

Revealing lexical bundles of non-native English essay at a private Islamic university: A corpus-based evaluation

Fauzul Etfita^{1*}, Sri Wahyuni², Arimuliani Ahmad³, Wasana Saranga Sudusinghe⁴, Chandima Kumara Walpita Gamage⁵

¹English Language Education, Faculty of Teachers Training and Education, Universitas Islam Indonesia, Pekanbaru, Indonesia; email: fauzuletfita@edu.uir.ac.id

²English Language Education, Faculty of Teachers Training and Education, Universitas Islam Indonesia, Pekanbaru, Indonesia; email: wahyuni@edu.uir.ac.id

³English Language Education, Faculty of Teachers Training and Education, Universitas Islam Indonesia, Pekanbaru, Indonesia; email: arimulianiahmad@edu.uir.ac.id

⁴Department of Language, Faculty of Management, Social Sciences & Humanities, General Sir John Kotewala Defence University, Sri Lanka; email: sudusinghews@kdu.ac.lk

⁵Department of Nursing and Midwifery, Faculty of Allied Health Sciences, General Sir John Kotewala Defence University, Sri Lanka; email: ckumara@kdu.ac.lk

ABSTRACT

***Corresponding Author:**
fauzuletfita@edu.uir.ac.id

Article history:

Received: 19 May 2025

Revised: 8 June 2025

Accepted: 11 August 2025

Published: 27 August 2025

Citation: Etfita, F., Ahmad, A., Sudusinghe, W.S., & Gamage, C.K.W. 2025. Revealing lexical bundles of non-native English essay at a private Islamic university: A corpus-based evaluation. *English Learning Innovation*, 6(2), 450-468. <https://doi.org/10.22219/englie.v6i2.40821>

© Copyright 2025: Etfita et al.
This is an open access article
under the CC-BY-SA license



P-ISSN 2723-7400

E-ISSN 2723-7419

Lexical bundles are a crucial element for Foreign Language Learners (FLL) as they facilitate the acquisition of new vocabulary and enhance fluency. Such multi-word statements confer a natural structure to language and enhance successful communication in both spoken and written forms. Although numerous studies address lexical bundles, research on their application in English writing or inside Indonesian EFL classes is notably scarce. The study employed a corpus methodology, utilizing AntConc software version 3.5.9 to categorize lexical bundles from student essays. The output comprised 17,100 varieties of N-gram lexical bundles, specifically three, four, five, and six-word bundles. The study revealed regular use of formulaic and repetitious bundles, including "at least one foreign language," "a foreign language," and "violent video games." Students' writing was dominated by these sentences, which suggested a reliance on memorized forms and a small range of scholarly discourse markers. Although three- and four-word bundles were most popular, longer five- and six-word bundles, though less common, showered better meanings and more formulaicity. These results imply that although students have come across lexical bundles, their usage is limited to known themes and lacks a more general academic background. To improve EFL students' academic writing abilities, the study underlines the requirement of explicit instruction on lexical bundles, especially those fulfilling argumentative and coherent purposes. The findings provide pedagogical ideas for including corpus-based tactics into writing education to support fluency, coherence, and lexical variation.

Keywords: antconc, corpus-based, essay writing, lexical bundles

INTRODUCTION

Proficiency in academic writing presents a significant challenge for English Foreign Language (EFL) learners, particularly those in higher education settings. It calls for not only mastery of difficult language structures but also the capacity to keep coherence, coherence, fluency (Alzahrani, 2025; D. Zhang et al., 2023; Prapobratnakul, 2024; Wang & Xie, 2022), and a formal tone consistent with academic conventions (Ahmad et al., 2024). For EFL students, who must concurrently negotiate foreign language, grammatical complexity, rhetorical requirements, and genre-specific elements (Cherry Chan, 2025; Nguyen & Vu, 2024), this multifarious character of academic writing sometimes proves intimidating. Among the several techniques available for academic writing, lexical bundles (LBs), defined as recurrent sequences of words that often show in a certain register, play a vital part in improving discourse fluency and coherence (Shirazizadeh & Amirfazlian, 2021). These bundles serve as formulaic language pieces that help authors to convey abstract concepts, build arguments, and arrange their works in expected patterns.

Most of the time, native speakers (NS) unintentionally pick up LBs from a lot of scholarly publications, spoken English, and casual conversation (Crossley, 2020; W. Wang et al., 2024). They generally utilise a wider array and higher frequency of LBs to establish logical connections, manage the progression of information, and express degrees of certainty or attitude, compared to non-native speakers (NNS) (Ahmad et al., 2024; Demiray Akbulut, 2020; Sanosi, 2022). Research demonstrates that NSs employ a higher frequency of discourse-organizing bundles, while NNS often rely on attitude expressions (Kashiha & Chan, 2015; Yunjung, 2025). Conversely, those who are NNS often find it difficult to identify, understand, and use these bundles with efficiency. This is so because they receive little guidance on standards unique to each given genre and little exposure to real input. The existence of this difference emphasizes the need for particular training and study into LBs as a strategy of instruction to enhance the writing capacities of students learning English as a foreign language (Birhan, 2021; Kaya, 2025; Y. Li & Lei, 2025).

LBs are word sequences frequently found in written academic discourse (Öztürk & Taşçı, 2023; Taşkaya & Özbay, 2023). They happen frequently and have several purposes, including hedging statements, indicating the writer's stance, connecting concepts, and marking the structure of speech (Cheng & Hsu, 2025; Gak et al., 2023). These pre-assembled pieces, which fulfill pragmatic and grammatical purposes across academic disciplines, are not a haphazard mix of words. Their formulaic character enables readers and writers to control the cognitive burden connected with text processing and comprehension, since they provide clear patterns that enable better reading and more fluid writing (Dan et al., 2024; Hosseinpur et al., 2023).

Further, LBs serve as rhetorical building blocks in academic writing that increase text cohesion and readability (Liu & Chen, 2020). For instance, bundles such as on the other hand should be emphasized that are commonly used to convey emphasis, contrast, or cause-and-effect relationships. These words enable authors to organize their arguments and lead readers through the logical sequence of ideas. By letting students match their production to the demands of academic discourse communities, LBs also help to contribute to the formality of academic writing (Gak et al., 2024).

LBs also help to achieve a formal tone and improve the readability and flow of the

text. They are defined as recurrent sequences of words, appear together and are crucial for writing (Dan et al., 2024; Kurniawan & Haerunisa, 2023; O'Flynn, 2022). LBs, defined as repeated sequences of three or more words that often co-occur in particular contexts, are essential tools for attaining fluency and professional academic writing (Dan et al., 2024; Kurniawan & Haerunisa, 2023; O'Flynn, 2022). The academic genre, discipline, and degree of language ability of the writer will affect the structure and purpose of these bundles. For instance, because of different rhetorical objectives and disciplinary standards, bundles used in the humanities could differ greatly from those used in scientific writing. Mastery of LBs helps students generate language that is not only grammatically accurate but also contextually suitable and rhetorically successful.

In this case, LBs represent the fundamental academic writing discourse structures—that of stance-taking, topic development, and argument organisation. Research has shown that LBs support text cohesiveness and signify rhetorical gestures prized in scholarly settings. LBs help to increase general writing performance (Hosseinpur et al., 2023; Nasrabady et al., 2020). by serving as processing units, therefore enhancing language fluency and idea creation. They also enable students to clearly, respectfully, and convincingly express ideas by helping them to fit academic norms.

Apart from enhancing the quality of the work, LBs help authors to follow disciplinary rules and communicate their views in a professional and consistent manner. Academic writers, especially those who want to publish or engage in worldwide academic dialogues, have to be skilled in using such bundles to generate work that satisfies academic criteria (Appel, 2022; Zhang et al., 2021). In other words, academic writers who wish to communicate clearly and meet the criteria of scholarly discourse must first master the use of LBs (Kim & Kessler, 2022; Nguyen, 2023).

Even so, past studies have mostly concentrated on their usage in expert academic writing generated by NS or advanced learners (Ardi et al., 2023; Kostromitina, 2022; Mashhadi & Qanavati, 2024; Shin, 2018). The research papers, dissertations, conference presentations, and other genres have seen LB frequency and function explored by these studies, therefore providing insights into how these bundles form professional discourse. But this focus has left a clear void in the literature: academic writing of undergraduate EFL students, especially in non-Western settings like Indonesia, still gets little attention.

The limited use of LBs in student writing could produce writings lacking academic sophistication, coherence, and clarity. Students' academic performance may suffer as a result, as well as their capacity for scholarly conversation. Dealing with these obstacles calls for a multifarious strategy combining explicit teaching of LBs, corpus-based learning, and scaffolded writing support. Good writing instruction should concentrate on discourse-level elements such as LBs, cohesion devices, and rhetorical organisation in addition to grammar and vocabulary (Li & Pei, 2024; Wei et al., 2024).

Accordingly, teachers can create focused activities, authentic writing models, and genre-specific practice assignments reflecting actual academic writing standards by means of corpus data. Including corpus-based approaches in writing classes also lets teaching materials be tailored to fit particular student requirements. For instance, teachers at the institutions can create corpora from academic literature or student essays pertinent to their disciplines, therefore guaranteeing that the LBs presented in class are pedagogically successful and contextually appropriate. By bridging the gap

between local educational environments and worldwide academic standards, this localised method makes writing instruction more relevant and powerful.

This oversight is significant because undergraduate learners, especially those in EFL contexts, often face unique challenges when acquiring academic writing proficiency. These include limited exposure to native-like academic input, inadequate training in genre-specific writing, and the influence of local academic cultures that may not emphasize the use of LBs. In Indonesian higher education institutions, these challenges are compounded by the need to meet international academic standards while maintaining local values and perspectives. Students in such settings are often required to write academic essays in English, yet many struggle to achieve fluency, coherence, and academic tone due to insufficient mastery of LBs.

Using corpus linguistics, the systematic study of vast collections of real-world texts to find patterns of language use, is a hopeful means of tackling this difficulty. Examining corpora helps researchers and teachers to find the LBs most often employed in academic writing and ascertain their function across many genres and disciplines (Agustina et al., 2024; Öztürk & Taşçı, 2023). Within the framework of language training, corpora are excellent tools for creating instructional resources reflecting real language use. They enable students to investigate actual instances of vocabulary, grammar, collocation, and conversation patterns, therefore fostering inductive, data-driven learning. Corpus analysis helps students to see how language is used, instead of how it is taught in textbooks. Students might look at frequency data to find which LBs are most frequently used in academic articles, understand idioms, and spot regular mistakes made by both NS and NNS (Ardi et al., 2023; Bal Gezegin, 2019; Bacha & Khachan, 2023). By use of this approach, students develop into more autonomous learners able to examine and implement language patterns in their work.

Examining the usage of LBs in student essays helps to close the current research gap, given the significance of LBs in academic writing and the particular difficulties experienced by Indonesian EFL learners, especially in private institutions. This study focuses especially on the frequencies, forms, and patterns of LBs used by undergraduate EFL students in essay writing at an Indonesian private Islamic university in Riau. It seeks to address the following study question: *What LBs are most often used in the academic essay writing of students?* Results from this study are expected to improve the efficacy of EFL writing education and assist students in developing competent, confident academic writers able to engage in worldwide scholarly communication.

METHODS

The current research used a corpus-based linguistic analysis using computer techniques to extract and examine LBs from student writings. Through corpus analysis, researchers can often find occurring word patterns and associated structural categories (McEnery & Brookes, 2024; Tracy-Ventura & Paquot, 2020). The corpus-based analysis is a linguistic method that examines a vast collection of texts (corpus) to investigate patterns, structures, and functions of language in real-world contexts. Researchers can use computer methods to assess the frequency and context of words, sentences, and grammatical structures across various texts. This approach enables the detection of language trends, variations, and changes throughout time, giving

empirical data to support linguistic theories. The corpus-based analysis is especially useful since it can offer insights into natural language usage.

The data for this study were drawn from a corpus of 35 argumentative essays written by third-year undergraduate students enrolled in the English Language Education at a private Islamic university in Riau. These students had recently completed a required course titled “Academic Essay Writing,” which is typically offered in the sixth semester of the program. This course is a capstone writing module designed to reinforce students’ academic writing competence by targeting intermediate-to-advanced English proficiency, corresponding to the B2 levels on the Common European Framework of Reference for Languages (CEFR).

The selection of these essays was purposive rather than random. The 35 samples were chosen to reflect students who had completed the essay writing course, ensuring that their texts were produced under comparable instructional conditions. The goal was to obtain authentic student writing that demonstrates the application of taught concepts, including the use of LBs in academic writing. The sample size of 35 essays was determined to provide a sufficient volume of language data for corpus analysis while maintaining manageability for manual annotation and software processing using AntConc.

During the course, students were systematically trained in academic discourse conventions, including how to develop coherent arguments, structure essays into well-organized paragraphs, and incorporate evidence to support their claims. LBs were introduced explicitly as essential components for achieving fluency and cohesion in writing. Instruction involved both deductive teaching (where common academic LBs were explained and practiced) and inductive learning (through students’ analysis of real texts). Students were exposed to sample essays from academic journals, model student writing, and a variety of scaffolding tasks aimed at enhancing their awareness and control of formulaic language.

In addition to in-class instruction, students engaged in peer review, received formative feedback from instructors, and revised their work over multiple drafts. Corpus-based resources and activities were also incorporated into classroom practice to help students notice frequent LB patterns and improve their use of academic phrases. The final essays submitted for grading, which are included in this study, represent the culmination of this instructional process and reflect students’ best efforts after multiple stages of revision.

Object of the study

The dataset for this study consists of 35 argumentative essays written by third-year undergraduates in the English Language Education at a private university in Riau. Each essay was composed as a final project for the “Academic Essay Writing” course, which required students to demonstrate their ability to develop coherent and well-structured arguments in English. The genre assigned for the task was argumentative writing, a core focus of the course, which emphasizes logical reasoning, critical thinking, and the use of academic register and conventions.

The students were instructed to write essays on digital technology in learning, and the impact of video games on children, using clear thesis statements, supporting arguments, counterarguments, and conclusions. Every essay was supposed to have a

five-paragraph format comprising an introduction, three body paragraphs, and a conclusion. In terms of length, the essays ranged from approximately 500 words per essay. All essays were written in English and submitted digitally. Students were instructed to produce original work without the use of machine translation tools. Before submission, they received general feedback on their outlines and drafts, but were responsible for revising and editing their final essays independently. This controlled and clearly defined context ensures that the essays analyzed in the study represent authentic learner language reflective of upper-intermediate to advanced EFL academic writing.

Figure 1 displays the corpus metadata from thirty-five essays used to create an English writing corpus.

Category	Description
full_name	my_corpus
short_name	my_corpus
file_count	35
token_count	18378
type_count	3095
encoding	utf_8_sig
token_definiti	[\p{L}]+
ignore_header	False
ignore_items	False
number_repla	False
format	raw_files
indexer_type	type
indexer	simple_word_indexer

Figure 1. Corpus Metadata

Data collection process

A corpus in language studies is defined as a collection of examples of language in use, which are selected and compiled in a principled way to be representative of a particular language or language variety. These collections can include written texts, transcriptions of spoken language, or multimodal data, and are stored in a machine-readable format for analysis (Huang & Yao, 2015; Lin & Adolphs, 2023). The size and scope of a corpus are particularly important, as a big and diverse corpus enables the examination of both frequent and uncommon linguistic events, offering robust and comprehensive statistical insights into language patterns and usage. These criteria ensure that a corpus is a solid base for empirical linguistic study. In short, to conduct this corpus research, the researchers gathered the students' essay writing as the first step, uploaded them to the AntCont corpus tool, and used N-gram to analyze the LBs.

Research instrument: Corpus tool

Corpus analysis is a computer-assisted study research method. AntConc version 3.5.9 was selected as the primary corpus analysis tool for this study due to its accessibility, reliability, and specialized functions tailored for LBs extraction. AntConc is a free, user-friendly concordance program widely used in corpus linguistics for analyzing large collections of texts. It offers powerful features such as N-gram generation, concordance lines, collocational analysis, and word frequency lists, which are essential for identifying recurring LB patterns in academic writing.

AntConc is appropriate for examining student-generated texts since it offers a

reasonable compromise between utility and simplicity of use without demanding sophisticated programming knowledge. Its N-gram function was very beneficial for this research since it allowed for the quick identification of 3- to 6-word LBs spread over all 35 student essays. Furthermore, AntConc increases result accuracy by supporting different text formats and letting researchers tailor stoplists and search criteria.

RESULTS

To guide the interpretation of the findings, it is important to restate the primary research question: *What lexical bundles (LBs) are most often used in the academic essay writing of the students?* In response to this question, the current section provides a detailed account of the most frequently occurring LBs identified in the students' essays. The analysis is presented in multiple forms, frequency tables, visual word clouds, and contextual sentence examples, to offer a comprehensive understanding of not only what LBs are used for but also how they function within the students' writing. The frequency table displays the lexical bundles in descending order, allowing for easy identification of the most dominant sequences. Meanwhile, the word cloud visually amplifies the most repeated items, emphasizing the lexical density and patterns favored by the students. Furthermore, example sentences illustrate how students employ these bundles in context, often overlapping several bundles in a single sentence to maintain flow and coherence.

A. Three-word lexical bundles

Three-word LBs are frequently occurring sequences of three words that typically appear together in texts, often surpassing what would be expected by chance. These bundles are crucial in language, as they improve the fluency and coherence of discourse. Table 1 displays the compilation of three-word LBs derived from the students' writings. The phrases most frequently used were "a foreign language," "violent video games," and "learning a foreign language".

Table 1. Three-word lexical bundling

Rank	Lexical Bundles	Freq
1	A foreign language	59
2	Violent video games	47
3	Learning a foreign	29
4	Of violent video	17
5	A second language	14
6	At least one	12
7	One foreign language	12
8	Least one foreign	11
9	Primary school homework	11
10	In the future	10
11	Of video games	10
12	Be able to	9
13	Foreign language can	9
14	One of the	9
15	That is why	9



Figure 2. Three-word lexical bundles word cloud in AntConc

The discovery revealed that the bundles served as effective foundational elements while exhibiting a reduced formulaic quality in conveying the complete meaning of the chunks. The subsequent sentences illustrate three-word bundles.

- (1) Maximizing children's brain development. Learning **a foreign language** can be beneficial.
- (2) Research on the impact of **violent video games** on children, adolescents, and adults is limited due to the lack of control for confounding variables like genetic predisposition, socioeconomic status, home violence, substance abuse, and psychiatric disorders.
- (3) The favorable impact on brain development is one of the key cognitive benefits of **learning a foreign** language during childhood.

B. Four-word lexical bundles

Four-word LBs are sequences of four words that often emerge in texts from distinct linguistic backgrounds. These bundles are significant because they reflect typical expression patterns and help maintain language coherence and fluency. These bundles frequently play crucial roles in conversation, such as offering examples, explaining concepts, expressing contrasts, and summarizing arguments. Their widespread recurrence in spoken and written language implies they are formulaic statements well ingrained in communication norms. Analyzing four-word LBs assists linguists and researchers in understanding how these patterns contribute to the structure and organization of speech, allowing for clearer and more effective communication in academic and everyday settings. Table 2 represents the highest frequency of the four LBs. They were "learning a foreign language", "of violent video", and "at least one foreign".

Table 2. Four-word lexical bundling

Rank	Lexical Bundles	Freq
1	Learning a foreign language	29
2	Of violent video games	17
3	At least one foreign	11
4	Least one foreign language	11
5	A foreign language can	9
6	So that is why	7
7	Can help children learn	6
8	Effect of violent video	6
9	Homework can help children	6
10	Learn a foreign language	6
11	Primary school homework is	6
12	To be able to	6
13	Violent video games and	6
14	Violent video games on	6
15	A foreign language from	5



Figure 3. Four-word lexical bundles word cloud in AntConc

The representations show that the significance of every element in this design may be completely grasped. Because of the difficulty in grasping the meaning of the three-word bundles, the four-word LBs sometimes coincide. Illustrations (4), (5), and (6) show how one sentence may include several four-word LBs, all with different starts and conclusions. The bundle contained "at least one foreign" and "at least one foreign language," with the former coming from the prior package. The examples showed syntagmatic overlaps to the right of the next bundle produced by the linear text arrangement.

- (4) Language is an essential component of culture, and learning a foreign language opens the door to comprehending diverse ways of living.
- (5) However, we can still prevent the effects of violent video games with support from our parents, or we can just avoid video games that have violent content.
- (6) As a result, the notion that children should learn at least one foreign language is more than just educational advice; it is a compelling requirement.

C. Five-word lexical bundling

Five-word LBs, which consist of sequences of five words that commonly appear together in texts, are observed to occur less frequently due to the dynamic and interactive nature of spoken communication. Longer LBs may be perceived as less spontaneous and can disrupt the interaction flow. Their examples are represented in [Table 3](#).

Table 3. Five-word lexical bundling

Rank	Lexical Bundles	Freq
1	At least one foreign language	11
2	Effects of violent video games	6
3	Learn at least one foreign	5
4	Learning a foreign language can	5
5	Learning a foreign language from	5
6	Least one foreign language is	5
7	Of learning a foreign language	5
8	Of violent video games on	5
9	Should learn at least one	5
10	Be replaced by modern technology	5
11	Cannot be replaced by modern	4
12	For children at this age	4
13	Have a bad impact on	4
14	Homework can help children learn	4
15	Impact of violent video games	4



Figure 4. Five-word lexical bundles word cloud in AntConc

Five-word and four-word bundles frequently overlap, indicating similarities in their structural composition and the contexts in which they appear. This variability underscores the complexity of LBs and highlights their role in shaping the coherence and efficiency of language use across different genres and communicative contexts.

- D. In conclusion, the case that students should learn at least one foreign language is based on numerous benefits that span cognitive, academic, and cultural realms
- E. The current Australian regulatory classification system for video games is similar to TV and film, but there is a need for high-quality research into the effects of violent video games and the issue of addictive internet use.

- F. The subject of whether children should **learn at least one foreign** language has received a lot of attention in today's linked and fast-changing world.

The examination of five-word LBs finds that such long sequences occur less frequently than three- or four-word bundles, a trend that corresponds with existing work on phrase length and formulaicity. The LBs usually include terms that are related to the topics of the students' essays, especially when it comes to learning a language and the effects of media. Five-word bundles are less common, but they have more complex grammatical structures and are often employed to wrap up arguments or make clear claims, especially at the end of an essay or thesis statement.

The five-word sequences in this study are more likely to show up in the more formal or high-stakes parts of the essay, including the introductions and conclusions, than the three- and four-word bundles. This backs up the concept that lengthier bundles are more likely to be used in planned, rehearsed academic writing and less likely to be used in casual speech or writing. The overall low frequency, nevertheless, shows that students may know about these constructions but haven't fully integrated them into their productive repertoire yet. This could be because they haven't had enough exposure to them or haven't had enough practice employing expanded lexical bundles.

D. Six-word lexical bundles

The examination of six-word LBs reveals their structural complexity, often comprising two or more embedded phrases and syntagmatic overlaps. Within a six-word sequence, overlapping syntactic structures typically exist to the left or right of the bundle, contributing to the complexity of these linguistic units. Six-word bundles, despite their complexity, are notably less common than shorter bundles. The rarity of six-word bundles indicates that they are specific and typically reserved for contexts that necessitate detailed or complex expression, such as academic writing or technical discourse. Analysis of these bundles reveals expression patterns that enhance our understanding of linguistic variation and communicative strategies across genres and domains. Table 4 presents the frequency of six-word LBs.

Table 4. Six-word lexical bundles

Rank	Lexical Bundles	Freq
1	At least one foreign language is	5
2	Learn at least one foreign language	5
3	Should learn at least one foreign	5
4	Cannot be replaced by modern technology	4
5	Master at least one foreign language	4
6	The effect of violent video games	4
7	Be an unnecessary burden on children	3
8	Benefit of learning a foreign language	3
9	Can be an unnecessary burden on	3
10	Can help children learn and strengthen	3
11	Children learn and strengthen their skills	3
12	Children should learn at least one	3
13	Help children learn and strengthen their	3
14	Homework can help children learn and	3
15	Impact of violent video games on	3

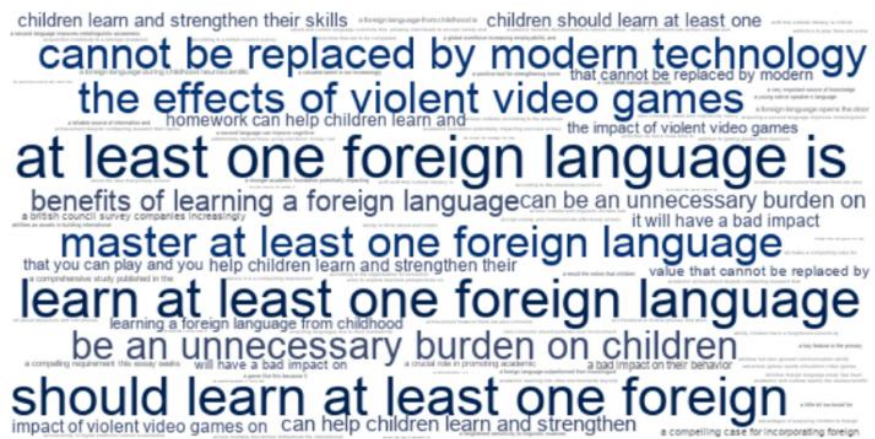


Figure 5. Six-word lexical bundles word cloud in AntConc

Even though this pattern featured rarely in students' essay writing, it was more formulaic than short ones because the entire meaning of the pieces was comprehended.

- D. Knowing **at least one foreign language is** very important for children. Foreign language skills can improve educational opportunities, careers and social relationships.
- E. Books remain a very important source of knowledge and have a value that **cannot be replaced by modern technology**.
- F. Children must **master at least one foreign language** in the current era of globalization; the ability to speak foreign languages is very important.

The study of six-word LBs in the students' academic writing shows how complicated their structures are and how particular their functions are. There are not as many six-word sequences as there are shorter ones. These show that students are trying to express more complicated and specific thoughts. Table 4 shows that these bundles usually showed up in high-stakes situations like making an argument or coming to a conclusion. This result backs up the idea that lengthier LBs are generally made on purpose instead of by chance, which shows that the language output is more rehearsed or planned. Their appearance also implies that students are starting to assimilate longer formulaic sequences that serve different rhetorical goals, especially when it comes to making conclusions, drawing implications, or making evaluations.

DISCUSSION

The analysis of four-word LBs backs up the trends identified in the three-word bundle results and gives us further information about how students build their discourse. The most common LBs show that student writers often use topic-driven lexical chunks to keep their writing together and make their points. These bundles generally come from common classroom themes or essay prompts, which suggests that students are using language they have memorized or learned rather than creating more abstract or discipline-specific academic frameworks. Many of these bundles are functionally referential, as seen in the example sentences. They can help youngsters learn, draw a contrast (so that is why), or explain cause and effect (as a result).

These findings align with those reported by [Ardi et al. \(2023\)](#) and [Kostromitina \(2022\)](#) who found EFL student authors commonly utilize LBs that are quite specialized in terms of vocabulary but not very functional. In both studies, students were likely to choose word sequences they were already familiar with that were based on common topics or classroom discussions. Similarly, [Mashhadi and Qanavati \(2024\)](#) noted that Iranian EFL students also relied extensively on bundles that help with direct explanation and topic development. However, the LBs found in this study show a narrower range of discourse functions, mostly focused on factual description and simple causal logic. This contrasts with [Shin \(2018\)](#) findings, which demonstrated that native writers used LBs in a wider and more rhetorically complex way.

Furthermore, the results concerning five-word bundles echo these trends. As reported by [Mashhadi and Qanavati \(2024\)](#), five-word bundles are more underused by intermediate EFL learners compared to prior studies. This is because they are cognitively demanding, and the learners don't know many collocations. Also, [Shin \(2018\)](#) discovered that NSs use longer bundles more often to show evaluative position and abstract links, while EFL writers commonly use shorter, topic-specific chunks. The current results support this, showing that while students can make five-word bundles when they are strongly related to the writing topic, their overall use is still limited in range and depth.

In terms of six-word LBs, the data confirm their rarity in EFL writing. This is in line with a study conducted by [Shin \(2018\)](#) and [Mashhadi and Qanavati \(2024\)](#), who found that learners generally have a hard time creating long lexical sequences without direct input or memorization. On the other hand, native authors are better at putting large bundles into different types of sentences and rhetorical gestures. The students in this study generally used these longer bundles in situations linked to the issue, including language education and media. This suggests that they used them more because of the content than because of their purpose or genre.

In general, the use of six-word lexical bundles by student writers is still low, but their presence indicates progress toward more advanced academic proficiency. These longer bundles suggest that learners are beginning to organize their ideas in a more complex and contextually appropriate manner. From a pedagogical perspective, incorporating corpus-informed writing instruction that emphasizes longer, functionally diverse lexical bundles can enhance students' fluency and rhetorical control, skills essential for advanced learners who must employ abstract reasoning, contrastive logic, and evaluative judgment in academic writing ([Dan et al., 2024](#); [Nasrabady et al., 2020](#); [Shin, 2018](#))

Looking at the variation of LBs in students' academic essays gives crucial information about their language proficiency and pragmatic choices. The results show that most student authors relied heavily on shorter LBs, particularly three and four-word sequences, which suggests dependence on familiar and memorized frames to maintain the surface level of cohesion and flow. This pattern indicates that students can produce coherent basic academic text, such as introducing themes, highlighting main ideas, or stating general facts, but they show limited experience with using bundles for more complex rhetorical purposes, such as contrast, hedging, or critical commentary. These findings are in line with [Ardi et al. \(2023\)](#) and [Kostromitina \(2022\)](#), who stated that EFL learners prefer lexically specific, topic-driven chunks with a narrow functional range, and with [Mashhadi and Qanavati \(2024\)](#), who observed similar topic-

bound uses among intermediate learners.

On the other hand, longer bundles, five- and six-word LBs less frequent but exhibited greater syntactic complexity and stronger propositional completeness. These longer bundles tended to appear in higher-stakes parts of essays, thesis statements, major argument claims, and conclusions, suggesting that students deploy them when a more formal or rehearsed expression is required. This pattern corroborates previous observations that learners underuse longer bundles because of their cognitive demands and limited collocational knowledge (Mashhadi & Qanavati, 2024). It also aligns with Shin (2018), who found that native writers integrate longer bundles more flexibly to express stance and to create abstract links. However, unlike the native-writing patterns described by Ahmad et al. (2024) and Shin (2018), the students used long bundles mainly in topic-related contexts rather than across varied rhetorical functions. In other words, while some students show emerging ability to use longer, more academicized sequences, their use remains topical and formulaic rather than rhetorically versatile.

Taken together, these comparisons show that our results both confirm and extend prior work. This study documented EFL writers' reliance on short, content-driven bundles and the relative scarcity of functionally diverse long bundles. At the same time, by focusing on Indonesian undergraduate essays after explicit classroom instruction, this study adds empirical evidence from an underrepresented, non-Western context and suggests that explicit teaching can help learners begin to acquire longer, more propositional sequences in a constrained, topic-dependent way. Pedagogically, these findings underscore the need for corpus-informed instruction that raises learners' awareness of the functional range of LBs, targets longer, discourse-organizing bundles in authentic contexts, and provides scaffolded practice that helps students generalize bundles beyond narrow topical uses. Such interventions would likely promote greater rhetorical flexibility and alignment with international academic discourse norms.

CONCLUSION

This study reveals that EFL students at a private Islamic university in Riau demonstrate frequent use of LBs in their essay writing, especially those related to common educational topics such as foreign language acquisition and the impact of video games. While students utilized a variety of three- to six-word LBs, their usage patterns were highly repetitive, lacking syntactic diversity and academic discourse markers. Most LBs were tied to familiar topics rather than broader academic contexts, indicating students' reliance on memorized or formulaic expressions rather than strategic language use.

The essays examined showed that students favored certain phrases like "a foreign language," "violent video games," and "at least one foreign language," which often reappeared with slight variations. However, more complex and discourse-organizing bundles such as metadiscoursal expressions were largely absent. This pattern suggests a gap in students' ability to structure academic arguments fluently and coherently, pointing to the need for targeted instruction.

Given these findings, educators at similar institutions can benefit from implementing corpus-based writing instruction that highlights not only the frequency but also the function of LBs in academic texts. Teaching students to recognize and use a wider

variety of LBs, especially those related to argumentation, contrast, and cohesion, can help improve the clarity and organization of their essays. This conclusion is drawn directly from the linguistic characteristics of the student corpus, demonstrating that practical, context-sensitive interventions are needed to support students' academic writing development.

This study only looks at the lexical bundles that third-semester students at a private Islamic university in Riau employ when they write English academic essays. The study's main goal was to find out how often, how they are structured, and what they do in a specific academic setting. This gives us useful information about how EFL students in this setting use language, but the results may not apply to students at different academic levels, from different locations, or at different types of schools. Also, the data only came from one type of writing: argumentative essays. This may not show the whole range of how students employ lexical bundles in other types of academic writing, such as papers, summaries, or reflections.

Another limitation of this study lies in its focus on written language and doesn't include other academic skills like speaking or listening. Since LBs are also important parts of spoken academic discourse, future studies might employ a multimodal approach to look into how students use formulaic language in oral presentations, discussions, or listening comprehension tasks. Also, this study mostly used frequency-based analysis and didn't look closely at how aware or intentional the learners were about using bundles. This may have been done through interviews, think-aloud protocols, or classroom observations.

Future studies can improve on this one by looking at more than one institution or by using longitudinal designs to see how the use of LBs changes over time. It would also be helpful to look into how explicit instruction or corpus-based interventions affect students' capacity to make bundles that are more varied and effective in terms of rhetoric. Also, using this study on older or graduate students could find out how their vocabulary is getting better and set goals for how their academic writing should improve.

In conclusion, this study gives a narrow look at how EFL students employ LBs in a certain academic and cultural setting. However, it also opens the door to more research on how formulaic language helps students become more literate. Future research can offer better teaching advice and help create more successful and inclusive models of academic writing instruction in EFL contexts by filling the gaps found here and using a wider range of methods.

ACKNOWLEDGEMENTS

The authors wish to convey their heartfelt thanks to the Directorate of Research and Community Service (DPPM) for the financial assistance given for this study.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interest.

AUTHOR(S) CONTRIBUTION

Etfita, F.: conceived the research idea, led the study design, and coordinated the

manuscript writing. Wahyuni, S.: contributed to data collection. Ahmad, A.: was responsible for data analysis and interpretation. Sudusinghe, W.S. and Gamage, C.K.W.: reviewed and edited the manuscript and provided critical feedback to improve the overall quality.

REFERENCES

- Agustina, Y., Widodo, P., & Margana, M. (2024). Creating an Islamic boarding school English corpus: corpus metadata, frequently used words, and unique words. *Journal of Education and Learning (EduLearn)*, 18(3), 804–816. <https://doi.org/10.11591/edulearn.v18i3.21095>
- Ahmad, M., Mahmood, M. A., Siddique, A. R., Imran, M., & Almusharraf, N. (2024). Variation in academic writing: A corpus-based investigation on the use of syntactic features by advanced L2 academic writers. *Journal of Language and Education*, 10(3), 25–39. <https://doi.org/10.17323/jle.2024.21618>
- Alzahrani, G. A. S. (2025). Academic writing difficulties of EFL learners at Najran University. *Journal of Educational and Social Research*, 15(3), 217–229. <https://doi.org/10.36941/jesr-2025-0093>
- Appel, R. (2022). Lexical bundles in L2 English academic texts: Relationships with holistic assessments of writing quality. *System*, 110, 102899. <https://doi.org/10.1016/j.system.2022.102899>
- Ardi, P., Oktafiani, Y. D., Widianingtyas, N., Dekhnich, O. V., & Widiati, U. (2023). Lexical bundles in Indonesian EFL textbooks: A corpus analysis. *Journal of Language and Education*, 9(2), 25–39. <https://doi.org/10.17323/jle.2023.16305>
- Bal Gezegin, B. (2019). Lexical bundles in published research articles: A corpus-based study. *Dil ve Dilbilimi Çalışmaları Dergisi*, 15(2), 520–534. <https://doi.org/10.17263/jlls.586188>
- Birhan, A. T. (2021). Effects of teaching lexical bundles on EFL students' abstract genre academic writing skills improvement: Corpus-based research design. *International Journal of Language Education*, 5(1), 585–597. <https://doi.org/10.26858/IJOLE.V5I1.14917>
- Cheng, S.-H., & Hsu, J.-Y. T. (2025). Structures and functions of lexical bundles in FinTech discourse. *Southern African Linguistics and Applied Language Studies*, 43(1), 30–44. <https://doi.org/10.2989/16073614.2024.2326500>
- Cherry Chan, S. Y. (2025). "Writing is re-writing": Developing language awareness through a process-oriented approach in an EFL academic writing course. In *English for Academic and Specific Purposes in the Classroom from Theory to Practice*. <https://doi.org/10.4324/9781003471899-14>
- Crossley, S. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11(vol. 11 issue 3), 415–443. <https://doi.org/10.17239/jowr-2020.11.03.01>
- Dan, C., Darmi, R. H., & Yusof, M. A. M. (2024). The use of lexical bundles in English language academic writing among university learners: A systematic literature review. *World Journal of English Language*, 14(3), 196. <https://doi.org/10.5430/wjel.v14n3p196>
- Demiray Akbulut, F. (2020). A bibliometric analysis of lexical bundles usage in native and non-native academic writing. *Dil ve Dilbilimi Çalışmaları Dergisi*, 16(3), 1146–1166. <https://doi.org/10.17263/jlls.803583>
- Gak, D., Bogdanović, V., & Bulatović, V. (2024). Structural analysis of lexical bundles in graduate students' research paper abstracts. *Folia Linguistica et Litteraria*.

<https://doi.org/10.31902/fli.46.2023.8>

- Hosseinpour, R. M., Behzadi, F., & Damavandi, A. A. (2023). Lexical bundles in EAP research articles of L1 and L2 English authors: A contrastive analysis. *The Journal of AsiaTEFL*, 20(2), 215–243. <https://doi.org/10.18823/asiatefl.2023.20.2.1.215>
- Huang, C.-R., & Yao, Y. (2015). Corpus linguistics. In the *International Encyclopedia of the Social Behavioral Sciences, Second Edition*. <https://doi.org/10.1016/B978-0-08-097086-8.52004-2>
- Kashiha, H., & Chan, S. H. (2015). A little bit about: Differences in native and non-native speakers' use of formulaic language. *Australian Journal of Linguistics*, 35(4), 297–310. <https://doi.org/10.1080/07268602.2015.1067132>
- Kaya, Ö. F. (2025). The use of prepositional phrase-based bundles in essays written by Turkish learners of English. In *Exploration of the Intersection of Corpus Linguistics and Language Science*. <https://doi.org/10.4018/9798369381465.ch006>
- Kim, S., & Kessler, M. (2022). Examining L2 English university students' uses of lexical bundles and their relationship to writing quality. *Assessing Writing*, 51, 100589. <https://doi.org/10.1016/j.asw.2021.100589>
- Kostromitina, M. (2022). Cross-linguistic transfer in academic journal writing: Evidence from lexical bundle analysis in Russian and English. *Research in Corpus Linguistics*, 10(2), 70–112. <https://doi.org/10.32714/ricl.10.02.05>
- Kurniawan, E., & Haerunisa, Z. F. (2023). A comparative study of lexical bundles in accepted and rejected applied linguistic research article introductions. *Studies in English Language and Education*, 10(2), 628–648. <https://doi.org/10.24815/siele.v10i2.28119>
- Li, M., & Pei, L. (2024). Exploring challenges in academic language-related skills of EFL learners in Chinese EMI settings. *Acta Psychologica*, 247, 104309. <https://doi.org/10.1016/j.actpsy.2024.104309>
- Li, Y., & Lei, H. (2025). Lexical bundles in L1 and L2 English academic writing: Convergent and divergent usage. *Sage Open*, 15(2). <https://doi.org/10.1177/21582440251333850>
- Lin, P., & Adolphs, S. (2023). Corpus linguistics. In *Routledge Handbook of Applied Linguistics*. <https://doi.org/10.4324/9781003082644-25>
- Liu, C.-Y., & Chen, H.-J. H. (2020). Analyzing the functions of lexical bundles in undergraduate academic lectures for pedagogical use. *English for Specific Purposes*, 58, 122–137. <https://doi.org/10.1016/j.esp.2019.12.003>
- Mashhadi, A., & Qanavati, M. (2024). Are phrases frozen? Investigating lexical bundles in non-native and native reports in MICUSP. *Language Related Research*, 15(3), 79–105. <https://doi.org/10.29252/LRR.15.3.4>
- McEnery, T., & Brookes, G. (2024). Corpus linguistics and the social sciences. *Corpus Linguistics and Linguistic Theory*, 0(0). <https://doi.org/10.1515/cllt-2024-0036>
- Nasrabady, P., Elahi Shirvan, M., & Ehsan Golparvar, S. (2020). Exploring lexical bundles in recent published papers in the field of applied linguistics. *Journal of World Languages*, 6(3), 175–197. <https://doi.org/10.1080/21698252.2020.1797992>
- Nguyen, A. N., & Van Vu, T. (2024). Typical linguistic errors committed by tertiary students in legal written outputs. *East European Journal of Psycholinguistics*, 11(2), 120–141. <https://doi.org/10.29038/eejpl.2024.11.2.ngu>
- Nola Bacha, N., & Khachan, V. (2023). A corpus-based lexical evaluation of L1 Arabic

- learners' English literary essays. *International Journal of Arabic-English Studies*, 23(2), 415–442. <https://doi.org/10.33806/ijaes.v23i2.470>
- O'Flynn, J. (2022). Lexical bundles in the academic writing of the arts and humanities: From corpus to CALL. *Yearbook of Phraseology*, 13(1), 81–108. <https://doi.org/10.1515/phras-2022-0006>
- Öztürk, Y., & Taşçı, S. (2023). A corpus-based analysis of lexical bundles in non-native post graduate academic writing and a potential L1 influence. *REFLectations*, 30(2), 488–505. <https://doi.org/10.61508/refl.v30i2.267463>
- Prapobratanakul, C. (2024). Thai EFL undergraduate engineering students' perspectives on academic writing: Challenges and strategies. *Pasaa*, 68, 1–31. <https://doi.org/10.58837/chula.pasaa.68.1.1>
- Sanosi, A. B. (2022). The use and development of lexical bundles in Arab EFL writing: A corpus-driven study. *Journal of Language and Education*, 8(2), 106–121. <https://doi.org/10.17323/jle.2022.10826>
- Shin, Y. K. (2018). The construction of English lexical bundles in context by native and nonnative freshman university students. *English Teaching*, 73(3), 115–139. <https://doi.org/10.15858/engtea.73.3.201809.115>
- Shirazizadeh, M., & Amirfazlian, R. (2021). Lexical bundles in theses, articles and textbooks of applied linguistics: Investigating intradisciplinary uniformity and variation. *Journal of English for Academic Purposes*, 49, 100946. <https://doi.org/10.1016/j.jeap.2020.100946>
- TAŞKAYA, H., & ÖZBAY, A. Ş. (2023). Corpus analysis of the use of lexical bundles by Turkish EFL learners at a tertiary level. *Hacettepe University Journal of Education*. <https://doi.org/10.16986/HUJE.2023.494>
- Tracy-Ventura, N., & Paquot, M. (2020). *The Routledge Handbook of Second Language Acquisition and Corpora* (N. Tracy-Ventura & M. Paquot, Eds.). Routledge. <https://doi.org/10.4324/9781351137904>
- Wang, W., Chen, W., Shi, S., Tong, L., & Duan, M. (2024). A Corpus-based study on cohesion development in writing by EFL novices. *Journal of Language Teaching and Research*, 15(3), 772–786. <https://doi.org/10.17507/jltr.1503.10>
- Wang, Y., & Xie, Q. (2022). Diagnostic assessment of novice EFL learners' discourse competence in academic writing: a case study. *Language Testing in Asia*, 12(1). <https://doi.org/10.1186/s40468-022-00197-y>
- Wei, Y., Sulaiman, N. A., & Ismail, H. H. (2024). Academic English writing challenges in a blended EFL and ESL learning environment: Insights from Chinese international high school students. *International Journal of Learning, Teaching and Educational Research*, 23(2), 275–293. <https://doi.org/10.26803/ijlter.23.2.13>
- Xuan Nguyen, N. (2023). Structures and functions of lexical bundles in student and expert timed argumentative writing: A corpus-driven study. *Pertanika Journal of Social Sciences and Humanities*, 31(3), 1015–1039. <https://doi.org/10.47836/pjssh.31.3.06>
- Yunjung (Yunie), K. (2025). A corpus analysis of prepositional phrase-lexical bundles in academic writing: L2 writers from Indo-European and Non-Indo-European languages. *Applied Corpus Linguistics*, 5(2). <https://doi.org/10.1016/j.acorp.2025.100128>
- Zhang, D., Tan, J. T. A., & Roy, S. S. (2023). A systematic review of interventions improving university students' EFL writing competence. *International Journal of Learning, Teaching and Educational Research*, 22(10), 93–112.

<https://doi.org/10.26803/ijlter.22.10.6>

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. <https://doi.org/10.3390/su13105553>