



<https://jrl.ui.ac.ir/?lang=en>

Journal of Research in Linguistics

E-ISSN: 2322-3413

18(4), 11-24

Received: 09.04.2026 Accepted: 26.05.2026

Research Paper

Structural Analysis of Lexical Bundles in Linguistics Research Article Abstracts: A Comparison of International and Local Arabic Journals

Alaa Alnajjar

Ph.D. student of Linguistics, Department of Linguistics, University of Isfahan, Isfahan, Iran
alaa7997@gmail.com

Rezvan Motavallian 

Associate professor, Department of Linguistics, University of Isfahan, Isfahan, Iran
r.motavallian@fhn.ui.ac.ir

Abstract

This study presents a structural analysis of three-word lexical bundles in English research article abstracts from the field of linguistics, comparing abstracts authored by native English-speaking scholars in international journals with those written by Arab scholars in local Arabic journals. Drawing on a specialized corpus of approximately 175000 words compiled from international and local Arabic journals, the study employed a corpus-driven and comparative approach to identify and categorize bundles according to the structural taxonomy of Biber et al. (1999). The findings revealed a macro-level convergence in the dominance of phrasal structures with both groups relying heavily on Noun Phrases (NPs) and Prepositional Phrases (PPs) to convey information density. However, certain micro-level divergences were observed. While the Non-Native Speaker (NNS) group demonstrated a slightly wider inventory of unique bundles, the Native Speaker (NS) corpus exhibited a significantly higher overall frequency and density of bundle usage. Structurally, the NNS group displayed a marked underuse of (DET) + NOUN + VERB + (SCONJ)/(DET) patterns compared to their NS counterparts ($\chi^2 = 7.32$, $p < 0.01$). While the analysis revealed no statistically significant differences between the two groups in the use of passive voice constructions or complex syntactic structures, a qualitative tendency was observed: The NNS group exhibited a somewhat greater reliance on passive voice and anticipatory it patterns. This tendency likely reflected the influence of prescriptive pedagogical norms commonly emphasized in English for Academic Purposes instruction. The study therefore advocates for English for Academic Purposes (EAP) pedagogy that extends beyond structural templates to foreground the acquisition of active voice and stance-taking formulaic sequences—explicitly targeting the underused (DET) + NOUN + VERB + (SCONJ)/(DET) pattern that marks authorial presence in native writing.

Keywords: Lexical Bundles, Structural Analysis, Biber et al.'s Taxonomy, Linguistic Article Abstracts, Native and Non-Native Writers.

1. Introduction

In the landscape of contemporary applied linguistics, the shift from a grammar-centric to a lexis-centric perspective has fundamentally altered our understanding of language fluency and production. Central to this paradigm shift is the concept of lexical bundles. Following Biber et al. (1999), this study adopted a purely corpus-driven, frequency-based definition of lexical bundles: recurrent three-word sequences that occur with statistically significant frequency and distribution across a representative corpus. This operationalization is distinct from the psychologically-oriented construct of formulaic sequences (Wray, 2002) and the broader term 'phraseology' (Granger & Meunier, 2008). Unlike single lexical items, bundles are theorized to be stored and retrieved holistically from memory as prefabricated chunks (Wray & Perkins, 2000), serving as critical resources for ensuring textual coherence and reducing cognitive load during production (Hyland, 2008). Grounded in the "lexical approach", which posits that language consists primarily of "grammaticalized lexis" (Lewis,

*Corresponding author



1993), this study viewed linguistic proficiency as the ability to manipulate these prefabricated chunks. While mastery of these sequences is a hallmark of proficient language command, acquiring these conventions presents a distinct challenge for L2 learners, who often struggle to produce native-like bundle frequencies and distributions (De Cock, 2002; Granger, 1998).

Methodologically, this inquiry draws upon the empirical rigor of corpus linguistics to facilitate the quantitative analysis of real-world language usage (Biber et al., 1999; Sinclair, 1991). While previous research has extensively documented bundle use across various disciplines, a focused structural analysis of three-word bundles in linguistics abstracts—specifically contrasting international English-medium journals with local Arabic journals publishing in English—remains under-explored. The research abstract serves as a "stand-alone" representation of a study (Swales, 1990) that demands specific rhetorical moves often realized through fixed expressions. To address this gap, the present study utilized a specialized corpus of linguistics abstracts to contrast the occurrence rates and grammatical constructions of the most frequent three-word bundles across Native Speaker (NS) and Non-Native Speaker (NNS) Arab scholars. By employing a descriptive and comparative framework, this research aimed to uncover potential divergences in abstract phraseology and identify the distinguishing structural characteristics of lexical bundles employed in these two distinct journal settings. Specifically, this study addressed the following research questions:

- (a) What are the structural patterns of three-word lexical bundles (Noun Phrase (NP)-based, Prepositional Phrase (PP)-based, Verb Phrase (VP)-based, and other) in linguistics abstracts from international versus local Arabic journals?
- (b) How do the two corpora differ in bundle frequency, normalized density, and type inventory?
- (c) Which specific grammatical patterns within NP, PP, and VP bundles distinguish the two writer groups?

2. Structural Taxonomy for Lexical Bundle Analysis

To facilitate a comprehensive analysis of lexical bundles within the selected corpus, this study employed a structural taxonomy that categorized these multi-word units based on their internal grammatical composition. Although lexical bundles rarely constitute complete syntactic clauses in isolation, they exhibit distinct internal grammatical patterns that reflect the underlying syntactic preferences of a discourse community (Biber et al., 1999). Consequently, a structural classification is essential for understanding the "building blocks" of academic writing that authors rely upon to construct complex arguments.

For the purpose of this structural analysis, the study adopted the robust classification system originally devised by Biber et al. (1999) (see Table 1). This analytical framework delineated recurrent patterns into three primary categories: NP-based, PP-based, and VP-based bundles—alongside a fourth residual category for structures that do not fit these patterns. Following Biber et al. (1999), the VP-based category included not only passive and active verbal constructions, but also clausal fragments (e.g., anticipatory *it* + verb/adjective phrase, *that*-clause fragments, and *there is* patterns) because these structures centrally involved verb-governed elements. The residual category, labeled 'Other structures' in Table 1, contains bundles like *as well as* that lack a clear phrasal or clausal verb core and are not classifiable as NP, PP, or VP bundles. By applying this taxonomy, the present study could precisely identify the grammatical formulations favored by NNS scholars compared to their L1 Arabic counterparts, thereby revealing divergences in syntactic fluency and rhetorical style.

Table1- Biber et al.'s (1999) structural model of lexical bundles

Category	Structure	Example
NP-based	Noun phrase + of	the use of
	Noun phrase with other post-modifier	the relationship between
	Other noun phrases	
PP-based	Prepositional phrase + of	in terms of
	Other prepositional phrases	in the present study
VP-based	Passive Voice Structures	can be used, is divided into
	Adverbial clause fragment	as it is
	Be + noun/adjective phrase	is part of
	Anticipatory <i>it</i> + verb/adjective phrase	it is possible
	(Verb phrase) + <i>that</i> -clause fragment	results showed that, we argue that, suggest that
	(Verb/adjective) + <i>to</i> -clause fragment	to examine how
	Pronoun/noun phrase + <i>be</i>	there is, the book is, this study is
Other structures		N/A as well as

3. Literature Review

Scholars have long recognized that language use is fundamentally formulaic, composed of recurrent multi-word expressions that constitute a significant portion of fluent communication. This perspective is rooted in Firth's (1957) seminal assertion that meaning is derived from the "company words keep", a principle that laid the groundwork for modern phraseology. The advent of corpus linguistics provided the empirical tools necessary to illuminate these patterns through large-scale reference corpora, facilitating a transition from intuition-based to data-driven analyses. A watershed moment in this field was the study by Biber et al. (1999), who coined the term "lexical bundles" to describe recurrent sequences of three or more words that extend beyond idiomaticity. Since then, investigation of lexical bundles has expanded into a vibrant domain of inquiry, with various frameworks emerging to analyze their structural forms and functional roles across diverse contexts.

Research comparing the structural characteristics of lexical bundles between native and non-native writers reveals significant divergences in syntactic preferences, often attributed to first language (L1) transfer and varying stages of interlanguage development. Foundational studies by Ädel and Erman (2012) and Wei and Lei (2011) highlighted that while non-native writers achieve a high level of lexical proficiency, they often fail to emulate the structural diversity found in native writing, particularly in the deployment of complex noun phrases and prepositional bundles. In the context of Persian L1 writers, Rahimi Azad and Modarres Khiabani (2018) found a distinct preference for PP-based structures and underutilization of VP-based bundles. Similarly, Pan, Reppen, and Biber (2016) observed that L2 writers tend to rely more heavily on clausal structures compared to the phrasal structures preferred by L1 writers, a shift interpreted as a compensatory strategy that often results in less dense academic prose. However, it is important to note that this pattern is not universal. Pan, Reppen, and Biber (2016) themselves found that in expert L2 writing (published research articles by Chinese scholars), the L2 writers actually used more NP bundles and fewer VP bundles than their L1 counterparts—directly contradicting the claim that NNS writers uniformly underuse NP-based structures. This contradiction suggests that proficiency level (student vs. expert) and genre (thesis vs. published article) may interact with L1 background in complex ways, a nuance that earlier descriptive accounts have often overlooked. Collectively, these studies underscore that non-native academic writing is frequently structurally distinct from native writing characterized by over-reliance on clausal and prepositional forms and underutilization of complex noun-based structures—though the direction and magnitude of these differences remain contested in the literature.

Beyond writer background, disciplinary context significantly influences bundle patterns. Foundational research by Hyland (2008) established that lexical bundle use varies according to epistemological constructs, distinguishing "hard" sciences (e.g., electrical engineering) from "soft" fields (e.g., business studies). Hyland observed that hard disciplines favor research-oriented and passive structures, whereas soft disciplines exhibit a higher frequency of text-oriented and stance bundles. This dichotomy is further supported by Cortes (2004) and Durrant (2017), the latter concluding that the vast majority of bundles are discipline-specific. Focusing specifically on applied linguistics, Nasrabad et al. (2020) provided a comprehensive baseline by identifying over 1,000 unique bundles, allowing for precise intra-disciplinary comparison. Research into register variation has also demonstrated that lexical bundle usage is systematically shaped by communicative purpose, with abstracts presenting a unique site for phraseological investigation due to their compressed, conventionalized structure (Swales & Feak, 2009). Studies like Qin (2014) and Bao and Liu (2022) have established that abstracts exhibit distinctive phraseological patterns closely associated with specific rhetorical moves. In summary, the literature confirms that lexical bundle use is highly sensitive to the interplay of linguistic proficiency, disciplinary norms, and generic constraints. While general patterns of L2 divergence and disciplinary variation are well-documented, the specific phraseological practices of Arab scholars writing in English—particularly within the genre of applied linguistics abstracts—remain insufficiently explored. The present study built upon this foundation by employing a comparative framework to address the under-explored phraseological practices of Arab scholars in applied linguistics research article abstracts—specifically, comparing their structural use of lexical bundles with that of NS scholars in the same genre.

4. Methodology

4.1 Research Design

The research employed a comparative corpus design with a corpus-driven approach to systematically examine phraseological similarities and differences. This allowed for robust quantitative comparison of bundle frequencies and structural patterns alongside qualitative interpretation. By ensuring analytical categories emerged inductively from the data, the study contrasted structural features in abstracts from international versus local Arabic journals, grounding findings in actual language usage to strengthen validity.

4.2 Corpus Design and Compilation

This section provides a transparent account of the corpus selection, processing, and preparation procedures. Following best practices in corpus design (Kennedy, 2014; Sinclair, 2005), each step was guided by the principles of representativeness, balance, and methodological rigor. To address the research questions, a specialized, purpose-built comparative corpus was compiled. Rather than being artificially constructed, the corpus was selected from published, peer-reviewed journals to ensure disciplinary authenticity and ecological validity.

The corpus comprised research article abstracts from the field of linguistics, which were categorized into two distinct



subcorpora: the NS subcorpus consisting of abstracts authored by native English scholars and the NNS subcorpus comprising abstracts written by Arab scholars publishing in English. Texts were collected from established journals indexed in international databases, specifically targeting articles published between 2015 and 2023 to ensure contemporary relevance.

The selection criteria were strictly defined to include only peer-reviewed journals in linguistics and applied linguistics where full abstracts were available. Representative international journals included in the study were the *Journal of Linguistics*, *Applied Linguistics*, *Linguistics and Education*, *TESOL Quarterly*, and the *Journal of Pragmatics*. The data collection process involved compiling a list of these high-impact international journals alongside reputable local journals from Arab institutions. International journals were selected from major publishers, such as Elsevier, Springer, and Wiley, while local journals were chosen based on their affiliation with reputable Arab universities and research centers (e.g., *Journal of the College of Languages*).

Operational criteria for assigning abstracts to the NS and NNS/Arab scholar subcorpora were as follows: Authors were assigned to the NS subcorpus if the authors were affiliated with a university in a predominantly English-speaking country (UK, USA, Canada) at the time of publication and bore an English name verified through institutional profiles or publication history. The NNS/Arab scholar subcorpus required the author affiliation with a university in an Arab country (e.g., Iraq, Saudi Arabia, Egypt, Jordan), an Arabic name verified via institutional profiles, and institutional or biographical confirmation of Arabic as L1. To avoid cross-group contamination, ambiguous cases (e.g., Arab name with US affiliation, English name with Arab university, or any indication of bilingual/bicultural identity) were also excluded. The finalized corpus was substantial in size, totaling approximately 175,000 words. Each of the NS and NNS subcorpus contained approximately 87,500 words.

4.3 Analytical Framework

4.3.1 Structural Taxonomy of Lexical Bundles

The study adopted Biber et al.'s (1999) structural taxonomy from the Longman Grammar of Spoken and Written English as the primary analytical framework (shown in Table 1).

4.3.2 Rationale for Focusing on Research Article Abstracts

The selection of abstracts was done by four factors. First, their generic stability and predictable rhetorical structures (Swales, 1990) allowed for confident attribution of differences to author-related factors rather than generic variation (Hyland, 2000). Second, the rhetorical density within strict word limits necessitated using lexical bundles as efficient shorthand to accomplish discourse tasks (Hyland & Tse, 2005). Third, as gatekeepers to publication, abstracts reflected writers' best strategic efforts to meet disciplinary expectations (Swales & Feak, 2009). Finally, abstracts offered methodological advantages; their bounded nature facilitated clean data extraction and functional mapping of bundles onto rhetorical moves, while their accessibility enhanced corpus construction and replicability (McEnery & Hardie, 2012).

4.3.3 Operational Definition of Lexical Bundles

This study focused exclusively on three-word lexical bundles to ensure the analysis remained manageable and allowed for direct comparison with recent work in the field (Bao & Liu, 2022). Furthermore, three-word combinations were of particular significance as Simpson-Vlach and Ellis (2010, p. 509) note that many important recurrent word combinations occur as three-word bundles.

Following established practices in lexical bundle research, the frequency threshold for identification was calibrated to the specific characteristics of the corpus. With each subcorpus containing approximately 87,500 words, a raw frequency threshold of 6 occurrences was adopted. This approach is consistent with the principle articulated by Biber et al. (1999) that frequency thresholds should be "somewhat arbitrary but consistently applied" (p. 992).

The frequency threshold was applied in conjunction with a dispersion criterion requiring occurrence in at least 5% of texts within each subcorpus. This followed Pan's (2020) finding that frequency and dispersion thresholds interact effectively in filtering bundles. This dual-criteria approach ensured that the identified bundles were both recurrent and representative across the corpus, addressing the concern raised by Cortes (2004) that "pure frequency thresholds may include sequences that are not psychologically real for language users" (p. 401).

4.4 Analytical Procedures

To ensure the corpus was clean, consistent, and suitable for lexical bundle extraction, a rigorous preprocessing workflow was applied to all abstracts. First, the texts were converted into plain text format. Non-linguistic content was then removed, including author names, institutional affiliations, journal headers and footers, page numbers, references, footnotes, tables, and figure captions. Sentence boundaries were cleaned to support accurate tokenization by separating punctuation from words and non-English characters were removed unless they formed part of meaningful content. Finally, all texts adhered to the same standardized formatting rules to ensure consistency.

Following preprocessing, three-word lexical bundles were extracted using the corpus analysis software AntConc and N-Gram tools while applying the frequency and dispersion thresholds defined in Section 4.3.3. The extraction adhered to

the following parameters: (a) overlapping bundles (e.g., in the field and the field of) were retained following standard practice in lexical bundle research (Biber et al., 1999; Chen & Baker, 2010); (b) bundles crossing sentence boundaries were excluded as sentence-final punctuation would disrupt the cohesion of formulaic sequences; and (c) only three-word sequences were extracted, consistent with the focus of the study as justified in Section 4.3.3.

The identified bundles were subsequently categorized into structural types according to Biber et al.'s (1999) taxonomy. This systematic approach ensured transparency, replicability, and objectivity, thereby enhancing the credibility of the findings and facilitating future comparative research.

5. Results

5.1 Quantitative Overview

In this study, a unique bundle (or bundle type) refers to a distinct three-word sequence (e.g., the role of counted once regardless of frequency), while a bundle occurrence (or token) refers to each individual instance of a bundle in the corpus. As shown in Table 2, the NNS corpus contains a slightly larger inventory of unique bundle types (165 vs. 155) representing 51.6% of the total unique types, whereas the NS corpus exhibits a substantially higher overall frequency of bundle tokens (1,983 vs. 1,653) and a higher average frequency per bundle type (12.8 vs. 10.0).

Table 2- Statistical Overview of 3-Word Bundles in NS and NNS Corpora

Corpus	N-Gram Length	Total Unique Bundle Types	% of Total Unique Types	Total Token (Occurrences)	Avg. Freq	Chi-square Test
Native	3-word	155	48.4%	1,983	12.8	$\chi^2 = 29.95$
Non-native	3-word	165	51.6%	1,653	10.0	$p < 0.001$
Total	3-word	320	100%	3,636	11.4	—

To determine whether the difference in total token frequencies was statistically significant, a chi-square goodness-of-fit test was conducted to compare the observed token frequencies (NS: 1,983; NNS: 1,653) against the expected frequencies under the null hypothesis of no difference (i.e., equal distribution of the 3,636 total tokens: 1,818 expected in each subcorpus). The test did not apply to the number of unique bundle types, which were reported descriptively. The result was highly significant ($\chi^2 = 29.95$, $df = 1$, $p < 0.001$), indicating that the observed divergence in token frequency was unlikely to have occurred by chance. Taken together, these results revealed a fundamental divergence in phraseological strategies: Native writers concentrated their usage on a smaller core set of bundle types with higher repetition, while non-native writers employed a more diversified repertoire but with lower overall token frequency and lower average frequency per bundle type.

The differences in both the distribution of bundle types and their total frequencies were highly statistically significant ($\chi^2 = 29.95$, $df = 1$, $p < 0.001$). This result underscored a fundamental divergence in phraseological strategies between the two groups. Native writers achieved a higher formulaic density—a characteristic strongly associated with textual fluency and idiomaticity—by concentrating their usage on a smaller core set of bundles. Non-native writers, conversely, employed a more diversified repertoire but with a lower overall frequency, suggesting that their formulaic language was less entrenched and more diffusely applied across the text.

5.2 Structural Analysis of Lexical Bundles

5.2.1 Class One: NP Structures

Table 3 presents the distribution of three-word noun phrase (NP) bundles across three major structural subclasses in the NS and NNS corpora. The total frequencies (NS: 879; NNS: 747) and the overall chi-square test ($\chi^2 = 2.85$, $df = 3$, $p = 0.241$) indicated that the two groups did not differ significantly in their overall reliance on NP bundles.

To examine whether the proportional distribution of each NP subclass differed between groups, a series of 2×2 chi-square tests of independence was conducted. For each subclass, the test compared, for that specific subclass, the observed frequency against the observed frequency of all other NP bundles combined across the two groups (NS vs. NNS). Under the null hypothesis of no association between group and subclass membership, the expected frequencies were calculated from the row and column totals. As shown in Table 3, none of the subclass-level tests have reached statistical significance (all $\chi^2 < 3.84$, $df = 1$, $p > 0.05$). A separate 2×4 chi-square test on the full contingency table (groups $\times$ 4 NP subclasses) also confirmed no overall association ($\chi^2 = 2.85$, $df = 3$, $p = 0.241$). Thus, the two groups of writers did not differ significantly in their structural distribution of NP-based lexical bundles.

Table 3- Distribution of NP Bundles in NS and NNS Corpora

Subclass	NS (Freq, %)	NNS (Freq, %)	Test (χ^2 , df, p)	Examples
DET NOUN ADP	560 (63.71%)	453 (60.64%)	1.49, 1, >0.05	the role of, the use of, the development of, the context of, the field of
Adjectival Modification Patterns (DET ADJ NOUN + ADJ NOUN NOUN)	152 (17.29%)	127 (17.00%)	0.01, 1, >0.05	the current study, second language acquisition, the present study, a foreign language, English language teaching
Other Noun-Based Patterns	167 (19.00%)	167 (22.36%)	2.59, 1, >0.05	ways in which, understanding of the, the ability to, research in the, the need to
TOTAL NP	879 (100%)	747 (100%)	2.85, 3, 0.241	

The DET NOUN ADP structure—comprising a determiner, a noun, and a preposition—emerged as the predominant bundle type in both corpora, accounting for 63.71% of NS bundles and 60.64% of NNS bundles. This structure represented the canonical form of nominal post-modification in English academic prose (Biber et al., 1999) and served as the primary mechanism for defining concepts, establishing entities, specifying relationships, and constructing conceptual frameworks through noun phrases with of-complements (Halliday & Martin, 1993). The convergence of NS and NNS writers on this pattern suggested that advanced L2 learners had successfully acquired the most prototypical NP bundle structure of academic English. The absence of a statistically significant difference ($\chi^2 = 1.49$, $df = 1$, $p > 0.05$) supported Sinclair's (1991) "idiom principle", indicating that both groups relied on these prefabricated chunks as core building blocks of academic discourse.

The subclass of Adjectival Modification Patterns—encompassing DET ADJ NOUN (e.g., the current study, the present study) and ADJ NOUN NOUN (e.g., second language acquisition, English language teaching)—showed near-identical usage rates between the groups (NS: 17.29%; NNS: 17.00%; $\chi^2 = 0.01$, $df = 1$, $p > 0.05$). This parity suggested that NNS writers had successfully acquired the specific terminology and collocational patterns required for complex nominal constructions within their field.

The subclass of Other Noun-Based Patterns—which encompassed diverse structures, such as NOUN ADP DET (e.g., use of the), NOUN CCONJ NOUN (e.g., racism and antiblackness), NOUN ADP PRON (e.g., ways in which), DET NOUN PART (e.g., the ability to), and NOUN ADP NOUN (e.g., point of view)—revealed proportional similarity between the groups. NS writers used 19.00% of their NP bundles in this category, while NNS writers utilized 22.36%. The chi-square test showed no statistically significant difference ($\chi^2 = 2.59$, $df = 1$, $p > 0.05$), indicating that both groups drew upon these more diverse structural frames with comparable frequency. The slightly higher proportional use by NNS writers might reflect a greater reliance on explicit framing devices, a pattern that aligned with observations of L2 writers' tendency toward explicitation (Klaudy & Károly, 2005; Pérez-Llantada, 2014).

In summary, the analysis of NP bundles revealed no statistically significant overall difference between NS and NNS writers ($\chi^2 = 2.85$, $df = 3$, $p = 0.241$). This suggested that advanced L2 writers had acquired the core NP-based formulaic repertoire of academic English to a level comparable with native speakers. The subtle divergence in lexical sophistication within adjectival patterns, however, pointed to an area where NNS writers might benefit from targeted instructional attention to expand their repertoire of discipline-specific modifiers and nominal compounds.

5.2.2 Class Two: PP Structures

The analysis of PP bundles in NS and NNS writer corpora revealed a high degree of structural convergence in advanced formulaic competence. Table 4 presents the distribution of PP bundles across three structural categories: Preposition + NP patterns, Preposition + Noun + Preposition patterns, and Other Prepositional patterns. The total PP frequencies (NS: 493; NNS: 412) and the overall chi-square test ($\chi^2 = 0.85$, $df = 2$, $p = 0.653$) indicate no statistically significant difference between the two groups in overall PP bundle deployment. As shown in this table, none of the three subclass-level tests have reached significance. However, non-significance does not demonstrate equivalence or convergence; it only indicates that the present data do not provide sufficient evidence to reject the null hypothesis of no difference. Within the limits of this corpus, the two groups exhibit similar proportional distributions across PP bundle types.



Table 4- Distribution of PP Bundles in NS and NNS Corpora

Subclass	NS (Freq, %)	NNS (Freq, %)	Test (χ^2 , df, p)	Examples
Preposition + Noun Phrase Patterns	301 (61.05%)	259 (62.86%)	0.24, 1, > 0.05	in this article, in applied linguistics, of language learning, in the field, in the context, of the study, in English language
Preposition + Noun + Preposition Patterns	113 (22.92%)	84 (20.39%)	0.70, 1, > 0.05	in terms of, as part of, in addition to, in relation to, in light of, in contrast to, by means of
Other Prepositional Patterns	79 (16.02%)	69 (16.75%)	0.04, 1, > 0.05	in order to, in the global, at the same, due to the, by examining the, in which the, in which they
TOTAL PP	493 (100%)	412 (100%)	0.85, 2, 0.653	

The Preposition + NP patterns constituted the largest category in both corpora, accounting for 61.05% of NS bundles and 62.86% of NNS bundles. This category included bundles, such as *in this article, in applied linguistics, in the field, in the context, of the study, and in English language*. The chi-square test confirmed no statistically significant difference between the groups ($\chi^2 = 0.24$, $df = 1$, $p > 0.05$), indicating that NNS writers had successfully acquired the most prototypical and frequent PP bundle structures of academic English.

The Preposition + Noun + Preposition patterns—sometimes referred to as “bare prepositional” patterns—included bundles, such as *in terms of, as part of, in addition to, in relation to, in light of, in contrast to, and by means of*. These structures served crucial logical and metadiscourse functions in academic prose, enabling writers to frame concepts, establish comparative relationships, enumerate arguments, qualify claims, and specify methodology (Hyland, 2005). NS writers employed these bundles slightly more frequently than NNS writers, both in absolute frequency (113 vs. 84) and proportional representation (22.92% vs. 20.39%). However, the chi-square test revealed no statistically significant difference between the groups ($\chi^2 = 0.70$, $df = 1$, $p > 0.05$). Both groups shared the same inventory of bundles, indicating that NNS writers had acquired the full lexical repertoire. The modest difference in frequency might reflect subtle variations in the density of metadiscourse marking, consistent with research suggesting that logical connectors and metadiscourse devices present persistent, albeit subtle, challenges for advanced learners (Paquot, 2010; Ädel & Erman, 2012).

The category of Other Prepositional Patterns encompassed diverse structural configurations representing less frequent, more specialized PP bundle types. This category included purposive (*in order to*), temporal (*at the same*), causal (*due to the*), methodological (*by examining the*), and relative clause-introducing (*in which the, in which they*) bundles. The proportional distributions were remarkably similar between groups (NS: 16.02%; NNS: 16.75%) and the chi-square test confirmed no significant difference ($\chi^2 = 0.04$, $df = 1$, $p > 0.05$). The comparable usage of syntactically complex patterns, such as *in which the* and *in which they* was particularly noteworthy. These structures required integration of prepositional phrases with relative clause constructions—a configuration often considered challenging for L2 learners.

In summary, the analysis of PP bundles revealed a high degree of structural convergence between NS and NNS writers. No statistically significant differences were observed in any of the three subcategories or in the overall distribution. This finding indicated that advanced NNS academic writers had achieved substantial alignment with native conventions in PP bundle use as in the deployment of metadiscourse and syntactically complex structures. These results contribute to theoretical understandings of formulaic competence in L2 academic writing by demonstrating that high-frequency, prototypical constructions are acquired robustly and that advanced L2 writers can achieve native-like levels of structural diversity even in complex syntactic configurations. The subtle proportional variations—such as the slightly lower use of Preposition + Noun + Preposition patterns by NNS writers—may point to areas where metadiscourse fluency continues to develop, but these differences do not reach statistical significance and should be interpreted with caution.

5.2.3 Class Three: VP Structures

This class encompassed bundles containing verb structures, including active and passive constructions, *that*-clauses, *to*-clauses, and other verb-related fragments. As illustrated in Table 5, VP bundles constitute a substantial component of the formulaic repertoire in both corpora though they are less frequent than NP and PP bundles. The total frequencies (NS: 463; NNS: 361) reflect the importance of verbal formulaic sequences in constructing authorial stance, reporting findings, and structuring argumentation.

Table 5- Distribution of VP Bundles in NS and NNS Corpora

Subclass	NS (Freq, %)	NNS (Freq, %)	Test (χ^2 , df, p)	Examples
1. (DET) + NOUN + VERB + (CONJ)/(DET)	240 (51.84%)	152 (42.11%)	7.32, 1, < 0.01(SIG)	we argue that, this study examines, the book is
2. (PART) + ACTIVE VERB + ...	77 (16.63%)	65 (18.01%)	0.18, 1, > 0.05	contribute to the, to understand the, suggest that the
3. Anticipatory it + verb/adjective phrase	6 (1.30%)	13 (3.60%)	3.82, 1, > 0.05	it is a, it is possible
4. Passive VP	87 (18.79%)	82 (22.71%)	1.68, 1, > 0.05	is known about, can be used, based on the
5. Dependent clause	53 (11.45%)	49 (13.57%)	0.66, 1, > 0.05	that there is, as it is
TOTAL VP	463 (100%)	361 (100%)	9.43, 4, 0.051	

The overall chi-square test comparing the distribution of VP bundles across the five subcategories between the two groups was not statistically significant ($\chi^2 = 9.43$, $df = 4$, $p = 0.051$). Thus, when considering all VP bundle types together, no significant overall difference was detected. Despite the non-significant overall result, subclass-level analyses revealed one notable exception. Class 1 ((DET) + NOUN + VERB + (CONJ)/(DET), e.g., we argue that, this study examines) was the most dominant category for both groups, accounting for 51.84% of NS bundles and 42.11% of NNS bundles. The chi-square test for this subclass was statistically significant ($\chi^2 = 7.32$, $df = 1$, $p < 0.01$). This divergence suggests that although NNS writers had acquired the fundamental structure of reporting and attribution, they had employed these formulaic sequences significantly less frequently than their NS counterparts. NS writers had utilized these structures as a primary vehicle for integrating sources and presenting findings, whereas NNS writers appeared to have employed a more restricted range of these high-frequency stance markers. No other subclasses reached statistical significance (all $\chi^2 < 3.84$, $df = 1$, $p > 0.05$). The non-significant overall result indicated that the difference observed in Class 1 was not large enough to affect the global distribution of VP bundles.

Class 2: (PART) + ACTIVE VERB + ... (e.g., contribute to the, to understand the, suggest that the) showed no statistically significant difference between the groups ($\chi^2 = 0.18$, $p > 0.05$). The proportional usage was nearly balanced (NS: 16.63%; NNS: 18.01%). This parity indicated that NNS writers had successfully acquired the structural patterns for active VPs with particles and infinitives, achieving a level of competence comparable to that of native speakers in this specific area. The similarity might reflect the high frequency and salience of these constructions in academic discourse, facilitating robust acquisition (Ellis, 2002).

Class 3: Anticipatory it + verb/adjective phrase (e.g., it is a, it is possible) showed a notable divergence in proportional terms (NNS: 3.60%; NS: 1.30%) though the chi-square value ($\chi^2 = 3.82$) did not reach statistical significance at the conventional level of $p < 0.05$. The higher frequency in the NNS corpus might reflect the influence of formulaic instruction, where the constructions of it is... and there is... are often taught as standard templates for introducing topics or adopting a stance (Biber et al., 1999).

Class 4: Passive VP (e.g., is known about, can be used, based on the) showed no statistically significant difference between the two groups (NS: 18.79%; NNS: 22.71%; $\chi^2 = 1.68$, $df = 1$, $p > 0.05$). Although the proportional difference (18.79% vs. 22.71%) was not significant and could be due to chance, it suggested a trend that can be explored in future research with larger corpora.

Class 5: Dependent clause—exemplified by that there is (noun clause fragment), as it is (adverbial clause fragment), and if there is (conditional clause fragment)—showed no statistically significant difference between the two groups ($\chi^2 = 0.66$, $df = 1$, $p > 0.05$). The proportional distributions were comparable (NS: 11.45%; NNS: 13.57%). This category included syntactically complex structures that required integration of clause-level elements. The comparable usage suggested that NNS writers had successfully acquired the structural patterns for dependent clause fragments, achieving a level of competence comparable to native speakers in this specific area.

In summary, the analysis of VP bundles revealed both similarities and differences between NS and NNS writers. A statistically significant difference was observed only in Class 1 (stance-taking constructions), underscoring the importance of this category as a site of ongoing development in L2 academic writing. For Classes 2, 4, and 5, no statistically significant differences were detected (all $p > 0.05$), suggesting that the two groups were comparable in their proportional use of these structures. Within the limits of this corpus, the two groups were comparable in their proportional use of active verb constructions, passive voice, and dependent clause fragments. The borderline finding for Class 3 (anticipatory it constructions, $p = 0.051$) suggested a trend that may warrant further investigation with larger samples.

5.2.4 Class 4: Other Expressions

This final class encompassed lexical bundles that did not conform to the standard structural categories of NPs, PPs, or VPs. These items consist of miscellaneous expressions, including fixed lexical collocations and incomplete fragments that defy easy grammatical classification (e.g., as well as, teaching and learning, the most important). The “Other” category accounted for 148 out of 1,983 (7.46%) of NS bundle occurrences and 133 out of 1,653 (8.05%) of NNS bundle occurrences.

To determine whether the proportion of “Other” bundles differed between the two groups, a 2×2 chi-square test of independence was conducted. The test compared the observed frequencies of “Other” bundles (NS: 148; NNS: 133) against the frequencies of all other bundles (NS: $1,983 - 148 = 1,835$; NNS: $1,653 - 133 = 1,520$) across the two subcorpora. The test contrasted the observed frequencies (NS: 148; NNS: 133) against the expected frequencies under the null hypothesis of no association between group and category membership (i.e., the proportion of “Other” bundles was the same in both subcorpora). As shown in Table 6, the result ($\chi^2 = 0.80$, $df = 1$, $p = 0.371$) indicates no statistically significant difference.

Table 6- Distribution of “Other” Lexical Bundles in NS and NNS Corpora

Subclass	NS (Freq, % of total NS)	NNS (Freq, % of total NNS)	Test (χ^2 , df, p)	Examples
Other Expressions	148 (7.46%)	133 (8.05%)	0.80, 1, 0.371	as well as, teaching and learning, the most important, according to the, ability to use, quantitative and qualitative, the most effective

This non-significant result indicates that the present data do not provide sufficient evidence of a difference between the groups in the proportional use of “Other” bundles.

An overview of the distribution across all four categories is presented in Table 7 (Section 5.3).

5.3 Overall Structural Distribution

Table 7 synthesizes the distribution of all identified lexical bundles across the four major structural categories. As shown, the overall chi-square test of independence comparing the full 2×4 contingency table (Groups $\times$ NP, PP, VP, Other) is not statistically significant ($\chi^2 = 1.45$, $df = 3$, $p = 0.694$). This indicates that the macro-level structural profile of lexical bundle use does not differ significantly between NS and NNS scholars.

Table 7- Comprehensive Structural Distribution of Lexical Bundles in NS and NNS Corpora

Major Class Structure	NS Freq	NS (%)	NNS Freq	NNS (%)
Noun Phrase (NP) Bundles	879	44.33%	747	45.19%
Prepositional Phrase (PP) Bundles	493	24.86%	412	24.93%
Verb Phrase (VP) Bundles	463	23.35%	361	21.84%
Other Bundles	148	7.46%	133	8.05%
TOTAL	1,983	100%	1,653	100%

The most salient finding across both corpora was the overwhelming predominance of phrasal structures—specifically NPs and PPs. These two categories accounted for 70.19% of the total NS bundles and 70.12% of the NNS total. The alignment with Biber et al.'s (2004) characterization of academic writing as relying heavily on nominalization confirmed that both groups had successfully internalized the core phrasal conventions of the register.

Although the overall structural distribution did not differ significantly, subclass-level analyses (Section 5.2.3) revealed a significant difference in stance-taking constructions (Class 1), where NS writers used these bundles more frequently. This difference, however, does not alter the global phrasal profile.

6. Discussion

The primary distinguishing structural characteristic of lexical bundles in linguistics abstracts written by Arab scholars (compared to NS scholars) was the statistically significant underuse of stance-taking verb phrase bundles of the type (DET) + NOUN + VERB + (CONJ)/(DET) (e.g., we argue that, this study examines). In contrast, no statistically significant differences were found between the two groups in the overall distribution of NP- or PP-based bundles, nor in the macro-level structural profile ($\chi^2 = 1.45$, $df = 3$, $p = 0.694$).

At the macro level, the two groups showed a similar dominance of phrasal structures (NP and PP bundles combined accounted for approximately 70% of all bundle occurrences in both subcorpora). This pattern supports usage-based accounts of language acquisition (Ellis, 2002; Tomasello, 2003), which posit that high-frequency, prototypical constructions are acquired earliest and most robustly. For advanced L2 writers, extensive exposure to academic texts appeared to have facilitated the internalization of these core formulaic patterns, resulting in proportional distributions that were comparable between the two groups. This finding diverged from earlier studies that reported lower formulaic competence in L2 academic writing (Chen & Baker, 2010; Ädel & Erman, 2012), suggesting that at the level of advanced academic writers, the fundamental structural repertoire of NP-based formulaic sequences may be equally available to both groups. However, the absence of a statistically significant difference did not imply equivalence; it merely indicated that the present data provided insufficient evidence of a macro-level difference. Accordingly, the term “convergence” was used descriptively to refer to the observed similarity in proportional distributions, not as a claim of statistical equivalence.

At the micro level, significant divergences emerged, particularly in the domain of VP bundles and stance-taking constructions. The substantial underuse of (DET) + NOUN + VERB + (CONJ)/(DET) bundles by NNS writers pointed to a persistent gap in the expression of authorial stance. This finding aligns with a robust body of research documenting L2 writers' reduced use of stance markers (Ädel & Erman, 2012; Hyland, 2005; Pérez-Llantada, 2014) and suggests that authorial presence remains a challenging aspect of academic writing even for advanced scholars.

The observed trend among NNS writers toward greater proportional use of passive voice and anticipatory it constructions, though not statistically significant, may be tentatively attributed to prescriptive pedagogical norms commonly emphasized in English for Academic Purposes (EAP) instruction; however, this interpretation requires further investigation. Such instruction often teaches passive voice as a marker of scientific objectivity and impersonal constructions as standard templates for stance expression (Swales & Feak, 2012). While these structures are indeed characteristic of academic writing, their overuse (if confirmed in larger samples) can result in a less varied syntactic profile compared to native writers, who tend to balance passives with active voice constructions (Pan, Reppen, & Biber, 2016). For anticipatory it constructions specifically, the higher proportional use by NNS writers (3.60% vs. 1.30%), though not statistically significant, aligns with observations that L2 writers sometimes overuse certain impersonal constructions as a compensatory strategy (Hinkel, 2004).

For Classes 2 (active verb complements), 4 (passive VP), and 5 (dependent clauses), no statistically significant differences were detected (all $p > 0.05$). Within the limits of this corpus, the two groups were comparable in their proportional use of these structures. Moreover, the comparable usage of syntactically complex patterns, such as in which the and in which they challenged the assumption that L2 writers universally avoid syntactic complexity (Ortega, 2003) and supported the view that formulaic sequences can serve as vehicles for acquiring and deploying complex syntax (Ellis, 1996; Myles et al., 1998).

The finding that NNS writers exhibited a slightly wider inventory of unique bundle types but a lower overall frequency and density of bundle usage is consistent with observations that advanced L2 writers often have a wider range of bundle types, yet fall short of native-like frequency and fluency (Ädel & Erman, 2012; Pan, Reppen, & Biber, 2016). This pattern suggested a proficiency gap not in the breadth of formulaic knowledge, but in its consistent, high-density deployment—a key marker of expert academic writing (Cortes, 2004; Hyland, 2008).

Taken together, these findings demonstrate that different structural categories follow distinct developmental trajectories in L2 academic writing. While phrasal conventions (NP and PP) are acquired robustly to a level comparable with native writers (contradicting earlier claims of general L2 deficiency in this area), stance-taking VPs remain a persistent challenge. The non-significant trends in passive voice and anticipatory it constructions, though not definitive, point to potential areas of pedagogical influence that warrant further investigation with larger samples.

7. Pedagogical Implications

The findings carry important implications for English for Academic Purposes (EAP) pedagogy targeting Arab scholars



and advanced L2 writers more broadly. However, these recommendations are preliminary and should be validated through classroom-based intervention studies as the corpus (175,000 words) may not fully generalize to all Arab scholars or all academic disciplines.

First, instruction should move beyond the teaching of basic structural templates to emphasize the acquisition of discipline-specific, stance-taking formulaic sequences. Explicit focus on bundles, such as we argue that, this study examines, and the findings suggest that can help learners develop a more pronounced authorial presence.

Second, while passive voice and impersonal constructions are legitimate features of academic prose, instructors should encourage a balanced use of active voice to achieve greater syntactic variety and express stance more directly. Contrastive analysis of native and non-native texts can raise learners' awareness of these stylistic differences.

Third, the successful acquisition of complex syntactic structures through formulaic sequences suggests that corpus-informed instruction can leverage bundles as "pedagogical vehicles" for introducing advanced grammar. Activities that involve noticing, analyzing, and practicing target bundles in context can facilitate both lexical and syntactic development (Nation, 2001; Ucar & Zarsaz, 2023). Specifically, instructors can use concordancing tools (e.g., AntConc, Sketch Engine) to highlight bundle frequencies and distributions in native vs. non-native corpora, allowing learners to observe phraseological patterns directly from authentic data.

Finally, the overall convergence in phrasal structures highlights the importance of sustained exposure to authentic academic texts. Instructors can complement such exposure with structured corpus-based tasks—such as identifying and categorizing NP/PP bundles in target journals or comparing bundle use across abstracts—to help learners consolidate the core grammatical and phraseological repertoire that characterizes expert academic writing.

8. Future Research Directions

The present study offers a foundation for several lines of future inquiry. First, longitudinal studies tracking Arab scholars' bundle use over time (e.g., from doctoral training to early-career publication) can reveal developmental trajectories and the impact of targeted instruction. Second, larger and more diverse corpora (e.g., expanding the current 175,000-word corpus to include multiple disciplines and sub-fields of linguistics) would enhance the generalizability of the findings. Third, functional analysis of lexical bundles—examining how bundles perform stance, discourse organization, and referential work—can complement the structural focus of the present study. Fourth, cross-L1 comparisons involving other mother tongue backgrounds (e.g., Turkish, Chinese, Spanish) would help disentangle universal L2 acquisition patterns from language-specific transfer effects. Such comparative work can further inform EAP pedagogy by identifying which bundle difficulties are shared across L1 groups and which are unique to specific linguistic backgrounds.

9. Conclusion

The distinguishing structural characteristic of lexical bundles in linguistics abstracts written by Arab scholars (compared to NS scholars) was the statistically significant underuse of stance-taking verb phrase bundles of the type (DET) + NOUN + VERB + (CONJ)/(DET) (e.g., we argue that, this study examines), while NP-based and PP-based bundles showed no significant differences between the two groups. Consistent with previous research characterizing the academic register (Biber et al., 1999, 2004), both NS and NNS writers demonstrated a strong reliance on phrasal bundles—specifically NPs and PPs—over clausal structures. This similar dominance of nominalization suggested that NNS writers had internalized the "nominal style" essential for conveying informational density and objectivity though the absence of a statistically significant difference did not imply equivalence.

The finding that NNS writers exhibited a slightly higher proportional use of frame-setting PP bundles and Other Noun-Based Patterns aligns with the explicitation hypothesis (Klaudy & Károly, 2005), suggesting that L2 writers may prioritize structural explicitness to ensure clarity. This tendency was corroborated by studies like those of Ädel and Erman (2012) and Pérez-Llantada (2014). Furthermore, the data underscored the impact of pedagogical norms on NNS writing. A non-significant trend toward higher use of passive voice and anticipatory it constructions was observed in the NNS corpus, which may reflect prescriptive pedagogical norms, though this interpretation requires caution (Swales & Feak, 2012). In contrast, NS writers demonstrated greater stylistic flexibility, alternating between passive and active voices and employing a wider range of stance-taking formulaic sequences.

The significant difference in VP bundles of Class 1 (stance-taking reporting structures) was particularly noteworthy. While NNS writers had acquired the basic structural patterns for attribution and stance expression, they had used these forms less frequently than their NS counterparts. This gap suggested that authorial presence and the strategic deployment of stance-taking bundles may be among the last aspects of formulaic competence to develop, consistent with research on the developmental trajectory of L2 academic writing (Li & Schmitt, 2009; Staples et al., 2013).

Ultimately, the analysis posited that advanced NNS proficiency requires a transition from reliance on safe structural templates to the flexible deployment of content-specific lexical bundles and the strategic use of stance-taking expressions. These findings have important implications for EAP pedagogy. Instruction should move beyond the basic teaching of structural frames (e.g., the aim of this study) to include a focus on lexical bundles and collocations specific to the discipline. Encouraging students to diversify their post-modification strategies beyond the *of*-genitive to incorporate active voice for stance-taking and expand their repertoire of reporting verbs and stance markers could help bridge the gap between structural accuracy and idiomatic fluency in high-impact academic writing.



10. Limitations of the Study

This study acknowledges several limitations. First, the exclusive focus on three-word lexical bundles excluded longer sequences (four-, five-, and six-word bundles) that may contribute to textual cohesion and reveal additional structural patterns (Nasrabady, Elahi Shirvan, & Golparvar, 2020). Second, the analysis was limited to structural classification and excluded systematic functional analysis, thereby not accounting for the potential multifunctionality of bundles or the complexity that a detailed move-based analysis would capture (Ädel & Erman, 2012). Finally, the corpus, while carefully compiled, might not fully represent the full range of disciplinary sub-fields within linguistics and findings may not generalize to other disciplines or other L1 groups. Future research may productively investigate longer bundles, incorporate functional and move-based analyses, and examine whether the patterns observed here characterize other academic disciplines and other L1 groups within the Arab world.

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(1), 81–92. <https://doi.org/10.1016/j.esp.2011.08.004>
- Bao, D., & Liu, M. (2022). A corpus study of lexical bundles used differently in dissertations abstracts produced by Chinese and American PhD students of linguistics. *Frontiers in Psychology* 13, 842965. <https://doi.org/10.3389/fpsyg.2022.842965>
- 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., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
- Chen, Y., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Journal of English for Academic Purposes* 22, 1–13. <https://doi.org/10.64152/10125/44213>
- 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.2004.01.001>
- De Cock, S. (2002). Recurrent phrases in English native speaker speech: A corpus-driven analysis of the spoken component of the British National Corpus (Unpublished doctoral dissertation). University of Louvain.
- Durrant, P. (2017). Lexical bundles and discipline variation in university students' writing: Mapping the territory. *Applied Linguistics* 38(5), 632–656. <https://doi.org/10.1093/applin/amv011>
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition* 18(1), 91–126.
- 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>
- Firth, J. R. (1957). *Papers in linguistics 1934–1951*. Oxford University Press.
- Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3–18). Addison Wesley Longman.
- Granger, S., & Meunier, F. (Eds.). (2008). *Phraseology: An interdisciplinary perspective*. John Benjamins.
- Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. University of Pittsburgh Press.
- Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. Lawrence Erlbaum.
- Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. Longman.
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum. London, UK.
- 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., & Tse, P. (2005). Hooking the reader: A corpus study of abstracting formulas. *Journal of English for Academic Purposes* 4(3), 193–213. <https://doi.org/10.1016/j.esp.2004.02.002>
- Kennedy, G. (2014). *An introduction to corpus linguistics*. Routledge.
- Klaudy, K., & Károly, K. (2005). Implication in translation: Empirical evidence for operational asymmetry in translation. *Across Languages and Cultures* 6(1), 13–28. <https://doi.org/10.1556/Acr.6.2005.1.2>
- Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. LTP.
- Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing* 18(2), 85–102. <https://doi.org/10.1016/j.jslw.2009.02.001>
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning* 48(3), 323–363. <https://doi.org/10.1111/0023-8333.00045>
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.



- Nasrabady, P., Elahi Shirvan, M., & Golparvar, S. E. (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>
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of the 1995–2002 period. *Language Learning* 53(s1), 47–84. <https://doi.org/10.1093/applin/24.4.492>
- Pan, F. (2020). Methodological differences matter: Identification thresholds and corpus composition in lexical bundle research. *Southern African Linguistics and Applied Language Studies* 38(4), 336–350. <https://doi.org/10.2989/16073614.2020.1858897>
- Pan, F., Reppen, R., & Biber, D. (2016). Lexical bundles in L1 and L2 academic writing. *Applied Linguistics* 37(5), 639–662. <https://doi.org/10.64152/10125/44213>
- Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. London: Continuum.
- Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes* 14, 84–94. <https://doi.org/10.1016/j.jeap.2014.01.002>
- Qin, J. (2014). Move analysis and lexical bundles in research article abstracts. In Proceedings of the 2014 International Conference on English, *Literature and Humanities* (pp. 12–18).
- Rahimi Azad, M., & Modarres Khiabani, S. (2018). Lexical bundles in English abstracts of research articles written by Iranian scholars: Examples from humanities. *Iranian Journal of Applied Language Studies* 10(2), 149–174. <https://doi.org/10.22111/ijals.2018.4634>
- Simpson-Vlach, A., & Ellis, N. C. (2010). An academic formulas list (AFL): New methods in phraseology research. *Applied Linguistics* 31(4), 487–512. <https://doi.org/10.1093/applin/amp058>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Sinclair, J. (2005). *Corpus and text – Basic principles*. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1–16). Oxford Text Archive.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes* 12(3), 214–225. <https://doi.org/10.1016/j.jeap.2013.05.002>
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Swales, J. M., & Feak, C. B. (2009). *Abstracts and the writing of abstracts*. University of Michigan Press.
- Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students: Essential tasks and skills* (3rd ed.). University of Michigan Press.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Ucar, S., & Zarsaz, E. (2023). A corpus-based teaching of lexical bundles to enhance writing skills of prospective Turkish EFL teachers. *International Journal of Learning and Teaching* 15(1), 42–54. <https://doi.org/10.18844/ijlt.v15i1.8577>
- Wei, M., & Lei, L. (2011). Lexical bundles in L1 and L2 academic writing. *International Journal of Corpus Linguistics* 16(3), 429–456. <https://doi.org/10.64152/10125/44213>
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language and Communication* 20(1), 1–28. [https://doi.org/10.1016/S0271-5309\(99\)00015-4](https://doi.org/10.1016/S0271-5309(99)00015-4)

