The deployment of English lexical bundles in applied linguistics research articles by Vietnamese researchers

Tran Quoc Thao[1*], Nguyen Minh Khoi[1]

[1]HUTECH University, Ho Chi Minh City, Vietnam
[*]Corresponding author: tq.thao@hutech.edu.vn

ARTICLE INFO

DOI:10.46223/HCMCOUJS.soci.en.12.2.2227.2022

Received: April 03rd, 2022

Revised: July 06th, 2022

Accepted: July 18th, 2022

Keywords:

applied linguistics; corpus; lexical bundle; research article; Vietnamese researcher

ABSTRACT

Lexical Bundles (LBs) have been considered an effective tool for not only academic writing but also spoken discourse. The use of these clusters has attracted a lot of researchers in different registers because of their characteristics and functions in constructing the texts. However, research into LBs seems to be quite rare in English Applied Linguistics (AL) Research Articles (RAs) written by non-native English researchers, especially Vietnamese researchers. This study, therefore, aims to scrutinise Vietnamese researchers' use of LBs in English AL RAs. A corpus of 77 AL RAs selected from three Journals was built. This study adapted the theoretical frameworks by Biber, Johansson, Leech, Conrad, and Finegan (1999), Biber, Conrad, and Cortes (2004) and Biber (2006) for data analysis. The findings revealed that Vietnamese researchers tended to deploy noun-based LBs more than verb-based and preposition-based ones in their AL RAs. It was further found LBs showing research-oriented functions were prevalent in AL RAs over text-oriented and participant-oriented ones. This research has pedagogical implications in terms of LB use for Vietnamese English as a Foreign Language (EFL) teachers and learners in the teaching and learning of academic writing.

1. Introduction

Lexical Bundles (LBs), which are considered a part of the formulaic expression (Byrd & Coxhead, 2010), are clusters of three or more words that co-occur frequently in a specific type of register (Biber et al., 1999). They are used to perform discourse functions for meaning construction in a text and differentiation in a particular register (Hyland, 2008). The use of word combinations is one of the most important factors for language learning, language processing, and language production (Wray, 2002). Frequently used LBs in daily conversation are *what I mean, it is important*, etc., while the examples of LBs in the academic register are as *in addition to, as the result of* and so on (Biber et al., 2004).

Scholars (e.g., Biber et al., 1999; Cortes, 2004; Hyland, 2008; Parvizi, 2011; Valipoor, 2010) have mentioned that LBs' grammatical structure is a distinct characteristic of registers, and the use of LBs across registers can share certain similarities and differences in terms of types and functions. For example, Hyland (2008) examined the use of LBs in Research Articles (RAs), dissertations, and theses (doctorate and master's level) within four academic disciplines, and he found a certain difference in the use of LBs in terms of frequency and functions across disciplines. Valipoor (2010) analyzed the use of LBs in chemistry RAs and found that LBs with particular

functions appear in different fragments of a RA. In the field of education, Parvizi (2011) found that writers employ uncommon LBs with different functions. This may imply that the LB use depends much on purposes, views, disciplines, and writers. Nonetheless, other studies have focused on LB use by writers with different language proficiency levels. For instance, Chen and Baker (2010) underscored that professional writers utilize LBs in terms of structures and functions differently from novice ones. It can be asserted that LBs used in different registers can differ in terms of frequency, structures, and functions.

It is observed that many scholars have analyzed the LB use across registers in one or various disciplines, while some others have analyzed the LB use in one section of an RA or the whole RAs. Additionally, RAs are different from one another in terms of disciplines to showcase the content and target audience (e.g., Tran & Duong, 2013; Tran & Tang, 2022). Nevertheless, there is a scarcity of studies on Vietnamese researchers' use of LBs in the English Applied Linguistics (AP) RAs. This study, therefore, aims at scrutinizing the LB use in terms of structural types and functions in AP RAs written by Vietnamese researchers.

By doing this study, the results can generate some theoretical and practical contributions. Theoretically, the study can add some further understanding to the body of literature review of LB. Practically, the study is hoped to provide an in-depth understanding of LB types and their functions, which can help writers write their RAs appropriately and effectively.

2. Literature review

The term LB can be considered a common expression, and it has standardized occurrence and distribution criteria (Biber et al., 1999). In a similar vein, LBs are groups of word combinations consisting of three or more words (e.g., *as seen in*, *on the one hand*, etc.) (Granger, 2014). LBs have characteristics that are different from other types of word combinations (e.g., collocation, idiom) since LBs are not idiomatic in meaning (Biber & Barbieri, 2007). The LBs' meaning is easy to understand in comparison to idioms' meaning as they are transparent in meaning Additionally, LBs appear at a higher frequency in a typical register than idioms (Cortes, 2004). Another characteristic is that LBs are fragmental units and parts of longer structures (Biber & Barbieri, 2007).

Regarding LBs' structure, Biber et al. (1999) classify them into three main categories, viz. noun-based, preposition-based, and verb-based LBs as seen in Table 1.

Table 1

LBs' Structural classification (Biber et al., 1999)

Structures	Examples
noun-phrase + of	*a number of, the role of*
other noun-phrases	*the fact that, the possible explanation*
prepositional phrase + of	*as the result of, in terms of, in the case of*
other prepositional phrases	*as a result, at the beginning*
passive + prepositional phrase	*is presented in, is observed from*
anticipatory it + verb/adj	*it is necessary, it is urgent*
be + noun/ adjectival phrase	*is a significant difference, is consistent with*
Others	*as seen from the Table, as presented in*

Source: The researcher's data analysis

Biber et al. (1999) divide LBs' functions into three main groups: participant-oriented LBs (also known as stance expressions), text-oriented LBs (or discourse organizers), and research-oriented LBs (or referential expressions) (Table 2). Research-oriented LBs function to describe research-related issues, and include five different sup-functions: (1) Locational LBs indicate time and place; (2) Procedural LBs describe the research process; (3) Quantitative LBs show size or scale; (4) Descriptive LBs describe the context; (5) Topical LBs present research focus. Text-oriented LBs function to construct the meaning in the text, and connect the proceeding and the following discourses (Biber & Barberi, 2007). There are four sub-functions: (1) Transition signals used to show additive or contrastive connection among textual components; (2) Resultative signals used to indicate the causative or inferential relations among textual elements; (3) Structuing signals used to show the text organization or direction to readers in the text; (4) Framing signals used to show limitation and arguments. Participant-oriented LBs function as a bridge to connect the writer and reader in the text, and they include two sub-types: Stance and Engagement which indicate the features of the writer and reader, and their interaction in the text (Hyland, 2005).

Table 2

LB's Functional classification (Biber et al., 2004; Biber, 2006)

Functions	Sub-types	Examples
Research-oriented LBs	Time and place	*at the end of, at the same time*
	Procedure	*the first step, in the process of*
	Quantification	*a big number of, a range of*
	Description	*the pattern of, a cohort of*
	Research related topic	*the research aim, to this end*
Text-oriented LBs	Transition signal	*apart from, what is more*
	Resultative signal	*as a consequence,*
	Structuring signal	*within this study, in the following section*
	Framing signal	*in respect of, in terms of*
Participant-oriented LBs	Stance	*a possible explanation, it seems possible*
	Engagement	*as observed in, as found in*

Source: The researcher's data analysis

A number of previous studies on LBs have interested many scholars and researchers. For example, Wei and Lei (2011) examined advanced Chinese EFL learners' use of LBs in academic writing. They analysed a corpus of published RAs written by professional writers and that of doctoral dissertations written by Chinese EFL learners, and found that there were more LBs in Chinese EFL learners' academic writing than in professional writers. They also found that professional writers used participated-oriented bundles more than Chinese EFL learners, but Chinese EFL learners employed more bundles with passive voice and less anticipatory it-structures than the professional authors. Qin (2014) analysed non-native English graduated writers' LB use in academic papers and published RAs in AL. She found out that non-native English graduate writers employed different LBs in their writing, and students at the higher levels of study used more LBs than those at the lower levels. In 2019, Gil and Caro (2019) did a study examining LB use in bachelor theses written by Spanish students. The results found the most common types,

structures, and functions used in students' theses. To sum up, it is noticed that LBs have been extensively examined in various contexts; nevertheless, not much research on LB use in English AP RAs written by non-native speakers of English has been conducted.

3. Methodology

3.1. Corpus

A corpus was built from a cohort of 77 RAs that have been published in three prestigious journals that are coded as Journal A (28 RAs), Journal B (26 RAs), and Journal C (23 RAs). These RAs are open-access and peer-reviewed, so they were easily downloaded. The RAs were chosen based on the following predetermined criteria: (1) the length of each paper is in the range between 4,000 and less than 8,000 words; (2) the topics of papers are relevant to the field of the AP; (3) the writers were Vietnamese authors; and (4) all the RAs were published from 2015 to 2021. As seen from Table 3, the total number of word tokens in the corpus is 450,193 in which there are 25,601 word types.

Table 3

Description of corpus

No.	**Journals**	**No. of RAs**	**Word tokens**	**Word types**
1	Journal A	28	156,894	8,584
2	Journal B	26	141,456	8,464
3	Journal C	23	151,843	8,553
Total		77	450,193	25,601

Source: The researcher's data analysis

3.2. Framework for data analysis

Prior to data analysis, the selected RAs in the corpus were converted into plain text format and coded as RA1, RA2 to RA77, and all the unnecessary materials (e.g., page numbers, references, figures, and tables, etc.) were omitted. The AntConc 4.0.5 version (Anthony, 2022) were employed seeking LB's frequencies and structural classifications with three- and 4-word LB based on Biber et al.'s (1999) structural classifications of LBs (Table 1). Additionally, Biber et al.'s (2004) and Biber's (2006) functional classfifications were adopted for LBs' funcation analysis (Table 2).

Regarding reliablity and validity, a pilot study was conducted with a sample of ten RAs so as to confirm the validity of the data analysis framework. After the main data analysis, two researchers in the same field were invited for double-checking. Ten pieces of data were randomly taken from the corpus for re-analysis, and the convergent findings were set from 95%.

4. Results and discussion

4.1. Results

4.1.1. Vietnamese researchers' deployment of LB in AP RAs

As demonstrated in Table 4, there are 440 types of eight structures of LBs with a frequency of 4,453 deployed in AP RAs written by Vietnamese researchers. Specifically, the structure *Noun phrase with of* was the most commonly and frequently used structure (Type: 232; F: 2,832) accounting for 63.60%, followed by the structures *Prepositional phrase + of* (Type: 48; F: 436),

Passive + prepositional phrase (Type: 46; F: 413), and *Other noun phrases* (Type: 31; F: 234) responsible for 9.79%, 9.27%, and 5.25% respectively. Meanwhile, the other structures accounting for low percentages in the corpus were *Anticipatory it + verb/adj* (Type: 27; F: 185) with 4.15%, *Other prepositional phrases* (Type: 21; F: 193) with 34.33%, and *Others* (Type: 19; F: 94) with 2.11%. The least common and frequently used structure was *Be + noun/adjectival phrase* and (Type: 16; F: 66) with 1.48%. It can be interpreted that Vietnamese researchers in the field of AL tended to use *Noun phrase of* in their RAs most, while they seemed to avoid using *Be + noun/adjectival phrase* in their RAs.

Table 4

The types of the structures of LBs in the corpus

No.	Structures	No. of types	F	%
1.	Noun phrase + of	232	2,832	63.60
2.	Other noun phrases	31	234	5.25
3.	Prepositional phrase + of	48	436	9.79
4.	Other prepositional phrases	21	193	4.33
5.	Passive + prepositional phrase	46	413	9.27
6.	Anticipatory it + verb/adj	27	185	4.15
7.	Be + noun/adjectival phrase	16	66	1.48
8.	Others	19	94	2.11
Total		**440**	**4,453**	**100**

4.1.2. LBs' functions in Vietnamese researchers' AP RAs

Table 5 shows that Vietnamese researchers utilized LBs for three functions in their RAs. The LBs with research-oriented functions had the highest frequency (F: 2,051) accounting for 68.21% of the total frequency, followed by LBs with text-oriented functions with a frequency of 717 times and 23.84% of the total frequency. The LBs with participant-oriented functions were the least used with a frequency of 239 times and 7.95% of the total frequency. This means that Vietnamese researchers had a tendency in using research-oriented LBs in their RAs more than text-oriented and participant-oriented ones.

Table 5

LBs' Functions in AL RAs by Vietnamese researchers

Functional classification		**Corpus**	
		F	**%**
Research-oriented LBs	Time and place	243	8.08
	Procedure	1,354	45.03
	Quantification	231	7.68
	Description	169	5.62
	Research related topic	54	1.80

Functional classification		**Corpus**	
		F	**%**
Text-oriented LBs	Transition signal	125	4.16
	Resultative signal	195	6.48
	Structuring signal	288	9.58
	Framing signal	109	3.62
Participant-oriented LBs	Stance	114	3.79
	Engagement	125	4.16
Total		**3,007**	**100**

Research-oriented LBs

Research-oriented LBs function to describe the research-related activities, and as seen in Table 5, there are five different sub-functions, namely *Time and Place*, *Procedure*, *Quantification, Description*, and *Research related topic*. LBs indicating *Procedure* were prevalent over the other types, accounting for 1,354 times (45.03%), followed by those of *Time and Place* with 243 times (8.08%), *Quantification* with 231 times (7.68%), and *Description* with 169 times (5.62%). The least common function was *Research related topic* to 54 times (1.80%). Some examples indicating research-oriented LBs are as follows:

(1) "The discourse marker only can be as a focusing adverb; however, focusing adverbs are not normally used *at the beginning of* a sentence." (Place-RA56)

(2) "Within the framework of RT, *a number of* discourse markers have been analyzed…." (Quantification- RA32)

(3) "Finally, the dual coding of information should be further stimulated through *the use of* images, pictorial elucidation, and mime to commit the target idiomatic expressions to their long-term memory." (Description- RA74)

In Example (1), the LB *at the beginning* was used to indicate the place where the discourse marker Only was examined. The LB *number* in Example (2) was employed to show the value of the discourse markers. With respect to Example (3), the LB *the use of* was used to indicate the description of how images, pictorial elucidation, and mime were used.

Text-oriented LBs

Text-oriented LBs are known as clusters that function to function to describe research-related issues. This type of LBs comprises four subtypes, viz. *Transition signals, Resultative signals, Structuring signals,* and *Framing signals.* It is seen from Table 5 that the *Structuring signals* were the most used LBs accounting for 288 times (9.58%), followed by *Resultative signals* responsible for 195 times (6.48%) and *Transition signals* making up 125 times (4.16%). The least used LBs were *Framing signals* appearing 109 times (3.62%). Below are some extracts from the corpus:

(4) "*On the other hand*, the fact that there was one out of three students said they rarely practiced translation at home is quite a serious situation that needs taking into consideration by the educators." (Transition signal-RA16)

(5) "Looking more closely at the stories told by the students, we can see that autonomy comes *as a result of* both their will and luck…." (Resultative signal- RA25)

(6) "As shown *in the present study* and other studies with Vietnamese EFL learners…there are a considerable number of Vietnamese EFL learners having insufficient knowledge of the most frequent 1,000 words." (Structuring- RA20)

(7) "With *respect to* the functions of phrasal verbs, Vietnamese M.A. students used different subcategories in syntax and semantics." (Framing- RA25)

The LB *On the other hand* in Example (4) presents the contrastive link between elements in the text, while the LB *as a result* in Example (5) shows the causative relations. In Example (6), the LB *in the present study* aims at directing readers to the intended context, and the LB *with respect to* Example (7) indicates a specific condition.

Participant-oriented LBs

Participant-oriented LBs, which are less used than the other two functional classifications, consist of two sub-types: Stance and Engagement. The former (F: 114; 3.79%) were deployed in RAs less than the latter (F: 125; 4.16%).

(8) "In addition, as mentioned in "Vietnam adds 05 new foreign languages" (2016), other foreign languages, including Chinese and French, *are likely to be* added to primary level teaching." (Stance- RA6)

(9) "*As can be seen*, there appear to be considerable similarities between the two journals in terms of hedging forms employed." (Engagement- RA5)

The LB *are likely to be* in Example (8) shows the writer's evaluation of the inclusion of foreign languages into primary level teaching. The LB *as can be seen* in Example (9) locates the intended context.

4.2. Discussion

This study has revealed some important findings. Firstly, it was found that Vietnamese researchers in AL tended to get more familiar with the structure *Noun phrase of* in their research writing than other ones. This structure can be used to reveal different meanings in academic discourse (Hyland, 2008) such as to show quantity, place, size, or qualities. This type of structure can be noun-based LB which accounts for a major proportion of RAs. Such a finding is aligned with that of previous studies (e.g., Hyland, 2008; Salazar, 2014). Meanwhile, the preposition-based and verb-based LBs seemed to be much less used by Vietnamese researchers in their RAs. It was further seen that the structure *Be + noun/adjectival phrase* (one of the verb-based LBs) was the least common type that Vietnamese researchers used in their RAs. This type is to indicate causative and comparative links among elements and to show writers' evaluations (Hyland, 2008), and Vietnamese researchers may tend to avoid stating the comparison and their evaluation in their RAs.

The second major finding is about the LBs' functional contribution. It was found that Vietnamese researchers in the field of AL preferred the *research-oriented LBs* to *text-oriented* and *participant-oriented ones*. Additionally, the LBs indicating *Procedure* were most used in RAs, accounting for nearly 50% of the total number of LBs in RAs. This may infer that the RAs in the corpus were empirical studies, so researchers tended to focus on the approaches and methods through which the studies were carried out.

Furthermore, *text-oriented LBs*, serve as a more discursive function of linking the textual elements (Biber & Barberi, 2007; Biber et al., 2004; Hyland, 2008). Vietnamese researchers were in favor of using this type of LBs to connect readers to the text content. This finding is supported by Hyland (2004) who has stated that *text-oriented LBs* can be of discursive and evaluative

functions in the field of soft science in which persuasion tends to be more explicitly interpretive. This may imply that the use of *text-oriented LBs* in AL is thanks to their interpretive and discursive features. However, this finding is opposite to that of Gil and Caro's (2019) study which has showed that *text-oriented LBs* were the most frequently used function in academic writing.

In terms of the distribution of sub-types of *text-oriented LBs*, Vietnamese researchers tended to deploy the *structuring signals* and the *resultative signals* in their AL RAs more than other types. According to Hyland (2008), these signals function to indicate inferential or causative relationships between textual elements. It can be said that Vietnamese researchers tended to use such LBs to get readers focused on the findings of the study as these signals can help to indicate the results of the study clearly to readers.

The least employed function was *participant-oriented LBs* which consist of two sub-types: *stance* and *engagement*. They refer to the connection and interaction between the writer and reader via texts (Hyland, 2005). Although stance LBs can help writers to reveal their claims or viewpoints (Salazar, 2014), Vietnamese researchers did not use these LBs much in their RAs. Moreover, it was found that the stance LBs made a smaller proportion than engagement ones. This may imply that Vietnamese researchers may not impose their judgments and evaluation on their readers, and tend to avoid using *participant-oriented LBs*.

5. Conclusion

This study has highlighted the importance of using LBs in RAs, and found that the LBs were variously deployed in AP RAs written by Vietnamese researchers, but focused much on noun-based LBs. Regarding the LB's functions, Vietnamese researchers tended to employ LBs to describe research-related issues more than signaling the text-related issues, raising their voices, and getting readers involved in the research. Accordingly, some pedagogical implications are drawn. This study provides some understanding of the LB use in AL RAs for Vietnamese EFL teachers and learners, and it may be assumed that the more learners are exposed to particular LBs, the more they will be able to produce such LBs in their speech or writing. Hence, it is recommended that teachers involved in teaching academic courses should help learners to be aware of the importance of LB use in academic writing and get learners exposed to commonly used LBs for different purposes. In addition, teachers should help learners to understand LBs' characteristics, structures, and functions so that learners can develop their discursive competence in using LBs. Furthermore, the significance of LBs and their application in AL RAs should be strongly highlighted because of their effectiveness and usefulness in meaning construction in texts. Teachers should help learners understand that, apart from forming coherence and cohesion in the texts, LBs can also function to indicate the reader-writer interaction via interpersonal LBs. Therefore, writers can improve their writing if they can use LBs appropriately.

Some limitations are found within this study. The first limitation of this study is that the number of 50 AL RAs is obviously not large enough, so it may be that the findings cannot be generalized. Another limitation is that this study just focused on analyzing LBs in terms of frequency, structure and function and did not analyze such LB use in separate sections of RAs. Therefore, it is recommended that the future study should explore the LBs with a larger scope involving more sources of data, and the LB use by gender (male vs. female) in various registers should be considered for exploration. What is more, the LB use between Vietnamese researchers with foreign counterparts should be also considered for further study.

References

Anthony, L. (2022). *AntConc* (Version 4.0.5) [Computer Software]. Tokyo, Japan: Waseda University.

Biber, D. (2006). *University language: A corpus-base study of spoken and written registers.* Philadelphia, PA: John Benjamins Publishing Company.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The longman grammar of spoken and written English.* London, UK: Pearson Education Ltd.

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.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing. Examples from history and biology. *English for Specific Purposes*, *23*(4), 397-423.

Gil, N. N., & Caro, E. M. (2019). Lexical bundles in learner and expert academic writing. *Bellaterra Journal of Teaching & Learning Language & Literature, 12*(1), 65-90.

Granger, S. (2014). A lexical bundle approach to comparing languages. Stems in English and French. *Languages in Contrast*, *14*(1), 58-72. doi:10.1075/lic.14.1.04gra

Hyland, K. (2004). *Disciplinary discourse: Social interactions in academic writing.* Ann Arbor, MI: University of Michigan Press.

Hyland, K. (2005). *Meta discourse: Exploring interactions in writing*. London, UK: Continuum.

Hyland, K. (2008). As can be seen: Lexical bundles and discipline variation. *English for Specific Purposes, 27*(1), 4-21.

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics, 11*(3), 283-304.

Parvizi, N. (2011). *Identification of discipline-specific lexical bundles in education* (Unpublished master's thesis). University of Kashan, Iran.

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *Systems, 42*(1), 220-231.

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Simpson, R. (2004). Stylistic features of academic speech: The role of formulaic expressions. In T. Upton & U. Connor (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 36-74). Amsterdam, Netherlands: John Benjamins.

Tran, T. Q., & Duong, M. T. (2013). Hedging: A comparative study of research article results and discussion section in Applied Linguistics and Chemical Engineering. *English for Specific Purposes World, 14*(41), 1-13.

Tran, T. Q., & Tang, T. B. (2022). Hedging in the results and discussion section of English applied linguistics research articles by Vietnamese and foreign writers. *Journal of Language Teaching and Research, 13*(1), 119-124. doi:10.17507/jltr.1301.14

Valipoor, L. (2010). *A corpus-based study of words and bundles in chemistry research articles*. (Unpublished master's thesis). University of Kashan, Iran.

Wei, Y., & Lei, L. (2011). Lexical bundles in academic writing of advanced Chinese EFL learners. *RELC Journal, 42*(2), 155-166.

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge, UK: Cambridge University Press.