

## Lexical Bundles in English Learners' Production: A Corpus-Based Study of University Students' Academic Writing in Togo

Akponi Tarno, Nassalénga Sampouma

Université de Lomé, Togo

[tarnoakponi@gmail.com](mailto:tarnoakponi@gmail.com)

How to cite:

Tarno, A., & Sampouma, N. (2026). Lexical Bundles in English Learners' Production: A Corpus-Based Study of University Students' Academic Writing in Togo. *International Journal of Linguistics and Translation Studies*, 7(4). 82-102. <https://doi.org/10.36892/ijlts.v7i4.794>

### ARTICLE HISTORY

Received:  
02/09/2026

Accepted:  
15/10/2026

### Keywords:

lexical bundles;  
corpus linguistics;  
English learner corpora;  
formulaic sequence;  
discipline-specific phraseology

### Abstract

This study investigates the use of lexical bundles in English as a foreign language through a corpus-based analysis of Master students' dissertations in African Literature, American Literature, and Applied Linguistics in Togo. Drawing on insights from Biber et al.'s Corpus-Based approach to lexical bundles and Gries's Phraseology approach, the study examines the frequency, structural patterns, and functional distribution of recurrent multi-word sequences in a purpose-built learner corpus of 225 academic essays. It selected only introductions and conclusions for the purpose of the analysis. The findings show that learners rely structurally on prepositional phrase-based bundles and functionally on referential bundles. Though the results show a structural preference for prepositional phrase fragments and a functional interest in referential bundles, it has found that Togolese EFL learners' phraseological competence remains underdeveloped and needs to be boosted through well-planned teaching programmes. The study contributes to ongoing debates in learner corpus research by highlighting the importance of formulaic language in academic proficiency.

## 1. INTRODUCTION

The increasing reliance on empirical data in linguistic research has led to the prominence of Corpus Linguistics as a central methodological and theoretical framework for the analysis of language in use. Through the examination of naturally occurring texts, corpus-based approaches enable systematic investigation of recurrent linguistic patterns across registers, genres, and speaker populations (Jassim & Joshi, 2023; Biber, 1996). The area of lexical bundles is a specific area that has received considerable attention within the paradigm of the study of formulaic language and that constitutes a fundamental component of fluent and coherent discourse.

Lexical bundles are typically defined as recurrent sequences of words that co-occur with high frequency in a given register (Biber et al., 2021). According to Richards and Schmidt (2010, p. 335), lexical bundle refers to "a type of fixed phrase consisting of a sequence of three or more words that co-occur frequently in a particular type of writing or register such as academic writing." Biber et al. (2004, p. 371) extend the definition of lexical bundles as "multi-words sequences." From the same perspective, Biber et al. (1999, p. 990) conclude that "lexical

bundles are simply sequences of word forms that commonly go together in natural discourse.” Consequently, unlike idioms, these sequences are often structurally incomplete and semantically transparent; however, they play a crucial role in discourse organization and meaning construction. In academic writing, lexical bundles serve a variety of functions, including structuring arguments, expressing stance, and guiding the reader through the text (Hyland, 2008). For example, bundles such as *in the case of*, *on the other hand*, *as a result of*, *this study intends to*, *the purpose of this study* contribute to cohesion and facilitate the articulation of complex ideas.

Previous research has demonstrated that the distribution and function of lexical bundles vary systematically across genres and levels of expertise (Yan & Hong, 2025; Nasrabady et al., 2020). Studies of expert academic writing reveal a high degree of phraseological sophistication, characterized by a wide range of structurally complex bundles, particularly noun phrase fragments that encode dense informational content (Biber et al., 2004; Hyland, 2008). In contrast, learners of English as a foreign language tend to rely on a more restricted repertoire of bundles. Consequently, they often overuse a limited set of high-frequency expressions while they underuse those associated with stance and disciplinary specificity (Kondos, 2025; Ranjbar et al., 2012; Chen & Baker, 2010; Granger, 2003). These differences express the development in phraseological competence and highlight the importance of formulaic sequences in foreign language learning.

Within learner corpus research, lexical bundles have been widely studied in various contexts such as Europe and Asia (Al-Khasawneh, 2024; Navarro Gil & Martínez Caro, 2019). These studies have led to significant insights into interlanguage development and learner variation. However, the lexical bundle area in the African context remains relatively underexploited in this body of research. This gap is particularly notable due to the increasing role of English in the academic communication across the continent and the whole world. In countries such as Togo, where English is learned as a foreign language within a predominantly Francophone educational system, students are required to engage with academic discourse in English despite the limited exposure to models that sound native. This situation raises important questions about the nature of learners’ phraseological competence and their ability to produce conventionalized academic discourse.

Previous studies in Togo have primarily focused on sociolinguistic variation (Essizewa, 2012; Maldja, 2018), phonological and morphological analysis of local languages (Ourso, 1989; Gawa, 2016), comparative studies between foreign and local languages and lexical innovation (Zato, 2012; Pere-Kewezima, 1994) with relatively little attention paid to corpus-based analyses of learners’ writing. Consequently, there is a lack of empirical evidence regarding the use of lexical bundles in English language learners’ production, particularly in specific domains such as academic writing in literature and applied linguistics. Addressing this gap is essential not only for advancing theoretical understanding of learner language but also for informing pedagogical practices in higher education.

The research aims to contribute to the area by providing a corpus-based analysis of lexical bundles in Master essays written by some students from the department of English at the University of Lomé in Togo (West Africa). By examining learners’ use of lexical bundles in the fields of literature and applied linguistics, the study seeks to explore how students construct

academic discourse and to what extent their usage aligns with established patterns in native academic writing.

## **2. LITERATURE REVIEW**

In the 1800s, English was introduced in Togo by merchants before colonisation. According to Lange (1998), the first merchants who settled down along the Atlantic coast founded schools where their children and some of the indigenous received instruction in English. Later on, some Christian missionaries followed the merchants, established and shaped the educational system of the Togoland where the main languages of instruction were English and Ewe, a local language. Lawrance (2000) declares that English was taught by Protestant missionaries both to spread Christianity and encourage local marketization for exploitation by the underwriters of missions. The advent of German colonisation did not change anything in the use of English in the Togolese educational system. Today, English is used in the educational system, where it coexists with French and local languages in a multilingual learning environment. As a foreign language, its teaching raises important questions about effective strategies for developing communicative competence. In this regard, research in Applied Linguistics highlights the role of formulaic sequences, particularly lexical bundles, in enhancing fluency in academic discourse. This section therefore reviews the literature on lexical bundles and examines their relevance for English language teaching and learning in the Togolese context. However, before dealing with literature review on lexical bundles, a quick look at Corpus Linguistics is important.

Indeed, the development of Corpus Linguistics has significantly reshaped linguistic inquiry by enabling the systematic analysis of authentic language use (Stefanowitsch, 2020; Biber et al., 1998/2006). According to McEnery & Wilson (2001, p. 1), “Corpus linguistics is perhaps best described for the moment in simple terms as the study of language based on examples of ‘real life’ language use.” Corpus linguistics is grounded in the principle that linguistic patterns emerge from frequency and distribution in naturally occurring data rather than from intuition alone (McEnery & Hardie, 2012; Sinclair, 1991). Through computational tools and large datasets, researchers can identify recurrent lexical and grammatical patterns, providing empirical evidence for theoretical claims (Boumediene, 2025; Özbay & Gürsoy, 2023).

One of the most influential contributions of corpus linguistics is the recognition of formulaic language as a central component of linguistic competence (Biber, 2009; Trklja & Grabowski, 2021; Brooke et al., 2015). Rather than viewing language as composed solely of individual words and rules, corpus-based research highlights the importance of multi-word units that recur across texts and contexts (Biber et al., 1998/2006). These units, which include collocations, idioms, and lexical bundles, play a crucial role in shaping fluent and idiomatic language use (Liu, 2012; Polić, 2020).

Initially considered as “sequences of words that tend to co-occur, irrespective of their idiomaticity and whether or not the sequence of words constitutes a grammatical unit” (Biber et al., 2021, p. 61), lexical bundles are typically defined as “a recurring sequence of three or more words” (Biber et al., 2021, p. 982) that co-occur in a particular register. Consequently, unlike idiomatic expressions, lexical bundles are multi-word sequences that are often semantically transparent and structurally incomplete (Biber, 2009); however, they are essential for discourse organization and communicative efficiency.

Research has shown that lexical bundles can be classified according to both structural and functional criteria (Laia et al., 2023; Long et al., 2024; Nguyen, 2023). Structurally, they include noun phrase fragments (e.g., *the role of the*), prepositional phrase fragments (e.g., *in the context of*), and verb phrase fragments (e.g., *is one of the*) (Biber et al., 2004; Dahunsi & Ewata, 2022; Hong, 2019). Functionally, Hyland (2008, p. 54) distinguishes between research-oriented bundles (used to describe activities or phenomena), text-oriented bundles (used to organize discourse), and participant-oriented bundles (used to express stance or engage the reader).

Importantly, the distribution of lexical bundles varies across genres and disciplines. Academic writing, for instance, tends to favour noun phrase bundles that encode dense informational content, reflecting the informational demands of the register (Biber et al., 2004). This variation underscores the role of lexical bundles as markers of register-specific competence.

The study of lexical bundles has been particularly prominent in research on academic discourse. Corpus-based analyses have demonstrated that expert writers employ a wide range of lexical bundles to structure arguments, signal relationships between ideas, and express nuanced positions (Neely & Cortes, 2009; Hyland, 2008; Hyland & Jiang, 2018). These bundles contribute to the cohesion and coherence of texts, facilitating reader comprehension and enhancing the clarity of communication.

According to Hyland (2008), the use of lexical bundles varies across disciplines, which reflects different epistemological practices and rhetorical conventions. Hyland found for example that while humanities writing often includes bundles that express evaluation and interpretation, scientific writing tends to employ bundles that describe procedures and results. Similarly, studies on disciplinary variation indicate that scientific and technical writing relies heavily on procedural and resultative bundles to explain methods and present experimental outcomes. Specifically, Navarro Gil and Martínez Caro (2019) note that scientific and linguistics texts favour “procedural bundles” and “resultative/inferential” bundles because writers need to explain how research is conducted and how conclusions are derived. Such findings highlight the importance of discipline-specific phraseology in academic writing.

The knowledge of lexical bundles is therefore a critical aspect of academic writing (Cortes, 2004; Navarro Gil & Martínez Caro, 2019; Lorenzo Salazar, 2011; Sánchez Hernández, 2020). The mastery of these recurrent sequences allows scholars and writers to produce texts that conform to the expectations of academic discourse communities. Conversely, limited knowledge of lexical bundles may hinder learners’ ability to express complex ideas effectively and to engage with disciplinary conventions.

In line with Hyland (2008), Hyland and Jiang (2018) investigated the use and variation of lexical bundles in academic writing across disciplines. They emphasized the important role of these multi-word expressions in shaping disciplinary discourse conventions. The two scholars demonstrate that lexical bundles are not randomly distributed; rather, they reflect a real part of users’ communicative experiences and epistemological preferences of specific academic communities. Their study reveals that writers in different fields employ distinct functional bundles to organize arguments, express stance, and construct disciplinary identity. Hyland and Jiang also argue that the mastery of these formulaic sequences contributes significantly to fluency, coherence, and academic credibility in scholarly writing.

Within the field of learner corpus research, lexical bundles have been widely studied as indicators of phraseological competence (Siu et al., 2024; Al-Khasawneh, 2024). Research consistently shows that learners of English as a Foreign Language (EFL) differ from native speakers in their use of lexical bundles, both quantitatively and qualitatively (Biber et al., 2004; Cortes, 2004; Hyland, 2005; De Cock, 2000). One of the most important findings is that learners tend to rely on a restricted set of high-frequency bundles, often overusing certain expressions while underusing others (Chen & Baker, 2010). They found for example that learners frequently overuse text-organizing bundles such as *on the other hand* and *in conclusion*. However, these students underuse stance bundles that are supposed to convey evaluation and argumentation. This imbalance suggests that learners may have limited exposure to the full range of functions performed by lexical bundles in academic discourse.

In addition, Granger (2003) found that English as a foreign language learners often produce non-native-like bundles or unconventional combinations that reflect incomplete mastery of phraseological patterns. Such patterns are indicative of interlanguage development, where learners' linguistic systems are influenced by both the target language and their first language.

The concept of phraseological competence, therefore, extends beyond knowledge of individual words to include the ability to use recurrent sequences appropriately in context (Paquot, 2018; Estaji & Hashemi, 2022). In English as a foreign language context, developing this competence is essential for achieving fluency and accuracy in academic writing.

Despite the extensive body of research, corpus-based studies remain underrepresented in African contexts. Most existing research has focused on European and Asian learner populations, leaving a significant gap in our understanding of African and especially Togolese learners. In Togo where English is learned as a foreign language within a multilingual and predominantly Francophone environment, learners often have limited exposure to authentic English input, particularly in academic contexts. This lack of exposure may affect learners' acquisition of lexical bundles. Furthermore, the language teaching and learning in such a context emphasizes grammatical mastery over discourse-level competence, leading to insufficient attention to formulaic language. The few studies that have examined Togolese learner English deal with learner motivation, challenges that learners face in learning English as a foreign language, and the teaching of different language skills. However, there is a lack of detailed corpus-based analysis focusing specifically on lexical bundles in this context. This gap is particularly significant given the increasing importance of English for academic and professional communication. The present study addresses this gap by providing a corpus-based analysis of academic writing produced by Master students in Togo. By examining the frequency, structure, and function of lexical bundles in this context, the study aims to contribute to both theoretical understanding and pedagogical practice in applied linguistics.

### **3. METHODOLOGY**

#### **3.1. Research Design and Learner Corpus**

This study adopts a corpus-based study to investigate the use of lexical bundles in Togolese learner English. The data were collected from Master students' dissertations. The data were collected from the following areas: African Literature, American Literature, and Applied Linguistics. All the collected texts have been defended in front of an examination board and a public made up of fellow students, family members, and guests. The introduction and the conclusion of each dissertation were selected for the purpose of data analysis. The analysis

focused on identifying the frequency, structural patterns, and functional distribution of recurrent multi-word sequences in learner corpus of academic writing. The introductions and conclusions were identified and stored separately according to the three fields of study. Since it was difficult to get the same number of words, 75 Master of Art dissertations were selected in each field of study. A total of 225 dissertations were used for this study. Information about the number of words in the introductions and the conclusions, the number of introductions and conclusions and the mean length of the introductions and the conclusions is offered by Table 1.

**Table 1. The learner corpus**

| Field of Study      | Variables               |                       |                                     |                                   |                                  |                                |
|---------------------|-------------------------|-----------------------|-------------------------------------|-----------------------------------|----------------------------------|--------------------------------|
|                     | Number of introductions | Number of conclusions | Number of words in the Introduction | Number of words in the Conclusion | Mean length of the introductions | Mean length of the conclusions |
| African Literature  | 75                      | 75                    | 166,581                             | 93,807                            | 2,221.03                         | 1,250.76                       |
| American Literature | 75                      | 75                    | 142,464                             | 87,591                            | 1,899.52                         | 1,167.88                       |
| Applied Linguistics | 75                      | 75                    | 124,971                             | 93,429                            | 1,666.28                         | 1,245.75                       |
| Total               | 225                     | 225                   | 434,016                             | 274,827                           | 1,928.94                         | 1,221.46                       |

### 3.2.Data Collection and Preparation

Though the researcher was a lecturer, the whole Master dissertations were collected with the consent of his students and then anonymized to ensure confidentiality. The corpus was converted into plain text (.txt) format to facilitate processing with the AntConc corpus analysis software. Prior to analysis, the corpus underwent a cleaning process. This process involved the removing of personal identifiers such as students' names and IDs, the correction of obvious typographical errors that could affect frequency counts and the standardization of formatting such as punctuation and spacing. However, no grammatical corrections were made because the study aims to capture authentic learner language.

### 3.3.Analytical Tool and Data Analysis Procedures

The analysis of the whole corpus was conducted with the use of AntConc (version 3.5.8), a widely used freeware tool for concordance and frequency analysis and the Statistical Package for the Social Sciences (SPSS) for statistical work. AntConc was adopted to identify 3, 4 and 5-word lexical bundles with different frequencies. However, the emphasis was mainly put on the 4-word lexical bundles. The tools of Cluster/N-gram, Concordance, Concordance Plot, and File View were frequently employed. The retained lexical bundles were classified according to their structural patterns, following the taxonomy proposed by Biber et al. (2004): noun phrase fragments (e.g., *the impact of the*), prepositional phrase fragments (e.g., *in the context of*), verb

phrase fragments (e.g., *is one of the*). This classification allowed the researcher to identify the structural preferences in learner writing. The lexical bundles were further categorized on the basis of their discourse functions by using the framework developed by Hyland (2008). Hyland developed the research-oriented bundles, the text-oriented bundles and the participant-oriented bundles. While the first category describes activities, quantities, or processes, the second organizes the structure of the text and the third one expresses stance or engages the reader. Each bundle was assigned to a functional category based on its most frequent usage, as determined through concordance analysis.

#### **4. THEORETICAL FRAMEWORK**

Though this study is grounded in the broad field of Corpus Linguistics, it adopts Biber et al.'s (1999, 2004) Corpus-Based approach to lexical bundles and Gries's (2008) Phraseology approach to analyse the data. From the broad standpoint, Corpus Linguistics perspective provides both the methodological and theoretical foundation for the analysis of repetitive linguistic patterns in authentic texts. It is based on the premise that language structure and use can be systematically described through the analysis of large collections of naturally occurring data (McEnery & Hardie, 2012; Sinclair, 1991). Instead of relying on introspection, Corpus Linguistic approach emphasizes frequency, distribution, and patterning as key indicators of linguistic knowledge (McEnery & Hardie, 2012).

A central tenet of corpus linguistics is that meaning is shaped by usage (Gries, 2009; Pollach, 2012). Sinclair (1991, p. 110) introduced the distinction between the 'open-choice principle' and the 'idiom principle'. The researcher argues that even though words do not occur at random, speakers do not construct language word by word in a purely combinatorial fashion but instead rely heavily on pre-constructed sequences. This insight has led to increased attention to formulaic language, including lexical bundles, as fundamental units of discourse.

The concept of formulaic language occupies a central position in contemporary linguistic theory, particularly in studies of second and foreign language learning and discourse analysis (Siyanova-Chanturia & Pellicer-Sánchez, 2019). Formulaic sequences are broadly defined as multi-word units that are stored and retrieved as single chunks rather than generated through rule-based processes (Wray, 2002). Within this broader category, lexical bundles represent a frequency-based approach through which recurrent word sequences in corpora are identified (Biber & Barbieri, 2007; Siyanova-Chanturia, 2015).

Unlike idioms, according to Biber et al. (2004), lexical bundles are not necessarily semantically opaque; instead, their significance lies in their functional contribution to discourse. From the Corpus-Based perspective, lexical bundles can be understood as manifestations of the idiom principle, reflecting the tendency of language users to rely on conventionalized patterns.

Moreover, according to Biber et al. (1999) and Hyland (2008), the structural and functional classification of lexical bundles further supports their role as key discourse-building units. Structurally, they are often incomplete fragments (e.g., noun phrase or prepositional phrase fragments), while functionally, they serve to organize text, express stance, and facilitate interaction. Their frequent occurrence in academic writing underscores their importance in constructing coherent and persuasive arguments.

On the other hand, Gries's (2008) approach to phraseology is within linguistics and in the methodological paradigm of corpus linguistics. Unlike traditional phraseology, which often treats idioms and fixed expressions as static lexical units, Gries argues that phraseological patterns are emergent, probabilistic, and usage-driven. His work integrates corpus linguistics, cognitive linguistics, and statistics to explain how multiword expressions function in real language use. Gries's phraseology advocates several key principles. According to this scholar, language is not simply a combination of individual words assembled according to grammatical rules. Instead, speakers rely heavily on recurrent multiword units that become entrenched through repeated use. Gries's phraseological framework is highly compatible with research on lexical bundles. It explains that lexical bundles emerge from repeated language use, become conventionalised through frequency, are psychologically entrenched, fulfil recurring discourse functions, and are identified objectively through corpus analysis.

Thus, Gries's approach complements Biber et al.'s corpus-based theory of lexical bundles. Biber focuses on identifying and classifying recurrent sequences by structure and function, whereas Gries provides a usage-based and statistical explanation for the reason why such sequences arise and become established.

## 5. RESULTS AND DISCUSSIONS

This section presents the findings of the corpus-based analysis of lexical bundles in the Togolese learner environment. The results are presented according to the overall frequency and distribution of lexical bundles and to Biber et al.'s (2004) structural patterns and functional classification.

A total of 9635 lexical bundles (3, 4 and 5-word sequences) met the frequency and range criteria. Of these, 6347 were 3-word bundles, 1757 were 4-word bundles and 813 were 5-word bundles. For the purpose of this study, only the 1,757 four-word bundles were used for further analysis as indicated in the methodology. Therefore, the percentages below are calculated primarily from the 1,757 tokens represented by the top 36 bundles. The corpus size of 708,843 words is used to calculate normalized frequencies per million words.

**Table 2. Top 36 Lexical Bundles in the Learner Corpus**

| Rank | Lexical bundle                   | Frequency | Structural classification | Functional classification |
|------|----------------------------------|-----------|---------------------------|---------------------------|
| 1    | <i>as a result of</i>            | 113       | PP-based                  | Discourse-organizing      |
| 2    | <i>in the present study</i>      | 106       | PP-based                  | Referential               |
| 3    | <i>in the selected texts</i>     | 101       | PP-based                  | Referential               |
| 4    | <i>on the other hand</i>         | 99        | PP-based                  | Discourse-organizing      |
| 5    | <i>to achieve this objective</i> | 96        | VP-based                  | Discourse-organizing      |
| 6    | <i>in the context of</i>         | 93        | PP-based                  | Referential               |
| 7    | <i>to shed light on</i>          | 90        | VP-based                  | Discourse-organizing      |
| 8    | <i>in the use of</i>             | 88        | PP-based                  | Referential               |
| 9    | <i>the aim of this</i>           | 80        | NP-based                  | Referential               |
| 10   | <i>a new form of</i>             | 77        | NP-based                  | Referential               |
| 11   | <i>one of the most</i>           | 76        | NP-based                  | Referential               |
| 12   | <i>in the light of</i>           | 70        | PP-based                  | Discourse-organizing      |
| 13   | <i>it is important to</i>        | 64        | Clausal                   | Stance                    |
| 14   | <i>in the same vein</i>          | 63        | PP-based                  | Discourse-organizing      |
| 15   | <i>in the world today</i>        | 58        | PP-based                  | Referential               |
| 16   | <i>it has shown that</i>         | 51        | Clausal                   | Stance                    |
| 17   | <i>in the same perspective</i>   | 49        | PP-based                  | Discourse-organizing      |
| 18   | <i>the problem of the</i>        | 42        | NP-based                  | Referential               |
| 19   | <i>we can see that</i>           | 38        | Clausal                   | Stance                    |
| 20   | <i>it is obvious that</i>        | 35        | Clausal                   | Stance                    |
| 21   | <i>in an attempt to</i>          | 31        | PP-based                  | Discourse-organizing      |
| 22   | <i>the different effects of</i>  | 28        | NP-based                  | Referential               |

| Rank | Lexical bundle                   | Frequency | Structural classification | Functional classification |
|------|----------------------------------|-----------|---------------------------|---------------------------|
| 23   | <i>it is clear that</i>          | 26        | Clausal                   | Stance                    |
| 24   | <i>based on the novels</i>       | 25        | PP-based                  | Referential               |
| 25   | <i>this work calls for</i>       | 23        | VP-based                  | Stance                    |
| 26   | <i>in the character of</i>       | 20        | PP-based                  | Referential               |
| 27   | <i>it is difficult to</i>        | 18        | Clausal                   | Stance                    |
| 28   | <i>of the data from</i>          | 17        | PP-based                  | Referential               |
| 29   | <i>it is necessary to</i>        | 15        | Clausal                   | Stance                    |
| 30   | <i>through the lenses of</i>     | 14        | PP-based                  | Discourse-organizing      |
| 31   | <i>as a matter of</i>            | 13        | PP-based                  | Discourse-organizing      |
| 32   | <i>the use of translation</i>    | 11        | NP-based                  | Referential               |
| 33   | <i>in the educational system</i> | 9         | PP-based                  | Referential               |
| 34   | <i>in this digital age</i>       | 7         | PP-based                  | Referential               |
| 35   | <i>the lack of technology</i>    | 6         | NP-based                  | Referential               |
| 36   | <i>there is need to</i>          | 5         | Clausal                   | Stance                    |

Table 2 has presented the overall frequency and distribution of lexical bundles and the raw structural patterns and functional classification. The most frequent bundle is *as a result of* (113 occurrences), followed by *in the present study* (106), *in the selected texts* (101), and *on the other hand* (99). This distribution suggests that learners have developed a repertoire of recurrent multi-word sequences that help them establish logical relationships, delimit the scope of their research, contrast ideas, and structure academic arguments. The following sections distinguish structural patterns from functional classification. The analysis of the Top 36 four-word lexical bundles in the corpus reveals a strong reliance on recurrent phraseological sequences for organizing information, establishing relationships between propositions, referring to the research process, and expressing authorial evaluation. This finding is consistent with the established view that lexical bundles are an important component of academic discourse because they contribute to fluency, coherence, and the construction of disciplinary meanings (Biber et al., 2004; Hyland, 2008).

### 5.1. Structural Classification

Biber et al. (2004) broadly distinguish structural categories such as noun-phrase-based, prepositional-phrase-based, verb-phrase-based, and clausal bundles. According to the results, the collected data are classified according to the following distribution.

**Table 3. Structural classification of the four-word lexical bundles**

| Structural category             | Frequency | Percentage % | Normalized per million words |
|---------------------------------|-----------|--------------|------------------------------|
| Prepositional-phrase-based (PP) | 976       | 55.55        | 1,376.89                     |
| Noun-phrase-based (NP)          | 320       | 18.21        | 451.44                       |
| Clausal                         | 252       | 14.34        | 355.51                       |
| Verb-phrase-based (VP)          | 209       | 11.90        | 294.85                       |
| Total                           | 1,757     | 100.00       | 2,478.69                     |

Assuming an equal expected distribution across the four structural categories, the total normalized frequency of the 36 bundles is approximately 2,478.69 occurrences per million words. The Pearson chi-square goodness-of-fit test gives  $\chi^2(3) = 888.785$ ,  $p < .001$ . The results in Table 3 indicate a statistically significant difference in the distribution of structural categories.

The most striking finding is the overwhelming dominance of prepositional-phrase-based bundles, which account for 55.55% of all bundles in the present dataset. This category includes expressions such as: *as a result of*, *in the present study*, *in the selected texts*, *on the other hand*, *in the context of*, *in the use of*, *in the light of*, *in the same vein*, *in the world today*, *in an attempt to*, *based on the novels*, *in the character of*, *of the data from*, *through the lenses of*, *in the educational system*, *in this digital age*.

The strong presence of PP-based bundles is consistent with the rhetorical demands of academic writing. Such bundles allow writers to establish relationships, locate information, frame arguments, identify contexts, express logical relations, and connect propositions. NP-based bundles constitute the second-largest structural category (18.21%). They include *the aim of this*, *a new form of*, *one of the most*, *the problem of*, *the different effects of*, *the use of translation*, and *the lack of technology*. These bundles help writers package concepts and information through nominal structures.

Clausal bundles account for 14.34%, including *it is important to*, *it has shown that*, *we can see that*, *it is obvious that*, *it is clear that*, *it is difficult to*, *it is necessary to*, and *there is need to*. These constructions are particularly important for expressing evaluation, evidentiality, certainty, difficulty, necessity, and authorial positioning. Another substantial structural pattern consists of the anticipatory *it* constructions. These structures are useful to express the evaluation and stance without placing the writer explicitly at the centre of the proposition.

VP-based bundles are least frequent (11.90%), represented principally by expressions such as *to achieve this objective*, *to shed light on*, and *this work calls for*. Their presence nonetheless indicates the use of recurrent verb constructions to express research purposes, analytical objectives, and recommendations.

One of the key features of the present corpus is the high frequency of discourse-organizing bundles. These expressions enable learners to signal causality, contrast, purpose, interpretation, and logical progression. This finding corroborates Biber et al.'s (2004) functional taxonomy according to which lexical bundles serve as discourse organizers when they explicitly connect a current proposition to preceding or subsequent discourse. The frequent occurrence of such bundles infers that the students have acquired a repertoire of formulaic devices that are relevant to make argumentative relations explicit.

The finding is also compatible with Hyland's (2008) observation that lexical bundles are central to the creation of academic discourse and contribute to textual coherence. In his analysis of research articles, doctoral dissertations, and master's theses, Hyland demonstrated that four-word bundles perform important rhetorical functions and vary according to disciplinary and genre conventions.

However, the present findings indicate a particularly strong dependence on explicit textual organization. Expressions such as *on the other hand*, *in the same vein*, and *in the same perspective* repeatedly tell the reader how one proposition should be related to another. Karaata et al. (2012) report that non-native student writers may overuse a limited set of discourse markers that they know well, partly because of their restricted knowledge of the wider repertoire of discourse markers. This may reflect a characteristic feature of learner academic writing: learners may rely on overt discourse markers to compensate for difficulties in constructing implicit textual relationships.

### **5.2.Functional Classification**

Biber et al. (2004) distinguish three major functional categories: Referential bundles, Stance bundles, and Discourse-organizing bundles. The classification of the 36 bundles gives the following distribution.

**Table 4. Functional classification of the four-word lexical bundles**

| Functional category  | Frequency | Percentage (%) | Normalized per million words |
|----------------------|-----------|----------------|------------------------------|
| Referential          | 844       | 48.04          | 1,190.67                     |
| Discourse-organizing | 638       | 36.31          | 900.06                       |
| Stance               | 275       | 15.65          | 387.96                       |
| Total                | 1,757     | 100.00         | 2,478.69                     |

The results in Table 4 indicate that the three functional categories are not equally distributed in the corpus. According to the results, referential bundles represent 48.04% of the data. This category is the most represented and consists of bundles that identify research objects, entities,

concepts, or circumstances. They include *in the present study* (106), *in the selected texts* (101), *in the context of* (93), *in the use of* (88), *the aim of this* (80), *a new form of* (77), *one of the most* (76), *in the world today* (58), *the problem of* (42), *the different effects of* (28), *based on the novels* (25), *in the character of* (20), *of the data from* (17), *the use of translation* (11), *in the educational system* (9), *in this digital age* (7), and *the lack of technology* (6).

The high frequency of *in the present study* and *in the selected texts* is particularly revealing. While the first bundle may be associated with general research-oriented discourse, the latter may exclusively be associated with literature research-oriented discourse. These bundles enable writers to delimit their analytical domain and explicitly connect claims to the study that is being conducted. Likewise, *in the context of* provides contextual framing, while *based on the novels* identifies the evidential or textual basis of an argument.

The occurrence of *the use of translation*, *in the educational system*, *in this digital age*, and *the lack of technology* indicates that the students use lexical bundles to construct discipline-specific and topic-specific meanings. Such bundles therefore contribute to the development of an academic register by repeatedly linking general grammatical structures to the content of the research. Also, the prominence of *in the present study* is particularly significant. It explicitly situates claims within the research being undertaken and therefore performs a strong research-oriented referential function. Similarly, *in the selected texts* identifies the textual material under investigation, while *in the context of* provides contextual framing.

This pattern corresponds with previous research showing that lexical bundles are closely connected to the communicative purposes of academic genres. Cortes (2004), for example, found that bundles differed according to disciplinary writing and the level of academic expertise, demonstrating that recurrent sequences help writers construct meanings appropriate to particular academic communities.

The present findings therefore reinforce the argument that lexical bundles should not be treated merely as frequent strings of words (Biber et al., 2004; Cortes, 2004). Their distribution reflects the rhetorical needs of academic writing: learners repeatedly use bundles to identify their research setting, describe their materials, establish research objectives, and frame their discussion. These bundles reflect the analytical practices associated with textual and literary interpretation rather than the experimental procedures commonly found in scientific disciplines. Consequently, some of the differences between the present findings and those reported in other studies may not necessarily reflect differences in proficiency. They may instead result from disciplinary, genre, or corpus-composition differences. Previous research confirms that informational and textual bundles can be strongly topic- and discipline-specific. Illustratively, Cortes (2004) compared lexical bundles in published academic writing in history and biology. The researcher found substantial differences in the bundles used by the two disciplines, both structurally and functionally.

Discourse-organizing bundles constitute a particularly prominent group in the corpus. The bundles in this category include *as a result of* (113), *on the other hand* (99), *to achieve this objective* (96), *to shed light on* (90), *in the light of* (70), *in the same vein* (63), *in the same perspective* (49), *in an attempt to* (31), *as a matter of fact* (13), and *there is need to* (5). These bundles perform important rhetorical functions. They helped Master students establish cause and effect relationships, contrast, addition, purpose, reformulation, and progression of ideas.

The very high frequency of *as a result of* indicates that students frequently mark causal relationships. Similarly, *on the other hand* is used to introduce contrastive information, while *to achieve this objective* and *in an attempt to* establish relationships between research purposes and actions. The recurrent use of such expressions indicates that students are aware of the importance of explicit discourse that is used in academic writing.

However, the frequent reliance on highly formulaic expressions such as *in the same vein*, *in the same perspective*, and *in the light of* may also indicate a tendency toward repetitive academic phrasing. This could reflect the influence of instructional materials, models of academic writing, or formulaic language acquired by learners.

As far as stance expressions, the corpus also contains several bundles that express the students' evaluation, certainty, importance, difficulty, or necessity. They include *it is important to* (64), *it has shown that* (51), *we can see that* (38), *it is obvious that* (35), *it is clear that* (26), *it is difficult to* (18), *it is necessary to* (15), and *this work calls for* (23). These expressions allowed Master students to position themselves toward propositions so as to guide readers toward particular interpretations. For example, *it is important to* expresses emphasis or evaluation, while *it is obvious that* and *it is clear that* communicate a high degree of certainty. *It is difficult to* introduces an evaluation of complexity, whereas *it is necessary to* expresses obligation or necessity. This finding is interesting in light of previous research on learner academic writing. Ädel and Erman (2012), comparing advanced non-native and native student writing, found that native writers employed a larger and more varied repertoire of lexical bundles, including more hedging and other stance-related sequences. The non-native writers, by contrast, displayed a more restricted repertoire.

The use of *we can see that* is particularly interesting because it explicitly constructs a relationship between writer and reader. Rather than simply presenting a proposition, the writer invites the reader to share an interpretation. Similarly, *this work calls for* enables the writer to formulate recommendations or implications arising from the research.

Nevertheless, some of these bundles may reveal a tendency toward strong or categorical authorial positioning. Expressions such as *it is obvious that* and *it is clear that* can sound overly assertive in academic writing when the evidence does not sufficiently justify such certainty. More cautious alternatives such as *the findings suggest that* or *the evidence indicates that* may sometimes be preferable.

From the above results, a Pearson chi-square goodness-of-fit test confirmed that the structural categories were significantly unevenly distributed,  $\chi^2(3) = 888.79$ ,  $p < .001$ . Functionally, referential bundles constituted the largest category, with 844 occurrences (48.04%), followed by discourse-organizing bundles (638; 36.31%) and stance bundles (275; 15.65%). This functional distribution was also statistically significant,  $\chi^2(2) = 283.42$ ,  $p < .001$ . The predominance of prepositional and referential bundles reveals that the texts rely heavily on recurrent linguistic sequences to specify research contexts, identify entities, and package information, while the substantial proportion of discourse-organizing bundles demonstrates the importance of lexical bundles in establishing logical relationships and guiding readers through academic arguments. Stance bundles, although less frequent, reveal the MA students' use of recurrent expressions to express evaluation, certainty, necessity, and authorial positioning.

The findings also support previous research concerning the restricted phraseological repertoire of non-native academic writers. Ädel and Erman (2012) found that advanced non-native university writers used fewer types of lexical bundles and a less varied repertoire than native speakers. Their conclusion is especially relevant to the present study because the corpus similarly shows repeated reliance on a relatively small number of highly conventional expressions. For example, the repeated use of expressions such as *in the light of*, *in the context of*, *in the present study*, *it is clear that*, and *one of the most* suggests that MA students have developed a set of preferred academic formulas. Such formulas can be beneficial because they provide ready-made linguistic resources for constructing academic prose. At the same time, excessive dependence on them can result in repetition and reduced phraseological diversity.

## 6. LIMITATIONS OF THE STUDY

Despite its contributions to understanding lexical bundle use in Togolese university students' academic writing, this study has several limitations that should be acknowledged when interpreting its findings. First, the corpus was drawn from a single university in Togo, which limits the generalisability of the results to students from other higher education institutions with potentially different academic cultures, instructional practices, and writing conventions. Second, the inclusion of African Literature, American Literature, and Applied Linguistics, only three academic fields from the same domain, restricts the extent to which the identified patterns can be representative of academic writing across a broader range of disciplines, particularly given that lexical bundle use may vary according to disciplinary conventions and communicative purposes. Third, the analysis focused on the introduction and the conclusion of dissertations rather than complete dissertations. The selection of these sections is a limitation to the generalisation of the results. Indeed, since lexical bundles may perform different discourse functions across sections such as introductions, literature reviews, methodology, results, and discussions, this sampling decision may have influenced the distribution and functional classification of the bundles that have been identified. Finally, the study did not report a comparison with a native-speaker reference corpus such as the British Academic Written English (BAWE) Corpus and the Michigan Corpus of Upper-Level Student Papers (MICUSP). Future research could address these limitations by developing larger, multi-institutional and multidisciplinary corpora, including a broader range of dissertation sections and academic genres, and incorporating comparable reference corpora of expert or native English academic writing. Such methodological extensions would provide a more comprehensive basis for investigating the influence of disciplinary conventions, academic experience, and first and mother language influence on lexical bundle use in EFL academic writing.

## 7. CONCLUSION

This study has examined the use of lexical bundles in African learner English through a corpus-based analysis of Master of Arts and language students' academic writing in Togo. Through Biber et al.'s Corpus-Based approach to lexical bundles and Gries's Phraseology approach, the study has analysed students' dissertations in African Literature, American Literature, and Applied Linguistics. For the purpose of the analysis, the study has only considered the introductions and the conclusions of the 225 dissertations. The study has shown that the Togolese English student writing is characterized by an overwhelming dominance of prepositional-phrase-based bundles in the structural category and the favour for the referential bundles in the functional classification. Hence, bundles such as *in the present study*, *as a result*

*of, in the selected texts, on the other hand, in the context of, in the use of, in the light of, in the same vein* are highly represented structural bundles. As far as functional bundles are concerned, the study has found that in the Togolese context, referential bundles constituted the largest category followed by discourse-organizing bundles and stance bundles in student academic writing. These patterns collectively indicate that learners rely on certain formulaic devices for textual organization but they have not yet fully developed the rhetorical resources that are important for stance-taking and disciplinary engagement. These findings have important implications for the teaching of academic writing. English language teachers should expose learners to alternative expressions so that they do not repeatedly rely on the same formulaic expressions. Teachers could introduce discipline-relevant stance bundles, such as *the results suggest that, the findings indicate that, and this study argues that*, to help students express interpretations, qualify claims, evaluate evidence, and establish an appropriate academic voice. These expressions should be taught in context through the analysis of authentic academic texts, guided noticing activities, and writing tasks that require students to justify their linguistic choices. Furthermore, the tendency to rely on familiar bundles can be addressed through corpus-based awareness-raising activities that use AntConc. Students could generate frequency lists and concordance lines from their own writing, identify recurrent bundles, and examine whether repetition reflects a legitimate rhetorical need or a restricted phraseological repertoire. They could subsequently compare their usage with that found in discipline- and genre-appropriate reference texts and revise their drafts to improve lexical variety, precision, and rhetorical effectiveness. Such activities would enable learners to develop a more flexible command of lexical bundles while recognising that variation should be guided by meaning, disciplinary conventions, and communicative purpose rather than by the avoidance of repetition at all costs. From this study, future research may further explore disciplinary variation and instructional interventions aimed at improving learners' mastery of formulaic academic discourse.

## REFERENCES

- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81–92. <https://doi.org/10.1016/j.esp.2011.08.004>.
- Al-Khasawneh, F. M. (2024). A corpus-based analysis of the lexical bundles use in academic discourse. *Explorations in English Language and Linguistics*, 2.2: 178-203
- Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26, 263–286.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D. (1996). Investigating language use through corpus-based analyses of association patterns. *International Journal of Corpus Linguistics*, Vol 1(2), pp. 171-197.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14:3, 275–311. doi 10.1075/ijcl.14.3.08bib.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405. <https://doi.org/10.1093/applin/25.3.371>.
- Biber, D., Conrad, S., & Reppen, R. (1998/2006). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.

- Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (2021). *Grammar of Spoken and Written English*. John Benjamins Publishing Company.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
- 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). <https://doi.org/10.26858/ijole.v5i1.14917>.
- Boumediene, H. (2025). Enhancing ESP Language Learning With Corpus Tools And Data-Driven Approaches. *Journal of Studies in Language, Culture, and Society (JSLCS)*. (08) 03, 2025. (PP. 37-48)
- Brooke, J., Hammond, A., Jacob, D., Tsang, V., Hirst, G., & Shein, F. (2015). Building a lexicon of formulaic language for language learners. In *Proceedings of the 11th Workshop on Multiword Expressions* (pp. 96–104). Association for Computational Linguistics. <https://doi.org/10.3115/v1/W15-0915>
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30–49.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397–423. <https://doi.org/10.1016/j.esp.2003.12.001>
- Dahunsi, T. N. & Ewata, T. O. (2022). An exploration of the structural and colligational characteristics of lexical bundles in L1–L2 corpora for English language teaching. *Language Teaching Research*, pp. 1–17. <https://doi.org/10.1177/136216882110665>.
- De Cock, S. (2000). Repetitive phrasal chunkiness and advanced EFL speech and writing. In C. Mair & M. Hundt (Eds.), *Corpus linguistics and linguistic theory* (pp. 51–68). Rodopi. DOI: 10.1093/elt/ccv045.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Essizewa, K. E. (2012). Diglossia/Triglossia and bilingualism attitudes of Kabiye speakers in Togo. *Geste et Voix*, No 16, pp. 33-47.
- Estaji, M., & Hashemi, M. (2022). Phraseological competence in IELTS academic writing task 2: phraseological units and test-takers' perceptions and use. *Language Testing in Asia*, pp. 2-25. <https://doi.org/10.1186/s40468-022-00180-7>.
- Gawa, D. (2016). *A Comparative Phonological and Morphological Analysis of Two Gur Languages: Moba and Nawdm*. Thèse de Doctorat Unique, Lomé, Université de Lomé.
- Granger, S. (2003). The International Corpus of Learner English: A new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly*, 37(3), 538–546. <https://doi.org/10.2307/3588407>.
- Gries, S. Th. (2008). Phraseology and linguistic theory: A brief survey. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective*. John Benjamins.
- Gries, S. Th. (2009). What is Corpus Linguistics? *Language and Linguistics Compass* 3(5):1225-1241. DOI:10.1111/j.1749-818X.2009.00149.x.
- Hong, J.-Y. (2019). Structural use of lexical bundles in the rhetorical moves of L1 and L2 academic writing. *English Teaching*, 74(3), 29–54.
- Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. Longman.

- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Hyland, K. & Jiang, F. K. (2018). Academic lexical bundles How are they changing? *International Journal of Corpus Linguistics*, 23:4, pp. 383–407.
- Jassim, H. J., & Joshi, J. S. (2023). Corpus linguistics: Analyzing language through large-scale textual data. *International Journal of Advanced Research in Science, Communication and Technology*, 3(3).
- Karaata, C., Çepik, Ş., & Çetin, Y. (2012). Enhancing the use of discourse markers in academic writing: The combination of incidental acquisition and explicit instruction. *Electronic Journal of Social Sciences*, 11(40).
- Kondos, S. (2025). Lexical Bundles: A Focused Framework for Enhancing Vocabulary and Syntax in English Composition Courses. *Journal of Education: Language Learning in Education*, Volume 13 – Issue 1, 171-193.
- Laia, Y. E., Widiastuti, D. P., & Christian, S. M. K. S. (2023). Structural and functional analysis of lexical bundles in articles of law and economics found ScienceDirect. In M. Umiyati et al. (Eds.), *Proceedings of the 2nd International Student Conference on Linguistics (ISCL 2022)* (pp. 299–310). Atlantis Press. [https://doi.org/10.2991/978-2-38476-014-5\\_27](https://doi.org/10.2991/978-2-38476-014-5_27)
- Lange, M-F. (1998). *L'école au Togo : Processus de scolarisation et institution de l'école en Afrique*, Paris : KARTHALA.
- Lawrance, B. N. (2000). Most Obedient Servants: The Politics of Language in German Colonial Togo. *Cahiers d'Études africaines*, n° 159, XL-3, 489-524.
- Liou, H.-C., & Chen, W.-F. (2018). Effects of explicit instruction on learning academic formulaic sequences for EFL college learners' writing. *Taiwan Journal of TESOL*, 15(1), 61–100.
- Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31(1), 25-35. DOI: 10.1016/j.esp.2011.07.002
- Long, B., Shao, H., Yao, J., & Zhang, J. (2024). Structural and Functional Usages of Lexical Bundle in Native and Non-Native English Sociology Academic Writings: A Comparative Study Based on MICUSP. *English Language Teaching and Linguistics Studies*, Vol. 6, No. 4, pp. 11-25.
- Lorenzo Salazar, D. J. (2011). *Lexical bundles in scientific English: A corpus-based study of native and non-native writing* (Doctoral dissertation, University of Barcelona).
- Maldja, K. (2018). *Sociolinguistic study of bilingualism in Moba speech community*. Doctoral dissertation, Université de Lomé.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction* (2nd ed.). Edinburgh University Press. <https://doi.org/10.1515/9781474470865>.
- Nasrabad, P., Shirvan, M. E., & Golparvar, S. E. (2020). Exploring lexical bundles in recent published papers in the field of applied linguistics. *Journal of World Languages*, Vol. 6, No. 3, 175–197 <https://doi.org/10.1080/21698252.2020.1797992>.

- Navarro Gil, N., & Martínez Caro, E. (2019). Lexical bundles in learner and expert academic writing. *Bellaterra Journal of Teaching & Learning Language & Literature*, 12(1), 65–90. <https://doi.org/10.5565/rev/jtl3.794>.
- Neely, E. & Cortes, V. (2009). A little bit about: analyzing and teaching lexical bundles in academic lectures. *Language Value*, Volume 1, Number 1 pp. 17-38.
- Nguyen, N. X. (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.
- Ourso, M. A. (1989). *Lama Phonology and Morphology*, Ph.D. dissertation, University of Illinois at Urbana-Champaign.
- Özbay, A. Ş., & Gürsoy, Z. (2023). Computerized Corpus as a Tool for Educational Technology and Learning in the Analysis of Four-Word Recurrent Expressions. *Journal of Educational Technology & Online Learning*, 6(1), 249-272.
- 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: Language Tests for Academic Enrolment and the CEFR*, Volume 15, Issue 1, pp 29-43 DOI:10.1080/15434303.2017.1405421.
- Pere-Kewezima, K., E. (1994). *Linguistic interference: the case of the interference of Kabiye into English*. Mémoire de Maîtrise, Lomé : Université de Lomé.
- Polić, T. (2020). Teaching Multi-Word Lexical Units in English for Specific Purposes. *International Journal on Studies in English Language and Literature (IJSELL)*, Volume 8, Issue 6, June 2020, PP 26-37.
- Pollach, I. (2012). Taming Textual Data: The contribution of Corpus Linguistics to Computer-aided Text Analysis. *Organizational Research Methods*, 15(2), 263-287.
- Ranjbar, N., Pazhakh, A., & Gorjian, B. (2012). The Effect of Lexical Bundles on Iranian EFL Learners Linguistic Production Fluency. *International Education Studies*; Vol. 5, No. 4, 243-251.
- Richards, J.C., & Schmidt, R. (2010). *Longman Dictionary of Language Teaching and Applied Linguistics* Fourth edition. Pearson.
- Sánchez Hernández, P. (2020). Lexical bundles and moves in research article abstracts of different disciplines. *Odisea*, 21, 181–198. <https://doi.org/10.25115/odisea.v0i21.4130>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Siu, B. W. Y., Afzaal, M., Aldayel, H. S., & Curle, S. (2024). Unlocking the mysteries of academic writing: A corpus-based analysis of lexical bundles in L2 English for engineering students. *SAGE Open*, 14(4), 1–12. <https://doi.org/10.1177/21582440241299997>
- Siyanova-Chanturia, A. (2015). On the 'holistic' nature of formulaic language. *Corpus Linguistics and Linguistic Theory*, 11(2), 285–301. <https://doi.org/10.1515/cllt-2014-0016>.
- Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2019). Formulaic language: Setting the scene. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 1–15). Routledge. <https://doi.org/10.4324/9781315206615-1>.
- Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology* (Textbooks in Language Sciences 7). Berlin: Language Science Press.

**Lexical Bundles in English Learners' Production**

- Trklja, A., & Grabowski, Ł. (eds.). (2021). *Formulaic language: Theories and methods*. (Phraseology and Multiword Expressions 5). Berlin: Language Science Press. DOI: 10.5281/zenodo.4727623.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Yan, L., & Hong, L. (2025). Lexical Bundles in L1 and L2 English Academic Writing: Convergent and Divergent Usage. *Sage Open*, 1-14. DOI: 10.1177/21582440251333850.
- Zato, S. (2012). *Lexical, syntactic and semantic approach to cloze test in Togolese high schools*. Mémoire de Maîtrise, Lomé : Université de Lomé.