 271 and 228 times respectively. The implication is that the above five formulaic formed at least 11.22% of the total reoccurring values of the formulaic in the introductions.

TABLE 3
FIVE MOST OCCURRING FORMULAIC IN THE CONCLUSIONS

S/N	Formulaic	Frequency	Percentage
1.	*It is concluded that*	206	13.32%
2.	*The result indicated that*	449	29.04
3.	*Based on the evidence*	291	18.83
4.	*In this research*	339	21.92
5.	*It has been discovered*	261	16.88
	Total	1546	100%

It is pertinent to reiterate that in the 150 conclusions drawn from various studies in language and literature, the discovered 80 formulaic occurred in a total of 9752, making it the part with the most frequently used formulaic in the corpus for this study. However, table 3 above indicates that the formulaic 'the result indicated that' occurred a total of 449 times in the conclusions, which is about 29.04% of the 1545 times the above five formulaic occurred in the conclusions examined. Surprisingly, this value is followed by the formulaic 'in this research', which is about 339 (21.92%). The data indicates that the above five most frequently occurred formulaic in the conclusions constitute above 15.86% of the total times formulaic occurred in the 150 conclusions drawn from various studies.

Structure of Formulaic Sequences in the Abstracts, Introductions, and Conclusions

From the 214 formulaic sequences discovered in the corpus, five major sequence structures are evident, including two-word structure, three-word structure, four-word structure, five-word structure, and six-word structure. The chart below provides a summary of the percentile values of each of the structures across the corpus.
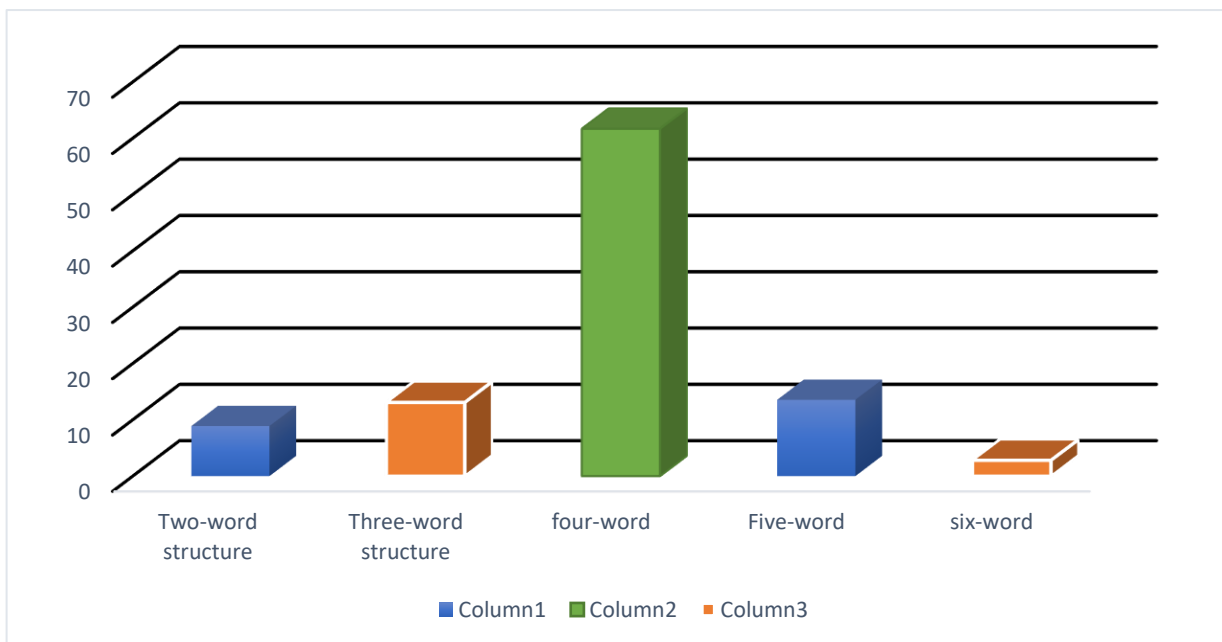
The above chart indicates that four-word structure formulaic sequences dominated the percentile value, standing at 61.68% of the 214 formulaic found in the corpus. This value further indicates that the four-word formulaic sequence is about 132 out o 214 lexical bundles.

Discussions

The presented data unveiled several findings, as listed below:

1. The most frequently used formulaic sequence is 'in this study', occurring more than 100 times in the abstract, more than 200 times in the introduction, and almost 300 times in the conclusion.
2. Four-word structure formulaic sequence dominated the corpus, occurring in about 132 out of the 214 formulaic sequences found in the corpus.
3. The sequences of the formulaic are repetitive, in a way that some formulaic are repeated severally in a particular work.
4. The data presented indicated that the 214 formulaic occurred a total of 21,004 times across the abstracts, introductions, and conclusions.

According to Biber and Barbieri (2007, p.247), lexical bundles or formulaic are significant tools that aid in the understanding and formation of discourse. As a result, the examination of various academic fields can uncover distinct word collocations that are specific to each discipline (Wray, 2002). The current investigation focused on the implementation of bundles in two fields through the utilization of a corpus-based methodology, as previously stated. The analysis presented in this study pertains to the utilization of formulaic sequences in the abstracts, introductions, and conclusions of research articles within the fields of linguistics and language literature. Antconc software was utilized to uncover a substantial number of four-word lexical bundles that do not exhibit further growth. Others expand their lexical bundles from three-word to four or five words. Upon conducting a more in-depth written inspection of the consistent lines that contained the aforementioned lexical sequence, it was observed that said sequence had not been amalgamated with any other words to create five-word sequences, with the potential to expand to six in select instances. Consequently, a multitude of lexical bundles comprising three words is extended to larger bundles consisting of four words. This assemblage of verbs, encompassing a passive construction, mainly denotes situations in which research subjects were required to undertake actions.

According to Bestgen (2020), the functions and grammatical groupings of the formulaic sequences are mainly based on functionality, but noun phrases usually dominate the list, followed by verb phrases. by the findings of the study conducted by Xia (2022), it is seen in this research that sequences of formulaic are formed from one order, growing into another. For instance, there are examples in Appendix A that are developed from smaller ones. The formulaic 'this study concludes that', may have been developed from three formulaic addition 'this study concludes', or two equal sequences 'this study', and 'concludes that'. This finding is premised on the observation that there are cardinal structures on which other formulaic sequences grow. For instance, '*this study*' is a formulaic structure other sequences may be based on; like 'this study investigates', *'this study concludes that', 'this study aims at', and 'this study revealed that'*, among others. As such, it could be seen that two formulaic sequences seem to be the basic order in some instances, from which other larger structures grow or expand.

CONCLUSIONS

This study has provided insights into formulaic sequences used in academic writing, focusing on abstracts, introductions, and conclusions of different articles in linguistics and language education. The findings of the study indicate that four-word formulaic sequences are prevalent in the corpus because they expand from two to three formulaic sequences. There are very few six-word formulaic sequences because the five-word formulaic rarely grows to six, and six is not found to expand to seven in any case. It can also be seen that academic writers use formulaic in different manners that can be explained within the purview of usage-based theory. It is thus concluded that writers repeat formulaic sequences more in the conclusions of studies than what is seen in the abstracts and introductions. However, the density of information contained in abstracts makes its way that almost five to eight formulaic are found in an abstract of about 280-290 words. The results of this investigation have the potential to assist educators who specialize in teaching "English for Academic Purposes (EAP)". This research endeavor has the potential to enhance student's awareness of lexical chunks, thereby facilitating more fluent language production. In addition, educators can acquaint students with the diverse operations of lexical bundles and illustrate the various

manners in which these linguistic units have been employed by distinct authors. EAP professionals have the potential to enhance the scholarly writing skills of students by modeling the linguistic structures employed by proficient writers. The primary constraint of this research pertains to the corpus size, which could have potentially impacted the study's outcomes. Conducting further studies employing larger corpora would yield more insightful outcomes.

ACKNOWLEDGMENTS

REFERENCES

AlHassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes*, *17*, 51–62.

Anthony, L. (2020). *AntConc (Version 3.5.9)* [Computer software]. Waseda University.

Bestgen, Y. (2013). Inadequacy of the chi-squared test to examine vocabulary differences between corpora. *Literary and Linguistic Computing*, *29*(2), 164–170.

Bestgen, Y. (2017). Beyond single-word measures: L2 writing assessment, lexical richness, and formulaic competence. *A System*, *69*, 65–78.

Bestgen, Y. (2020). Comparing lexical bundles across corpora of different sizes: The Zipfian problem. *Journal of Quantitative Linguistics*, *27*(3), 272–290.

Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing*, *26*, 28–41.

Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics*, *14*(3), 275–311.

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

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

Bybee, J.L., & Beckner, C. (2015). Usage-based theory. In B. Heine, & H. Narrog (Eds.), *The Oxford Handbook of linguistic analysis*. Oxford University Press.

Candarli, D. (2021). A longitudinal study of multi-word constructions in L2 academic writing: The effects of frequency and dispersion. *Reading and Writing*, *34*(5), 1191–1223.

Casal, J.E., & Kessler, M. (2020). Form and rhetorical function of phrase-frames in promotional writing: A corpus- and genre-based analysis. *System*.

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

Chen, Y.H., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2, and C1. *Applied Linguistics*, *37*(6), 849–880.

Cunningham, K.J. (2017). A phraseological exploration of recent mathematics research articles through key phrase frames. *Journal of English for Academic Purposes*, *25*, 71.

Demir, C. (2017). Lexical collocations in English: A comparative study of native and non-native scholars of English. *Journal of Language and Linguistic Studies*, *13*(1), 75–87.

Ellis, N., 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.

Ellis, N.C. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned-attention, and the limited L2 endstate. In P. Robinson, & N.C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 372–405). Routledge.

Ellis, N.C., & Ogden, D.C. (2017). Thinking About Multiword Constructions: Usage-Based Approaches to Acquisition and Processing. *Topics in Cognitive Science*, *9*(3), 604–620.

Garner, J. (2020). The cross-sectional development of verb–noun collocations as constructions in L2 writing. *International Review of Applied Linguistics in Language Teaching*, *1957*, 1–27.

Garner, J., Crossley, S., & Kyle, K. (2019). N-gram measures and L2 writing proficiency. *A System*, *80*(2), 176–187.

Garner, J., Crossley, S., & Kyle, K. (2020). Beginning and intermediate L2 writer's use of N-grams: An association measures study. *IRAL - International Review of Applied Linguistics in Language Teaching*, *58*(1), 51–74.

Garner, J.R. (2016). A phrase-frame approach to investigating phraseology in learner writing across proficiency levels. *International Journal of Learner Corpus Research*, *2*(1), 31–67.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. In A.P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (pp. 145–160). Clarendon Press.

Granger, S. (2018). Formulaic sequences in learner corpora: Collocations and lexical bundles. In A. Siyanova-Chanturia, & A. Pellicer-Sánchez (Eds.), *Understanding Formulaic Language: A Second Language Acquisition Perspective* (pp. 228–247). Routledge.

Granger, S., & Bestgen, Y. (2018). Tracking L2 writers' phraseological Development Using Collgrams: Evidence from a Longitudinal EFL Corpus. In H. Sebastian, S. Andrea, A.-L. Sabine, & D. Lisa Marie (Eds.), *Corpora and Lexis* (Issue 81, pp. 277–301). Brill.

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger, & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 27–49). John Benjamins.

Gries, S.T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, *13*(4), 403–437.

Gries, S.T. (2010). Useful statistics for corpus linguistics. In A. Sánchez, & M.A. Sánchez (Eds.), *A mosaic of corpus linguistics: Selected approaches*.

Gries, S.T., & Ellis, N.C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, *65*(S1), 228–255.

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

Lei, L., & Liu, D. (2018). The academic English collocation list. *International Journal of Corpus Linguistics*, *23*(2), 216–243.

Paquot, M. (2014, August). Cross-linguistic influence and formulaic language. *EUROSLA Yearbook*, *14*, 240–261.

Paquot, M. (2018). Phraseological competence: A missing component in university entrance language tests? Insights from a study of EFL Learners' use of statistical collocations. *Language Assessment Quarterly*, *15*(1), 29–43.

Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second Language Research*, *35*(1), 121–145.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–149.

Paquot, M., Naets, H., & Gries, S.T. (2019). Using syntactic co-occurrences to trace phraseological complexity development in learner writing: Verb + object structures in Longdale. *Second Language Acquisition and Learner Corpora*, pp. 1–19.

Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Construction*, *3*(1), 95–119.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, *12*(3), 214–225.

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University.

Yoon, H.J. (2016). Association strength of verb-noun combinations in experienced NS and less experienced NNS writing: Longitudinal and cross-sectional findings. *Journal of Second Language Writing*, *34*, 42–57.

APPENDIX 1: FORMULAIC SEQUENCES FROM ABSTRACT

1. The findings obtained from the
2. For the study of a different language
3. The results obtained from the
4. Within the realm of
5. There was a substantial amount of
6. Conclusions are drawn from the research
7. The findings pointed to the fact that
8. Taking part in the research as a subject
9. The findings of the research
10. The purpose of this study was to examine.
11. The purpose of this study was to try
12. The results of the investigation showed that
13. The purpose of the current investigation was to
14. The current investigation was a
15. The findings demonstrated that
16. It has been discovered that
17. Regarding the current investigation
18. Being counted among the
19. Make an effort to look into the subject.
20. To what extent this is the case
21. The current investigation looked at the
22. At the conclusion of
23. In light of the fact that
24. The evidence suggested that there was
25. People who took part in the research
26. Distinct and important distinctions between the
27. The conclusion of the
28. The results of this investigation
29. The findings pointed to the fact that
30. was performed based on the findings of
31. With regard to their position
32. Examine the effects of
33. Examine the connection between
34. Of the research that was done
35. In this research,
36. As opposed to that,
37. The results indicate that
38. to ascertain the effect
39. were chosen based on
40. the control group although
41. There is a strong positive link
42. a control group too.
43. Data were examined using
44. In the light of
45. In the control group
46. no change that is statistically significant
47. of the present research
48. shown that there was
49. the two have a strong positive association.
50. study aimed to ascertain
51. The results indicate that
52. The examination of the text
53. The information was studied
54. The research showed that
55. We can deduce with this investigation
56. to make a data analysis
57. To research the connection
58. if there is an identifiable variation

APPENDIX 2: FORMULAIC SEQUENCES FROM CONCLUSIONS

1. Regarding the current investigation
2. study's results suggest that...
3. The results obtained from the investigation.
4. study's findings have indicated that
5. there are significant correlations between
6. the variables under investigation.
7. The findings of this investigation.
8. study reveal the results
9. obtained in the current research.
10. Regarding the present investigation,
11. with respect to the ongoing research.
12. The current investigation
13. study's findings are indicative of.
14. During the course of the procedure.
15. conducted to investigate the current research topic.
16. that a conclusion can be drawn
17. based on the available evidence.
18. principles of this particular framework,
19. we will proceed with the proposed methodology.
20. It can be inferred that.
21. It has been discovered that.
22. In addition to the aforementioned.
23. It has been disclosed that.
24. the subject matter pertains to
25. a specific area of study.
26. According to the results obtained from the study,
27. they align with
28. The current findings were analyzed and evaluated.
29. Within the given context,
30. as it pertains to the subject.
31. It is important to note.
32. This study's findings can be analyzed.
33. The findings of this study indicated.
34. The outcomes of the present investigation.
35. It is important to acknowledge that.
36. Regarding the matter of
37. In a certain manner
38. The results indicated that.
39. The results indicate that.
40. that a conclusion may be drawn.
41. This study comprises.
42. conducted to investigate the topic further
43. gain a deeper understanding of the subject matter.
44. study's findings have indicated that
45. there is a significant correlation between the
46. the given statement implies that.
47. The results obtained from the investigation.
48. principles of this particular framework,
49. Based on the evidence,
50. one may draw the conclusion.
51. study reveal the results obtained from the current research.
52. The findings indicated that.
53. It is important to note.
54. The observation that the
55. Empirical evidence suggests that
56. it has been discovered that.
57. Regarding the utilization of.
58. This phenomenon is observable
59. On the contrary to the aforementioned.
60. In addition to the aforementioned.
61. To facilitate or enable the process of making
62. Owing to the circumstance that
63. It has been disclosed that.
64. According to the results obtained
65. from the research conducted
66. it can be inferred that...
67. Are consistent with.
68. Simultaneous with the
69. the significance of this matter
70. The degree to which
71. There exists a possibility that...
72. Due to the outcome of.
73. It is noteworthy that.
74. Regarding the matter of
75. In order to have the capability to.
76. During the duration of.
77. One the one hand,
78. it can be argued
79. The manners in which
80. conducted to investigate the topic at hand.

APPENDIX 3: FORMULAIC SEQUENCES IN INTRODUCTIONS

1. arrays of studies have
2. different researches have
3. in the present study
4. the focus is to
5. accounting for the
6. it will be pursued
7. in the language of the
8. in the area of
9. it is argued here that
10. the following objectives are
11. the rationale for the
12. The purpose of
13. in a wide range of
14. purpose of the current
15. A field of study of
16. research was done to assess
17. of this research were
18. most often employed
19. was to assess the
20. to ascertain whether the
21. research was done to look at
22. In our observation
23. The purpose of this
24. a crucial function in
25. to evaluate the research
26. connected to an increased
27. have been connected to
28. been linked to an increase
29. been discovered to
30. investigated the impact of
31. is said to be
32. for the purpose of
33. was to evaluate the
34. been employed to
35. research was done to ascertain
36. Recent research
37. is the most typical
38. The objective of the
39. examined the impact of
40. Objectives of the Investigation
41. is believed to be
42. There are little facts concerning
43. It is essential to
44. the purpose of this investigation
45. was to look into the
46. under the administration of
47. among the most crucial
48. studies have emphasized
49. have accustomed to
50. Research has shown that
51. the extent that
52. information about the
53. The connection between the
54. There was a theory that
55. Various research have
56. These investigations were to
57. due to the growth of
58. several studies
59. crucial function in
60. is connected to increased
61. a major factor in
62. a key contributor to
63. research aimed to contrast
64. exhibited to have
65. a deeper comprehension
66. been shown to
67. understood to be
68. and the employing
69. been discovered to
70. Not much is known.
71. with the help of
72. In this essay
73. among the most typical
74. thought to be
75. research aimed to look at
76. Its distinguished by a