

Vol. 15, No. 3  
pp. 79-105  
July &  
August 2024

Received: 29 November 2022  
Received in revised form: 21 February 2023  
Accepted: 5 March 2023

## Are Phrases Frozen? Investigating Lexical Bundles in Non-native and Native Reports in MICUSP

Amir Mashhadi\*<sup>1</sup>  & Marjan Qanavati<sup>2</sup>

### Abstract

As one of the most prominent components of native-like language use, lexical bundles play a crucial role in academic texts. Considering the amount of written language produced by writers, this study explored lexical bundles in non-native and native academic reports. To that end, 100 sample reports of non-native and native writers taken from the Michigan Corpus of Upper-level Student Papers (MICUSP) were selected and analyzed to find non-native and native writers' use of lexical bundles regarding frequency, structure, and function. Results indicated that native writers generally used a more extensive variety of lexical bundles in their reports. Regarding the structure of bundles, both non-native and native writers tended to employ more noun phrases with of-phrase fragments. In terms of the functions of bundles, both groups of writers tended to use more research-oriented bundles, followed by text-oriented and participant-oriented bundles, respectively. Findings highlight the value of fixed expressions in producing coherent academic texts by proficient writers, hence their importance in materials design and second language instruction.

**Keywords:** academic writing, lexical bundles, report writing, non-native and native writers of English, MICUSP

<sup>1</sup> Corresponding Author, Department of English Language and Literature, Faculty of Letters and Humanities, Shahid Chamran University of Ahvaz, Ahvaz, Iran; Email: [Mashhadi.scu@gmail.com](mailto:Mashhadi.scu@gmail.com); ORCID: <https://orcid.org/0000-0001-9261-3151>

<sup>2</sup> Department of English Language and Literature, Faculty of Letters and Humanities, Shahid Chamran University of Ahvaz, Ahvaz, Iran

## **1. Introduction**

A key measure of academic success is proficiency in academic writing (Kellogg & Raulerson, 2007). Academic writing plays a significant role in helping students fulfill their academic requirements, such as doing papers, assignments, and exercises, necessitating due attention to its development in formal educational settings (Maniati et al., 2020). Lexical bundles, as fundamental components of academic texts, are mainly represented as recurring combinations of three or more words fulfilling specific functions in a corpus of natural language (Wood, 2010). Such recurring combinations have already been considered under a variety of rubrics, including formulaic language, fixed expressions, pre-fabricated patterns, lexical phrases, multi-word units, phrasal vocabulary units, formulas, frozen phrases, clusters, routines, or n-grams (Adel & Erman, 2012; Biber, 2006; Biber & Barbieri, 2007; Hinkel, 2016). Basically, lexical bundles consist of combinations of words following each other more recurrently in a register than expected by chance, enabling writers to shape their texts and create a sense of naturalness and distinctiveness of text meanings (Hyland, 2009). Moreover, the use of lexical bundles as highly prevalent and significant word clusters in academic discourse is deemed to be a crucial element of fluent linguistic production and native-like proficiency. Likewise, Sykes (2017) argued that learners and teachers of English need to identify and use lexical bundles properly to replicate the appropriate language in the academic genre.

Lexical bundles play an indispensable part in discourse and language use, but their variation across different disciplines makes them even more significant for academic writing purposes (Hyland, 2012). It is thus essential that writers use lexical bundles typical of an academic discipline. The absence or inappropriate use of such expressions might indicate inadequate language experience. Lexical bundles are indeed building blocks of native and native-like language fluency which has been considered to be a significant factor in learner development (Ellis et al., 2008; Wray, 2002). As maintained by Hinkel (2017), non-native and native speakers of many languages acknowledge that using lexical bundles in formulaic sequences is more fitting than putting words in arbitrary sequences.

## **2. Literature Review**

Lexical bundles have gained particular attention in academic literacy research as target constructs for developing fluency and whole language competence (e.g., Boers et al., 2006; Hunston, 2002; Nesselhauf, 2005; Serrano et al., 2015; Staples et al., 2013; Wood, 2010). There is also some evidence that lexical bundles have a processing advantage for advanced and intermediate second language (L2) speakers, suggesting a direct correlation between knowledge of lexical bundles and L2 proficiency (e.g., Conklin & Schmitt, 2008; Yeldham, 2018). Therefore, advanced and intermediate L2 speakers are more likely to possess an internalized formulaic sequence knowledge than low-proficient speakers (Serrano et al., 2015; Northbrook & Conklin, 2019; Valsecchi et al., 2013).

More specifically, to discover the patterns and provide the source of evidence for characterizing the nature, structure, and language use, corpus linguistics considers the compilation and analysis of collections of spoken and written texts in real-life discourse (Schmitt & Rodgers, 2020). As one of the early studies in corpus linguistics, De Cock et al. (1998) analyzed the formulaic competence of advanced French learners of English as a foreign language (EFL) in a corpus of informal speech. The findings of this study and some subsequent studies showed that advanced EFL learners tended to employ multi-word combinations, even more than native speakers did (Ahmadi et al., 2013; Amirian et al., 2013; Öztürk & Köse, 2016; Bychkovska & Lee, 2017; Vo, 2019; Appel & Murray, 2020). Conversely, other studies have indicated that native writers employed a larger variety of bundles compared to non-native writers (Chen & Baker, 2010; Bal, 2010; Ädel & Erman, 2012; Ucar, 2017). Alternatively, several corpus-based studies have examined how lexical bundles were used across disciplines, registers, and among groups of writers with different first languages (L1s) and writing proficiency (Ghafar Samar et al., 2015; Hassanzadeh & Tamleh, 2023). In a similar research strand, the use of lexical bundles was examined in Ph.D. and Master's theses (Hyland, 2008), research article introductions (Cortes, 2013), WhatsApp conversations between L1 and L2 speakers (Fitriati & Wahyuni, 2018), L1 and L2 English teachers' talk (Lorenna et al., 2020), EFL students' interactions (Neno & Agustien, 2016) and quite recently, applied linguistics and pharmaceutical sciences (Ren, 2021), L2 business emails (Xia et al., 2022) and academic writing of L2 English learners in a testing context (Kim & Kessler, 2022).

As an overarching area of academic writing, report writing at higher education provides students with the foundation skills in writing various types of technical documents such as business letters, memos, resumes, advertisements, posters, minutes of meetings, proposals, reports and manuals (Gould & Losano, 2008). A good report communicates ideas, data, and conclusions efficiently; therefore, an understanding of technical writing principles can benefit students who intend to go into any field of work (Raus et al., 2019). Writing a report also serves as a primary means of communication at work, so learning to write a report prepares the students for future practical writing tasks they are likely to encounter as part of their careers (Riordan & Pauley, 2002). Thus, report texts need to be mastered by the students so that they can report on their observations of the environment (Estes, 2004; Pedersen & Liu, 2003).

Concerning the significance of report writing in academic contexts, some recent studies have also investigated how academic report writing can be adapted to workplace needs. The results indicate that the key to improving writing products in the classroom lies in implementing effective instructional methods (Bourelle, 2012; Cilliers, 2012; Linsdell & Anagnos, 2011; Tazl et al., 2012). Despite the significance of report writing in academic contexts, and the large number of reports written by non-native and native writers, little attention has already been paid to lexical bundles in this text type, and further studies seem indispensable. The ability to use formulas appropriately in written reports is considered to be crucial to achieving idiomatic production (Prodromou, 2009). Hence, the question arises as to what lexical bundles learners require to take an effective shortcut into their respective discourse communities. However, it is generally held that the methods used to identify lexical sequences are not entirely satisfactory (Wray, 2008). For this reason, Biber (2009) suggested that researchers adopt various methodological approaches to identify lexical bundles of academic discourses.

Accordingly, this study was intended to investigate the degree of similarities and differences between the non-native and native uses of lexical bundles in English academic reports in the Michigan Corpus of Upper-level Student Papers (MICUSP). Quantitative and qualitative analyses were thus carried out on two datasets in order to detect the recurrent word combinations in the students' reports. In order to address the research objectives, the following research questions stand out:

1. What kinds of lexical bundles are used in the sample reports, taken from

MICUSP, produced by non-native and native writers of English?

2. What particular functions and structures are associated with the lexical bundles in these sample reports?

### 3. Methodology

#### 3.1. Data Source

The data for this study came from MICUSP, a compilation of 829 papers with about 2.6 million words from 16 different disciplines (<https://elicorpora.info>). The dataset comprises seven academic written genres of creative writing, argumentative essays, critique/evaluation, proposals, reports, research papers, and response papers. All papers in this dataset were written by final-year non-native and native English undergraduate and graduate students whose papers received an A grade. As the focused materials for this study, two sub-groups of 50 non-native and 50 native writers' report texts with slightly over 300000 words were selected. This study aimed at selecting both non-native and native texts included in MICUSP; however, the total number of non-native reports was only 57, of which seven contained only tables and non-textual content, leaving 50 non-native texts for the next stage of the study. These texts were written by writers of diverse nationalities and in 12 disciplines (Education, Linguistics, Psychology, Sociology, Political Science, Economics, Biology, Nursing, Natural Resources and Environment, Civil and Environmental Engineering, Industrial and Operations Engineering, and Mechanical Engineering). In a similar research strand, the native report texts were selected from the same disciplines. This size of language was assumed to be small enough to manage, though not large enough to allow for reliable generalizations, as this was not the intention behind the study (see Table 1).

**Table1**

*Description of the Selected Dataset*

|                  | No. of selected texts | No. of words | No. of reports in MICUSP |
|------------------|-----------------------|--------------|--------------------------|
| Non-native texts | 50                    | 144372       | 57                       |
| Native texts     | 50                    | 156441       | 307                      |
| Total            | 100                   | 300753       | 364                      |

### 3.2. Procedure

Given that the whole dataset in this study comprised almost one-third (over 300000 words) of the criterion suggested by Biber et al. (1999), the researchers primarily set the standard on any word clusters appearing at least three times in at least three texts. However, as acknowledged in other studies, among word clusters, four-word clusters fulfill a greater range of functions than three-word clusters and occur more often than five-word clusters (Biber et al., 1999). Hence this study focuses on four-word clusters as the basis of analysis of the report texts.

AntConc 3.5.8 (Anthony, 2019), a concordance program developed at the Centre for English Language Education, Waseda University (Japan) was used to retrieve lexical bundles from the plain texts, not including graphics, formulas, references, tables, and figures. The function Clusters/N-Grams of this software was employed to identify word clusters in the dataset, using frequency as the first identification criterion, which was considered to be at least three times in this study. In addition to frequency, the range as the second criterion, characterizing the distribution of bundles in the report files, was considered in order to avoid idiosyncratic use of language (Biber & Barbieri, 2007). Given the size of the selected data, the range in this study was also set on three texts.

After the clusters were identified, bundles were searched and categorized structurally and functionally according to the structural taxonomy suggested by Biber et al. (1999), and the functional taxonomy proposed by Salazar (2014). Although several functional taxonomies were developed by some other researchers (e.g., Biber et al., 2004; Hyland, 2008a), this study followed Salazar's (2014) taxonomy which is an advanced version of the taxonomy suggested by Hyland (2008a).

Concerning the functions of bundles, the researchers assigned each bundle the specific function it served. There were, however, bundles that served multiple functions in various contexts (Conrad & Biber, 2005), making it difficult to assign each a single function. In order to address the issue of multifunctionality, we followed Biber et al. (2004) and Hyland (2008a), who investigated the concordances of multifunctional bundles and classified them according to their most frequent function. For instance, *can be seen in* which can function either as structuring (I) or framing (II) was classified as a structuring signal due to its more frequent occurrence in this dataset.

(I) Results *can be seen in* Tables 2.1 and 2.2. As in study 1, participants with

at least one status-quo response rated the questions more complex ( $t = -1.536$ ,  $df = 47$ ,  $p = .131$ ).

(II) When the brain is organized by hormones early in life, the effects are permanent and *can be seen in adults*.

To establish intercoder agreement, each author with the help of the research assistant categorized the bundles, and reached approximately 95% agreement for structural types, and 88% for functional types. The remaining discrepancies were discussed until complete agreement was reached.

#### 4. Results and Discussion

Quantitative and qualitative analyses were carried out on the obtained data to identify the possible similarities and differences in recurrent word combinations in terms of their frequency, structure, and function, as explained below.

##### 4.1. Frequency of Bundles

The frequency of bundles strongly depends on the definition of bundles, since different frequency cut-off regarding a particular corpus size would result in a different number of bundles (Conrad & Biber, 2005). Having set the cut-off frequency in Antconc, the researchers identified a list of 459 total bundles of varying lengths; 249 in the native and 210 in the non-native datasets. However, the first inspection of these bundles indicated that some of the bundles could not be included in the dataset due to their specific characteristics. Following Salazar (2014), topic-specific bundles (e.g., the reaction mixture was), bundles consisting of function words only (e.g., to that of the), and meaningless bundles (e.g., i x m going) were excluded from the extracted list. Exclusion criteria were then applied to narrow the original list to 192 in the native and 153 in the non-native corpora.

**Table 2**  
*Frequency of Bundles in the Dataset*

|                         | Non-native | Native | Total  |
|-------------------------|------------|--------|--------|
| Number of texts         | 50         | 50     | 100    |
| Number of words         | 144372     | 156441 | 300753 |
| Number of bundle types  | 153        | 192    | 345    |
| Number of bundle tokens | 729        | 967    | 1696   |

The comparison between the two datasets considering the frequency of bundles revealed the greater frequency and diversity of bundles in the native dataset, as shown in Table 2. This means that native writers generally employ a wider range of bundles. The first ten most frequent bundles used in the native texts were *in the U S, in the United States, as a result of, the university of Michigan, as well as the, it is important to, has the potential to, at the same time, at the university if, in addition to the, in the form of, on the other hand, and the end of the*. However, non-native writers used *on the other hand, in the United States, at the end of, as well as the, in the case of, to the fact that, at the same time, in the context of, the end of the, in the process of, it is important to, and it was found that* more frequently than other bundles. *On the other hand* was the most frequent cluster, occurring 46 times in 28 texts per 300000 words in the whole data. These findings, consistent with those of Adel and Erman (2012), Chen and Baker (2010), and Erman (2009), suggest that non-native speakers of English employ far less diverse bundles than native speakers. However, as not all the studies in the related literature agree with the above studies, some other studies have demonstrated that the non-native groups produced a larger and broader number of bundles than their native counterparts (Hyland, 2008a & b; Pérez-Llantada, 2014; Pan et al., 2016; Römer, 2009). This calls for careful consideration of the results of the current and earlier studies.

#### 4.1.1. Shared Bundles

Within the native and non-native reports, 36 bundles occurred in both, as displayed in Table 3. Generally, 18.75% of the bundles in the native reports were also present in the non-native reports. These bundles occurred with different frequencies and over a different range of texts. A comparison of the shared four-word bundles in this study with those in Adel and Erman (2012) and Chen and Baker (2010), who worked on university students' essays, revealed seven shared four-word bundles: *on the other hand, is one of the, one of the most, as well as the, as a result of, in the form of, in the case of*.

**Table 3**  
*Shared Bundles Found in the Dataset*

|   | Bundles                           | F- Non-native | F- Native |
|---|-----------------------------------|---------------|-----------|
| 1 | <i>in the United States</i>       | 25            | 24        |
| 2 | <i>as a result of</i>             | 6             | 19        |
| 3 | <i>the university of Michigan</i> | 6             | 17        |
| 4 | <i>as well as the</i>             | 13            | 16        |



| Bundles                                | F- Non-native | F- Native |
|----------------------------------------|---------------|-----------|
| 5 <i>it is important to</i>            | 9             | 14        |
| 6 <i>at the same time</i>              | 10            | 12        |
| 7 <i>in addition to the</i>            | 6             | 12        |
| 8 <i>in the form of</i>                | 7             | 12        |
| 9 <i>on the other hand</i>             | 34            | 12        |
| 10 <i>the end of the</i>               | 10            | 12        |
| 11 <i>the beginning of the</i>         | 6             | 11        |
| 12 <i>the rest of the</i>              | 6             | 11        |
| 13 <i>in the case of</i>               | 13            | 10        |
| 14 <i>to the fact that</i>             | 11            | 10        |
| 15 <i>one of the most</i>              | 6             | 9         |
| 16 <i>the fact that the</i>            | 8             | 9         |
| 17 <i>at the end of</i>                | 6             | 8         |
| 18 <i>due to the fact</i>              | 6             | 8         |
| 19 <i>in terms of the</i>              | 6             | 7         |
| 20 <i>the role of the</i>              | 5             | 7         |
| 21 <i>a part of the</i>                | 7             | 6         |
| 22 <i>are more likely to</i>           | 8             | 6         |
| 23 <i>at the beginning of</i>          | 8             | 6         |
| 24 <i>the nature of the</i>            | 5             | 6         |
| 25 <i>the needs of the</i>             | 5             | 5         |
| 26 <i>it is difficult to</i>           | 7             | 4         |
| 27 <i>the details of the</i>           | 3             | 4         |
| 28 <i>the relationship between the</i> | 7             | 4         |
| 29 <i>to the lack of</i>               | 3             | 4         |
| 31 <i>a large number of</i>            | 6             | 3         |
| 32 <i>has to do with</i>               | 4             | 3         |
| 33 <i>in the process of</i>            | 9             | 3         |
| 34 <i>is one of the</i>                | 8             | 3         |
| 35 <i>the development of a</i>         | 6             | 3         |
| 36 <i>the purpose of this</i>          | 7             | 3         |

#### 4.2. Structural Analysis

The second research question examined the structural categories used in lexical bundles in the report texts in MICUSP (see Table 4).

**Table 4**  
*Structural Analysis of Four-Word Bundles*

| Structural categories                             | Non-native |       | Native |       |
|---------------------------------------------------|------------|-------|--------|-------|
|                                                   | F          | %     | F      | %     |
| 1 Noun phrase with of-phrase fragments            | 27         | 17.65 | 39     | 20.31 |
| 2 Noun phrase with other post-modifiers fragments | 6          | 3.92  | 7      | 3.65  |
| 3 Prepositional phrase with of-phrase fragments   | 27         | 17.65 | 26     | 13.54 |
| 4 Other prepositional phrases (fragments)         | 22         | 14.38 | 33     | 17.19 |
| 5 Anticipatory it + verb phrase/adjective phrase  | 10         | 6.53  | 8      | 4.16  |

| Structural categories |                                              | Non-native |       | Native |       |
|-----------------------|----------------------------------------------|------------|-------|--------|-------|
|                       |                                              | F          | %     | F      | %     |
| 6                     | Passive verb + prepositional phrase fragment | 9          | 5.88  | 9      | 4.68  |
| 7                     | Copula be + noun phrase/adjective phrase     | 14         | 9.15  | 12     | 6.25  |
| 8                     | (Verb phrase +) that-clause fragment         | 2          | 1.30  | 5      | 2.60  |
| 9                     | (Verb/adjective +) to-clause fragment        | 9          | 5.88  | 19     | 9.89  |
| 10                    | Adverbial clause fragment                    | 1          | 0.65  | 2      | 1.04  |
| 11                    | Pronoun/noun phrase + be(+...)               | 2          | 1.30  | 1      | 0.52  |
| 12                    | Other expressions                            | 23         | 15.03 | 31     | 16.14 |

Despite bundles being fragments of phrases and clauses and structurally incomplete (Biber et al., 1999; Cortes, 2004), Biber et al. (1999) have suggested a taxonomy based on several basic structural types. Bundles in the two groups showed slight variations in relation to their grammatical types. As revealed by Table 4, both groups of writers incorporated more noun phrases with of-phrase fragments (20.31% and 17.65% of bundles in native and non-native lists, respectively). In the native texts, this is followed by other prepositional phrase fragments and other expressions, which cover 17.19% and 16.15% in that order. However, the non-native writers employed noun phrases with of-phrase fragments and prepositional phrases with of-phrase fragments with the same frequency (17.65%). Other structures were also variably used in the reports. The frequent use of noun and prepositional phrase bundles in this dataset is consistent with some other studies (e.g., Biber et al., 1999; Biber et al., 2004; Byrd & Coxhead, 2010; Hyland, 2008a; Simpson-Vlach & Ellis, 2010) showing academic writing as being noun-centric (Byrd & Coxhead, 2010). Hyland (2008) claims that bundles in academic writing commonly occur as parts of noun or prepositional phrases ending with prepositions, articles, and complimentizers such as *that*. The structural category of *prepositional phrases with of fragments* (e.g., *as a result of*, *in addition to the*, *at the beginning of the*, *at the end of*, *in the form of*) makes up 13.54% and 17.65% of native and non-native bundles, respectively. As suggested by Biber et al. (1999) and Hyland (2008a), these phrases are commonly employed by academic writers to denote logical relationships between propositional elements.

Other prepositional-phrase fragments account for 17.19% ( $n = 33$ ) in the native and 14.38 ( $n = 22$ ) of non-native lists. They typically refer to different sections within a text providing text-reflexive explanations, which facilitate comprehension and guide readers through the text (*In this paper*); many bundles with the preposition *in* function as framing devices (*in the sense that*). Others are used to describe procedure (*in the process of*), quantification (*with the number of*), to identify place (*in the middle*

of), and to signal the objectives (*to fully understand the*).

Other structural categories with smaller frequencies are employed to communicate the writer's appraisal of possibility (*it is likely that, more likely to have*) or importance (*it is necessary to*). They serve to describe things (*is the lack of*), to indicate categories (*is one of the*), to draw the attention of readers (*is interesting to note*), to express objectives (*to make sure that*), and those with predicative adjectives preceding the to-clause to convey engagement with the readers (*important to note that, as can be seen*), to make comparisons (*as compared to the*), and to indicate reason (*this is because the*).

#### 4.3. Functional Analysis of Bundles

In terms of functions, the extracted bundles were classified based on Salazar's (2014) modified version of Hyland's (2008a) functional taxonomy. Table 5 provides the functional description of the bundles identified in the data and their corresponding frequencies.

**Table 5**  
*Functional Analysis of Four-Word Bundles*

| Function                            | Non-native |              | Native     |              |
|-------------------------------------|------------|--------------|------------|--------------|
|                                     | F          | %            | F          | %            |
| <b>Research-oriented bundles</b>    | <b>74</b>  | <b>48.36</b> | <b>106</b> | <b>55.20</b> |
| Location                            | 17         | 11.11        | 25         | 13.02        |
| Procedure                           | 16         | 10.45        | 11         | 5.72         |
| Quantification                      | 6          | 3.92         | 19         | 9.89         |
| Description                         | 22         | 14.37        | 40         | 20.83        |
| Grouping                            | 12         | 7.84         | 11         | 5.72         |
| <b>Text-oriented bundles</b>        | <b>64</b>  | <b>41.83</b> | <b>59</b>  | <b>30.72</b> |
| Additive                            | 7          | 4.57         | 8          | 4.17         |
| Comparative                         | 0          | 0.00         | 1          | 0.52         |
| Inferential                         | 7          | 4.57         | 14         | 7.29         |
| Causative                           | 3          | 1.96         | 3          | 1.56         |
| Structuring                         | 11         | 7.18         | 2          | 1.04         |
| Framing                             | 12         | 7.84         | 10         | 5.20         |
| Citation                            | 4          | 2.61         | 6          | 3.12         |
| Generalization                      | 3          | 1.96         | 1          | 0.52         |
| Objective                           | 18         | 11.76        | 16         | 8.33         |
| <b>Participant-oriented bundles</b> | <b>15</b>  | <b>9.80</b>  | <b>25</b>  | <b>13.02</b> |
| Stance                              | 11         | 7.18         | 21         | 10.93        |
| Engagement                          | 4          | 2.61         | 3          | 1.56         |
| Acknowledgment                      | 0          | 0.00         | 1          | 0.52         |

Table 5 shows that the proportion of using the three main functional categories is almost similar in both groups. This, in turn, indicates that both native and non-native writers tended to use more research-oriented bundles, which covered almost half of the bundles in each dataset, with 55.20% in the native, and 48.36% in the non-native texts. These numbers support the findings by Hyland (2008a), who showed the domination of research-oriented bundles in the corpus of research articles. Text-oriented and participant-oriented functions were in second and third place, respectively. The predominance of research-oriented bundles in the sample reports taken from MICUSP, similarly to the findings from other science corpora, stems from the necessity of relaying in-depth information about the research procedures in order to make any methodology replicable and to convince the reader of the accuracy and validity of the findings reported (Allen, 2009; Hyland, 2008a).

As regards the ranking of the functional subcategories, however, slight differences were identified between the native and non-native texts. The five most frequent functions of native bundles are description (40 types, 20.83%), location (25 types, 13.02%), stance (21 types, 10.93%), quantification (19 types, 9.89%), and objective (16 types, 8.33%). In comparison, the five most frequent functions of the non-native bundles are description (22 types, 14.37%), objective (18 types, 11.76%), location (17 types, 11.11%), procedure (16 types, 10.45%), and the two subcategories of grouping and framing (12 types, 7.84%) with similar numbers. These findings affirm the results of previous research (e.g., Chen & Baker, 2010; Salazar, 2010), showing that the bundles used by native writers were more research-oriented and less participant-focused. In Salazar's study (2010), 51.3% of the bundles were research-oriented, 42.4% text-oriented, and 6.3% participant-oriented, close to the findings of this study. The referential, discourse, and stance bundles in Biber's taxonomy comprised 60%, 21%, and 19%, respectively. As a different finding, even though Biber's referential bundles, corresponding to research-oriented bundles in Hyland's taxonomy, have been described as the most frequent functional category in academic texts in several studies (e.g., Biber, 2009; Biber & Barbieri, 2007; Chen & Baker, 2010; Salazar, 2010), this category was followed by stance bundles conforming participant-oriented bundles in Hyland's taxonomy in other studies (e.g., Biber, 2009; Biber & Barbieri, 2007; Juknevičienė, 2009). On the contrary, text-oriented bundles were identified as the most frequent functional category in other studies (e.g., Pan et al., 2016). In what follows, the functional characteristics of the retrieved bundles with their structural correlates will be discussed.

#### 4.3.1. Research-Oriented Bundles

Among research-oriented bundles, description and location bundles constituted the lion's share, accounting for 20.83% ( $n = 40$ ) and 13.02% ( $n = 25$ ) of bundles in the native reports. Similarly, these two functions were the most frequently used bundles in the non-native reports, with 14.37% ( $n = 22$ ) and 11.11% ( $n = 17$ ), respectively. Description bundles are used to describe the research entities (*the lack of the*) or contexts (*the existence of the*), specifying the aspects of models, equipment, and materials (*the development of, the existence of, the nature of the, the capacity of the, to the development of, in the role of, is the lack of, there is a lack, of the existence of, the details of the, the quality of the*), and location bundles indicate time, place, extremity and directions (*at the same time, at the end of the, at the beginning of the, in the world and, of the university of, at the age of, for the first time, at the very beginning, the top of the, the middle of the, at the time of, by the end of*). Both description and location bundles generally take the form of noun phrases + preposition fragments, and prepositional phrases.

The other subcategories of research-oriented bundles that occur in smaller numbers in comparison to description and location, are quantification (*a great deal of, the majority of the, a large number of, a large proportion of, with the number of, the extent to which, that many of the, the vast majority of, and to a lesser, by the number of, is one of the, to a lesser extent*), procedure (*in the process of, through the use of, the role of the, to account for the, is a way to, as a means of*), and grouping (*a part of the, in a variety of, the rest of the, is one of the, one of the ways*). Even though they are less frequent, they contribute to documenting the research process accurately by identifying procedures, and indicating measures, quantities, amounts, and proportions.

In brief, the use of research-oriented bundles reflects writers' preoccupation with producing an unbiased account of the research procedures, and the subsequent data analysis in a verifiable and reproducible manner. This supports Hyland's (2008a) argument that the use of research-oriented bundles "emphasizes the empirical over the interpretive, minimizing the presence of researchers and contributing to the strong claims of the sciences" (p. 15).

#### 4.3.2. Text-Oriented Bundles

Placing the second of the three main functional categories in this data analysis is

text-oriented bundles. These bundles, corresponding with the text organization and its meaning as a message or argument, were used to create additive links, compare and contrast elements, signal inferences, mark cause and effect relations, situate arguments by specifying restrictive conditions, and cite sources and show writers' goals (Salazar, 2014). Results indicated that the functional distribution of text-oriented subcategories is not similar in the two text groups. As stated earlier, the text-oriented bundles ranked second, accounting for 30.72% of native and 41.83% of non-native bundles types. This finding contradicts Salazar (2014) in that text-oriented bundles make the most widely used category among the three main functional categories, constituting nearly half of bundle types and tokens. This is probably attributed to the differences in the text types in the two corpora representing each register.

The most frequent bundles in this category are objective bundles used to demonstrate writers' aims. They constitute the predominant text-oriented bundles in the two groups making up 8.33% ( $n = 16$ ) of native and 11.76% ( $n = 18$ ) of non-native bundles. Objective signals are realized by passive structure. However, there are instances of noun phrase + of, prepositional phrases, and noun phrases + post modifier fragments as the following (e.g., *are expected to be*, *the purpose of this*, *in order to avoid*, *in this paper I*, *to make sure that*, *for the purpose of in order to avoid*, *this paper will focus*, *to keep up with*, *to look at the*, *to be able to*, *to fully understand the*, *for the purposes of*).

The next most frequently used functional category across the native reports refers to the inferential bundles associated with bundles that are used to indicate conclusions of the study and inferences the readers are asked to draw from the arguments (Hyland, 2008a). Inferential bundles constituted the second most frequent text-oriented bundles, accounting for 6.77% ( $n = 14$ ) of native bundles. By comparison, they occur less frequently in the non-native bundles, with 4.57% ( $n = 7$ ). The structural analysis revealed that inferential bundles take the form of prepositional phrase structures, noun phrases, to clause fragments (e.g., *the relationship between the*, *the results of the*, *be attributed to the*, *is a result of*, *was found to be*). Another subcategory of text-oriented bundles refers to framing signals which accounted for 5.20% ( $n = 10$ ) of native bundles and 7.84% ( $n = 12$ ) of non-native bundles. Framing bundles "are used to focus readers on a particular instance or to specify the conditions under which a statement can be accepted" (Hyland, 2008a, p. 16). They frequently take the form of prepositional phrase structures and

passive structures (e.g., *from the perspective of*, *are based on the*, *in the form of*, *in terms of the*, *in the sense that*, *to the fact that*).

Additive signals constitute the following subcategory of text-oriented bundles, making up 4.17% ( $n = 8$ ) of native and 4.57% ( $n = 7$ ) of non-native bundles. Additive signals are used to establish additive (*in addition to the*) or contrastive links between elements (*on the other hand*) and clarification (*in other words*, *there*). A comparison of bundles in both corpora shows that these signals generally take the form of prepositional phrases. Structuring signals, which constituted 7.18% ( $n = 11$ ) of non-native bundles, formed the next most frequent bundles after objectives. By comparison, they rarely occurred in the native bundles, with only 1.04% ( $n = 2$ ). According to Salazar (2014, p. 102), these bundles “work to facilitate comprehension by providing text-reflexive explanations and guiding readers through the text” (e.g., *in the previous section*, *in this paper I*, *of this study was*, *the purpose of this*, *are included in*). The less frequent use of structuring bundles by native writers can be very revealing in that non-native writers employ bundles that are characteristic of non-native language use.

The other text-oriented categories including additive (*on the other hand*, *as well as the*, *in addition to the*), citation (*is defined as*, *they also found that*, *studies have shown that*, *have been shown to*, *research has shown that*), structuring (*can be seen in*, *in this paper the*), causative (*as a result of*, *this is because the*, *have an effect on*, *play a role in*), comparative (*as compared to the*), and generalization (*little is known about*) appeared in lower frequencies than the others. However, they fulfilled important functions and were used to signal cause and effect relations, cite research resources and support data, compare and contrast elements, and also indicate facts or statements that were generally accepted.

#### 4.3.3. Participant-Oriented Bundles

As said earlier, the participant-oriented bundles are located in the third place of the three main functional categories. Writers can communicate their views on findings and conclusions while also creating a proper interaction with their readers through the use of such types of bundles (Hyland, 2008a). A dialogic interaction is indeed established between the writer and the reader through the participant-oriented bundles. Compared to the two other functional categories, participant-oriented bundles appear less frequently, characterizing only 13.02% of native and 9.80% of



non-native bundles. This finding agrees with other studies (e.g., Cortes, 2004; Hyland, 2008; Salazar, 2014), indicating that the participant-oriented bundles are associated with the lowest proportion of the three main functional categories.

Consistent with Salazar (2014) and Cortes (2006), most participant-oriented bundles in the reports include lexical sequences used to express stance. Stance markers making up 10.93% ( $n = 21$ ) of native and 7.18% ( $n = 11$ ) of non-native bundles denote meanings such as probability, possibility, certainty, and importance, and assist writers in communicating their degree of confidence (Salazar, 2014) (e.g., *it is possible that, is very different from, more likely to have, are more likely to, it is clear that, important to note that, it is likely that, it is interesting to*). The second subcategory of the participant-oriented bundles refers to engagement markers which occur much less frequently than stance markers. These bundles “seek to involve readers in the developing argument by addressing them directly, requesting them to focus on specific points and to see things in a particular way” (Salazar 2014, p. 106) (e.g., *it is difficult to, it is important to*). Engagement markers make up 1.56% ( $n = 3$ ) of the native bundles and 2.61% ( $n = 4$ ) of the non-native bundles (e.g., *it can be seen, it should be noted, it is necessary to, as can be seen*). Acknowledgment is the last subcategory of participant-oriented bundles. Bundles of this subcategory enable writers “to thank individuals or entities for financial assistance or the provision of experimental materials” (Salazar, 2014, p. 106). Acknowledgments, with the least frequency in the native reports with just one occurrence in the whole data, account only for 0.52% ( $n = 1$ ) in native reports (e.g., *I would like to*).

This result corresponds to Chen and Baker (2010), and Ädel and Erman's studies (2012), in which participant-oriented bundles accounted for the least frequently used functional category. The current study, however, contradicts the results obtained in some previous research (e.g., Biber, 2009; Biber & Barbieri, 2007; Juknevičienė, 2009) in which the participant-oriented bundles appeared more often than the text-oriented or discourse organizing bundles.

## 5. Conclusion and Pedagogical Implications

Lexical bundles, as building blocks of native-like language discourse, have been viewed as a significant index of learner development (Ellis et al., 2008; Wray, 2002). Considering the significance of lexical bundles in learner writing, and the



amount of written language produced by non-native writers, this study explored the use of lexical bundles in non-native and native academic report texts. Sample reports of writers taken from MICUSP were analyzed to find the similarities and differences between non-native and native writers' use of lexical bundles regarding their frequency, structure, and function.

Overall, the findings of this study support the idiom principle proposed by Sinclair that words do not occur in isolation and come together to make meaning (Sinclair, 1991, 2004). The idiom principle states that language users have at their disposal a large variety of semi-preconstructed phrases that signify single choices, despite looking as if they could be separated into segments (Sinclair, 1991). The existence of lexical bundles in academic texts indicates that "words are co-selected by speakers and writers which lead to collocation and other features of idiomaticity" (Cheng et al., 2006, p. 411).

The results of this study can improve pedagogy, empowering both students and teachers in their choice of methodology in using lexical bundles. Language researchers are seemingly missing the boat in addressing the challenges encountered with this essential yet complex feature of language. They also assist EFL teachers in materials development for writing courses as fundamental components for the construction of discourse in all university registers (Biber, 2006), having the potential to construct students' expertise in a particular field (Hyland, 2013; Manchon, 2011; Ortega, 2012). Given the importance of lexical bundles in academic texts, the findings of this study assist EFL teachers, writing instructors, and materials developers in identifying what kinds of lexical bundles students apply in their academic reports. As such, teachers should make learners aware of the important role of frequent lexical bundles like *as a result of*, *on the other hand*, *in the case of*, *at the same time*, *it is important to*, and *in terms of the* in the production of coherent texts. In addition, the results highlight the value of lexical bundles in helping language users to maintain their identity in a disciplinary community, and also provide insight into how proficient writers use lexical bundles in writing coherent texts.

Like many other studies, this research is not without its shortcomings. Since this study collected both non-native and native reports from MICUSP and the number of non-native reports in this database was low, the material was limited to only 100 report texts. Further studies can overcome this limitation and extend the research

using a larger corpus size. Besides, since the number of three-word bundles is very high, this study focused only on four-word bundles to have a manageable investigation; future studies can also examine five-word and six-word bundles to make the coverage more comprehensive. Researchers may further investigate the correlation between corpus size, cut-off frequency, range, and bundle length. Additionally, this study relied only on report texts; however, an expanded study using corpora from different genres would respond to the more diverse needs of local students. What is more, according to the list of high-frequency bundles identified in many studies, developing a corpus of various academic genres, especially for non-native learners of academic English, would provide a better understanding of the ways discipline-specific bundles and shared bundles among various disciplines can be taught to EFL students based on their needs.

This study assigned only one function to each bundle to avoid confusion; however, according to Biber et al. (2004), a single bundle is likely to serve more than one function, even in one context. Employing lexical bundles in different disciplines and exploring how each bundle type serves one or more functions in the same context of that particular discipline require further investigation in the future. This study could also be extended by including other academic registers, both spoken and written to broaden the scope of the description.

**Conflict of interest:** The authors declare they have no conflict of interest.

**Ethical Approval:** This article does not contain any studies with human participants or animals performed by any of the authors.

**Acknowledgments:** We would like to express our deep gratitude to Prof. Alireza Jalilifar at Shahid Chamran University of Ahvaz for his immense contribution to the completion of this research project.

## References

- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81–92. <https://doi.org/10.1016/j.esp.2011.08.004>
- Ädel, A., & Römer, U. (2012). Research on advanced student writing across disciplines and levels: Introducing the Michigan Corpus of Upper-level Student Papers. *International Journal of Corpus Linguistics*, 17(1), 3–34. <https://doi.org/10.1075/ijcl.17.1.01ade>
- Ahmadi, H. S., Ghonsooly, B., & Fatemi, A. H. (2013). An analysis of lexical bundles in research article abstracts by Iranian and native English-speaking authors of applied linguistics articles. *The Asian ESP Journal*, 9(1), 5–25.
- Allen, D. (2009). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education*, 1, 105–127.
- Amirian, Z., Ketabi, S., & Eshaghi, H. (2013). The use of lexical bundles in native and non-native post-graduate writing: The case of applied linguistics MA theses. *Journal of English Language Teaching and Learning*, 5(11), 1–29.
- Anthony, L. A. (2019). *Antconc (version 3.5.8)* Computer software. Tokyo: Waseda University. Retrieved from <https://www.laurenceanthony.net>.
- Appel, R., & Murray, L. (2020). L1 differences in L2 English academic writing: A lexical bundles analysis. *Journal of English for Academic Purposes*, 46, 1–15. <https://doi.org/10.1016/j.jeap.2020.100873>
- Bal, B. (2010). *Analysis of four-word lexical bundles in published research articles written by Turkish scholars*. Unpublished master thesis. Georgia State University. <https://doi.org/10.57709/1665591>
- Biber, D. (2004). Lexical bundles in academic speech and writing. In B. Lewandowska-Tomaszczyk (Ed.), *Practical applications in language corpora* (pp. 165–178). Peter Lang.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English:

- Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3), 275–311. <https://doi.org/10.1075/ijcl.14.3.08bib>
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263–286. <https://doi.org/10.1016/j.esp.2006.08.003>
- Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: An initial taxonomy. In A. Wilson, P. Rayson & T. McEnery (Eds.), *Corpus linguistics by the lune: A festschrift for Geoffrey Leech* (pp.71–93). Peter Lang.
- 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.
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language teaching research*, 10(3), 245–261. <https://doi.org/10.1191/1362168806lr19>
- Bourelle, T. (2012). Bridging the gap between the technical communication classroom and the internship: Teaching social consciousness and real-world writing. *Journal of Technical Writing and Communication*, 42(2), 183–197. <https://doi.org/10.2190/TW.42>
- Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, 30, 38–52. <https://doi.org/10.1016/j.jeap.2017.10.008>
- Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, 5(5), 31–64.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30–49. <https://doi.org/10.125/44213>
- Cheng, W., Greaves, C., & Warren, M. (2006). From n-gram to skip-gram to conc-gram. *International Journal of Corpus Linguistics*, 11(4), 411–433. <http://dx.doi.org/10.1075%2Fijcl.11.4.04che>

- Cilliers, C. B. (2012). Student perception of academic writing skills activities in a traditional programming course. *Computers & Education*, 58(4), 1028–1041. <https://doi.org/10.1016/j.compedu.2011.12.001>
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than non-formulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72–89. <https://doi.org/10.1093/applin/amm022>
- Conrad, S., & Biber, D. (2004). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica* 20, 56–71. <https://doi.org/10.1515/9783484604674.56>
- Cortes, V. S. (2002). *Lexical bundles in academic writing in history and biology*. Unpublished doctoral dissertation. Northern Arizona University. Flagstaff, Arizona
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing Examples from history and biology. *English for Specific Purposes*, 23, 397–423. <https://doi.org/10.1016/j.esp.2003.12.001>
- Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, 17(4), 391–406. <https://doi.org/10.1016/j.linged.2007.02.001>
- Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, 12(1), 33–43. <https://doi.org/10.1016/j.jeap.2012.11.002>
- Cowie, A. P. (1994). Phraseology. In R. E. Asher (Ed.), *The encyclopedia of language and linguistics* (pp. 3168–3171). Oxford University Press.
- Cowie, A.P. (1998a). Phraseological dictionaries: some East-West comparisons. In Cowie, A.P. (ed.), *Phraseology: Theory, analysis and applications* (pp. 209–228). Oxford University Press.
- Cowie, A.P. (1998b). Introduction. In Cowie A.P. (ed.), *Phraseology: Theory, analysis and applications* (pp. 1–20). Oxford University Press.
- Coxhead, A. (2008). Phraseology and English for academic purposes: Challenges and opportunities. In F. Meunier & S. Granger (Eds.), *Phraseology in language learning and teaching* (pp.149–161). John Benjamins.
- De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated

- approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on computer* (pp. 67–79). Longman.
- De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures*, 2(1), 225–246.
- Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375–396. <https://doi.org/10.1002/j.1545-7249.2008.tb00137.x>
- Erman, B. (2009). Formulaic language from a learner perspective: What the learner needs to know. In B. Corrigan, H. Quali, E. Moravcsik, & K. Wheatley (Eds.), *Formulaic language* (pp. 27–50). John Benjamins.
- Estes, C. A. (2004). Promoting student-centered learning in experiential education. *Journal of Experiential Education*, 27(2), 141–160. <https://doi.org/10.1177/105382590402700203>
- Fitriati, S. W., & Wahyuni, S. (2018). Lexical bundles in Whatsapp conversation between native and non-native speakers of English. In *UNNES International Conference on English Language Teaching, Literature, and Translation (ELTLT 2018)* (pp. 314-320). Atlantis Press.
- Ghafar Samar, R., Shirazizadeh, M., & Kiany, G R. (2015). Backgrounds, prospects, applications and challenges in studying lexis in academic texts: The growing need for particular attention to Persian and Persian speaking language learners. *Language Related Research*, 6(4), 153–181.
- Gould, J. R., & Losano, W. A. (2008). *Opportunities in technical writing*. The McGraw-Hill Companies.
- Granger, S., & Meunier, F. (2008). Phraseology in language learning and teaching. Where to from here? In Meunier, F. & Granger, S. (Eds.), *Phraseology in foreign language learning and teaching* (pp. 247–252). John Benjamins.
- Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, 18(1), 109–136. <https://doi.org/10.1075/ijcl.18.1.08gra>
- Gray, B., & Biber, D. (2015). Phraseology. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 125–145). Cambridge

University Press.

- Hassanzadeh, M., & Tamleh, H. (2023). The use of lexical bundles by native English authors in applied linguistics: A corpus-driven study. *Language Related Research*, 13(6), 541–569. <https://doi.org/10.29252/LRR.13.6.17>
- Hinkel, E. (2016). Practical grammar teaching: Grammar constructions and their relatives. In E. Hinkel (Ed.), *Teaching English grammar to speakers of other languages* (pp. 171–99). Routledge
- Hinkel, E. (2017). Teaching idiomatic expressions and phrases: Insights and techniques. *Iranian Journal of Language Teaching Research*, 5(3), 45–59. <https://doi.org/10.30466/ijltr.2017.20304>
- Hunston, S. (2002). *Corpora in applied linguistics*. Oxford University Press.
- Hyland, K. (2008a). 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. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41–62. <https://doi.org/10.1111/j.1473-4192.2008.00178.x>
- Hyland, K. (2009). *Academic discourse: English in a global context*. Continuum.
- Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32, 150–169. <https://doi.org/10.1017/S0267190512000037>
- Hyland, K. (2013). 5 ESP and writing. In B. Paltridge & S. Starfield (Eds.), *The handbook for English for Specific Purposes* (pp. 95–113). Wiley-Blackwell.
- Jalilifar, A., Ghoreishi, S. M., & Emam Roodband, S. A. (2016). Developing an inventory of core lexical bundles in English research articles: A cross-disciplinary corpus-based study. *Journal of World Languages*, 3(3), 184–203. <https://doi.org/10.1080/21698252.2017.1301279>
- Juknevičienė, R. (2009). Lexical bundles in learner language: Lithuanian learners vs. native speakers. *Kalbotyra*, 61, 61–72. <https://doi.org/10.15388/Klbt.2009.7638>
- Kellogg, R. T., & Raulerson, B. A. (2007). Improving the writing skills of college students. *Psychonomic Bulletin & Review*, 14(2), 237–242. <https://doi.org/>



[10.3758/BF03194058](https://doi.org/10.3758/BF03194058)

- Kim, S., & Kessler, M. (2022). Examining L2 English University students' uses of lexical bundles and their relationship to writing quality. *Assessing Writing*, 51, 100589. <https://doi.org/10.1016/j.asw.2021.100589>
- Linsdell, J., & Anagnos, T. (2011). Motivating technical writing through study of the environment. *Journal of Professional Issues in Engineering Education and Practice*, 137(1), 20–27. [https://doi.org/10.1061/\(ASCE\)EI.1943-5541.0000032](https://doi.org/10.1061/(ASCE)EI.1943-5541.0000032)
- Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31(1), 25–35. <https://doi.org/10.1016/j.esp.2011.07.002>
- Lorenna, M., Fitriati, S. W., & Widhiyanto, W. (2020). The comparison of lexical bundles in EFL Teachers' Talk between Non-native and Native English Teachers. *English Education Journal*, 10(1), 69–75. <https://doi.org/10.15294/eej.v10i1.33830>
- Maniati, M., Jalilifar, A., Mashhadi, A., & Validy, M. (2020). Transforming textual meaning during the revision process of research articles written by Iranian scholars. *Journal of English Language Teaching and Learning*, 12(25), 191–222. <https://doi.org/10.22034/elt.2020.10681>
- Neno, H., & Agustien, H. I. (2016). The use of formulaic expression in EFL students' interaction. *English Education Journal*, 6(1), 39–44.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins.
- Northbrook, J., & Conklin, K. (2019). Is what you put in what you get out? Textbook-derived lexical bundle processing in beginner English learners. *Applied Linguistics*, 40(5), 816–833. <https://doi.org/10.1093/applin/amy027>
- Ortega, L. (2012). Epilogue: Exploring L2 writing–SLA interfaces. *Journal of Second Language Writing*, 21(4), 404–415. <https://doi.org/10.1016/j.jslw.2012.09.002>
- Öztürk, Y., & G. D. Köse (2016). Turkish and native English academic writers' Use of Lexical Bundles. *Journal of Language and Linguistic Studies*, 12 (1), 149–165.
- Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications



- research journals. *Journal of English for Academic Purposes*, 21, 60–71. <https://doi.org/10.1016/j.jeap.2015.11.003>
- Pedersen S, & Liu M. (2003). Teachers' beliefs about issues in the implementation of a student-centered learning environment. *Educational Technology Research and Development*, 51(2), 57–76. <https://doi.org/10.1007/BF02504526>
- 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>
- Prodromou, L. (2009). English as a lingua franca: A corpus-based Analysis. *ELT journal*, 63(1), 78–81. <https://doi.org/10.1093/elt/ccn064>
- Raus, T. Z. M., Yusoff, M., & Subatira, B. (2019). Technical report writing in engineering studies. *LSP International Journal*, 6(2), 45–59. <https://doi.org/10.11113/lspi.v6n2.91>
- Ren, J. (2021). Variability and functions of lexical bundles in research articles of applied linguistics and pharmaceutical sciences. *Journal of English for Academic Purposes*, 50, 100968. <https://doi.org/10.1016/j.jeap.2021.100968>
- Riordan, D. G., & Pauley, S. E. (2002). *Technical report writing today* (8th ed.). Houghton Mifflin Company.
- Römer, U. (2009). English in academia: Does nativeness matter? *Anglistik. International Journal of English Studies*, 20(2), 89–100.
- Salazar, D.J.L. (2010). Lexical bundles in Philippine and British English. *Philippine Journal of Linguistics*, 41, 94–109.
- Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. John Benjamins Publishing Company.
- Schmitt, N. & Rodgers, M. (2020). *An introduction to applied Linguistics* (3rd Ed). Routledge.
- Serrano, R., Stengers, H., & Housen, A. (2015). Acquisition of formulaic sequences in intensive and regular EFL programmes. *Language Teaching Research*, 19(1), 89–106. <https://doi.org/10.1177/1362168814541748>
- Simpson-Vlach, R. & Ellis, N. C. (2010). An academic formulas list: New methods

- in phraseology research. *Applied Linguistics*, 31(4), 487–512. <https://doi.org/10.1093/applin/amp058>
- Sinclair, J. (1991). *Corpus, concordance, collocation. Describing English language*. Oxford University Press.
- Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge.
- 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>
- Sykes, D. (2017). *An investigation of spoken lexical bundles in interactive academic contexts*. Doctoral dissertation, Carleton University. <https://doi.org/10.22215/etd/2017-12110>
- Tazl, D., Hassler, W., Messnarz, B., & Fluhr, H. (2012). The development of a project based collaborative technical writing model founded on learner feedback in a tertiary aeronautical engineering program. *Journal of Technical Writing and Communication*, 42(3), 279–304. <https://doi.org/10.2190/TW.42.3.f>
- Ucar, S. (2017). A corpus-based study on the Use of Three-Word Lexical Bundles in the Academic Writing by Native English and Turkish Non-Native Writers. *English Language Teaching*, 10(12), 28–36. <https://doi.org/10.5539/elt.v10n12p28>
- Valsecchi, M., Künstler, V., Saage, S., White, B. J., Mukherjee, J., & Gegenfurtner, K. R. (2013). Advantage in reading lexical bundles is reduced in non-native speakers. *Journal of Eye Movement Research*, 6, 1–15. <https://doi.org/10.16910/jemr.6.5.2>
- Vo, S. (2019). Use of lexical features in non-native academic writing. *Journal of Second Language Writing*, 44, 1-12. <https://doi.org/10.1016/j.jslw.2018.11.002>
- Wood, D. (2010). Lexical clusters in an EAP textbook corpus. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 88–106). Continuum.
- Wood, D. (2015). *Fundamentals of formulaic language*. Bloomsbury Academic.
- Wray, A. (2002). Formulaic language in computer-supported communication: theory meets reality. *Language Awareness*, 11(2), 114–131.

<https://doi.org/10.1080/09658410208667050>

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

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, 20(1), 1–28. [https://psycnet.apa.org/doi/10.1016/S0271-5309\(99\)00015-4](https://psycnet.apa.org/doi/10.1016/S0271-5309(99)00015-4)

Xia, D., Ai, H., & Pae, H. K. (2022). “Please let me know”: Lexical bundles in business emails by business English learners and working professionals. *International Journal of Learner Corpus Research*, 8(1), 1–30. <https://doi.org/10.1075/ijlcr.20019.xia>

Yeldham, M. (2018). The influence of formulaic language on L2 listener decoding in extended discourse. *Innovation in Language Learning and Teaching*, 12(2), 105–119. <https://doi.org/10.1080/17501229.2015.1103246>

### **About the Authors**

**Amir Mashhadi** is an Assistant Professor of Applied Linguistics at Shahid Chamran University of Ahvaz, Iran, specializing in computer-assisted language learning. He has presented at regional and international conferences and published in international and national refereed journals in his area of research. He is currently the managing editor of *Journal of Research in Applied Linguistics (RALs)*, and also serves on the editorial board of *Journal of Language and Discourse Practice*.

**Marjan Ghanavati** received her M.A. in English Language Teaching from Shahid Chamran University of Ahvaz. She is interested in corpus-based studies and has presented at regional and international conferences in her area of research interest.