

Article

Understanding the Sustainable Growth of EFL Students' Writing Skills: Differences between Novice and Expert Writers in Their Use of Lexical Bundles in Academic Writing

Shaojie Zhang ¹, Hui Yu ^{1,*} and Lawrence Jun Zhang ² 

¹ School of Foreign Languages and Literature, Beijing Normal University, Beijing 100875, China; zhangshaojie@mail.bnu.edu.cn

² Faculty of Education & Social Work, The University of Auckland, Auckland 1023, New Zealand; lj.zhang@auckland.ac.nz

* Correspondence: yuh@bnu.edu.cn

Abstract: Lexical bundles, as building blocks of discourse, play vital roles in helping members from the same academic community achieve successful communication and disseminate sustainable disciplinary knowledge. However, little attention has been paid to lexical bundles in postgraduate writing. Drawing on Biber et al.'s (1999) structural taxonomy and Hyland's (2008a) functional taxonomy, we identified and compared lexical bundles in two self-built corpora, an EFL student writing corpus and an expert writing corpus. The results indicate considerable structural differences between the two groups: the student writers used verb phrase-based bundles more frequently and prepositional phrase-based and noun phrase-based bundles less frequently. In terms of function, although the two academic groups showed similar distributions of the three main functional categories, as student writers they exhibited insufficient reader-awareness and incomplete knowledge of stance expressions. It is hoped that the findings will shed light on future pedagogical practices to help novice writers improve their academic writing competence as a sustainable goal in enhancing their academic scholarship.

Keywords: lexical bundles; academic writing; expert writer; postgraduate student writer



Citation: Zhang, S.; Yu, H.; Zhang, L.J. Understanding the Sustainable Growth of EFL Students' Writing Skills: Differences between Novice and Expert Writers in Their Use of Lexical Bundles in Academic Writing. *Sustainability* **2021**, *13*, 5553. <https://doi.org/10.3390/su13105553>

Academic Editor: Fuhui Tong

Received: 23 April 2021

Accepted: 13 May 2021

Published: 16 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The past two decades have witnessed an increasing interest in lexical bundles because of their significance in formal writing, especially in academic contexts. The term "lexical bundles" was defined as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status" [1] (p. 990). As is well documented, lexical bundles not only contribute to fluent linguistic production [2] but also form essential building blocks of discourse [3]. A good command of lexical bundles could be indicative of a proficient and professional academic writer and is thus considered a pivotal skill for student writers, especially EFL student writers, for achieving sustainable growth of writing competence. Appropriate use of lexical bundles in academic writing helps writers from an academic community demonstrate their research writing ability. To date, previous research has examined lexical bundles in different academic registers, including university textbooks, student–student academic interactions, and professional academic writing [3–5], in various academic disciplines (e.g., electronic engineering; biology; business; applied linguistics; history, etc.) [6–12], and in different groups of writers (e.g., native vs. nonnative English writers; writers of different proficiency levels; novice vs. expert writers, among others) [13–19].

Novice writers constructing academic writing often encounter difficulties regarding the proficient use of disciplinary language. Previous studies on writer groups tended to focus on how native and nonnative writers deploy lexical bundles. However, very few studies have compared lexical bundles between expert writers and student writers and

even fewer have studied MA students, who are really novice writers. For MA student writers, it should be of great interest and importance to know what lexical bundles are frequently used by expert writers and how they are used. Such knowledge can assist these students in making choices regarding which lexical expressions to employ and how to employ them.

Therefore, the current study aims to compare the use of lexical bundles in MA students' writing and expert writing in terms of frequency and distribution. It is hoped that the findings may provide insight into how novice academic writers can be guided more effectively for improving their academic writing ability as a long-term goal for sustainable development of academic scholarship.

2. Review of Literature

2.1. Lexical Bundles

As Biber et al. [4] posit, lexical bundles are the most frequent, recurrent, multiword sequences in a register, which are defined "strictly on the basis of frequency" (p. 399) rather than intuitive criteria. Even though the identification of lexical bundles is solely based on frequency without considering structural or functional features, Biber and associates think that these multiword sequences are "interpretable in both structural and functional terms" (p. 399).

The first structural classification was proposed by Biber et al. [1], in which the prevalent lexical bundles were grouped into fourteen structural categories in conversation and twelve categories in academic prose. Concerning structural analysis of lexical bundles, their framework has been used as a major reference.

Functionally, two taxonomies are widely adopted. Biber et al. [4] distinguished three main functions: (1) stance expressions for displaying "attitudes or assessments of certainty", (2) discourse organizers that "reflect relationships between prior and coming discourse", and (3) referential bundles that "make direct reference to physical abstract or single out some particular attribute of the entity as especially important." (ibid., p. 384). Inspired by Biber et al. [4], Hyland [11] proposed his functional taxonomy of lexical bundles, including (1) research-oriented bundles that "help writers to structure their activities and experiences of the real world", (2) text-oriented bundles that are "concerned with the organization of the text and its meaning as a message or argument", and (3) participant-oriented bundles that are "focused on the reader or writer of the text" (ibid., p. 13).

2.2. Studies of the Structural and Functional Analyses of Lexical Bundles

Comparative studies of lexical bundles are largely carried out along three dimensions. They are lexical bundle use across registers, across disciplines, and across writer groups. We review these studies below.

2.2.1. Lexical Bundle Use across Registers

One of the important variables influencing the use of lexical bundles is register variation. Based on the comparison of two types of registers, i.e., conversation and academic prose, Biber et al. [1] found that, in terms of structure, bundles were more clausal in conversation but more phrasal in academic prose. In their studies of university classroom teaching and university textbooks, Biber et al. [4] concluded that the use of lexical bundles in classroom teaching reflects a mixture of characteristics typical of both conversation and academic prose. Similarly, Nesi and Basturkmen [20] found that academic lectures are also featured by combined use of oral and literate bundles. In terms of function, classroom teaching combines functional characteristics of both spoken (by using stance and discourse organizing bundles) and written registers (by using referential bundles) [4]. Biber and Barbieri [3] further examined lexical bundles in a broader range of spoken and written university registers. They concluded that both spoken/written register differences and communicative purposes influence the use of lexical bundles.

2.2.2. Lexical Bundles across Disciplines

Discipline is also a crucial variable influencing the use of lexical bundles [6,7,10,11]. Of these studies, lexical bundles in soft science and hard science are often compared. In terms of structure, two main structural types are found in history, including noun phrases and prepositional phrases, whereas more structural types are found in biology [6]. It is further found that social science texts make use of a large number of bundles with an embedded *of* phrase to identify the logical relations in the argument. By contrast, hard science texts make more use of formulaic passive constructions and *anticipatory it* patterns to disguise the personal role of writers in the interpretation of data [11]. In terms of function, soft-science texts use more text-oriented and participant-oriented bundles, whereas hard science texts are dominated by research-oriented bundles [11,21]. Disciplinary variation is also explored in student writing [7]. The results suggest that research-oriented bundles are used for assertion of importance in soft science but for physical descriptions in hard science. Stance-oriented bundles are used to evaluate the importance of the topic in soft science but to state findings in hard science. Furthermore, soft science writing is characterized by text-oriented bundles indicating relationships or differences. Hard science writing, by contrast, contains text-oriented bundles that guide readers' attention to data presented in figures and tables. Some studies have further investigated the relation between lexical bundles and rhetorical moves in a given discipline [22–24]. Previous studies mostly compare lexical bundles from a macrodiscipline level, such as soft/hard science distinctions and humanities/natural sciences distinctions. It would be more helpful to investigate disciplinary-specific lexical bundles, which may help writers express stances more appropriately in their research community.

2.2.3. Lexical Bundles across Writer Groups

The third influential factor regarding the use of lexical bundles is the background of different writers, such as between L1 English and L2 English writers [13,14,16,18,25,26], between student writers across different proficiency levels [19,27], or between expert writers and novice writers [6,28].

Most of the research indicated structural and functional differences between L1 and L2 writings. In terms of structure, for instance, L1 Swedish student writers are found to use a higher number of *anticipatory it* (e.g., *it is difficult to*) and *attended this* (e.g., *in this essay I*) constructions than L1 English student writers [13]. It is also found that L1 English writers produce more verb phrase (with a passive verb) lexical bundles, whereas L1 Persian writers use more noun phrase bundles [16]. However, L1 Chinese writers, including both student writers and expert writers, use more verb patterns, whereas L1 English writers use a slightly more extensive range of noun sequences and prepositional sequences [18,25]. In terms of function, L1 English writers are found to use a higher proportion of stance bundles than Swedish writers [13] but a smaller proportion than L1 Chinese writers [25]. Chinese writers are also found to use lexical bundles of description, transition and structure more frequently than English writers, whereas English writers employ more quantification and framing bundles than Chinese writers [18]. Persian writers overused statistical markers compared to English writers [16]. Other studies, however, reported no significant differences between L1 and L2 writing. For instance, Chen and Baker [14] found that lexical bundles in L1 and L2 student writing are surprisingly similar. This finding is consistent with Shin [26], who found that both L1 and L2 student writers heavily use clausal bundles.

Comparisons have also been made between student writings of different proficiency levels [19,27,28] and between student writers and expert writers [6,11,14].

Regarding student writings of different proficiency levels, previous research has demonstrated a mixture of divergent and even contrasting results. Whereas lower proficiency student writings are reported to feature a higher number of NP-based lexical bundles [27], Chen and Baker [28] found that the lowest level has the lowest proportion of NP-based bundles. Similarly, Vo [27] reported a higher frequency of stance bundles in lower-level writing, whereas Staples et al. [19] found different proficiency groups have a

similar distribution of stance bundles and discourse organizing bundles. Such diversity in research results may be largely due to the different criteria for determining the proficiency level of different students.

Römer [29] and Chen and Baker [14] conducted a three-way comparison: L1 English expert writer versus student writers of both L1- and L2-English backgrounds. It was argued that novice/expert distinction is more important than L1/L2 distinction based on the findings that few differences existed between the L1 and L2 student writers. It was found, though, that many lexical bundles frequently used by expert writers are rarely found in student writing [6,11,14,29], whereas student writing features more VP-based bundles [14] and bundles commonly found in the spoken register [30]. Nonetheless, the findings, useful as they are, may not reflect the whole picture of novice writers' discourse features. Many previous studies focus on how bundles identified in expert writing are used by student writers. Such comparisons generate insightful findings but provide insufficient understanding of lexical bundles that are unique to student writing. In addition, most studies on student writers focus on writings by undergraduate writers, including, for example, argumentative essays [25,26,31], research papers [6], and writing examination papers [27]. Very little attention has been paid to the use of lexical bundles in MA student thesis writing. One of the few existing relevant studies was Hyland [11], which compared published article bundles to those identified in Master theses and PhD dissertations. However, he treated MA theses in his corpus as highly proficient texts and explained the features from the perspective of genre variation rather than novice/expert distinction. Therefore, it offers limited pedagogical guidance for student writers in developing sustainable linguistic resources to express their stance in more mature way. Another relevant study is by Pan and Liu [32], who compared L1-L2 differences in bundles in masters' theses and research articles. Although their findings indicated that both L1 background and the level of expertise affect the bundle employment, Pan and Liu [32] did not compare the student bundles directly with expert bundles and they mainly focused on comparison between L1-L2 students and between L1-L2 experts. Despite the fact that their research was among one of few attempts to investigate how postgraduate students employ lexical bundles, comparing MA student writing to expert writing can provide useful information on expert writers' linguistic choices.

Therefore, the current study seeks to focus on this understudied writer group by comparing the use of lexical bundles between Chinese English-major MA theses and expert writers' published articles. Informed by the previous literature, research articles published in leading international journals such as those covered in the SSCI can generally be considered as samples of expert writing. MA theses can be viewed as unique pieces of student writing at the level between argumentative essays/course papers and published research articles. They are written by apprentice academic writers who are under the pressure to display their extensive knowledge in one discipline as well as the ability to conduct independent research appropriately. It is hoped that the present study will contribute to existing knowledge of bundle research on MA writers, especially on Chinese EFL learners. The study aims to provide further insights into pedagogical implications for teaching academic writing.

Specifically, this study addresses the following questions:

1. Are there any differences in the structural types of lexical bundles used by MA student writers and expert writers?
2. Are there any differences in the functional types of lexical bundles used by MA student writers and expert writers?

3. Materials and Methods

3.1. Corpora

Data for the present study consist of two corpora of academic writing texts. Pan and Liu [32] did not provide additional information on their master theses corpus. In the present study, the student writing corpus (SWC) comprises 24 MA theses on linguistics

from two prestigious Chinese universities, with 12 theses from each university submitted during 2015–2017. Ranked among the A+ and A level Foreign Linguistics disciplines in the China Discipline Evaluation (Round 4 in 2017), graduate program in these two universities represents the highest level of linguistic education and academic training. MA theses from these two universities represent a high level of academic writing among novice writers in China.

There are a few ways in which our data might be different from that of Pan and Liu [32]. As for the research article corpus, Pan and Liu [32] selected articles from both world's leading journals and English-medium journals published in China. Although they reported that the selected journals are somewhat comparable, our selection of journals in the current study was based on the journal impact factor and rank statistics retrieved from the SSCI database. The expert writing corpus (EWC) comprises 66 research articles published in 2018 from two leading journals in the field of linguistics. Both are peer-reviewed, with a five-year impact factor of above 2.0. The detailed information is presented in Table 1.

Table 1. Corpora Information.

Corpus	Representation	Word Count	Mean Length	No. of Texts
SWC	Student writing	456,243	19,010	24
EWC	Expert writing	451,581	6842	66

One criterion for selection is to ensure that the size of both corpora is roughly comparable, since lexical bundles are sensitive to the word counts in a corpus [33] rather than the number of texts from which they are extracted [6]. Due to the variations of text lengths of theses and research articles, balancing the size of subcorpora inevitably leads to a small number of longer texts in the student corpus being compared to relatively large numbers of short texts in the expert corpus. It would seem that a difference in average text length could suggest a difference in the token frequency of bundles. There would be an increased possibility of repetition of the same bundle in longer texts. To mitigate this potential impact, we report information on type (the number of different bundles), token (the total occurrence of the same bundle), and type/token ratio. The proportional distribution of lexical bundles will mainly be based on type frequency.

For both corpora, only the body part of the writings was selected. All the tables, figures, notes, footnotes, references, mathematical equations, and appendices in each text were manually excluded, in accordance with advice in Cortes [6]. The plain texts were numbered so that the first document in the student writing corpus and the expert writing corpus were labeled S1 and E1, respectively.

3.2. Lexical Bundle Extraction

One major difficulty for the current study is the selection of lexical bundles. Three criteria, including length, frequency, and dispersion, are used to pick out appropriate lexical bundles.

First, in terms of length, all four-word bundles are chosen “because they are far more common than 5-word strings and offer a clearer range of structures and functions than 3-word bundles” [11] (p. 8). Second, in terms of frequency, the cut-off frequency thresholds used in lexical bundle studies range from 10 [1] to 20 [6,10,11] to 25 [13,14] to 40 [4] times per million words. The lack of agreed-upon criteria might be a result of different corpus sizes which vary from 50,000 to more than one million. The present study decided to follow Ädel and Erman [13] to set the cut-off frequency at 25 times per million words since one sub-corpus in their study is 417,777 words, which is similar to our corpus size. Third, dispersion criterion is set to avoid idiosyncrasies introduced by individual writers [4]. Previous studies have not agreed on the number of texts in which a lexical bundle should occur. The average requirement is that lexical bundles need to occur in three texts [14] or at least five different texts of the corpus [1,6,18] or in at least 10% of all texts [10,11]. Considering that student writing corpus contains 24 texts in total,

adopting the 10% criteria [10] is not strict enough to ensure that there are no idiosyncratic employment of a few authors. Biber and Barbieri [3] suggested an optimal threshold of 3 for 50,000 sub-corpora, 4 for 100,000 sub-corpora, and 5 for 200,000 corpora. The present study also set the strict dispersion to at least five different texts.

Antconc, a software program [34], was used to generate lists of qualified lexical bundles from the two corpora according to the aforementioned requirements for length, frequency, and dispersion. After running the software, two lists of bundles were extracted from SWC and EWC, with the type number being 148 and 99, respectively. As the software cannot automatically delete unqualified lexical bundles, manual filtering of lexical bundles is necessary. The lexical bundles identified by *Antconc* were checked for cases misidentified by the software. For example, a structure like “*the other hand, the*” would be identified as bundles. We used punctuation as a stopping point (the sequences of words must be uninterrupted to be treated as lexical bundles) [3]. Eight and five types of bundles were excluded from the extracted bundle lists in SWC and EWC.

There are two exceptions to lexical bundles thus generated. First, following Chen and Baker [14], the current research does not consider context-dependent bundles, which are directly related to specific topics in the articles (e.g., *as a foreign language*). These bundles are usually considered terminological (e.g., *of systemic functional linguistics* in SWC). The reason for this exclusion is to “guard against idiosyncrasies introduced by those topics that happen to be represented in the material” [13] (p. 84).

The second exception is the overlapping lexical bundles. We identified cases of what Chen and Baker [14] called “complete subsumption”, which refers to a situation where two (or more) bundles overlap and one is subsumed within the other. For instance, *as a matter of* occurs 38 times, and *a matter of fact* occurs 35 times in the SWC. According to concordance analyses, all occurrences of *a matter of fact* are preceded by the word *as*. The lower frequency bundle was combined into the higher frequency one: *as a matter of (fact)*. The two lists of bundles were then reduced to 113 and 84 in SWC and EWC.

3.3. Lexical Bundle Classification

Structural and functional comparisons are based on the lexical bundles retrieved and refined. First, in terms of structure, the classification was based on the structural categories proposed by Biber et al. [1]. The broad structural classification includes phrasal bundles and clausal bundles. Phrasal bundles include NP-based bundles featuring noun phrases and PP-based bundles featuring a prepositional element plus a noun element. Clausal bundles involve VP-based bundles with verb components.

Second, in terms of function, the classification was based on Hyland’s [10] framework, which is chosen over that of Biber et al. [4] because Hyland’s taxonomy was established using academic writing, whereas Biber et al.’s classification is based on a corpus covering both spoken and written registers. There were no bundles in the subcategory “topic” in the current study since topic-specific bundles were excluded in the filtering process. Functional classification in previous studies were consulted in order to label Lexical bundles. Besides, bundles were checked in the concordance lines to find out their primary functions. To ensure consistency, two well-trained researchers are invited to join in the process and reached approximately 98% agreement for structural types and 90% for functional types. The remaining discrepancies were discussed until complete agreement was achieved.

4. Results and Discussion

This section will firstly present an overall quantitative comparison between the two groups; the number of bundle types and tokens will also be discussed. Then, a detailed discussion will be based on a structural and functional comparison of lexical bundles between student writers and expert writers.

4.1. Lexical Bundles Identified in Total

Altogether 113 lexical bundles are identified in the SWC and 84 in the EWC. The number of occurrences in each corpus is illustrated in Table 2, with further distinction between type and token.

Table 2. The number of occurrences of lexical bundles.

Corpus	No. of Types	No. of Tokens	Type/Token Ratio
SWC	113	2514	1/22
EWC	84	1685	1/20

The results show that student writing contains more types and tokens of lexical bundles than expert writing. This finding is in line with previous findings that novice writers and less proficient students tend to use more lexical bundles [10,18,25,35]. For novice academic writers, the employment of greater numbers of bundles may be a way of demonstrating their academic competence [10]. By using these fixed expressions, student writers are making an effort to construct what they understand as expert-like writings [25]. Through the greater use of frequently recurring lexical phrases, student writers attempt to avoid the expressions that might be considered uncommon or inappropriate in academic contexts.

The type/token ratio of expert writers was slightly higher than that of the student writers, which indicated that fewer bundles were repeatedly employed in the EWC. Although the repetition of lexical bundles by learners is in line with Granger's [36] and Cortes's [6] finding that learners tend to overuse sequences, it should be noted that writing a short passage (published articles) without overlap is likely easier than completing a longer writing (MA theses) with no overlap.

Among all bundles, 80 bundles are exclusive to student writing and 51 to expert writing, and 33 lexical bundles are shared by both corpora. Eight out of ten most frequent bundles in EWC are also used in SWC. This suggests that student writers are aware of the most typical expression characteristic of academic written register. Lexical bundles, such as *on the other hand* and *at the same time*, are used with high frequency and range widely in both groups. Other shared bundles are not used equally frequently, for example, student writers do not opt for bundles like *at the end of* and *the ways in which* as frequently as expert writers. Among the non-shared bundles, several bundles in student writing are more commonly used in conversations such as *that is to say* and *we can see that*. In the following sections, structural and functional analyses will be presented.

4.2. Comparison of the Structural Categories of Lexical Bundles

Overall, VP-based bundles account for 35.4% and 19.0% of the bundles in student writing and expert writing, respectively, whereas phrasal bundles, which include NP-based bundles and PP-based bundles, amount to 58.4% in student writing and 79.8% in expert writing. It shows that student writers rely more on VP-based bundles and less on NP-based and PP-based bundles than expert writers. Previous studies suggest that the lexical bundles most frequently used in academic writing are parts of noun or prepositional phrases [11], whereas clausal bundles are more typical in spoken registers. Biber et al. [4] concluded that 70% of lexical bundles in academic prose are phrasal bundles and 90% of bundles in conversation are clausal bundles. Student writing in the current study uses phrasal bundles much less than expert writing, reflecting a mixed style of academic prose and conversational spoken register.

Table 3 shows the distribution of each subcategory together with the log-likelihood test results (The present study used the log-likelihood calculator created by Jiajin Xu (<http://corpus.bfsu.edu.cn/TOOLS.htm>, accessed on 6 December 2020); * = significant at $p < 0.05$; ** = significant at $p < 0.01$; *** = significant at $p < 0.001$ throughout the paper. The raw frequency is shown in brackets. "+" means that the token frequency in Corpus 1 (SWC) is higher than in Corpus 2 (EWC), i.e., "+" means overused, "-" means underused through-

out the paper.). In the following section, lexical bundles in the structural subcategories will be compared.

Table 3. Distribution of structural categories of lexical bundles.

Categories	Subcategories	Type		Token	
		SWC	EWC	SWC	EWC
NP-based	NP with of-phrase fragment	20.4% (23)	20.2% (17)	13.5% (339)	19.3 (325)
	NP with other postmodifier fragment ***-	3.5% (4)	4.8% (4)	1.8% (45)	5.9% (99)
	Other NP ***-	0.9% (1)	6.0% (5)	1.6% (40)	4.7% (80)
	Total **-	24.8% (28)	31.0% (26)	16.9% (424)	29.9% (504)
PP-based	PP with embedded of-phrase ****	21.2% (24)	27.4% (23)	28.8% (721)	31.2% (526)
	Other PP fragment	12.4% (14)	21.4% (18)	16.3% (408)	22.9% (386)
	Total ****	33.6% (38)	48.8% (41)	45.1% (1129)	54.1% (912)
VP-based	Anticipatory it +VP/adj phrase ****	6.2% (7)	4.8% (4)	4.9% (123)	4.3% (72)
	Passive verb + PP ****	9.7% (11)	4.8% (4)	8.5% (214)	3.3% (55)
	Copula be +NP/adj phrase ****	2.7% (3)	2.4% (2)	3.1% (77)	1.7% (28)
	(VP) +that-clause fragment ****	5.3% (6)	1.2% (1)	4.9% (122)	0.9% (15)
	(Verb/adjective) to-clause fragment	3.5% (4)	3.6% (3)	2.3% (57)	2.5% (42)
	Pronoun/NP + be fragment	1.8% (2)	-	1.1% (28)	-
	Adverbial clause fragment ****	3.5% (4)	2.4% (2)	3.8% (96)	1.4% (23)
	VP with active verb	2.7% (3)	-	1.9% (47)	-
Total ****	35.4% (40)	19.0% (16)	30.5% (764)	13.9% (235)	
Others	As well as ****	6.2% (7)	1.2% (1)	7.9% (197)	2.0% (34)

The present study used the log-likelihood calculator created by Jiajin Xu (<http://corpus.bfsu.edu.cn/TOOLS.htm>, accessed on 6 December 2020); ** = significant at $p < 0.01$; *** = significant at $p < 0.001$.

Student writers employ a smaller number and lower proportion of NP-based bundles than expert writers. The two groups differ most in the use of *NP with other postmodifier fragment* bundles. Expert writers use twice as many tokens as student writers. The findings presented in Table 4 support Chen and Baker's [14] study that *NP with other postmodifier fragment* bundles are usually part of relative clauses.

Table 4. Lexical bundles of NP with other postmodifier fragment.

	NP with Other Postmodifier Fragment	Type	Token
SWC	the ways in which ***- (12), the fact that the (11), the previous studies on (11), the relationship between the (11)	4	45
EWC	the ways in which (42), the extent to which (23), the fact that the (21), the relationship between the (13)	4	99

The shared bundles are marked in bold throughout the paper. *** = significant at $p < 0.001$.

Student writers use *the ways in which* significantly less often than expert writers. Another frequently used bundle by expert writers, *the extent to which*, is not found in student writing. Although *the fact that the* and *the relationship between the* are shared bundles, student use is significantly less than expert use. The infrequent use of embedded relative clauses as post modifiers by Chinese MA student lent support to Chen and Baker [14] and Pan and Liu [32].

In addition, student writing is notably different from expert writing in the *Other NP* subcategory. As shown in Table 5, expert writers use five times as many types and twice as many tokens in bundles of this subtype than students. It seems that this structure is used by expert writers to highlight their research methods and remind readers of their research questions. Student writers do not use this category as frequently as experts do, and they only use this structure to summarize their findings according to the concordance lines.

Table 5. Lexical bundles of *Other NP*.

	Other NP	Type	Token
SWC	the similarities and differences (40), quantitative and qualitative data (22), the second research	1	40
EWC	question (17), the first research question (16), the first research question (16), a mixed methods approach (12)	5	80

Lexical bundles of the *NP with of-phrase fragment* subcategory can be mostly grouped into the frame “*the + noun + of + the/a*”, which was considered a fixed frame by Biber et al. [37].

Table 6 presents the nouns collocating with this frame. Student writers use a slightly more extensive range of noun types but fewer tokens than expert writers. Lexical bundles of this frame are described as “extremely productive frames” [37] (p. 78). However, their importance is underestimated by student writers. The underuse of “*the + noun + of + the/a*” bundles is also found in Chen and Baker’s [14] study, which concluded that neither British students nor Chinese students use this frame like experts. This underuse may be regarded as a feature of student writing due to their underdeveloped writing proficiency rather than the L1 background.

Table 6. The “*the + noun + of + the/a*” frame.

	the + Noun + of + the/A * ⁻	Type	Token
SWC	basis (21), beginning (18), end *** ⁻ (18), form (18), frequency (14), meaning (14), structure (14), development (13), analysis (12), validity (12), majority (12), significance (12), results (11), use (11)	14	200
EWC	end (45), use (28), nature (25), analysis (23), meaning (23), role (21), quality (18), focus (15), beginning (13), use (13), context (13), basis (11)	12	248

* = significant at $p < 0.05$; *** = significant at $p < 0.001$.

Another notable finding is that, as shown in Table 7, the majority of these bundles begin with *the* in both groups.

Table 7. Distribution of NP with different initial words.

	Type		Token	
	SWC	EWC	SWC	EWC
NP beginning with <i>a/an</i> **** ⁺	5	2	75	32
NP beginning with <i>the</i> ** ⁻	15	14	214	281
NP beginning with <i>one</i>	1	1	18	12

** = significant at $p < 0.01$; *** = significant at $p < 0.001$.

Expert writers use more tokens of bundles beginning with *the* than student writers, while student writers overuse bundles beginning with *a/an*, including bundles such as *a better understanding of*, *a wide range of*, *a summary of the*, and *a larger number of*, whereas expert writers used two kinds: *a high level of* and *a wide range of*.

One more observation is that both groups use only one bundle ending with the indefinite article *a* (*the form of a* in SWC, *the use of a* in EWC) and more bundles end with the definite article *the*. This might be a result of disciplinary characteristic. For example, *the use of a* indicates any individual case or generic set, whereas *the use of the* refers to specific cases, which are more characteristic of soft science [7]. The reliance on the bundles ending with the definite article *the* in student writings may reflect students’ awareness of the disciplinary features of academic writing.

The most frequently employed types by the expert group are PP-based bundles. Expert writers use a smaller number but higher proportion than student writers.

The *Prepositional Phrase with Embedded Of-Clause* structure takes up the largest number of bundles (both types and tokens) among the 14 categories in both corpora. Many bundles

of this structure appear to be the most frequently employed ones (e.g., *from the perspective of* in the SWC, and *in the case of* in the EWC). Many *Prepositional Phrase with Embedded Of-Clause* bundles fill the frame “*in the + noun + of*”, which is another “extremely productive frame” proposed by Biber et al. [37]. Table 8 shows such lexical bundles.

Table 8. The “*in the + noun + of*” frame.

	<i>in the + Noun + of + the/A</i> *–	Type	Token
SWC	Use ****(80), process ***(44), field (39), form (26), analysis (15), development (12)	6	216
EWC	Case (58), context (40), form (39), use (23), process (20), production (19)	6	199

*=significant at $p < 0.05$; **=significant at $p < 0.01$; *** = significant at $p < 0.001$.

The two groups both used six types of lexical bundles in this frame. The number of tokens is slightly higher in student writing than in expert writing, which may result from the repetitive employment of *in the use of* (80 tokens). Three bundles in this frame are shared by both groups but are used in different ways. For example, student writers use *in the form of* as an adverb, as shown in example (1), but expert writers mostly use it as a post nominal modifier as shown in example (2).

- (1) Their realization patterns are to be summarized *in the form of* tables to see whether certain categories are used more often than others. (S15).
- (2) Qualitative assessment *in the form of* feedback or written comment is more appropriate for novice assesses. (E66).

The results also suggest that the two groups fill this frame for different functions. Expert writers use *in the case of* and *in the context of* to provide research background information, whereas student writers use *in the use of* and *in the process of* significantly more frequently to introduce the research procedure.

As for *other prepositional phrase fragments* bundles, student writers use a higher number but smaller proportion than expert writers. Many bundles in student writing are used to direct readers’ attention to certain positions in the text, such as *in the following table* and *in the above example*. Expert writers are found to use more lexical bundles indicating logical relations, such as *in relation to the* and *in line with the*. These bundles are among the top 30 frequent ones in expert writing but can hardly be found in student writing.

The verb phrase-based lexical bundles are the most frequently used category in student writing. Student writers use significantly more VP-based bundles than expert writers: more than twice as many types and three times as many tokens. The results show that experts use no bundles of the *Pronoun/NP + be fragment* and *VP with active verb* structure. Of other VP-based categories, student writers use more types and tokens of bundles than expert writers. Consistent with Chen and Baker’s [28] study, which concluded that students with lower proficiency employ more VP-based lexical bundles, our results suggest that novice academic writers tend to rely too much on VP-based lexical bundles.

Student writers use more *Anticipatory it + verb phrase/adjective phrase* bundles than expert writers. Pan and Liu [32] reported infrequent use of this structure by L2 expert writers and explained that it might due to the lack of its counterpart in Chinese. However, L2 student writers in their study used much more *anticipatory-it* bundles than L2 experts. It seems that MA students are aware of this pattern and use it effectively in theses writing. According to Hyland [10], this structure can downgrade the personal role in interpretation without identifying the source of evaluation. The frequent occurrence of this structure may partly be explained by genre difference between student theses and research articles. MA students appears to realize the possible risks of explicitly attributing the source of evaluation to themselves in a high stakes genre where students are under the pressure of assessment. Both groups use this structure to highlight significance, provide explanation, and report findings. Although student writers seem to realize the importance of distancing themselves from judgment, they show a preference for different adjectives such as

necessary, obvious, and clear. Expert writers used only two adjectives to fill this structure (i.e., *important, possible*).

Another structure student writers significantly overuse is the *V + that* frame, which is more commonly found in conversation [38]. Student writers frequently use human subjects such as *we can see that* to indicate findings as shown in (3). When expert writers use *V + that* frame, they are more willing to express similar meanings in an objective way with impersonal subjects as in (4). This is a strategy to strengthen the objectivity of their findings and interpretations.

(3) *We can see that* the latter genre is expanding the previous one by specifying in detail the very focus of it and thus elaborates the previous genre. (S12)

(4) *The results showed that* the learners portrayed a high level of pragmatic awareness in the three languages even though their L1 and L2 languages were still developing. (E59)

Overall, these findings demonstrate that student writers use bundles more frequently and show a strong preference for clausal bundles. Expert writers generate fewer bundle types and tokens, and they rely more on phrasal bundles than student writers. Phrasal bundles cannot be acquired naturally [39]; it can be assumed that such linguistic resources can be incorporated into academic writing courses. The above structural analysis offers only a partial picture of bundle use. In the following section, a comparison of functional categories will be presented.

4.3. Comparison of Functional Categories of Lexical Bundles

Text-oriented lexical bundles rank as the largest category in both corpora; EWC contains a higher proportion of 53.6%, whereas SWC contains 46.0%. SWC contains a higher proportion of research-oriented bundles (44.2%), and this category is less frequent in EWC (41.7%). Participant-oriented bundles are the smallest ones in both corpora, accounting for 9.7% in SWC and 4.8% in EWC.

Table 9 shows the proportional distributions of subcategories in the two corpora. The results reveal that the type and token frequencies in all three main categories and most subcategories are higher in student writing. In terms of proportion, student writers rely more on research-oriented and participant-oriented bundles and less on text-oriented bundles than expert writers.

Table 9. Distribution of functional categories of lexical bundles.

Categories	Subcategories	Type		Token	
		SWC	EWC	SWC	EWC
Research-oriented	Location	6.2% (7)	7.1% (6)	7.7% (192)	10.5% (177)
	Procedure **	11.5% (13)	13.1% (11)	9.1% (229)	10.7% (181)
	Quantification ****	12.4% (14)	11.9% (10)	10.8% (270)	9.9% (166)
	Description ****	14.2% (16)	9.5% (8)	10.6% (266)	9.6% (161)
	Total ****	44.2% (50)	41.7% (35)	38.2% (957)	40.7% (685)
Text-oriented	Transition signals ****	7.1% (8)	6.0% (5)	10.5% (264)	9.6% (161)
	Resultative signals	9.7% (11)	13.1% (11)	9.1% (229)	11.7% (197)
	Structuring signals ****	14.2% (16)	13.1% (11)	11.0% (276)	9.0% (151)
	Framing signals ****	15.0% (17)	21.4% (18)	22.9% (575)	25.3% (426)
	Total ****	46.0% (52)	53.6% (45)	53.6% (1344)	55.5% (935)
Participant-oriented	Stance features ****	7.1% (8)	3.6% (3)	5.3% (134)	3.1% (53)
	Engagement	2.7% (3)	1.2% (1)	3.2% (79)	0.7% (12)
	Total ****	9.7% (11)	4.8% (4)	8.5% (213)	3.9% (65)

* = significant at $p < 0.05$; *** = significant at $p < 0.001$.

Research-oriented lexical bundles can be used to express real-world activities and experiences. In the current study, student writers use quantification and description bundles more frequently and expert writers use location and procedure bundles more frequently.

Location bundles are used to indicate time and place. Although this category does not contain many different types, location bundles turn out to be the most commonly used ones, and most of them are shared between the two corpora, such as *at the same time* and *at the end of*. A close examination reveals differences in the use of shared bundles. For example, student writers often use *at the end of* to indicate a particular place in their writing, as in (5), and expert writers tend to use it to refer to research stage, as in (6):

- (5) *At the end of* the chapter, there is a description of the procedures of the research. (S18)
- (6) It is only *at the end of* the activity that the instructor combines the students' ideas and briefly explains the conceptual meaning. (E42)

Both groups use description bundles that fit the *noun + of the* structural pattern. These different nouns reveal the different functions they serve in student writing and expert writing. Expert writers use this pattern for more abstract functions, with nouns such as *nature*, *meaning*, *quality*, and *context* describing quality and property. Student writers, on the other hand, use this frame to provide more primary information with nouns such as *basis*, *structure*, and *form*.

Student writers use quantification bundles more frequently than expert writers. This finding is inconsistent with Pan et al. [18], whose conclusion is that novice writers produce fewer quantification bundles than expert writers. A further examination reveals that the two groups employ quantification bundles for different purposes. It seems that student writers attach more importance to detailed quantitative information by using bundles like *the frequency of the* and *more than half of*. Expert writers, on the other hand, use bundles describing more generalized and abstract information such as *the extent to which* and *a high level of*. Hyland [11] concluded that student writers are under pressure to demonstrate their ability to handle research and their familiarity with the subject content. That difference may explain why student writers employ more research-oriented bundles than expert writers and why they use this category to provide more basic and detailed information.

Text-oriented lexical bundles are concerned with the organization of texts and comprise the most substantial proportion in both corpora. Student writers use a significantly higher number but a smaller proportion of text-oriented bundles than expert writers. This massive concentration of text-oriented bundles is in line with previous studies [16,18] and indicates the discursive and evaluative characteristics of soft science language [11].

Framing signals appear to be the most frequent group of text-oriented bundles in both corpora. These bundles are heavily employed to frame arguments by highlighting limitations and specifying cases. There are, respectively, 10 and 9 types of framing bundles among the Top 30 in the SWC and EWC. Bundles such as *on the basis of* and *in the form of* are shared and frequently used by both groups. In line with Hyland's observation [10], many framing signals found in the current study were *preposition + of* structures, e.g., *in terms of the*, *on the basis of*. Both student writers and expert writers use a large number of PP-based bundles beginning with *in*.

Student writers use framing signals to introduce helpful resources when describing their research procedure, such as the software/tool they used or the experienced researcher who helped them make a decision. It seems that such information makes their methodologies sound convincing, as in (7) and (8):

- (7) Then *with the help of* other two teachers who have been teaching English for 20 years, the researcher identifies the relative clause errors. (S19)
- (8) The data are analyzed *with the aid of* AntConc 3.3.5. (S18)

Expert writers use more bundles to set detailed criteria or limitations to their arguments, as in examples (9) and (10):

- (9) *With the exception of* a small-scale study with low-proficiency EFL students (Shehadeh, 2011), previous research has not compared the longer-term effects of writing practice...(E53)
- (10) There are also signs, at least in the case of universities, of national proclivities in the choice of adjectives. (E14)

These expressions specify the special cases in which the argument can be accepted, thus protecting them from direct contradiction with other research findings. In a word, framing signal bundles helps writers make their research methods and conclusions more convincing.

The second most frequently employed subcategory is structuring signals in both corpora. The use of structuring signals makes arguments well managed and organized. Some structuring signals are used to point to additional material to make it more salient [40], and student writers in the present study frequently use bundles like *as shown in table* and *in the following examples*. It is common that further explanations do not immediately follow the tables, figures, and examples due to page layout. Writers need to remind their readers of where to find expected information. These expressions help readers follow their analysis. Student writers also use structuring bundles to summarize their findings, such as *based on the above*. The nouns usually following this bundle include *discussion*, *analysis*, and *results*. Some structuring signals can announce discourse goals [11], and such bundles are connected with different *attended this* structures, allowing writers to build interaction with readers [41]. Student writers use the *noun + active verb* pattern (*this thesis aims to*), whereas experts prefer *noun + of pattern* (*the aim of this*). Another observation is that expert writers use structure bundles to scaffold the text. For example, they use *in the next section* to introduce what they are going to talk about in the next step. Such guiding expressions function as a road map that helps readers follow their writing in the way the writer expects. However, it seems student writers have not realized the importance of stating the purpose of the subsequent sections. Other bundles frequently used in the EWC but rarely found in the SWC are those containing “*research question*”. The use of such bundles is a reflection of reader awareness. Expert writers tend to remind their readers of research questions at different stages in the text. They present the question right before their explanation, helping readers know what can be expected in the following sections, as in (11):

- (11) In order to address *the second research question* of the study, that is, the effect of sociocultural adaptation on production of routines, a first analysis was focused on the cultural congruity factor (RQ2a). (E50)

Transitional signals are used to build additive or contrastive links between elements. These bundles help to maintain the cohesion and coherence of the writing. Transitional bundles have comparable proportions in the two groups, and many of them are the most frequently employed ones, such as *on the other hand* and *as well as the*. Student writers and expert writers both use transitional bundles to establish connections in the text. However, student writers sometimes misuse the bundle *on the other hand*. Instead of using it to indicate a contrary situation or alternative viewpoint (see example 13), students sometimes use it to link two sentences but fail to achieve cohesive writing as they hoped, such as in (12):

- (12) On the one hand, word list, key word list, or adjective list can only offer us a little information about words expressing opinions in reports. *On the other hand*, although concordances analysis is an efficient way to identify opinions in news, it falls short of dealing with larger text. (S3)
- (13) Translingual scholars such as Makoni and Pennycook (2007), Blommaert (2005), Canagarajah (2017a), or Li (2017), *on the other hand*, offer an alternative perspective. (E7)

Resultative signals are used to establish inferential or causative relations between elements. Both groups use bundles containing *result*. Student writers prefer the bundle *the result of the*, and usually use *show*, *indicate*, and *present* after it. Expert writers, on the other hand, use *the results show that* more frequently. Another finding is that both student writers and expert writers favor verbs *find* and *show*, but student writers also use bundles containing *point out*. Furthermore, student writers often use *it is found that* while expert writers use *it was found that*. The use of past tense indicates that the result or finding is reasonable in the specific case, thus opening a space where writers feel free to challenge the conclusion.

Participant-oriented lexical bundles comprise the smallest proportions in the two corpora. Student writers use a more significant number and a higher proportion of bundles than expert writers. Stance bundles in the current study are expressed impersonally and show a connection to the *anticipatory-it* structures, such as *it is important that*, *it is clear that*, and others. Expert writers employ participant-oriented bundles to convey a reluctance to express full commitment. The use of hedges helps writers express opinions with a degree of uncertainty, thus protecting them from the potential disagreement with others. When they offer result interpretation, expert writers seem to express their opinions in a more tentative and cautious way, as illustrated in examples (14):

- (14) *It is possible that* this was due to stronger semantic connections with word groups (animals, food) that were more familiar to the children... (E47)

However, student writers have not realized the importance of hedging bundles and they often use bundles to indicate necessity as illustrated in examples (15)–(16):

- (15) *It is necessary to* study the effects of enhanced model on high school students' incidental noticing. (S7)
 (16) In the two examples above, *it is obvious that* students do not know which conjunctions can lead relative clauses. (S19)

Engagement bundles are employed to engage readers at a certain point in the text. The majority of engagement bundles in the present study used modal words to express the writer's attitude of absolute necessity or importance, e.g., *it should be noted* and others. Student writers use bundles beginning with personal pronouns (e.g., *we can see that*), whereas expert writers do not. Pan and Liu [32] reported that MA writers frequently use *we can see that* to present a proposition based on information from table or figure. Findings from our corpora demonstrated different patterns. Student writers mostly use this bundle to report findings based on reviewing previous studies, as in example (17)–(18):

- (17) As reviewed above, *we can see that* this essential perspective has apparently been embedded within SFL linguists' conception of genre. (S12)
 (18) From the research results, *we can see that* the majority of the respondents are bilingual and diglossic speakers. (S1)

Taken together, student writers and expert writers demonstrate comparable functional proportions. The two groups both employ lexical bundles to help construct propositions, unfold the text, and engage readers in a reader-friendly way. However, student writers tend to focus on detailed information and usually shape findings and conclusions with a high degree of certainty. Expert writers often develop their writing in a way such that readers understand the text as the writer expects. They also carefully express opinions and interpret results with hedging bundles, which creates a space for readers to argue with them.

5. Conclusions

This study conducted a comparison of lexical bundles between Chinese EFL MA theses and expert writers' published articles. Our study extends prior knowledge on learner employment of lexical bundles, in that it investigated theses by the understudied group MA students, comparing it to writings by expert writers. The two academic groups display comparable features in their use of functional categories of lexical bundles. However, differences are more prominent in their structural use.

In line with findings of previous studies on student writing, the result shows that MA theses writing contains significantly more types and tokens of lexical bundles than expert writing. The widespread use of lexical bundles in student writing offers novice writers rich linguistic resources to manage their discourse. In terms of the structural variation of lexical bundles, student writers rely more on clausal bundles and less on phrasal bundles than expert writers. This finding is consistent with Biber et al.'s [39] hypothesis that novice writers need to learn to shift from clausal style to phrasal style. The excessive use of VP-based bundles (35.4% all types, 30.5% of all tokens) reflects how student writers construct

discourse with conversational features. In terms of functional categories, student writers use more types of lexical bundles in all categories than expert writers. The two groups show similar proportions, with text-oriented bundles constituting the most substantial proportion and participant-oriented ones the smallest proportion. Student writers show heavier reliance on research-oriented bundles and participant-oriented bundles than expert writers.

Taken together, the study seems to reveal that although MA students have gone some way towards acquiring conventions of academic writing, they still exhibit many features of less skillful writers. One of the most sophisticated features of MA students is their stance expression. Like the expert writer, they are aware of the importance of *anticipatory-it* pattern and use it to distance themselves from the source of judgment and claims. However, when they emphasize their own findings, the students use themselves as subject (e.g., *we can see that*); in comparison, expert writers largely use nonhuman subjects such as *results*. Another notable difference is that although both groups use bundles to indicate importance, necessity and possibility, student writers offer greater certainty (e.g., *it is obvious that*) while experts are reluctant to overstate and employ hedging expression (e.g., *it is possible that*). Furthermore, there is evidence that student writers reveal insufficient reader awareness in these writing. Expert writers frequently use strategies such as repeating the research questions and introducing contents of next section to help readers better follow the long discourse.

The results of the current study, hopefully, could shed light on future pedagogical practice. Extensive reading and studying the linguistic choice of expert writers can offer student writers rich resources to construct these writing. Empirical studies have shown that explicit instruction could sustainably benefit student writers' writing quality [42]. The expert bundles could be incorporated into writing lessons that might help students promote their expression in a more disciplinarily accepted way. Explicit instruction in the structures, functions, and preferred patterns of lexical bundles in expert writing could be useful for students to identify the appropriate writing style. In addition, incorporating corpus tools like *Antconc* into academic writing lessons may allow students to find out how bundles are used in the surrounding co-text. The large amounts of data provide them with authentic examples of language use. Instructors can guide students to notice features such as stance and encourage them to compare the effects of different expressions.

Despite the useful findings outlined above, future studies may benefit from investigating five-word and six-word lexical bundles if the size of corpora can be enlarged. Longer lexical bundles could be related to different moves across writing; their special function in the sustainable growth of EFL students' academic writing skills deserves further investigation with more data.

Author Contributions: Writing—original draft preparation, S.Z.; writing—review and editing, H.Y. and L.J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The data were collected from the available written documents accessible by the general public. So, this requirement is not relevant.

Informed Consent Statement: The data were collected from the available written documents accessible by the general public. So, this requirement is not relevant.

Data Availability Statement: The data were collected from the available written documents accessible by the general public.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Biber, D.; Johansson, S.; Leech, G.; Conrad, S.; Finegan, E. *Longman Grammar of Spoken and Written English*; Pearson Education: Harlow, UK, 1999.
2. Hyland, K.; Jiang, F.K. Academic lexical bundles: How are they changing? *Int. J. Corpus Linguist.* **2018**, *23*, 383–407. [[CrossRef](#)]
3. Biber, D.; Barbieri, F. Lexical bundles in university spoken and written registers. *Engl. Specif. Purp.* **2007**, *26*, 263–286. [[CrossRef](#)]

4. Biber, D.; Conrad, S.; Cortes, V. If you look at...: Lexical Bundles in University Teaching and Textbooks. *Appl. Linguist.* **2004**, *25*, 371–405. [[CrossRef](#)]
5. Biber, D.; Gray, B. Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *J. Engl. Acad. Purp.* **2010**, *9*, 2–20. [[CrossRef](#)]
6. Cortes, V. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *Engl. Specif. Purp.* **2004**, *23*, 397–423. [[CrossRef](#)]
7. Durrant, P. Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories. *Appl. Linguist.* **2017**, *38*, 165–193. [[CrossRef](#)]
8. Grabowski, Ł. Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *Engl. Specif. Purp.* **2015**, *38*, 23–33. [[CrossRef](#)]
9. Hardy, J.; Römer, U. Revealing disciplinary variation in student writing: A multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora* **2013**, *8*, 183–207. [[CrossRef](#)]
10. Hyland, K. Academic clusters: Text patterning in published and postgraduate writing. *Int. J. Appl. Linguist.* **2008**, *18*, 41–62. [[CrossRef](#)]
11. Hyland, K. As can be seen: Lexical bundles and disciplinary variation. *Engl. Specif. Purp.* **2008**, *27*, 4–21. [[CrossRef](#)]
12. Nesi, H.; Gardner, S. *Genres Across the Disciplines: Student Writing in Higher Education*; Cambridge University Press: Cambridge, UK, 2012.
13. Ädel, A.; Erman, B. Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *Engl. Specif. Purp.* **2012**, *31*, 81–92. [[CrossRef](#)]
14. Chen, Y.H.; Baker, P. Lexical bundles in L1 and L2 academic writing. *Lang. Learn. Technol.* **2010**, *14*, 30–49.
15. De Cock, S. Preferred sequences of words in NS and NNS speech. *Belg. J. Engl. Lang. Lit.* **2004**, *2*, 225–246. Available online: <http://hdl.handle.net/2078.1/75157> (accessed on 1 October 2020).
16. Esfandiari, R.; Barbary, F. A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *J. Engl. Acad. Purp.* **2017**, *29*, 21–42. [[CrossRef](#)]
17. Karabacak, E.; Qin, J. Comparison of lexical bundles used by Turkish, Chinese, and American university students. *Procedia Soc. Behav. Sci.* **2013**, *70*, 622–628. [[CrossRef](#)]
18. Pan, F.; Reppen, R.; Biber, D. Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *J. Engl. Acad. Purp.* **2016**, *21*, 60–71. [[CrossRef](#)]
19. Staples, S.; Egbert, J.; Biber, D.; McClair, A. Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *J. Engl. Acad. Purp.* **2013**, *12*, 214–225. [[CrossRef](#)]
20. Nesi, H.; Basturkmen, H. Lexical bundles and discourse signalling in academic lectures. *Int. J. Corpus Linguist.* **2006**, *11*, 283–304. [[CrossRef](#)]
21. Omidian, T.; Shahriari, H.; Siyanova-Chanturia, A. A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts. *J. Engl. Acad. Purp.* **2018**, *36*, 1–14. [[CrossRef](#)]
22. Abdollahpour, Z.; Gholami, J. Embodiment of rhetorical moves in lexical bundles in abstracts of the medical sciences. *S. Afr. Linguist. Appl. Lang. Stud.* **2019**, *37*, 339–360. [[CrossRef](#)]
23. Cortes, V. The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *J. Engl. Acad. Purp.* **2013**, *12*, 33–43. [[CrossRef](#)]
24. Qi, H.; Pan, F. Lexical bundle variation across moves in abstracts of medical research articles. *S. Afr. Linguist. Appl. Lang. Stud.* **2020**, *38*, 109–128. [[CrossRef](#)]
25. Bychkovska, T.; Lee, J.J. At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *J. Engl. Acad. Purp.* **2017**, *30*, 38–52. [[CrossRef](#)]
26. Shin, Y.K. Do native writers always have a head start over nonnative writers? The use of lexical bundles in college students' essays. *J. Engl. Acad. Purp.* **2019**, *40*, 1–14. [[CrossRef](#)]
27. Vo, S. Use of lexical features in non-native academic writing. *J. Second Lang. Writ.* **2019**, *44*, 1–12. [[CrossRef](#)]
28. Chen, Y.H.; Baker, P. Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Appl. Linguist.* **2016**, *37*, 849–880. [[CrossRef](#)]
29. Römer, U. English in academia: Does nativeness matter? *Anglistik Int. J. Engl. Stud.* **2009**, *20*, 89–100.
30. Wang, Y. As Hill seems to suggest: Variability in formulaic sequences with interpersonal functions in L1 novice and expert academic writing. *J. Engl. Acad. Purp.* **2018**, *33*, 12–23. [[CrossRef](#)]
31. Jaworska, S.; Krummes, C.; Ensslin, A. Formulaic sequences in native and non-native argumentative writing in German. *Int. J. Corpus Linguist.* **2015**, *20*, 500–525. [[CrossRef](#)]
32. Pan, F.; Liu, C. Comparing L1-L2 differences in lexical bundles in student and expert writing. *S. Afr. Linguist. Appl. Lang. Stud.* **2019**, *37*, 142–157. [[CrossRef](#)]
33. Bestgen, Y. Comparing lexical bundles across corpora of different sizes: The zipfian problem. *J. Quant. Linguist.* **2020**, *27*, 272–290. [[CrossRef](#)]
34. Anthony, L. AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available online: <https://www.laurenceanthony.net/software> (accessed on 1 October 2020).
35. Paquot, M.; Granger, S. Formulaic language in learner corpora. *Annu. Rev. Appl. Linguist.* **2012**, *32*, 130–149. [[CrossRef](#)]

36. Granger, S. Prefabricated patterns in advanced EFL writing: Collocations and formulae. In *Phraseology: Theory, Analysis and Applications*; Cowie, A., Ed.; Oxford University Press: Oxford, UK, 1998; pp. 145–160.
37. Biber, D.; Conrad, S.; Cortes, V. Lexical bundles in speech and writing: An initial taxonomy. In *Corpus Linguistics by Lune: A Festschrift for Geoffrey Leech*; Wilson, A., Rayson, P., McEnery, T., Eds.; Peter Lang: Frankfurt, Germany, 2003; pp. 71–93.
38. Biber, D.; Gray, B. *Grammatical Complexity in Academic English: Linguistic Change in Writing*; Cambridge University Press: Cambridge, UK, 2016.
39. Biber, D.; Gray, B.; Poonpon, K. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Q.* **2011**, *45*, 5–35. [[CrossRef](#)]
40. Hyland, K. Bundles in academic discourse. *Annu. Rev. Appl. Linguist.* **2012**, *32*, 150–169. [[CrossRef](#)]
41. Jiang, F.K.; Wang, F. 'This is because . . . ': Authorial practice of (un)attending this in academic prose across disciplines. *Aust. J. Linguist.* **2018**, *38*, 162–182. [[CrossRef](#)]
42. Zhang, L.; Zhang, L.J. Fostering stance-taking as a sustainable goal in developing EFL students' academic writing skills: Exploring the effects of explicit instruction on academic writing skills and stance deployment. *Sustainability* **2021**, *13*, 4270. [[CrossRef](#)]