English Language & Literature Teaching, Vol. 30, No. 1 Spring 2024
https://doi.org/10.35828/etak.2024.30.1.1

Lexical Bundles in the Argumentative Essays of Vietnamese and American Student Writers

Kevin Parent[*]
(Korea Maritime and Ocean University)
Thuy Quynh Bui[**]
(Korea Maritime and Ocean University)

Parent, Kevin, & Bui, Thuy Quynh. (2024). Lexical bundles in the argumentative essays of Vietnamese and American student writers. *English Language & Literature Teaching*, *30*(1), 1-24.

This study examines lexical bundles used in argumentative writings by Vietnamese EFL university students-L2 learners and American university students-L1 learners to see how two groups use bundles and identify problems which L2 learners face. Two corpora including writings by L2 and L1 learners are collected. By using Biber et al.'s structural taxonomy (1999, 2004) and Hyland's functional framework (2008a), we analyse the structures and functions of the four-word bundles used by the two group writings to see any differences. Compared with L1 learners, L2 learners use a wider range of bundles. Regarding structures, L2 learners overuse VP-based bundles, while L1 learners use VP, NP and PP-based bundles flexibly. For functions, L1 learners prefer to use participant-oriented bundles, while L2 learners tend to employ research-oriented bundles. The text-oriented bundles are the least commonly occurring ones in the two groups' essays. To understand how L2 learners overuse bundles in their writings, a qualitative method is used to examine each essay. Some underlying explanations for the overuse might be side effects of teaching and lack of knowledge of the bundle. A pedagogical implication of teaching L2 learners to use bundles in writing class is suggested to help their writing more convincingly and academically.

[lexical bundles/argumentative writings/L1 learners/L2 learners/corpora]

[*] First author [**] Corresponding author

I. INTRODUCTION

Formulaic language has attracted a large number of researchers in recent decades. Many researchers have shown that prefabricated sequences, such as *on the one hand, at the same time,* or *as a result of,* play an important role in discourse and natural language, both spoken and written (Biber & Barbiere, 2007; Hyland, 2008a, 2008b; Sinclair, 1991; Wray, 2002). One avenue of research has taken a frequency- or dispersion-based approach to identify types of formulaic sequences (Shin, 2018). Biber et al. (1999) pioneered corpus study by investigating these frequent word combinations and proposed structural and functional categorisations to describe the combinations in both written and spoken discourse. These word combinations are considered lexical bundles which are defined as "recurrent expressions that usually co-occur in natural language use, regardless of their idiomaticity and lexical status" (Biber et al., 1999, p. 990). To avoid writers' idiosyncratic tendencies, the lexical bundles must occur at least 20 times per million words and across five texts (Biber et al., 2004; Cortes, 2004, 2006).

Despite their frequency and importance, lexical bundles are not easy to master since they are complex and not fixed (Biber et al., 1999). Non-native English speakers therefore struggle to use the lexical bundles in their productive output and may need to become more familiar with their structures, functions, and meanings. Also, Allen (2009) emphasises the knowledge and use of bundles in language learners' writings as an important indicator to determine the success of language learners in discourse communities. Numerous studies about the structural and functional features of lexical bundles in L2 learners' writing have therefore been conducted in the recent decade (Ädel & Erman, 2012; Byrd & Coxhead, 2010; Chen & Baker, 2010; Hyland, 2008a, 2008b; Pan & Liu, 2019; Shin, 2019). Regarding the Vietnam setting, few studies investigated lexical bundles in Vietnamese EFL learners' writings and none of them compared lexical bundles used by Vietnamese university students and American university students. Thus, this study would be more useful for Vietnamese EFL learners to acknowledge the importance of bundles and deploy them accurately. In EFL classes, among numerous types of academic writing, argumentative writing, which is used to express ideas with supporting evidence and examples, might be one of the most common genres (Mei, 2006). However, to most EFL learners, writing an efficient essay is a challenging task (Kaur, 2015). In particular, for Vietnamese undergraduate students, composing a good argumentative essay not only allows them to gain higher academic achievement at university but is also useful for those who have the intention of studying in English-speaking countries such as England, America, etc. as this writing type is used in Standard English tests, including the Test of English as a Foreign Language Exam (TOEFL) or the International English Language Testing System Exam (IELTS) to examine English competence of non-native English

speakers. To be eligible to pursue higher education in these developed countries, these Vietnamese students must take these tests to get high band scores. Therefore, an examination of argumentative essays in terms of lexical bundles somehow helps Vietnamese students reach their goals. The following research questions guide this study:

1. What are the differences in lexical bundles used by L1-English and L2-Vietnamese undergraduate students regarding structural types?
2. What are the differences in lexical bundles used by L1-English and L2-Vietnamese undergraduate students regarding functional types?

II. LITERATURE REVIEW

Phraseological patterns or formulaic sequences are considered one of the crucial aspects in both spoken and written registers, which might improve coherence, and writing fluency (Gray & Biber, 2015) and help language users present their identity within the discourse community (Wray, 2002). Since these formulaic sequences are important building blocks of discourse in both conversations and writings (Biber & Barbieri, 2007), various studies have been conducted to examine these formulaic sequences under numerous umbrella terms, such as recurrent word combinations (De Cock, 1998), clusters (Hyland, 2008a) or lexical bundles (Biber et al., 1999; Cortes 2002, 2004, 2006). A study by Schmitt and Carter (2004) shows that formulaic sequences account for a large proportion of discourse. Similarly, in a study by Erman and Warren (2000), these formulaic sequences make up more than 50% of written and spoken discourses. The widespread appearance of these sequences in discourse might imply that to have proficient competence in linguistic production, language users should have the knowledge and master how to use these sequences. In line with this, studies by Cortes (2004, 2006) also emphasize the frequency of use of lexical bundles in writings might reflect language users' proficiencies. Although these bundles are important in certain registers, it seems that the acquisition of these bundles and the way to employ them accurately is not a natural process (Cortes, 2006). Therefore, the investigation of lexical bundles in terms of syntactic features and discourse functions in different registers or different groups of writers is essential to have a better understanding of how these bundles work in real discourse.

Biber et al. (2004) investigated lexical bundles of two university registers, classroom teaching, and textbooks, then compared them to those found in conversation and academic prose. Based on the structural characteristics of lexical bundles extracted from each register, they categorize these bundles into three main types: 1) type I consists of verb phrase fragments (VP-based bundles) e.g. *it's going to be*; 2) type II includes dependent

clause fragment e.g. *what I want to*; 3) type III bundles are phrasal, which contain noun phrase fragment (NP-based bundles) e.g. *the end of the*, or prepositional phrases (PP-based bundles) e.g. *on the other hand*. Another feature that gained attention in this study is the function of bundles in contexts. Depending on examining the meaning and uses of bundles, authors distinguish three main functions: 1) stance expressions are used to present the author's attitudes about other propositions, e.g. *I think it was*, 2) discourse organizers refer to the relationship of the previous and current text, e.g. *on the other hand*; 3) referential expressions refer to an attribute of the entity as being important, e.g. *one of the most.* The result of this study revealed that textbooks and academic prose showed more extensive use of NP and PP-based bundles, while VP-based phrases accounted for the majority of conversations and classroom teaching. Regarding the function, classroom teaching employed more stance-oriented bundles than in conversation, and it also showed a more common use of referential bundles than in academic prose. Following Biber et al.'s (1999) investigation of lexical bundles in academic prose and conversation, Cortes (2002) analysed the grammar and function of four-word bundles found in a corpus of freshman composition. The findings illustrated that though, at first glance, structures of lexical bundles (e.g., *in front of the, in the middle of the*) used by the freshmen looked similar to those in academic writing, closer examination revealed that the functions of these bundles were not found in academic prose. This might indicate that these students attempted to produce lexical bundles similar to those used in academic writing instead of conversation. Additionally, Cortes (2004) examined lexical bundles utilized in published articles and student papers in the disciplines of history and biology, finding that students rarely used the lexical bundles employed in the published articles; even when they did, their functions did not resemble how professional writers used them.

Other noticeable lexical bundle studies in academic writing include Hyland's (2008a, 2008b) studies on research articles, PhD dissertations, and Master's theses from engineering, microbiology, business, and applied linguistics, examining the forms, structures, and functions of four-word bundles. The results revealed variations which differentiate writings from different disciplines.

The above studies show that register, discipline, and writers' proficiency can impact the use of lexical bundles in different ways; for example, many specific bundles employed in published research articles were not commonly used in students' dissertations, such as *the extent to which, the degree to which, etc.* (Hyland, 2008b). Although the analysis of texts by professional writers might be crucial to creating a taxonomy for the structure and function of lexical bundles, these might not be the best models for L2 learners to follow because these students and professional writers write different genres for different audiences (Bychkovaska & Lee, 2017). Therefore, a few studies investigating lexical bundles used in L1 and L2 undergraduate students' texts have emerged. Chen and Baker

(2010) analysed the structural patterns and functions of lexical bundles in English essays by L2-Chinese and L1-English university students and expert writing. A comparison revealed similarities in using lexical bundles regarding structures and discourse functions by non-native and native students. Both were more dependent on VP-based bundles and discourse organisers, while expert writers showed a larger range of NP-based bundles and referential bundles. Adel and Erman (2012) also investigated the use of lexical bundles in writing by L1-Swedish and L1-English undergraduate students. The findings illustrated that the native student writers employed more types and more varied ones than non-native speakers. This can be shown in the examples of the unattended *this* without a head noun: *this can be seen, this may be because, this would suggest that, etc.* The pattern with the unattended *this* used to explain or clarify ideas with various degrees of writers' certainty was one of the most striking patterns in L1 but did not occur in L2 writing.

In academic writing, argumentative writing may be one of the most common genres of undergraduate writing (Mei, 2006; Shin, 2018). To compose a good argumentative essay, writers need to present ideas and develop reasons logically with evidence or examples to persuade readers. According to Ferris (1994), second language learners have difficulties in accomplishing this type of writing because of lexical deficiencies as well as differences in rhetorical patterns between L1 and L2. However, it has yet to receive much research on lexical bundles. Thus, more studies on L1 and L2 argumentative writing are needed to investigate lexical units and their usage, which might assist L2 learners in learning more lexical expressions used in L1 writing. Among those studies that have examined this junction, Pang (2010) conducted a study comparing the use of lexical bundles by Chinese learners of English and native English speakers at the university level. The non-native speakers used lexical bundles more frequently than the native speakers; the Chinese learners employed more VP-based bundles, NP-based bundles, and stance bundles, third person plural, while the native speakers used an extensive impersonal stance bundle. Shin (2018) conducted a dissertation on lexical bundles in argumentative essays written by native and non-native freshmen. Especially noteworthy in her study is her analysis of the functions of lexical bundles regarding semantic preference and prosody, making her dissertation more informative as well as helping readers understand the functions of lexical bundles used by different language groups.

Similarly, in a study by Bychkovska and Lee (2017), a comparison of lexical bundles in argumentative essays by L1-English and L2-Chinese undergraduate students was conducted and showed a similar finding that L2-Chinese students used substantially more bundles than L1-English students, for example, *a lot of time, a lot of people, more and more people*, etc. According to the authors, the possible reason for quantitative differences between the two groups' writings might be that L2 learners translated the L1 language directly from the L2 language. Additionally, it was found that L2 students misused a

certain number of bundles, which related grammatical mistakes with articles and prepositions. This finding was similar to the result in Joharry's (2021) study, which showed Malaysian students' overuse of a limited number of lexical bundles in argumentative writing.

Regarding the Vietnam context, several studies on lexical bundles in argumentative writing have been investigated (Dang & Vo, 2020; Nguyen, 2023). The study by Dang and Vo (2020) analysed the structures and functions of bundles in argumentative writing by Vietnamese L2 majors of English. Their findings showed that these learners employed a large number of 3, 4 and 5-word bundles, and overused the clausal structure, and stance function. Another study by Nguyen (2023) compared bundles used in argumentative essays by Vietnamese first-year students and English native professional experts. The result revealed that in terms of structures, clausal bundles are highly preferred in two group writings; concerning function, discourse organizers were most commonly used in student writing, and stance bundles were prominent in expert writing. While these studies mainly focus on writings by Vietnamese students or compare them with native professional writers, none of these compare bundles used by Vietnamese student writers and native student writers. Therefore, this study examines lexical bundles in argumentative essays written by L2-Vietnamese university students and those written by L1-American university students to determine the similarities and differences in the usage between the two groups as well as reveal the potential problems in using the lexical bundles by L2 learners, which might raise Vietnamese student writer's awareness of the importance of bundles in the genre of argumentative writing and deploy them smoothly.

III. METHODOLOGY

1. Data

Two corpora are used in this study. The first corpus (VE) includes 264 essays written by Vietnamese EFL learners. These EFL learners are second-year students in the English Department at a national university in Vietnam. They took the institutional IELTS test at the end of the second semester and got band 5.5-6.5 scores, indicating an intermediate English level. On the test day, they were given several topics to choose from and then wrote an argumentative essay of about 230 words within 45 minutes. Five different topics were provided to the students to discuss and present their viewpoints as shown in Table 1.

The second corpus is LOCNESS (Louvain Corpus of Native English Essays), which contains British and American university students' essays. LOCNESS consists of argumentative and literary essays on various topics. To meet the compatibility with

Vietnamese university students' corpus in terms of genre, 156 argumentative essays (AE) written by American students are extracted. The general features of the two corpora are presented below.

TABLE 1
General Features of Two Corpora

Corpus	Demographics	Topics	Word Count	No. of Texts
VE	Vietnamese learners	5 topics 1. Disruptive students should be grouped together and taught separately. 2. As online learning becomes more common and more resources are converted to digital form, some people suggest that public libraries should be shut down, and everyone should be given an iPad with an e-reader subscription. 3. A big salary is much more important than job satisfaction. 4. Some people think that an international sports event is good for the country, while other people think it is bad. 5. Children should begin their formal education at a very early age (before 7 years of age).	81148	264
AE	American learners	A number of topics, such as 1. Controversy in the classroom 2. Animal testing 3. Capital punishment 4. Yoga 5. Gender roles in our society 6. Money is the root of all evil	121642	156

2. Identification of Lexical Bundles

To retrieve the lexical bundles by using an automated corpus tool, there are three criteria: frequency, dispersion, and the length of the word combination. The first one is the frequency threshold, which is somewhat arbitrary (as in Biber et al., 2004; Hyland, 2008b). In previous studies, based on the research aims or corpus size, researchers set frequency thresholds variably: at 10 times per million words (Biber et al.,1999), 20 times per million words (Byrd & Coxhead, 2010; Cortes, 2002, 2006; Hyland, 2008a, 2008b), 25 times per million words (Chen & Baker, 2010), 30 times per million words (Ruan, 2017; Yoo & Shin, 2022) or 40 times per million words (Biber & Barbieri, 2007; Corad & Biber, 2005). The second criterion concerns the dispersion threshold, which refers to the number of texts

in which a word combination occurs. In the previous study, to avoid idiosyncratic uses by individual authors, the dispersion thresholds are set variably at 10% of the text (Hyland, 2008a, 2008b) or 3-5 texts (Adel & Erman, 2012; Biber & Barbieri, 2007; Chen & Baker, 2010; Cortes, 2004), or 20 texts (Biber & Barbieri, 2007). The last criterion is the length of word combinations, which may range from three to six words or longer. Among these, four-word bundles have been investigated the most frequently in previous studies since four-word bundles could hold the three-word bundles, and they present a range of structures and functions to analyse (Cortes, 2004). Following this line, in this present study, four-word lexical bundles were automatically retrieved by the software tool Antconc (Anthony, 2020). Regarding the frequency threshold, the cut-off point was set at 30 times per million words, and the dispersion threshold was set at four different texts to retrieve four-word lexical bundles in the two corpora.

After the retrieval, the authors used a concordance tool to manually check whether word clusters derive from the same longer sequences or contain topic-specific bundles, such as *an international sports event*, which was extracted from the essay topic "Some people think that an international sports event is good for the country, while other people think it is bad", then excluded these clusters from the bundle lists.

Additionally, the authors manually checked for bundle overlapping which might inflate the number of types and tokens via concordance analyses. According to Chen and Baker (2010), there were two types of overlaps, including complete overlap and complete subsumption. The first one, complete overlap, referred to a case that two 4-word bundles were derived from a 5-word bundle, such as two bundles *will support my opinion* (10 times) and *support my opinion with* (10 times) coming from one 5-word combination *will support my opinion with*. The second type of overlap was complete subsumption, referring to two or more bundles overlapping and the occurrence of one bundle subsuming others. For example, *my point of view* appeared 16 times, while *from my point of* view occurred 10 times, these bundles occurred as a part of the word combination *from my point of view*.

Finally, the lexical bundles were categorised according to Biber et al.'s (2004) structural and Hyland's (2008a) functional framework.

IV. RESULTS AND DISCUSSION

As can be seen in Table 2, there are 224 types of four-word bundles in VE and 96 types in AE, whose larger data generates fewer lexical bundles. This result seems to indicate that EFL learners employ lexical bundles more frequently than American students. A possible explanation for this finding might be due to a discrepancy in corpus size. A large corpus has a broader distribution, which could lead to retrieving fewer word combinations

(Chen & Baker, 2010). Another reason might be the difference in the number of topics in the two corpora. While VE consists of five topics, AE includes various topics, for example; yoga, prayer in schools, capital punishment, animal testing, etc. To guarantee the compatibility of the two corpora, the topic-specific bundles were deleted as mentioned in Section III.

TABLE 2
Number of Lexical Bundles in Two Corpora (Types & Tokens)

Corpus	Number of Bundles (types)	Number of Bundles (tokens)
VE	224	1645
AE	96	610

1. Structure of Lexical Bundles

The structural categorization is based on the taxonomy devised by Biber et al. (1999) for describing structural patterns of lexical bundles in academic writing. This taxonomy has 12 major structural categories distinguished in academic prose. However, the subcategorisation of passive verb+prepositional phrase fragment *(*e.g., *is shown in the)* is excluded from this study because this is not found in the two corpora. Instead, the verb phrase with an active verb subcategory (e.g., *disagree with this idea*) from the conversation register (Biber et al., 1999) is added to show all patterns occurring in the data. Additionally, to have a broad view of bundle structures, these categories are grouped into three more general categories: NP-based bundles, PP-based bundles, and VP-based bundles (Chen & Baker, 2010). NP-based bundles consist of noun phrases with *of* or any noun phrases with post-modifier fragments, for example, *my point of view* or *my opinions with example.* PP-based bundles refer to phrases that start with a preposition, then noun phrases added, such as *as a result of*. VP-based bundles indicate any word combination with a verb (Chen & Baker, 2010), such as *it is true that* or *believe that they are*. Table 3 presents the distribution of lexical bundles across structural categories in VE and AE.

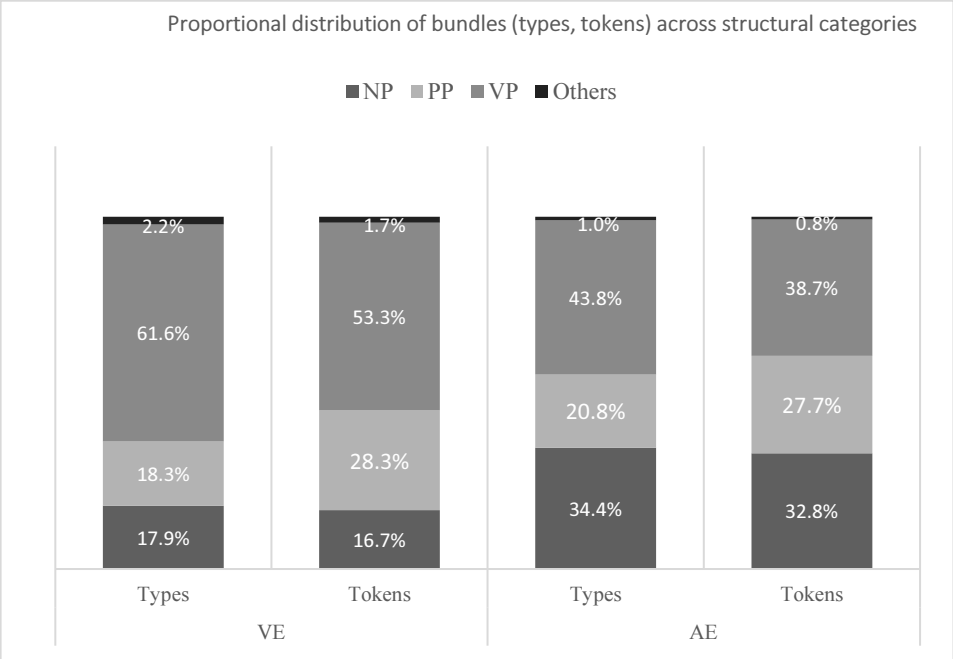
TABLE 3
Distribution of Structural Categories Across Corpora

Categories	Subcategories	Types		Tokens(pmw)	
		VE	AE	VE	AE
NP-based	1a. NP with *of* (*my point of view*)	24	23	169 (2082.6)	140 (1150.92)
	1b. NP with other post-modifiers (*my opinions with examples*)	6	8	40 (492.9)	50 (411.04)
	1c. Other NP (*a second major reason*)	10	2	65 (801.01)	10 (
PP-based	2a. PP with *of* (*as a result of*)	9	9	52 (640.80)	77 (633.01)
	2b. Other PP (*in the long run*)	32	11	414 (5101.7)	92 (756.32)
VP-based	3a. Anticipatory it +VP/ADJ (*it is true that*)	21	5	149 (1836.2)	26 (213.74)
	3b. Copula be +NP/ADJ (*is a valid point*)	12	4	82 (1010.5)	27 (221.96)
	3c. (VP) + that clause (*believe that they are)*	11	5	56 (690.1)	24 (197.30)
	3d. (VP/ADJ+) to-clause (*do not need to*)	5	13	32 (394.3)	79 (649.45)
	3e. Adverbial clause (*when it comes to)*	4	3	20 (246.5)	18 (147.98)
	3f. Pronoun/NP + be/VP (*I would like to*)	52	8	350 (4313.1)	40 (328.83)
	3g. VP with active verb (*support my opinion with*)	33	4	188 (2316.8)	22 (180.86)
ther	3h. Others (*more and more popular*)	5	1	28 (345.1)	5 (41.10)
	Total	224	96	1645 (20271.6)	610 (5014.7)

pmw = per million words

Since the two corpora sizes are different, to see differences in using three structural categories of bundles between two groups, the raw number of types and raw tokens of bundles were normalised as percentages as shown in Figure 1. As seen in Figure 1, both L1 and L2 prefer using VP-based bundles to other phrasal bundles. However, compared to AE, there is a denser use of VP-based bundles in VE, 61.6% of bundle types, and 53.3% of bundle tokens, which suggests that L2 learners frequently employ this structure. In AE, although the VP-based bundles are still the largest structural type by L1 learners, the total number of this structural bundle is less than both NP- and PP-based bundles. According to Biber et al. (1999, 2004), clausal bundles consist of VP-based bundles, while phrasal bundles include mainly NP or PP. As mentioned in the previous study (Biber et al.,1999), while academic prose mainly uses phrasal bundles, clausal bundles are common in conversation. In this study, AE consists of more phrasal bundles (55.2% of types and 60.5% of tokens) than clausal bundles (43,8% of types and 38.7% of tokens), while lexical bundles in VE are clausal than phrasal. This result is consistent with the previous study by Pan, Reppen, and Biber (2016). The following sections will elaborate on the differences in each structural category of the two groups.

FIGURE 1
Distribution of Lexical Bundles in Structural Categories in Two Corpora



1) Comparison of NP-based Bundles

Looking at Table 4, there is not much difference between the two groups using NP-based bundles with *o*f, over 60%. While L2 learners prefer utilising NP plus *of* to describe quantity, such as *a large amount of, a large number of, a great deal of, a great amount of, a huge amount of,* etc., L1 learners tend to use this structure to indicate a location or time, for example, *the end of the, the beginning of the, the rest of the, a part of the*, etc. This result might indicate that using the preposition *of* is considered the most productive frame to produce noun phrases in writing.

TABLE 4
Proportional Distribution of Subcategories Within NP-based Bundles

NP-based	VE		AE	
	Types (%)	Tokens (%)	Types (%)	Tokens (%)
1a. NP with *of*	60	61.7	69.7%	70
1b. NP with other post-modifiers	15	14.6	24.2%	25
1c. Other NP	25	23.7	6.1%	5

Regarding NP with other post-modifier fragments, as seen in Table 3, though there are six types in VE and eight types in AE, both groups share three bundle types, including *an important role in, the best way to, the fact that the.* Converging with the finding by Biber et al., (1999), the subcategory noun +complement clause combination, such as *the fact that the,* recurs frequently in academic prose. In other words, these bundles might be crucial in writing academic essays.

2) Comparison of PP-based Bundles

As shown in Table 5, in the AE, two PP-based fragments occur with a proportion of around 50%, meaning L1 learners are more flexible in employing PP-based bundles in their writings. In the VE, other PP phrase accounts for the most PP-based structures, 78% of types, and 88.8% of tokens. Detailed investigation of this pattern in two corpora shows that two groups share several other PP-based bundles, *on the other hand, at the same time, in the long run, to the fact that*. The bundle *on the other hand,* accounts for the highest proportion, around 1/5 of the total tokens of other PP patterns in both corpora. Though this bundle has idiomatic meanings and is used as linking adverbials to compare two events (Biber et al., 1999), there is a slight difference in use between the two groups; while L1 learners utilise it as an independent adverbial, L2 learners tend to co-occur with *on the one hand* as pair bundles.

TABLE 5
Proportional Distribution of Subcategories Within PP-based Bundles

PP-based	VE		AE	
	Types (%)	Tokens (%)	Types (%)	Tokens (%)
2a. PP with *of*	22.0	11.2	45.0	45.6
2b. Other PP	78.0	88.8	55.0	54.4
Total	100	100	100	100

3) Comparison of VP-based Bundles

Scrutiny of all VP-based patterns in each corpus indicates that the (verb/adjective) to-clause, pronoun/NP+ be/VP, and VP with active verb patterns merit noteworthy attention. The Verb/adjective to clause constitutes the highest proportion of all VP-based patterns in the AE (31% of types and 33.5% of tokens), approximately three times higher than in VE. However, L1 learners favour the *to*-clause bundle more than L2 learners; both groups of writers use this structure mainly to indicate possibility and ability (Biber et al.,1999). Such examples include *are more likely to, not be able to, will be able to, are not able to, would be able to*, etc. This might indicate that L1 learners are inclined to use a *to*-clause structure

as an efficient way to express their stances on information, which is illustrated in the below extracts from AE:

> *It would be able to explain why the money is not being spent on things which are much more important. (AE)*

> *These generations will be able to look in retrospect and see what we thought was important. (AE)*

TABLE 6
Proportional Distribution of Subcategories Within VP-based Bundles

VP-based	VE		AE	
	Types (%)	Tokens (%)	Types (%)	Tokens (%)
3a. Anticipatory it +verb phrase/adjective phrases (*it is true that*)	15.2	17.0	11.9	11.0
3b. Copula be +noun phrase/adjective phrase (*is a valid point*)	8.7	9.4	9.5	11.4
3c. (Verb phrase) + that clause fragment (*believe that they are*)	8.0	6.4	11.9	10.2
3d. (verb/adjective+) to-clause fragment (*do not need to*)	3.6	3.6	31.0	33.5
3e. Adverbial clause fragment (*when it comes to*)	2.9	2.3	7.1	7.6
3f. Pronoun/noun phrase + be/VP (+…) (*I would like to*)	37.7	39.9	19.0	16.9
3g. VP with active verb (*support my opinion with*)	23.9	21.4	9.5	9.3
Total	100	100	100	100

As shown in Table 6, the presence of the pronoun/noun phrase +be/VP stands out in both corpora. This structure is the most commonly employed in VE (37.7% of types and 39.9% of tokens) and ranks second in the AE (19% of types and 16.9% of tokens). Examination of this subcategory in both corpora, it is clear that both groups of writers use the first person pronoun plus a stative main verb, for example, *I disagree with the, I firmly believe that, i agree with the*, *I would like to* and *I feel that if, etc.* The majority of these bundles are located at the beginning of sentences, which enables writers to present their thoughts about the ideas in the following clause, for example:

> *I would like to say to the editors and publishers of today and to the newscasters, please get off this merry-go-round that you are on (AE)*

> *I firmly believe that this view is not completely true and will discuss the reasons why in this essay. (VE)*

The VP with active verb ranks the second in the VE, around doubling those in AE. Examining this structure closely in VE reveals that a substantial number of verbs are stative verbs. This finding reinforces the above findings, in which L2 learners tend to use personal pronouns plus the stative verbs to express their stance. The finding also highlights the differences in how the two group writers express their ideas. While L1 learners prefer using *to*-clause, L2 learners consider using pronoun/noun phrases plus stative verbs as an accessible and straightforward method. However, according to Biber et al. (1999), the subcategory of the pronoun/noun phrase +be and VP with active verb is the most prevalent type of bundle in conversation. Thus, this might indicate that using these bundles frequently produces a more colloquial writing style.

2. Function of Bundles

Biber, Conrad, and Cortes (2004) propose a functional taxonomy of lexical bundles based on their meanings and uses in the text. This taxonomy was revised and extended, providing three primary functions: (1) stance expressions, (2) discourse organisers, and (3) referential expressions.

> Stance bundles express attitudes or assessments of certainty that frame other propositions. Discourse organisers reflect the relationship between prior and coming discourse. Referential bundles make direct reference to physical or abstract entities or the textual context itself, either to identify the entity or to single out some particular attribute of the entity as especially important. (Biber et al., 2004, p.384)

This taxonomy has been adopted and extended in several studies (Adel & Erman, 2012; Chen & Baker, 2010; Cortes, 2006; Cortes, 2013; Pang, 2010). Hyland (2008a) created a new framework representing the functions of lexical bundles in academic writing from four disciplines. His framework comprised three major categories: research, text, and participant-oriented bundles. While Biber et al. (2003, 2004)'s functional framework was produced by analysing lexical bundles in both conversation and academic prose, Hyland (2008a)'s taxonomy focused on academic writing; this study analyses lexical bundles in academic writing; therefore, Hyland (2008a)'s taxonomy is more appropriate to be employed to analyse the functions of lexical bundles. The three main categorisations are retained, and the subcategorisations are modified. In categorising the research-oriented bundles, the topics related to the field research are eliminated as the topic-related specific bundles had been

omitted. Table 7 presents the distributions of bundles across functional categories. As shown in Table 7, there are differences in the use of functional patterns of bundles between the two groups. The log-likelihood test comparing tokens illustrates that these functional subcategories are more frequently used by L2 relative to their English counterparts.

TABLE 7
Functional Distribution of Lexical Bundles Across Two Corpora

Categories	Subcategories	Types		Tokens (pmw)	AE	LL
		VE	AE	VE		
Research-oriented	A1 Location, time (*in the long time*)	9	17	88 (1084)	119(978)	0.53
	A2 Quantification (*a great deal of)*	23	13	153 (1885)	89(731)	52.88*
	A3 Description (*one of which is*)	6	1	29(357)	4(33)	32.83*
	A4 Procedure (*first of all it*)	12	5	77(948)	23(189)	56.7*
Text-oriented	B1 Transition signals (*on the one hand)*	18	5	218 (2686)	45(370)	204.62*
	B2 Resultative signals (*as a result of)*	15	6	119 (1466)	38(312)	83.05*
	B3 Structuring (*this essay will discuss*)	4	0	72(887)	-	-
	B4 Framing (*in the case of)*	25	18	136 (1675)	107(879)	25.1*
Participant-oriented	C1 Stance features (I *would like to*)	97	29	664 (8182)	174 (1430)	538.05*
	C2 Engagement features *(if you do not)*	13	1	81(998)	5(41)	115.33*
Others	*he or she is*	2	1	8(98)	6(49)	1.67
	Total	224	96	1645 (20271)	610 (5014)	1004.04*

* = significant at $p < 0.0001$ level, pmw = per million words

TABLE 8
The Standardised Residuals (R) in Achi-Square Contingency Table for Functional Distribution (Tokens)

$\chi^2 = 80.6158$, $df = 2$, $p = .000$ Cramer's $V = 0.189$		Research-Oriented Bundles	Text-Oriented Bundles	Participant-Oriented-Bundles
VE	Observed	347	545	745
	Expected	425.1	536.9	675
	R	**-3.8**	**0.3**	**2.7**
AE	Observed	235	190	179
	Expected	156.9	198.1	249
	R	**6.2**	**-0.6**	**-4.4**

To see which functional categories make the most significant contribution to the difference between VE and AE, the standardised residuals (R) in a chi-square contingency

table is conducted as shown in Table 8. It is clear that there is a significant difference in the functional distribution of tokens between the two groups. Additionally, the R-values show that L1 learners employ significantly more research-oriented bundles and fewer participant-related bundles than expected. By contrast, L2 use considerably more participant-oriented bundles than expected. These results will be investigated further in the following sections.

1) A Comparison of Research-Oriented Bundles Across Two Corpora

According to Hyland (2008a), the primary function of research-oriented bundles is to help the writer organise their activities and experiences of the real world. The research-oriented bundles can be applied in argumentative essays to present detailed information on time, location or evidence to persuade the readers of the validity of their arguments. As shown in Table 9, within the VE corpus, a comparison across subcategories showed that bundles functioning as quantification and location were used more than other categories of discourse functions. The L2 learners used bundles relating to quantification more frequently than the L1 learners. This result is in line with the finding of Bychkovska and Lee (2017) but not similar to the findings of several previous studies (Cortes, 2004; Pan & Biber, 2016), which showed that L1 learners employed more quantification bundles than L2 learners. Under this pattern, there was a prominent use of the frame *a* +adj + N+ *of*, for example, *a large number of, a large amount of, a wide range of, etc*., supporting Hyland's (2008a) finding that several functional categories were connected to the structural patterns, with noun phrase plus of prominent in research-oriented functions. In interviews with L2 learners of this study, they explained the reasons why they took advantage of these lexical bundles since in English writing class, their instructors provided them with these bundles for describing quantity during the writing process.

TABLE 9
Proportional Distribution of Subcategories Within Research-Oriented Bundles

Research-Oriented Bundles	VE		AE	
	Types (%)	Tokens (%)	Types (%)	Tokens (%)
A1. Location: indicating times/place	18	25.4	47.2	50.6
A2. Quantification	46	44.1	36.1	37.9
A3. Description (one of which is)	12	8.4	2.8	1.7
A4. Procedure	24	22.2	13.9	9.8
Total	100	100	100	100

Regarding location-related bundles, L1 learners used twice as many as L2 learners. Two bundles used by two groups of writers are *all over the world* and *at the same time*. This finding is partly similar to the findings of Bychkovska and Lee (2017). The explanation for the abundance of *all over the world* in L2 writing might be due to their general tendency

(Chen & Baker, 2010).

In both VE and AE, the two remaining functions of research-oriented bundles, description and procedures, are identified by structural patterns of NP and PP, such as *the development of the, an example of this, in the form of, in a way that,* etc. Though these bundles occur in a lower proportion, they still contribute to the clarity of ideas presented in the essays, as shown in the below extracts:

> *Firstly, it will promote the development of the economy because the host country may receive investment and financing from funds, ticket sales, and many other sources. (VE)*
>
> *In this depiction, he chose to communicate in a way that would last through the ages. (AE)*

2) A Comparison of Text-Oriented Bundles

As shown in Table 10, bundles of framing signals make up the majority of text-oriented bundles in AE, at 62.1% of types and 56.3% of tokens. In VE, although framing signal bundles still account for the most extensive functional type (40.3 % of types), the number of tokens of these framing bundles (25%) is smaller than those of transition signals (40% of tokens). This suggests that L2 learners do not use the framing bundles as consistently as L1 learners. In AE, the framing bundles are used to specify cases and refer to a specific point in time or space as below:

> *In the case of Dr. Kevorkian the legal issues were shaky. (AE)*
>
> *With the end of the Cold War, there have been more peaceful relations throughout the world. (AE)*

TABLE 10
Proportional Distribution of Subcategories Within Text-Oriented Bundles

Text-Oriented Bundles	VE		AE	
	Types (%)	Tokens (%)	Types (%)	Tokens (%)
B1. Transition signals	29.0	40.0	17.2	23.7
B2. Resultative signals	24.2	21.8	20.7	20.0
B3. Structuring signal	6.5	13.2	0.0	0.0
B4. Framing signals	40.3	25.0	62.1	56.3
Total	100	100	100	100

At the same time, L2 learners prefer employing bundles of framing signals to orient readers to factual information or a framework for the subsequent content, as shown in the

below examples:

The fact that the divorce rates are going up at any moment (VE)

There are many reasons why divorces happen. (VE)

Another group of bundles used frequently in the text-oriented category is transition signals. These bundles help organise the flow of information and establish relationships between parts of discourse. In both corpora, a majority of these bundles express contrasting viewpoints or introduce new topics, helping readers follow and understand text with ease:

When it comes to the element of choosing a career, people are split on whether being well-paid or feeling pleased with work is more important. (VE)

On the one hand, many people say that unruly pupils would lose gradually their motivation in studying. (VE)

On the other hand, proponents for biological parents could easily use this point to their advantage by stating this ability of young children. (AE)

The bundles of resultative signals in both groups occur with a similar proportion (around 20%). These bundles are used to express cause-effect relationships and highlight that writers's claims are supported by reasons and evidence. This makes writers' arguments more credible and convincing, which is one of the crucial elements in argumentative writing. Among resultative bundles, though some lexical bundles consist of the same word, the two groups constructed them differently. An example can be taken from the usage of lexical bundles including *as a result*. In AE, this bundle is performed with PP plus of and a specific noun, while in VE, this takes the form of other PP to refer to the result of action in the previous clause or sentence.

As a result, they will never want to break their shell and feel vulnerable. (VE)

As a result of their racial discriminations, the African American black community has withdrawn to form their own University system. (AE)

The last group of bundles in the text-oriented category is structural signals found in VE, but not in AE. The majority of these provide readers with a frame in which new arguments or content will be discussed in the following sections of discourse:

This essay will discuss the reasons why I partly disagree with the above statement (VE)
I completely disagree with this idea and in this essay, I will support my opinion with examples. (VE)

3) A Comparison of Participant-Oriented Bundles

Participant-oriented bundles were predominantly used by two groups. They are persistent in VE, ranking as the highest category among functions. According to Hyland (2008a), participant-oriented bundles include stance and engagement; stance refers to the way writers explicitly express their voices, judgements or evaluations of their claims or statements; engagement refers to the engagement of readers, which writers consider readers as participants in the discourse. Looking at the distribution of subcategories in the participant-oriented bundles, it is clear that stance bundles were highly preferred by both groups of writers, around 90%, as shown in Table 11.

TABLE 11
Proportional Distribution of Subcategories Within Participant-Oriented Bundles

Participant-Oriented Bundles	VE		AE	
	Types (%)	Tokens (%)	Types (%)	Tokens (%)
Stance features	88.2	89.1	96.7	97.2
Engagement features	11.8	10.9	3.3	2.8
Total	100	100	100	100

Interestingly, several stance bundles shared by the two groups are impersonal; for example, *it is difficult to, it is obvious that, it is clear that, etc*. According to Hyland (2005), these bundles can be categorised as boosters, enabling writers to express their viewpoints with certainty and seek readers' solidarity. At the same time, unlike L1 learners, L2 learners prefer using the first person pronoun *I* in their writing, such as *I completely disagree with, I believe that it, I agree with this*, etc. These bundles show the writers's intention of mentioning themselves and their attitude to propositions. Hyland (2005) also mentions that employing the first person in discourse is to identify oneself with a particular claim and gain credit for writers' perspectives. Having a closer look at VE, we recognised that in most essays in VE, the introduction part consists of these bundles as below:

I strongly disagree with the statement because I believe that a child's school years are the first important stage in their social development stage. (VE)

From my point of view, I disagree with this opinion and I believe that job satisfaction is incredibly important. (VE)

Some argue that high salary is much more important than job satisfaction. In my opinion, I disagree with this view because I believe that job satisfaction plays a vital role in keeping an individual motivated at work. (VE)

In interviews with the participants and their instructors, instructors explained that they intended to teach these bundles in writing class since L2 learners have struggled to write the introduction part, so mastering a certain number of these bundles might facilitate L2 learners to express their viewpoints directly in a short time. However, this teaching method also has side effects. In the writing process, the learners did not provide enough evidence or examples to support their views, making their arguments more subjective than objective. Therefore, when teaching L2 learners how to use lexical bundles in writing, EFL instructors should explain the functions of bundles, helping L2 learners understand comprehensively and use them effectively.

V. CONCLUSION

This study compared lexical bundles used by L1 and L2 learners to see the differences between the two groups using four-word bundles. Generally, the findings show that, compared to L1 learners, L2 learners utilise a larger number of four-word bundles in almost all structural patterns and functional types. In particular, though both groups of writers prioritise VP-based bundles, L2 writers heavily rely on VP-based bundles with the structure of pronoun/noun phrase +be/VP to present the writers' stance. Also, L1 writers are found to be more flexible in employing PP-based bundles utilising different types of PP-based bundles in balance; L2 writers prefer using other PP-based bundles without *of*. It is also found that there is a similarity in the use of NP-based bundles between the two groups, favouring NP with of. Regarding research-oriented bundles, though both groups of writers, especially L2 learners, employ many bundles of quantification, L1 learners are highly dependent on bundles of location. In terms of text-oriented bundles, two groups of writers show their frequent use of framing bundles; however, L1 learners use these framing bundles to specify cases in detail, and L2 learners tend to present factual information. For participant-oriented bundles, stance bundles are commonly utilised by two groups to express their judgment or evaluation; L1 learners employ a great number of lexical bundles, starting with the personal pronoun *I* and booster, to show their commitment to their proposition.

The findings in this study provide insights into how non-native learners and English native learners use lexical bundles in argumentative essays, as well as present some existing problems in the use of lexical bundles by non-native English speakers. To help learners address problems, and master these bundles, formal instruction of lexical bundles should be

implemented in writing classes at universities (Cortes, 2004, 2006). Firstly, EFL instructors can introduce the definition of lexical bundles, structures and functions with specific examples. Secondly, to assist learners in having a deeper understanding of how these bundles are used in real discourse, the instructors can show the difference in using bundles between native and non-native speakers's writings. At the same time, instructors can show existing issues which non-native speakers have and then explain them explicitly. Lastly, beyond the scope of this study, besides using a corpus by native student writers, the instructors can take advantage of corpora by native professional writers to show how bundles are used, which might allow them to expose more bundles in the contexts and deploy them more correctly.

Finally, it has to be admitted that this study has some limitations, one of which is the differences in a number of topics in two corpora. The VE includes essays from 5 different topics, while AE contains writings on various topics, which might affect the number of bundles in two corpora. To mitigate the impacts on the findings of this study, the authors deleted all topic-specific bundles. The compilation of texts in the study is insufficient, focusing on writings by Vietnamese and American student writers, so no general rules are drawn. Further research should be conducted on a larger number of writings from different language backgrounds to further examine the questions raised by this study.

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.

Allen, D. (2009). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education*, *1*(4), 105-127.

Anthony, L. (2020). *AntConc* (Version 3.5.9). Tokyo, Japan: Waseda University. http://www.laurenceanthony.net/software.

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. (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. 71–92). 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.

Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. Longman.

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.

Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and 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*, 30-49.

Conrad, S. M., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica, 20*, 56-71.

Cortes, V. (2002). Lexical bundles in freshman composition. *Using Corpora to Explore Linguistic Variation*, *9*, 131-145.

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

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, *17*(4), 391−406.

Dang, P., & Vo, N. (2020). Vietnamese majors of English: The use of lexical bundles in argumentative essays. *Proceedings of the International English Language Teachers and Lectureture conference* (pp.79-85). Malang, Indonesia: Universitas Negeri Malang.

De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics, 3*(1), 59-80.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk*, *20*(1), 29-62.

Ferris, D. R. (1994). Rhetorical strategies in student persuasive writing: Differences between native and non-native English speakers. *Research in the Teaching of English*, *28*(1), 45-65.

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.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, *7*(2), 173-192.

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

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18*(1), 41-62.

Joharry, S. A. (2021). Repetitive bundles in Malaysian learner writing. *Australian Review of Applied Linguistics*, *44*(1), 82-108.

Kaur, S. (2015). Teaching strategies used by Thai EFL lecturers to teach argumentative writing. *Procedia-Social and Behavioral Sciences*, *208*, 143-156.

Mei, W. (2006). Creating a contrastive rhetorical stance: Investigating the strategy of problematization in students′ argumentation. *RELC Journal, 37*(3), 329-353.

Nguyen, N. X. (2023). Structures and functions of lexical bundles in student and expert timed argumentative writing: A corpus-driven study. *Pertanika Journal of Social Sciences & Humanities, 31*(3), 1015-1039.

Pan, F., & Liu, C. (2019). Comparing L1-L2 differences in lexical bundles in student and expert writing. *Southern African Linguistics and Applied Language Studies*, *37*(2), 142-157.

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.

Pang, W. (2010). Lexical bundles and the construction of an academic voice: A pedagogical perspective. *Asian EFL Journal, 47*(1), 10-11.

Ruan, Z. (2017). Lexical bundles in Chinese undergraduate academic writing at an English medium university. *RELC Journal*, *48*(3), 327–340.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 1-22). John Benjamins.

Shin, Y. K. (2018). *Lexical bundles in argumentative essays by native and non-native English-speaking novice academic writers* [Unpublished doctoral dissertation]. Georgia State University.

Shin, Y. K. (2019). Do native writers always have a head start over non-native writers? The use of lexical bundles in college students′ essays. *Journal of English for Academic Purposes, 40*, 1-14.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

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

Yoo, I. W., & Shin, Y. K. (2022). English lexical bundles in a learner corpus of argumentative essays written by Korean university students. *Corpora*, *17*(Supplement), 23-42.

Examples in: English

Applicable Languages: English

Applicable Levels: Secondary

Kevin Parent (professor)
Department of English Language and Literature
Korea Maritime and Ocean University
727 Taejong-ro, Dongsam-dong, Yeongdo-gu, Busan, South Korea
Parent@kmou.ac.kr

Thuy Quynh Bui (student)
Department of English Language and Literature
Korea Maritime and Ocean University
727 Taejong-ro, Dongsam-dong, Yeongdo-gu, Busan, South Korea
Quynhbui@g.kmou.ac.kr

Received: January 1, 2024
Reviewed: February 29, 2024
Revised version received: March 21, 2024