Combining a Corpus-Based Approach and Qualitative Analysis to Create a Pedagogically Useful List of Multiword Combinations for Research Publication

Attapol Khamkhien

Abstract

University lecturers and graduate students have been under increasing pressure to publish their research work in prestigious journals. Among other things, multiword sequences seem to be challenging for novice writers and graduate students struggling to make their manuscripts convincing and interesting for potential readers in their research communities. This study reports on the frequency and use of phraseological patterns in applied linguistics research articles. Using corpus linguistics methodology in combination with qualitative analysis, the present study identified phraseological patterns frequently co-occurring in research articles, their structural types, and pragmatic functions in contextual environments. To begin with, two, three, and four-word n-grams were extracted from a corpus of the four internal sections of 50 research articles (equating to approximately 205,187 running words) to identify the suitable n-grams to focus on. Based on their semantic and grammatical relations, three-grams appeared to be pedagogically interesting, were further explored. Five EAP instructors were invited to select the strings identified, contributing to a pedagogically useful list of 289 multiword combinations. The results of the structural analysis indicated that the majority of writers employed amount of noun-based structures most, followed by other structures, verb-based and prepositional-based structures. Functional analysis clearly revealed that the use of certain phraseological patterns was strongly associated with the communicative functions of a given section of the texts. Pedagogically, this study sheds light on language use in this particular writing research genre and provides more evidence-based instructional practices, especially advanced language courses targeted at scholarly reading and writing.

Keywords: Academic word list, TCI database, English for Academic Purposes, Writing for publication

บทคัดย่อ

อาจารย์มหาวิทยาลัยและนิสิตนักศึกษาระดับบัณฑิตศึกษาได้รับการผลักดันให้เผยแพร่ผลงานวิจัยในวารสารที่ได้รับการยอมรับและมีชื่อเสียง ทั้งนี้ การใช้กลุ่มคำศัพท์เพื่อให้ผู้อ่านบทความเกิดความสนใจและการเขียนบทความให้มีความน่าเชื่อถือ ในชุมชนวิจัยนั้นดูเหมือนจะเป็นสิ่งที่ท้าทายสำหรับผู้เริ่มต้นเขียนและนิสิตนักศึกษาระดับบัณฑิตศึกษา งานวิจัยนี้ เป็นการศึกษาความถี่และการใช้รูปแบบกลุ่มคำศัพท์ที่ปรากฏขึ้น

บ่อยครั้งในบทความวิจัย ด้านภาษาศาสตร์ประยุกต์ โดยใช้วิธีการทางภาษาศาสตร์คลังข้อมูลร่วมกับการวิเคราะห์เชิงคุณภาพการศึกษา รวมทั้งจำแนกประเภทโครงสร้าง และหน้าที่ของกลุ่มคำศัพท์เมื่อในไปใช้ในบริบทต่างๆ การวิจัยเริ่มจากการวิเคราะห์กลุ่มคำศัพท์ (n-grams) ขนาดสองคำ สามคำ และสี่คำ ตามลำดับจากคลังข้อมูลที่ได้จากสี่ส่วนหลักของบทความวิจัยจำนวน 50 บทความ (205,187 คำ) เพื่อระบุขนาดของกลุ่มคำที่เหมาะสม ในการวิเคราะห์ เบื้องต้นพบว่ากลุ่มคำศัพท์ขนาดสามคำมีความน่าสนใจและเป็น ประโยชน์ต่อการเรียนการสอนโดยพิจารณาจากความสัมพันธ์ของความหมายและไวยากรณ์ จากนั้น อาจารย์มหาวิทยาลัย จำนวน 5 ท่านได้ทำการเลือกกลุ่มคำศัพท์ขนาดสามคำดังกล่าวทำให้ได้รายการกลุ่มคำศัพท์ขนาดสามคำที่มีประโยชน์ และเอื้อต่อการเรียนการสอนจำนวน 289 กลุ่มคำ ผลจากการวิเคราะห์ทางโครงสร้างชี้ให้เห็นว่าผู้เขียนบทความส่วนใหญ่ ใช้โครงสร้างกลุ่มคำนามมากที่สุด ตามด้วยกลุ่มโครงสร้างอื่นๆ โครงสร้างคำกริยา และโครงสร้างคำบุพบท การวิเคราะห์อย่างเป็นระบบยังเผยให้เห็นว่า การใช้กลุ่มคำศัพท์มีความสัมพันธ์อย่างยิ่งกับหน้าที่การสื่อสารของแต่ละส่วน ของบทความที่กลุ่มคำนั้นๆ ปรากฏ งานวิจัยนี้แสดงให้เห็นความสำคัญของการใช้ภาษาในการเขียนงานวิจัยและเป็นประโยชน์ต่อการเรียนการสอนภาษาโดยเฉพาะอย่างยิ่งการนำผลไปใช้กับการพัฒนาหลักสูตรภาษาขั้นสูงที่มุ่งเน้นการอ่านและการเขียนเชิงวิชาการ

คำสำคัญ: ชุดคำศัพท์ทางวิชาการ, ฐานข้อมูล TCI, ภาษาอังกฤษเพื่อวัตถุประสงค์ทางวิชาการ, การเขียนเพื่อตีพิมพ์ในวารสาร

Introduction

Currently, writing a research article and getting acceptance to publish it in a peer-reviewed journal are vital. There is increasing pressure on Thai novice scholars and graduate students to publish in prestigious journals, particularly those indexed in the TCI (Thai-citation index) database. Most universities require their graduate students to publish at least one refereed article in refereed journals as part of their graduation. The task, nonetheless, may be difficult for them because being competent in writing for publication requires knowledge of the genre and its associated textual features (Feak & Swales, 2011; Hyland, 2004b). Macro- and micro-structures of the stretch of the texts have gained attention and prominence in writing instruction based on Swales' (1990; 2004) discoursal work on academic and research articles written in English. As a macro-structure, phraseological patterns are claimed to be helpful as it is "pre-packaging of information or of the structures used to present information" (Reppen, 2004, p. 83), which helps writers by reducing the processing load.

With current technology, phraseological patterns have established themselves as an important field of study within corpus linguistics. Implications for the teaching and learning of language can be made, in writing academic works and in several areas related to academic communication (e.g. Biber & Barbieri, 2007; Conrad & Biber, 2004; Cortes, 2004; 2006; 2013; Hyland, 2008a; 2008b; Li & Schmitt, 2009; Martinez & Schmitt, 2012). It is commonly agreed that this knowledge play a facilitative role in

learning and using a language because they represent fluent linguistic production, particularly in spoken language (Pawley & Syder, 1983), and academic texts (e.g. Biber, Johansson, Leech, Conrad & Finegan, 1999; Conrad & Biber, 2004; Cortes, 2013; Hyland, 2008a, 2008b; Nesi & Basturkmen, 2006).

Given the umbrella of formulaic language, scholars differently define and use various terms in research on phraseology. For instance, Altenberg (1998), in his seminal work, multiword units referred to the term "*recurrent word-combinations*" in investigating word patterns typically recur in spoken English. The term "lexical bundles" is used in many studies (e.g. Biber et al., 1999; Biber & Barbieri, 2007; Biber, Conrad & Cortes, 2004; Conrad & Biber, 2004; Chen & Baker, 2010; De Cock, Granger, Leech & McEnery, 1998; Hyland, 2008a; 2008b). Schmitt (2004) preferred the term "formulaic sequences", whilst "phraseology" and "phraseological patterns" introduced by Charles (2006) and Granger and Meunier (2008) were generally used to refer to sets of recurrent word combinations. Moreover, the terms referring to phraseological patterns include "lexical clusters" (Hyland, 2008a), "recurrent word combinations" (Altenberg, 1998), "phrasicon" (De Cock, Granger, Leech & McEnery, 1998), "n-grams" (Stubbs, 2007). Among these distinctive terminologies, Erman and Warren (2002) stated that multiword combinations denote "combinations of at least two words favored by native speakers in preference to an alternative combination which could have been equivalent had there been no conventionalization" (p. 31), which is quite similar to Biber et al. (1999) who discuss that, based on our intuition, they can be fixed expression or so-called idiomatic phrases that have fixed meanings and can be understood by language speakers cannot be included because lexical bundles are distinct from those. They are also semantically transparent. Wray (2002) preferred to use the "word sequences" and explained that they are stored in the mental lexicon. These strings are nevertheless glued together in everyday discourse.

Even though several terms are used to refer to phraseological patterns, it can be seen that these terms and definitions share some common ground as they are a type of word combination or phrases frequently recurring in text, performing particular discourse functions (Cortes, 2004), helping to shape meanings and contributing to a sense of coherence in a text (Hyland, 2008a). Biber (2006) asserted that the identification of phraseological patterns is a resolutely quantitative activity, based solely upon frequency and distribution criteria obtained from computer programs. Cortes (2004) suggested that basic techniques employed to identify lexical bundles are word frequency counts, while concordance lines, lexico-grammatical profiles, and keyword analysis are techniques that can be done with multiword combinations once they have been identified, in order to identify their functions in context and their occurrence in the texts.

Considerable attention has been paid to the role of phraseological patterns in language teaching and learning (e.g. Cortes, 2004; 2006; Li & Schmitt, 2009). For example, Cortes (2004) compared the frequency and function of lexical clusters in the written production of professional authors and student writing in history and biology. The study showed that some lexical bundles were extremely frequent in corpora of

research articles, while rarely found in the students' works. Cortes confirmed that the exposure to the use of lexical bundles in reading materials did not transfer directly into students' active production of lexical bundles in writing process. Therefore, the acquisition and appropriate use of lexical bundles does not seem to be a natural process, and, as such, corroborate Jones and Haywood's (2004) work. The study revealed that after a 10-week instruction period focusing on the production of lexical bundles, university students found that knowledge of word strings could assist them technically to express complex ideas in writing tasks, to facilitate the structure of different writing stages, and to show the necessary level of formality. Moreover, in Li and Schmitt's (2009) study, the development of students' formulaic sequence repertoire over the course tends to be slow, even though these students were language majors. Some studies (Durrant, 2015; Liu, 2012; Martinez & Schmitt; 2012; Simpson-Vlach & Ellis, 2010) are further complicated by the fact that the researchers have tried to create lists of academic words that are frequently used in different registers such as basic conversation, reading and writing (Nation, 2001), different academic journals and university textbooks (Coxhead, 2000), medical texts (Wang, Liang & Ge, 2008), academic writing in disciplinary differences (Durrant, 2015; Liu, 2012). These researchers have congruently stated that each subject discipline has its own way of argumentation, preferred forms, meanings and syntactical patterns (Martinez et al. (2009) because lexical items included in the lists may result from the shaping of the disciplines, text selection, and 'the particular ways of representing experience' (Yang, 2015, p. 30).

As can be seen, a number of academic word lists have been created in the literature to serve different purposes. Corpora used in each study are varied according to research purposes and interests, mostly taken from several disciplines. Therefore, to provide scaffolding support for Thai novice writers and graduate students to enhance the opportunities for scholarly publication, especially in journals indexed in the TCI database, knowledge of phraseology could provide them with a head start in embarking on an academic research writing task. The frequency of occurrence of the phraseological patterns identified may facilitate fluency in language production, especially how to start crafting their own articles effectively (Bybee, 2002). It is believed that, if learning to use the phraseological patterns of a discipline can contribute to gaining a communicative competence, identifying these lexical clusters may be advantageous so as to help these people acquire the specific rhetorical practices of the texts they have to produce (Hyland, 2008b). To accomplish this, this study sought to shed some light on the way language is directly experienced in academic articles by identifying the most frequent phraseological patterns and their pragmatic functions found in published research articles in applied linguistics. This study attempts to investigate the use of phraseological patterns in the research articles indexed in the TCI database and tries to answer the following questions:

1) What phraseological patterns do the article authors use in their academic writing for publication?
2) What are the structural types and functions of these phraseological patterns identified?

3) How are the pragmatic functions of these phraseological patterns identified associated to the functions of the given sections?

Related Studies on the development of academic word lists

Researchers acknowledge the usefulness of phraseological studies, and they aim to use the findings for pedagogic purposes, especially creating classroom materials to enhance student competence. Coxhead (2000), for example, created an Academic Word List (AWL) which contains 570 word families taken from a corpus of 3.5 million from four areas of arts, commerce, law and natural science. Each word included in the AWL list has to occur 10 times or more in each of the four main areas of the corpus, in at least 15 of the 28 subject areas, and more than 100 times in the entire corpus. With these criteria, the list can be then divided into 10 sub-categories based on frequency of occurrence. Due to its high coverage in academic text, both from research articles and textbooks, the AWL list has been claimed to be relevant for L2 learners with an academic purpose.

Simpson-Vlach and Ellis (2010) argue that frequency alone in identifying multiword combinations "does not necessarily imply either psycholinguistically salient sequence or pedagogical relevance; common sequences of common words, such as '*and of the*,' are expected to occur frequently" (p. 490). Their study combined measures of mutual information (MI) and frequency to investigate target corpora of academic discourse: MICASE, BNC files of academic speech, Hyland's (2004) research article corpus, selected BNC files, and genre categories for the BNC. The study first used 10 words per million (WPMs) as a cut-off criterion to identify three-, four- and five-word bundles. The MI score was then performed to assess the degree to which the words in a phrase occurred together more frequently than would be expected by chance. Qualitatively, 20 experienced instructors were invited to rate the formulas to determine whether the phases found were a formulaic expression, a phrase, or expression. Correlation analysis was performed with qualitative judgment data and the quantitative statistics to reassure the reliability and validity of these instructor insights. All of the processes made it possible to select multiword formulas and three functional categories were reached: referential expressions, stance expressions, and discourse organizers as suggested by Biber et al. (2004). Their study added contrast and comparison category into referential expressions group. Moreover, Simpson-Vlach and Ellis collapsed the categories of desire and intention/prediction into one called volition/intention. Finally, the discourse organizer category was modified.

Liu (2012), in parallel with Simpson-Vlach and Ellis' (2010) study, argued that some lexical bundles identified by previous research contained "structural/semantic incompleteness" (p. 27) and seemed neither "terribly functional nor pedagogically compelling" (Simpson-Vlach & Ellis, 2010, p. 493), for example, *and this is*, and *this is the*. Liu thus identified the most frequently-used multiword constructions (MWCs), covering a variety of fixed or semi-fixed expressions, in general academic writing across the academic divisions of the sub-corpora in Corpus of Contemporary American

English and British National Corpus. To determine the most common MWCs and examine their usage patterns, a cut-off value of 20 tokens PMWs was adopted. All MWCs identified that end with a/the, or any other incomplete NP (e.g. one of the) are represented in the form ending with "det+NP" (e.g. one of det+N). With this approach, Liu claimed that MWCs with "partially-filled constructions are more productive than the structurally-incomplete ones" (p. 28), facilitating students' learning process. A list of the 228 most common MWCs in general academic written English across the academic divisions, categorized into three bands according to frequency and semantic function, respectively, was finally created. It would be argued that the list generated may assist students in effectively grasping the constructions in their writing, raising their awareness of these common MWCs and encouraging them to use these MWCs in their writing. However, functions of the word list generated by the study are not provided, and thus work remains to be done on how best to make use of this resource.

While the description of common recurrent word combinations can help us understand something of the feature of academic writing, Durrant (2009) has argued that previous studies listing essential words for EAP students failed to include 'positionally-variable collocations' (p. 157). His study explored high-frequency collocations across academic discipline which learners need to acquire. He identified two-word collocations commonly found in research articles taken from written texts from five faculties: Life sciences, Science and Engineering, Social-Psychological, Social-Administrative, and Arts and Humanities. Using WordSmith Tools (Scott, 2004), the total frequency of all collocations found were compared with their frequency in an 85 million word of the BNC corpus. To arrive at the list of collocations used, these collocations needed to meet a minimum mutual information score of at least four in all five of the subject groupings. The analysis yielded a list of distinctively academic collocations comprising the 1000 most key items. A striking point is that most of the word pairs were grammatical collocations, containing at least one non-lexical word such as "prepositions, determiners, primary and modal verbs, conjunctions, subordinating adverbs, pronouns and numerals and ordinals other than *one* and *first*" (p. 163). Moreover, key collocations were reporting pattern verbs such as argue, assume, conclude, confirm, and demonstrate, to name a few, with that-clause construction. Unfortunately, like Liu (2012), the study looked at the grammatical forms, not at the functions of collocation, which limits the pedagogic usefulness of the list.

Similarly, Martinez and Schmitt (2012) recognized that most of the vocabulary list such as the Academic Word List (AWL) by Coxhead (2000), provides only individual words or "only the tips of phraseological icebergs" (p. 302), which might not be comprehensive enough for pedagogic purposes. Martinez and Schmitt, therefore, combined frequency and qualitative criteria in selecting individual words and phrasal expressions. The BNC corpus was selected as the corpus source, and Wordsmith Tools (Scott, 2004) was used to search for any two- and four-words strings repeated in the corpus at least five times. Then, a series of "Core criteria" and "Auxiliary criteria" (p. 308-310) were taken into account as a needed guidance to help justify intuitions regarding what may or may not be formulaic when selecting a multiword expression to be included on the list. Exhaustively, a random sampling technique was applied to the

process of searching derived multiword lexical items line-by-line to check whether these items were phraseological polysemy. The final list called the PHRASE List ended up consisting of 505 multiword items, which was claimed to be "useful for pedagogic materials including more multiword items, such as textbooks, graded readers, and language tests" (p. 316). However, the benefits of applying the PHRASE List are still open to question because, similar to Durrant (2009) and Simpson-Vlach & Ellis (2010), the crux of the matter is that the functions of these multiword items were not included in the list, which might at first seem difficult, particularly for lower-proficiency learners.

The studies outlined above all proved very different approaches to create the pedagogic useful list, enabling us to see the importance and application of the research on corpus-based studies. These researchers and scholars have therefore increasingly come to see phraseological patterns as important 'building blocks of coherent discourse and as characteristics of language use in particular settings' (Hyland, 2008b, p. 8). It is interesting to note, however, the size and types of corpora compiled and the approaches adopted for investigation and a wide variety of texts as well as sources of corpus taken in these studies should be taken into consideration.

Corpus and Methods

The corpus of this study was systematically collected from 50 research articles taken from nine journals indexed in the TCI database. The main and practical reason for the selection of these articles was that, based on the results of journal quality evaluation of the TCI database, all the journals were classified in tier 1 which would be further included in the ASEAN Citation Index (Svasti & Asavisanu, 2007). Also, Thai graduate students and writers have been pushed to publish their research work in these journals. The sampling of the journals was restricted to the year 2013 and 2014 only to control for potential changes within the discipline and across time and to enhance the coherence and validity of the results of the study. Regarding the size of the corpus, Bowker and Pearson (2002) stress that "there are no hard and fast rules that can be followed to determine the ideal size of a corpus" (p. 45). We therefore decided to analyse 50 applied linguistics research articles as this corpus size is manageable and relatively suitable for the kind of analysis and objectives of the study, and it is possible to get much useful data and in-depth information from this corpus.

It is acknowledged that some other factors e.g. style of writing used in the articles, peer-review process and copy editing process would remain in these articles. These variations, however, are not taken into account in the study. Given the focus of this study on the investigation of the four internal sections (Introduction, Methods, Results, Discussion or IMRD) of the articles, other sections of the articles were excluded from the analyses, including all the tables, figures, notes, abstracts, references, and appendices in each text. These systematic procedures resulted in the corpus of approximately 205,187 running words. All of the selected 50 articles were then assigned numbers from 1 to 50 together with abbreviation representing the name of journal (e.g. [JES1] to [MNY 50]) to ease further reference and analysis in the study.

To investigate frequency statistics for sequences of words in the corpus, n-grams were generated with the use of SketchEngine (SkE) software (Kilgarriff et al, 2004). Since the SkE requires plain text to identify the formulaic patterns occurring in the text, the procedure began with first cleaning all of the texts by removing non-textual content. All the edited files were then saved, corresponding to the IMRD sections. We used the wordlist option to investigate two-, three-and four-word n-grams, which we refer to as high frequency formulaic expressions in the corpus. A number of issues need to be considered when identifying phraseological units based solely on frequency of occurrence. First, as n-grams are defined essentially by their frequency of occurrence, the cut-offs of frequency are somewhat arbitrary (Hyland, 2008b). The frequency threshold was set; each frequent n-gram reported occurs with the minimum of five times in the entire corpus. Second, to make comparisons of the findings of the study with previous studies possible, Biber et al. (1999) suggest a convenient formula for normalizing frequencies. Based on the length and number of words of the corpus, the choice of norming to 1,000 words was appropriate to use in the present study. Although using n-grams recurring at least five times per thousand words as selection criteria remains somewhat arbitrary, this cut-off point is determined by the total number of words and by the goals of this research as to examine the usage of phraseological patterns in the corpus. Besides, we carefully checked all the n-grams generated to ensure that they occurred in at least three files in the corpus, representing the occurrence of such n-grams in at least three articles. This practice was necessary to guard against the subjectivity and idiosyncratic expressions introduced by individual writers.

Four-word bundles are claimed to be more phrasal in nature (Biber & Barbieri, 2007; Biber et al., 2004; Chen & Baker, 2010; Cortes, 2004; 2006; Grabowski; 2015; Hyland, 2008a; 2008b). However, in the analysis, we started generating the list of two-word n-grams and found that they appeared ungrammatically complete (e.g. *of the, in the, to the*). According to Simpson-Vlach & Ellis (2010), the incomplete bundles are "neither terribly functional nor pedagogically compelling" (p. 493). Taken together, most of the four-word n-grams (e.g. *simple past tense form, intrinsic motivation of English*) were content-based lexical items in relation to a particular subject-matter, which are simply an artefact of what the writers are writing about. With regard to a pedagogic implication, they might not provide many implications for the whole context and the register in which they are written, compared with those n-grams which are pragmatically and grammatically complete units.

The three-word n-grams identified in the analysis appeared to be more interesting than the others because many of them constitute complete syntactic units as being independent meaningful phrases, including a number of grammatical items, expressing semantic relations (e.g. *in order to, as well as*), which are not simply content-based items. Even if most of them do not represent complete structural units (e.g. *the use of, the results of*), they are still seen as "important building blocks in discourse" (Biber & Barbieri, 2007, p. 270).

As for qualitative criteria, we removed from the list content-based strings or noun groups (e.g. *language learning strategies, teaching and learning*). By application of this

qualitative criterion, we arrived at a first list of 476 potential n-grams. We then applied a further set of selection criteria. We went down the n-gram list item-by-item, looking for 'plausibly formulaic' multiword strings (Wray, 2009, p. 41) which realise meanings or pragmatic functions. To assure a high degree of reliability of the list, five EAP experienced English instructors were invited to select the items that appear to be pedagogically useful for article reading and writing. These instructors, at the time of study, were pursing their PhD at different universities in the United Kingdom. They were advised that the purpose of the exercise was the construction of the list of phraseological patterns worth learning and teaching when writing an article for publication. From the instructor insights, each of the potential three-word n-grams carefully chosen by at least three instructors was included in the final list. The selected multiword combinations were further explored in order to investigate to what extent they are used by article writers. They were then categorized and structurally according to their grammatical types and functionally in terms of their pragmatic meaning in each section of the texts they appeared.

Findings

1. Frequency and Structures of the Phraseological Patterns Identified

Through the process and these criteria, 289 different multiword combinations were categorized structurally in terms of their grammatical types. We used a mixture of a priori classification and inductive classification. That is, we initially looked through the list using Biber et al. (1999; 2004)'s classification consisting of three major types: 1) verb phrase fragments, 2) dependent clause fragments, and 3) noun or prepositional phrase fragments, but found that not all selected complete structural units seemed to nicely fit the taxonomy. We therefore attempted a more bottom up classification structure, with a resulting set of categories as follows: noun-based, verb-based, preposition-based structures, and other structures, with specific sub-types under each category.

Table 1 summarizes the structural types of the selected multiword items with the corresponding type, frequencies and percentages.

Table 1 Structural classification of the selected multiword items

Structure	Examples	Fre-quency	% of all structures
Noun structure			
Noun phrase with of-phrase fragment	a corpus of, a lot of, a number of, a result of, a variety of, all of the ,an analysis of, analysis of the, development of the, each of the, findings of the, frequency of the, good	74	25.61

	level of, knowledge of the, large number of, majority of the, most of the, number of the, one of the, part of the, some of the, the context of, the development of, the effectiveness of, the effects of, the findings of, the frequency of, the importance of, the number of, the part of, the percentage of, the process of, the purpose of, the result of, the role of, the study of, the success of, the use of		
Noun phrase with other post-modifier fragment	criteria based on, participants in the, results from the, the findings from, The results from	8	2.77
Noun/ pronoun phrase + be	finding is also, participants were able, research has been, the data were, the participants were, the questionnaire was, there is a, there is no, This can be, This finding is	17	5.88
Other noun phrase	an important role, study aims to, The above table, the current study, the present study	5	1.73
Verb structure			
Passive + prepositional phrase fragment	are presented in**,** are shown in**,** based on the**,** be seen from**,** be seen in, compared to the, considered as a**,** focused on the**,** followed by the**,** found in the**,** given by the**,** is obtained for**,** is used as**,** presented in Table**,** related to the**,** seen in Table**,** shown in Table**,** used as a**,** used by a**,** was based on**,** was carried out**,** was divided into**,** were used in	33	11.42
Anticipatory it + verb or adjective phrase	it can be, it could be, It is also, it is necessary, it is possible, It must be, It should be, it was found, it would be	9	3.11
Copula be + noun phrase/adjective phrase	are likely to, be able to, be aware of, is consistent with, is not only, is one of, is possible that, is similar to, were consistent with	12	4.15
Verb phrase with noun/pronoun	data were analyzed, the participants had, the respondents had, The results show	5	1.73
Other verbal fragment	can be seen, can be used, contribute to the, focus on the, is in line, should be conducted, should be noted	10	3.46
Prepositional-phrase fragments			
Prepositional phrase with embedded of-phrase fragment	by means of, in terms of	2	0.69
Prepositional phrase expressions	as a means, by the participants, for further research, from this study, in the study, of the findings, of the participants, of the respondents, of this study,	22	7.61

Other prepositional phrase	according to the, As can be, As shown in, at the beginning, At this stage, because of the, due to the, In addition to, in agreement with, in line with, in order to, In other words, in the following, On the other, with regard to	20	6.92
Other structures			
Verb or adjective to-clause fragment	are expected to**,** be related to**,** considered to be**,** found to be**,** found to exist**,** is used to**,** need to be**,** seems to be**,** was employed to**,** was found to**,** was used to**,** were required to	19	6.57
verb phrase or noun phrase + that-clause fragment	agreed that the, be concluded that, be seen that, findings show that, indicate that the, is suggested that, point out that, result shows that, the fact that, This indicates that, This means that, This suggests that	25	8.65
Adverbial-clause fragment	as a result	1	0.35
Other expressions	above table showed, as well as, aware of the, below illustrates the, consistent with that, likely to be, might not be, similar to that, such as the, the other hand, This is because, to be able, to be aware, to complete the, to determine the, to find out	27	9.34
Total		**289**	**100**

As seen from Table 1, most of the multiword combinations are parts of noun-based structures (104 strings, or 35.99%) of the entire list. The noun phrase with *of*-phrase fragment is the most dominating sub-category, comprising more than half of all forms in the list. Most noun phrase with *of*-phrase fragment structure is a result of a combination of *a/the+noun+of*, where the noun slot can be filled by a wide range of words, such as characteristics, purpose, part, type, frequency, etc. The second position is occupied by noun/ pronoun phrases + be, followed by noun phrase with other post-modifier fragments, and other types of noun phrases, comprising 8 and 5 items, or about 2.77% and 1.73%, respectively, of all the items on the list. The sub-category of noun or pronoun phrase with verb *be* (17 items or 5.88% of the entire list), which is not reported in the Biber et al.'s (1999; 2004) structural classification, was also observed in the present study. Given the high number of the strings classified into the noun-based group, the finding is consistent with what would be expected. It is in agreement with the works by Byrd and Coxhead (2010) and Hyland (2008a; 2008b), indicating that most of formulaic sequences in the view of academic writing are "noun-centric" (Swales, 2008, p. V). This finding also coincides with the notion of Biber and Barbieri (2007) observing that academic prose tends to be constituted from word combinations that are more phrasal than clausal.

Verb-based phrases account for 68 items (23.53%). The majority of combinations in the verb-based category are composed of a verb in the passive with prepositional phrase construction, which typically indicates a locative or logical relation (e.g. *are presented in, based on the*). In terms of passive construction with prepositional

phrase, there is also a prevalence of passive structures with past tense verbs. They generally tend to be activity verbs which refer to specific research procedures such as *was based on, was carried out*. They also carry the sense of accomplishment of research activities. According to Hyland (2008a), this structural combination is used to convey the objective discussion of methodology used in a study being reported, and justification of claims. Some of the verb-based items can serve the function of referring to the findings in a study (*based on the*), comparing the findings with previous studies (*compared to these*), interpreting the findings (*highly related to*) and labelling data presented in tables (*shown in Table*). The multiword combinations in the other verbal fragment construction and the structure of verb phrases with noun or pronoun still highlight the importance of the active and passive construction usage in research writing (e.g. *can be seen, contribute to the, should be conducted*). In addition, anticipatory *it* with verb or adjective phrases, and copula *be* with noun or adjective phrases, could be classified. Although these two structural types were not frequent in the list, this finding coincides with Hyland's (2008b) study suggesting that these clauses are characteristic of academic writing rather than spoken data. Hyland (2008b) also claims that this structure introduced extraposed structures or anticipatory *it* and functions as a means of disguising authorial interpretations or to foreground the writer's evaluation of the findings without explicitly identifying its source (e.g. *it can be, it is possible, it should be*).

The third structural category, preposition-based structure, falls into three sub-types as follows: prepositional phrase with embedded *of*-phrase fragment, prepositional phrase expressions, and other prepositional phrase. The largest proportion of this category is prepositional phrase expressions (22 items or about 7.61%), followed by other prepositional phrases (20 useful items or 6.92% of the entire list). Most of the multiword combinations in the other prepositional phrase category are grammatically complete strings (e.g. *according to the, due to the, in addition to, in other words*). Some of them contain figurative meanings (*with regard to*). In the prepositional phrase expressions category, most of the items are characterized by specific meanings depending on the noun associated, for example, referring to the study (*from this study, in this study*). Only two items, *by means of* and *in terms of*, classified in the group prepositional phrase with embedded *of*-phrase, were found.

The other structure category can be divided into the following sub-categories: verb or adjective *to-clause* fragment; *verb* phrase or noun phrase with *that-clause* fragment; adverbial-clause fragment; anticipatory *it* with verb or adjective phrase; copula *be* with noun or adjective phrase; and other expressions. Other expression group is the most dominating sub-category containing 27 multiword combinations (9.34%) of the entire list. The second position is occupied by verb or adjective with *that*-clause fragment (25 items or about 8.65%), followed by verb phrase or noun phrase with *to*-clause fragment (19 strings or about 6.57%). Only one item can be found in the sub-category of adverbial-clause fragment. With regard to the structure of verb or adjective with *to*-clause fragment, most of the recurrent words are frequently used to interpret or discuss the findings as hedging statements (e.g. *seemed to be, considered to be, found to be*), whereas some of them are used to indicate research methodology (e.g. *were used*

to, were required to). Likewise, verb phrases followed by *that*-clause fragments are used to report the findings (e.g. *indicate that the, revealed that the, show that the*) or to preface inferences drawn from those of other previous studies (*agreed that the*). It is worth noting that these two sub-categories can be accomplished by the use of the present and past tense verbs. The multiword items in the other expression category do not fit any previously classified categories (Biber et al. 1999; 2004; Hyland, 2008a; 2008b), they can be formed by different grammatical words such as infinitives (*to determine the, to find out, to identify the*), which might express the objectives of the study or research procedures, adjectival fragments (*likely to be, to be aware*), when reporting the findings of the study. Grammatically complete units, relating to semantic and pragmatic functions (e.g. *as well as, such as the*) can be also found in this sub-category.

2. Pragmatic Functions of the Phraseological Patterns Identified

We further looked at whether certain n-grams appeared more frequently in a particular section of the articles, while others did not, being more evenly distributed over several sections. The distribution of the phraseological patterns suggests that, some of them can perform multiple functions according to the contextual environments in which they are used. However, out of 289 phraseological patterns identified, 236 lexical items seem to perform their pragmatic function associated with the distinctive communicative purposes of a particular section of the articles (See Appendix).

In the following, we will focus on a small selection of phrases for a more detailed analysis of their contextual environments, using concordance lines in order to further elucidate a potential clue the way these multiword units are used in the text in terms of salient pragmatic functions related to the primary purpose of the section identified.

Introduction section

The Introduction section provides background information of the topic and writers' evaluation of existing research. It helps readers get a contextual fix on the article. The section also establishes a niche in previous research by either indicating a gap or raising a question to justify the need for the current study being reported. The use of *the importance of,* and *the purpose of* which seemed likely to be associated with the function of this particular section. The text examples below illustrate the use of these strings in the Introduction section.

[1] *Great attention is paid to grammatical aspects, whereas* ***the importance of*** *collocations is overlooked; likewise, lexical choices in a second language structure are often arbitrarily and improperly produced.*

LEA 8

[2] *However, since English has become an international language, this leads to an increased awareness of* ***the importance of*** *cultural aspects represented in*

English language textbooks that should incorporate more multicultural aspects in order to enhance learners' awareness of cultural diversity in English language teaching and learning.

VEJ 2

The instances above indicate that *the importance of* is used as a lexical unit to highlight the importance of the topic being presented, alerting readers to see the need for conducting the present study. This string is usually used with the present verb tense. As Swales (1990) stated, present tense used in a professional genre has two major pragmatic functions: 1) to situate a particular event in the present tense, and 2) to mark a particular proposition as a generalization. In the latter case, the use of present tense indicates that the propositional information is valid regardless of time, reflecting the topic being discussed is still of interest in the field. The use and function of *the importance of* in this context is therefore strongly associated with the function of this section which provides a much more specific description, particularly with regards to the needs for conducting a research study being reported.

[3] *Given the importance now placed upon the development of lifelong learning skills, inclusive of self-regulatory efficacy for learning, as well as the ever increasing use of the Internet as a learning resource,* ***the purpose of*** *this small-scale research study was to assess the validity, in the context of an international English-medium Thai university… .*

SCOJ 1

The string *the purpose of* allows writers to prepare readers for the current study after introducing the topic of the study and pointing out the importance of, and/or commenting on some problematic issues needed to be addressed. Shortly afterwards, the cluster *the purpose of* is used to explicitly foreground the purpose(s) of the study, which is seemingly associated with the communicative purpose of introducing the study.

The pragmatic functions of these strings are quite similar to those of description bundles in the research-oriented function category proposed by Hyland (2008a; 2008b), and those called descriptive bundles in the referential bundles category in Biber et al.'s (2004) taxonomy. The finding is also consistent with Hyland's (2008a) assumption that the structure 'noun phrase + of' is prominent in research-oriented functions.

Methods section

The Methods section basically provides the description and detailed information regarding the research methodology, including participants, research instruments, data collection and analysis used in the study being reported. Multiword items salient in this section include, for instance, *in this study, this study was, the participants were/ the questionnaire was, to determine the,* and *was used to/were used to/were asked to*. The

examples below drawn from the Methods section of the corpus illustrate the use of these strings in this particular section.

[4] *To investigate the commodification of English through media discourses **in this study**, I will adopt a critical approach.*

NET 5

[5] *The instrument employed in **this study was** a LLS questionnaire. It focused on two sections: background information and SILL.*

LEA 1

[6] *In-depth interviews were firstly conducted as a major source of data collection because this method provides in-depth and specific data. At this stage, **the participants were** individually interviewed at a time and place convenient to them.*

LEA 3

[7] *Upon the final revision and improvement of the piloted questionnaire (see 3.3), **the questionnaire was** administered to the participants for a period of one week, in the middle of the second semester of the 2012 academic year.*

JES 2

The strings above were commonly used when writers give the information about the research samples and instruments used in the study. The use of demonstrative (*this*) allows the writers to frame and pinpoint specific information of the research methodology of the present study being reported, featuring how the methodology of the present study is different from previous research with regard to setting, participants or instruments used in the study to serve the specific objective of the study. The function of these strings is therefore seen to be related to the function of the section in offering a description of participants and research instruments used in the study being reported. According to Biber et al. (2004) and Hyland (2008a; 2008b), these strings were classified as a location marker. Although the case of singular and plural verbs (*was, were*) was found in the analysis, the verb cases and forms here are not a great deal of attention as they depend on the number of research tools used in the study referred to and mentioned in the text.

[8] ***To determine the** reliability of part II of the survey questionnaire, the researcher conducted a try-out of the survey. The tryout-revised items were conducted for improving the instruments' wording and language.*

SCOJ 3

[9] *The first questionnaire consisting of 20 statements was designed **to find out** what factors the students perceived as affecting their ability to select appropriate meanings of homonyms and homographs.*

JLA 2

As observed, *to determine the* can occur either at the beginning or in the middle of the sentence. In relation to their form, the strings *to determine the* and *to find out* are regarded as infinitives. As Biber (1995) points out, infinitives can be used with four specific functions: 1) to frame points in a discussion when *to*-clauses are used as predicates, 2) to introduce an aim, objective, plan, goal, purpose, strategy, task, or idea, 3) to introduce a methodology, and 4) to introduce a complement and as an adverbial purpose clause. These observations closely correspond to the instances above generated from concordance lines, demonstrating that *to determine the* and *to find out* are used to express specific and definite purpose of using research tools to accomplish the stage of analysis of the study and to explain the research methodology.

[10] *Chi-Square **was used to** analyze for fitted correspondence between the model and the empirical data.*

VEJ 1

[11] *The two tests were administered to the two groups of participants after the regular class. They **were asked to** complete the multiple-choice test and the translation test without any interaction with their classmates.*

LEA 8

The strings *was used to, were used* to and *were asked to*, have the functions of expressing purposes of the use of statistics for data analysis, and of the methodology used in the study. The instances above also evidence the importance of the past tense verbs with passive construction in the Methods section. The form of '*be+verb+to* infinitive' is the typical and effective feature used when the emphasis is put on the action, the description of research instruments used and procedure performed. The function of the co-occurrence of passives and past tense verbs (e.g. *was/were used to* and *were asked to*) is thus regarded to be associated with dealing with the research instrument and data analyses.

Hyland (2008a) states that bundles in the procedure sub-category clearly puts the writers "under some pressure to showcase their ability to handle research methods appropriately and to demonstrate the familiarity with the subject content of the discipline" (p. 55). As presented above, we would argue that a certain set of strings (e.g. *the participants were, the questionnaire was, to determine the, to find out, was used to, were asked to*) strongly associated with the function of the Methods section is parallel with what Hyland (2008a; 2008b) would call procedure bundles in the research-oriented function category because they have the main function of showing the ways that research is conducted, and to explain why something is done.

Results section

Major findings are reported in the Results section. Sometimes, most experienced writers not only report, but also comment on the results and compare them with those of previous studies. The following text examples illustrate the use of some recurrent multiword units in their context in the Results section.

[12] *The **above table showed** the gender of sample ABAC students by their nationality. The majority of them were Thai Male 73.3% (178) and 4.9% (12) was from other country.*

SCOJ 4

[13] *Table 2.1 **below illustrates the** results of the oral proficiency score (out of 25) of low-proficiency students performed in three information gap tasks.*

NET 2

The multiword units *above table showed* and *below illustrates the* are used to report findings. These strings, based upon Hyland's (2008a; 2008b) functional classification, are called structuring signals in the text-oriented bundle category, whilst text deixis bundles in Biber et al. (2004). The main function of these strings is to help organize stretches of discourse, pointing to data presented in tables and figure in the text. These strings usually tend to co-occur with present and past tense verbs, with pointers (*Table 2.1*). Pointers or visual representations of data can be an efficient way of presenting information clearly and concisely, often allowing the readers to get an at-a-glance picture or sense of data before going through the illustration followed. The writers, therefore, used pointers together with prepositions such as *above* and *below* to facilitate readers to find the location where the information is being presented in the text.

[14] *The triangulation of the data **revealed that the** students' perceptions of coherence and the teaching were, to a large extent, related. Interestingly, the term `coherence' was mentioned neither in the class nor in the documents, but the activities and the exercises addressed four main components of coherence.*

LEA 4

[15] *The scores gained for each group **showed that the** Thai version favored the low-score group the most, followed by the mid-score group, with the high-score group not benefitting at all.*

JES 1

[16] *The **results of the** analysis show significant differences in five out of nine speaking factors and one out of six listening factors (see Table 6).*

LEA 7

The instances demonstrate that reporting verbs (*reveal, show*) were commonly used when the findings are reported in the Results section. These strings can denote an integration of reporting verbs and past tense when research findings of the study are presented, and no claim is made in these contexts of use. Definite articles (*the*) in the text indicate that the shared knowledge between the writers and the readers is established based on the preceding context (Biber et al., 1999). These strings *c*ould be called resultative signals in the text-oriented function of Hyland's (2008a; 2008b) functional classification. These strings mark inferential or causative reflation between elements, which is a key function in the rhetorical presentation of research findings (Hyland, 2008a). As expected, these strings show writers' interpretations and understandings or research processes and findings, which coincides with the primary function of the Results section.

[17] *Such elements of 'small talk' in conversation seem to be supported by* ***the findings from*** *the interviews with the cabin crews. 85% of the interviewees still commented that correct grammar and structures were important in conversation with the passengers.*

NET 4

The findings from is another string generally used to present the results of the study being reported. This string allows writers to frame claims based on the findings of the study when used in the context of '*seems to be supported by the findings from*'. This co-occurring feature illustrates the main function of the Results section, which the writer can make comments on the basis of data. This typical practice can be found in this particular section before the writers can go beyond the results to situate the findings again in the Discussion section.

[18] *This was explained later in the interviews in which he acknowledged that the strategy instruction helped him know what strategies to use and how to use them;* ***as a result****, it improved his reading comprehension.*

SPPJ 1

[19] *The causes of these limitations appear to be due to linguistic and processing differences, unsupportive L2 learning environments (covering individual and experiential differences, and socio-cultural institutional differences), and poor reading skills* ***as a result*** *of not reading enough (Ref).*

JES 1

A fixed-form or grammatical phrase *as a result* functions a conjunctive adverbial phrase to indicate cause-and-effect relationship. As in [18], this string is used as a logical connection to the preceding sentence. However, [19] reveals that this string is originally from *as a result of* which grammatically must be followed by a noun or noun phrase. The writers actually use these two forms of string simply to indicate that the subsequent text describes a finding of the study, and thus reflecting that the writers can

objectively present the findings, and then interpret them based on the further details from the results obtained as in [18], or by using references or citations *(Ref)* as in [19]. The use of this string may be seen as a convincing approach to support the findings presented in the Results section.

[20] *The findings related to the specific motivation of participants shows that the extrinsic motivation-identified regulation* ***seemed to be*** *the most predominant motivation of all participants in learning to speak English in CLT classroom.*

LEA 3

[21] *The outcome of strategies report could lead to the assumption that the use of e mail writing strategies was somewhat ineffective. The ineffectiveness* ***seems to be*** *affected from factors involving linguistic and background knowledge.*

NET 6

With the likelihood verb (*seem*), either in present tense or past tense, it indicates the writers' expression of probability and interpretation of the findings of the study. Looking more closely at [20], the string *seemed to be* was tied with attitudinal adjectives (*the most predominant*) expressing evaluation, the meaning of this string can show the writers' expression of degree of certainty or commitment when the research results or activities are reported. Consequently, this string can be employed to convey a level of their agreement, opposition, evaluation, and interpretation of particular entities when interpreting the results. According to Hyland (2008a), verbs expressing probability and modality are called 'downtoners' which help qualify the status of knowledge as speculative and help mitigate tone to the statement. Through the use of downtoners, the writer's counter-argument is expressed and modulated.

[22] *Noticeably, although the participants' perceptions of achievement were highly* ***related to the*** *social context, the participants did not reveal any positive attitudes towards the English community or the local people.*

LEA 3

[23] *However, it didn't score well in the reading rubric was that most of them didn't connect their personal experience in their reading responses. In their conversational interviews, they revealed that it was* ***due to the*** *time constraint given in class.*

SPPJ 1

The writers use the strings *related to the,* and *due to the* when they not only wish to present the results of the study, but also comment upon them. The string *related to the* is usually used when the writers interpret or evaluate the results by mentioning related results found in the study, while syntactically, 'due to' is a prepositional phrase

used to give a reason for why something is true. Concerning Hyland's (2008a; 2008b) taxonomy, we would argue that the functions of the strings *as a result, seems to be, related to the,* and *due to the* could be considered stance features in the participant-oriented function category because they refer to the writer-focused bundles that the writers convey epistemic and affective judgment, evaluation and degrees of commitment to their claims. Therefore, these strings again underline the pragmatic function of the Results section, which not only presents the findings or discoveries, but sometime with a particular evaluative connotation to the findings.

Discussion section

In the Discussion section, writers can express their own evaluation of the study, acknowledging that their own study is not perfect. Future studies are also suggested as they might be worth further investigation. The following text examples illustrate the use of recurrent word combinations (*of the present/of the study/ of this study, the findings of, and for further research*) in contexts, performing the pragmatic functions associated with the purposes and functions of the section.

[24] *Second, and more importantly, the results* ***of the present*** *study elucidated understanding of coherence and, at the same time, causes of incoherence from the students' points of view.*

LEA 4

[25] *Regarding assertiveness, the results* ***of the study*** *confirm (Ref) and (Ref) explorations on the expediency of assertiveness in successful SLA.*

LEA 5

[26] *The findings* ***of this study*** *support the evidence from previous research that showed students wanted to do more service learning activity.*

SCOJ 3

The strings above are frequently used, especially when the results of the study are being discussed. Considering from the above instances, given the grammatical function of articles (*the*) and demonstrative adjectives (*this*), the writers used these multiword combinations as prepositional phrases to emphasize that they are discussing the results generated by the current study to draw readers' attention to the issue discussed. In written academic discourse, demonstrative adjectives mark referential cohesion directly referring to the immediate preceding text or the immediate textual context as a referent (Biber et al., 1999). This device allows shared knowledge between readers and the writers to be established as referring to the present study being reported. The function of these strings is in line with what Hyland (2008a; 2008b) would call the

research oriented function of location, while referential bundles based on Biber et al.'s (2004) taxonomy.

This group of phraseological patterns often tend to co-occur with the past tense in this particular section because the past tense is principally used to report completed actions at the particular time frame, and to express newly achieved evidentiality or knowledge in the field, reflecting limited degree of certainty and reliability of the new information. Since the major results are already reported in the Results section, the past tense used in this context in the Discussion section is appropriate.

[27] *To wrap up the discussion, on the whole,* ***the findings of*** *the current study demonstrated L2 upper-intermediate learners with higher levels of L2 metaphorical competence possessed more assertiveness, intrapersonal, self-actualization, and independence and contrariwise.*

LEA 5

[28] *Although* ***the findings of*** *this study have shed some light on the role of the training, some limitations yield suggestions for future research.*

JLA 2

The string *the findings of* is also prevalently used in the Discussion section. This cluster can co-occur with common nouns (*study*), and thus contributes to the longer string *the findings of the current study* or *the findings of this study*, which were usually used at the initial position of the clause. The sting can occur when writers summarizes the study [27], and indicating some limitations of the study [28] which is quite similar to the pragmatic function of the string *in this study* as previously discussed.

[29] *The recommendations* ***for further research*** *are presented as follows: 1. As this study was undertaken at two universities in Banteay Meanchey Province, Cambodia, generalizations can be made if more universities from other settings are included in future studies.*

LEA 1

The above represents some suggestions for further research which commonly occurs at the very end of the Discussion section. The string may be used after some of the limitations of the study are expressed in the text, indicating that the statements of findings remain to be substantiated or validated by future studies. According to Hyland (2008a; 2008b), this string is performing as a text framing signal in the text-oriented function category.

Implication and Conclusions

This study has looked at applied linguistics research articles through the lens provided by repeated frequent three-word sequences. The overall goal of this research was to create a pedagogically useful list of the phraseological patterns. Using the corpus-based approach, qualitative approach, and with assistance from five instructors, a list of 289 three-word multiword combinations for teaching research article writing in English could be generated. We also assessed the extent to which phraseology contributes to article writing by investigating the structural types and pragmatic functions of the lexical clusters identified. Based on Biber et al.'s structural taxonomy (Biber et al., 1999; 2004), the findings reveal that three-word multiword combinations frequently used by the article writers identified could be structurally classified into four structures. Moreover, four new categories, including other noun phrase, noun or pronoun phrase with *be*, other prepositional phrase expressions, and other verbal fragments were added to the original taxonomy. The corpus-based approach also showed that the greatest range of content words can be seen in the corpus, but excluded from the final list concerning their pedagogic purposes. Based on the context dependency, some strings appeared to have multi-functionality because a section of an article contained several move types performing different communicative purposes. However, a set of multiword combinations could be identified as their pragmatic functions were closely associated with communicative functions of certain sections of the articles. This bottom-up perspective is seen to be an insightful account of identifying and matching the functional types in context and the occurrence in the text.

Regarding the pragmatic functions of the phraseological patterns, based on Hyland's (2008a; 2008b) and Biber et al.'s (1999; 2004) classification, the multiword units found were grouped into functional categories. The findings also confirmed results of previous studies such as Hyland (2008a; 2008b), Wei (2007), and Durrant (2015), in that passive structures were used and anticipatory it structures and participant-oriented bundles were rarely employed. This may be due to the writers' preference for the impersonality in their article writing (Grabowski, 2015). However, it should be noted that some phraseological patterns appeared in more than one section of the article.

Taking the pedagogical perspective, Csomay (2013) suggests that students would seldom think in terms of grammatical patterns and multiword items can indicate changes in text types within discourse. Hence, students and novice writers should be aware of types of lexical items and how they relate to the structuring of information and/or function of discourse. The findings and the list generated from the present study can be seen particularly as the skeleton of proficient academic writing. In this respect, when designing a course for academic writing for publication, instructors could make full use of the list and integrate a description of this study to their instruction. As suggested by Hyland (2008b), writers are expected to both abide by the linguistic rules of language and comply with the intended readers' expectations by implementing the potential lexical clusters of the discourse in question. Students and novice article writers, thus, should be aware of knowledge of the use and pragmatic functions of multiword combinations frequently used in a given section when preparing their

research manuscript. Instructors could apply the knowledge and the list of multiword items of this study by implementing some activities that are replete with different types of lexical clusters with an emphasis on fostering students' expressive skills and how to use the clusters to serve their communicative purposes. Practically, instructors may draw students' attention to the words covered in the list and encourage them to use these words in their writing assignments. It is also believed that introducing different lexical clusters that serve various pragmatic functions and raising students' awareness of the importance of this language phenomenon in academic contexts can be advantageous for them to meet an acceptable level in academic communities (Coxhead & Byrd, 2007; Martinez & Schmitt, 2012).

Caution is needed, however, in the application of the findings and the list to pedagogy as the corpus of this study was taken from a single discipline: applied linguistics. The results and the list obtained from this study should be considered only illustrative and have some limitations because they applied only to the applied linguistics research articles published in English indexed in the TCI database, rather than to the English in different disciplines. Other multiword combinations may be inconclusive and have not been included in the pedagogically useful list in this study. Furthermore, methodology and scope of this study should be also taken into account. First, the number of articles analyzed in this study is considered relatively small and specific. In order to generalize the findings of this study, we acknowledge that a bigger size of the corpus could yield and represent a better global picture of multiword combinations used in the article in this field. Second, it is possible to consider taking other eminent statistic criteria such as MI-score and formula teaching worth (Simpson-Vlach and Ellis (2010), rather than careful selection from EAP instructors to support an identification task of multiword units useful for pedagogic purposes, and to obtain a more refined of pedagogically-useful list of three-word n-grams (Salazar, 2011). Next, this study sought to find out high-frequency words, leaving those low-frequency items. It is important to encourage and teach students to consult other reliable resources when they encounter such multiword items and experience difficulty in their reading and writing. Despite the possibilities for further research, to some extent, the descriptive results presented here is still of crucial importance for EAP instructors in developing instructional materials in teaching writing for scholarly publication. They may facilitate novice writers and graduate students the task of preparing manuscripts for publication.

References

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A.P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 101-122). Oxford: Oxford University Press.

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

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

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

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Bowker, L., & J. Pearson (2002). *Working with specialized language: A practical guide to using corpora*. London: Routledge.

Bybee, J. (2002). Phonological evidence of exemplar storage of multiword sequences. *Studies in Second Language Acquisition, 24*, 215-221.

Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes, 25*, 310-331.

Chen, Yu-Hua., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30-49.

Conrad, S., & Biber, D. (2004). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*, 20, 56-71.

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

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

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.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-238.

Csomay, E. (2013). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423.

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). London: Longman.

Durrant, P. (2015). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, 1-30.

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes, 23*(3), 157-169.

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

Feak, C., & Swales, J. (2011). *Academic writing for graduate students: Essential tasks and skills*. Ann Arbor: University of Michigan.

Grabowski, L. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes, 38*, 23-33.

Granger, S., & Meunier, F. (2008). *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins.

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

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

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

Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences. In N. Schmitt, (Ed.), *Formulaic sequences* (pp. 269-300). Philadelphia: John Benjamins.

Kilgarriff, A., Pavel, R., Pavel, S., & David, T. (2004). The Sketch Engine. Proceedings of EURALEX, Lorient, France, pp. 105-116.

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing, 18*(2), 85-102.

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31, 25-35.

Martinez, R., & Schmitt, N. (2012). A Phrasal Expressions List. *Applied Linguistics, 33*(3), 299-320.

Martinez, I.A., Beck, S.C., & Panza, C.B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes, 28*, 183-198.

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics, 11*(3), 147-168.

Pawley, A., & Syder, F.H. (1983). Two puzzles for linguistic theory native like selection and nativelike fluency. In J.C. Richards & R.W. Schmidt (Eds.), *Language and communication* (pp. 191-230). London: Longman.

Reppen, R. (2004). Academic language: An exploration of university classroom and textbook language. In U. Conner & T. Upton (Eds.), *Discourse in the Professions: Perspectives from corpus linguistics* (pp. 65-86). Amsterdam: Benjamins.

Salazar, D. (2011). *Lexical bundles in scientific English: A corpus-based study of native and non-native writing*. Unpublished dissertation, University of Barcelona.

Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins.

Scott, M. (2004). *WordSmith Tools Version 4.0*. Oxford: Oxford University Press.

Simpson-Vlach, R., & Ellis, N.C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512.

Stubbs, M. (2007). An example of frequent English phraseology: Distribution, structures and functions. In R. Facchinetti (Ed.), *Corpus linguistics 25 years on* (pp. 89-105). Amsterdam: Radopi.

Swales, J.M. (1990). *Genre analysis: English in academic and research setting*. Cambridge: Cambridge University Press.

Swales, J.M. (2004). *Research genre: Explorations and applications*. Cambridge: Cambridge University Press.

Swales, J.M. (2008). Foreword. In D. Belcher & A. Hirvela (Eds.), *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions* (p. v-viii). Ann Arbor: University of Michigan Press.

Wang, J., Liang, S.L., & Ge, G.C. (2008). Establishment of a medical academic wordlist. *English for Specific Purposes, 27*, 442-458.

Wei, N.X. (2007). Phraseological characteristics of Chinese learners spoken English: Evidence of lexical chunks from COLSEC. *Modern Foreign Languages, 30*, 280-291.

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

Wray, A. (2009). Identifying formulaic language: Persistent challenges and new opportunities. In R. Corrigan, E.A. Moravcsik, H. Ouali, and K.M. Wheatley (Eds.): *Formulaic language volume 1: Distribution and historical change* (pp. 27-51). John Benjamins Publishing Company.

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

Yang, M.N. (2015). A nursing academic word list. *English for Specific Purposes, 37*(1), 27-38.

Appendix A: Corpus of research articles included in the study

JES Journal of English Studies

JLA Journal of Liberal Arts, Prince of Songkla University

SPPJ Sripatum Review of Humanities and Social Sciences

NET The New English Teacher

LEA Language Education and Acquisition Research Network (LEARN) Journal

VEJ Veridian E-Journal

SJSH Silpakorn University Journal of Social Sciences, Humanities, and Arts

SCOJ Scholar

MNY MANUSYA: Journal of Humanities

Appendix B: Pedagogically useful list of multiword combinations in applied linguistics research articles

No.	Multiword combinations	Freq. of occurrence in a particular section	Sections associated with	Structural types	Pragmatic function
1	a corpus of	5	Methods	noun phase with of-phrase fragment	Quantification
2	a lot of	9	Results	noun phase with of-phrase fragment	Quantification
3	a number of	5	Results	noun phase with of-phrase fragment	Quantification
	a number of	8	Discussion	noun phase with of-phrase fragment	Quantification
4	a part of	7	Discussion	noun phase with of-phrase fragment	Quantification
5	a result of	5	Results	noun phase with of-phrase fragment	Resultative signal
6	a sense of	5	Results	noun phase with of-phrase fragment	Intangible framing
7	A total of	6	Methods	noun phase with of-phrase fragment	Quantification
8	a variety of	5	Introduction	noun phase with of-phrase fragment	Quantification
	a variety of	5	Results	noun phase with of-phrase fragment	Quantification
9	above table showed	5	Results	other expressions	Structuring
10	according to the*	9	Methods	Other prepositional phrase	Structuring
	according to the	21	Results	Other prepositional phrase	Structuring
	According to the	11	Results	Other prepositional phrase	Structuring
	According to the	7	Discussion	Other prepositional phrase	Structuring
11	agreed that the*	5	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
	agreed that they	5	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
12	all of the	8	Results	noun phase with of-phrase fragment	Quantification
13	an analysis of	7	Results	noun phase with of-phrase fragment	Procedure
14	an important role	5	Introduction	Other noun phrase	Intangible framing
15	an investigation of	18	Methods	noun phase with of-phrase fragment	Procedure
16	analysis of the	17	Results	noun phase with of-phrase fragment	Procedure
17	are expected to	5	Introduction	verb or adjective to-clause fragment	Procedure
18	are likely to	6	Discussion	verb or adjective to-clause fragment	Stance feature
19	are more likely	5	Results	Copula be + noun/ adjective phrases	Stance feature
20	are presented in	5	Results	Passive+prepositional phrase fragment	Structuring

21	are shown in	7	Results	Passive+prepositional phrase fragment	Structuring
22	as a means	5	Introduction	Prepositional phrase expressions	Procedure
23	as a part	5	Discussion	other expressions	Quantification
24	as a result*	5	Results	Adverbial-clause fragment	Transition signal
	As a result	6	Discussion	Adverbial-clause fragment	Transition signal
	as a result	5	Discussion	Adverbial-clause fragment	Transition signal
25	As can be	15	Results	Other prepositional phrase	Structuring
26	As for the	7	Results	Other prepositional phrase	Framing signal
27	As shown in*	15	Results	Other prepositional phrase	Structuring
	as shown in	8	Results	Other prepositional phrase	Structuring
28	as well as	16	Introduction	other expressions	Transition signal
	as well as	20	Methods	other expressions	Transition signal
	as well as	15	Results	other expressions	Transition signal
	as well as	22	Discussion	other expressions	Transition signal
29	at the beginning	6	Methods	Other prepositional phrase	Location
	at the beginning	5	Results	Other prepositional phrase	Location
30	At this stage	6	Methods	Other prepositional phrase	Location
31	aware of the	12	Results	other expressions	Intangible framing
	aware of the	18	Discussion	other expressions	Intangible framing
32	based on an*	7	Results	Passive+prepositional phrase fragment	Structuring
	based on the	6	Introduction	Passive+prepositional phrase fragment	Structuring
	based on the	26	Methods	Passive+prepositional phrase fragment	Structuring
	based on the	21	Results	Passive+prepositional phrase fragment	Structuring
	Based on the	14	Results	Passive+prepositional phrase fragment	Structuring
	based on the	6	Discussion	Passive+prepositional phrase fragment	Structuring
	Based on the	5	Discussion	Passive+prepositional phrase fragment	Structuring
33	be able to	8	Introduction	verb or adjective to-clause fragment	Resultative signal
	be able to	9	Results	verb or adjective to-clause fragment	Resultative signal
	be able to	20	Discussion	verb or adjective to-clause fragment	Resultative signal
34	be aware of	11	Discussion	Copula be + noun/ adjective phrases	Stance feature
35	be concluded that	6	Results	verb phrase or noun phrase+that-clause fragment	Stance feature
36	be noted that	12	Results	verb phrase or noun phrase+that-clause fragment	Engagement feature
37	be related to	6	Results	verb or adjective to-clause fragment	Stance feature
38	be said that	5	Discussion	verb phrase or noun phrase+that-clause fragment	Stance feature

39	be seen from*	5	Results	Passive+prepositional phrase fragment	Structuring
	be seen in	12	Results	Passive+prepositional phrase fragment	Structuring
40	be seen that	6	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
41	because of the*	6	Results	Other prepositional phrase	Stance feature
	because of their	5	Results	Other prepositional phrase	Stance feature
42	below illustrates the	8	Results	other expressions	Structuring
43	by means of	6	Methods	Prepostitional phrase with embedded of-phrase fragment	Procedure
44	by the participants	5	Results	Prepositional phrase expressions	Resultative signal
45	can also be	5	Results	Other verbal fragment	Stance feature
46	can be seen	25	Results	Other verbal fragment	Stance feature
	can be seen	5	Discussion	Other verbal fragment	Stance feature
47	can be used	6	Discussion	Other verbal fragment	Procedure
48	compared to the*	9	Results	Passive+prepositional phrase fragment	Stance feature
	compared to those	7	Results	Passive+prepositional phrase fragment	Stance feature
49	considered as a	5	Introduction	Passive+prepositional phrase fragment	Stance feature
50	considered to be	5	Results	verb or adjective to-clause fragment	Stance feature
51	consistent with that*	6	Discussion	other expressions	Resultative signal
	consistent with the	5	Discussion	other expressions	Resultative signal
52	contribute to the	6	Introduction	Other verbal fragment	Stance feature
53	criteria based on	6	Results	noun phrase with other post-modifier fragment	Description
54	data from the	5	Results	noun phrase with other post-modifier fragment	Resultative signal
55	data were analyzed	6	Methods	verb phrase with noun/pronoun	Procedure
56	Data were collected	5	Methods	verb phrase with noun/pronoun	Procedure
57	developed by the	6	Methods	Passive+prepositional phrase fragment	Other
58	development of the	5	Discussion	noun phase with of-phrase fragment	Intangible framing
59	differed from the	5	Results	Passive+prepositional phrase fragment	Resultative signal
60	due to the*	12	Results	Other prepositional phrase	Stance feature
	due to their	5	Discussion	Other prepositional phrase	Stance feature
61	each of the	11	Results	noun phase with of-phrase fragment	Quantification
62	employed by the	7	Discussion	Passive+prepositional phrase fragment	Resultative signal
63	exist at the	7	Results	Other verbal fragment	Other
64	fact that the	6	Results	verb phrase or noun phrase+that-clause fragment	Other

65	finding is also	5	Discussion	Noun/pronoun phrase+be	Resultative signal
66	findings of the*	8	Discussion	noun phase with of-phrase fragment	Resultative signal
	findings of this	14	Discussion	noun phase with of-phrase fragment	Resultative signal
67	findings show that	5	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
68	focus on the*	5	Discussion	Other verbal fragment	Structuring
	focused on the	5	Introduction	Passive+prepositional phrase fragment	Structuring
69	followed by the*	12	Results	Passive+prepositional phrase fragment	Resultative signal
	followed by them	6	Results	Passive+prepositional phrase fragment	Resultative signal
70	for further research	7	Discussion	Prepositional phrase expressions	Framing signal
71	found in the	14	Results	Passive+prepositional phrase fragment	Structuring
	found in the	6	Discussion	Passive+prepositional phrase fragment	Structuring
72	found that the*	6	Introduction	verb phrase or noun phrase+that-clause fragment	Resultative signal
	found that the	7	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
	found that the	6	Discussion	verb phrase or noun phrase+that-clause fragment	Resultative signal
	found that these	5	Discussion	verb phrase or noun phrase+that-clause fragment	Resultative signal
73	found to be	7	Results	verb or adjective to-clause fragment	Resultative signal
	found to be	5	Discussion	verb or adjective to-clause fragment	Resultative signal
74	found to exist	9	Results	verb or adjective to-clause fragment	Resultative signal
75	frequencies of the*	7	Results	noun phase with of-phrase fragment	Quantification
	frequency of the	9	Results	noun phase with of-phrase fragment	Quantification
76	from the context	7	Discussion	Prepositional phrase expressions	Framing signal
77	from this study	7	Discussion	Prepositional phrase expressions	Location
78	given by the	8	Results	Passive+prepositional phrase fragment	Resultative signal
79	good level of	7	Results	noun phase with of-phrase fragment	Intangible framing
80	have shown that	5	Discussion	verb phrase or noun phrase+that-clause fragment	Resultative signal
81	highly related to	5	Discussion	Passive+prepositional phrase fragment	Stance feature
82	identified according to	7	Results	Passive+prepositional phrase fragment	Other
83	illustrates the results	7	Results	Other verbal fragment	Structuring
84	in a text	6	Methods	Prepositional phrase expressions	Location
85	In addition to	5	Results	Other prepositional phrase	Transition signal

86	in agreement with	7	Discussion	Other prepositional phrase	Transition signal
87	in line with	6	Results	Other prepositional phrase	Transition signal
	in line with	6	Discussion	Other prepositional phrase	Transition signal
88	in order to*	27	Introduction	Other prepositional phrase	Procedure
	in order to	21	Methods	Other prepositional phrase	Procedure
	in order to	18	Results	Other prepositional phrase	Procedure
	In order to	7	Results	Other prepositional phrase	Procedure
	in order to	21	Discussion	Other prepositional phrase	Procedure
89	In other words	9	Results	Other prepositional phrase	Transition signal
	In other words	15	Discussion	Other prepositional phrase	Transition signal
90	in relation to	7	Results	Other prepositional phrase	Framing signal
	in relation to	5	Discussion	Other prepositional phrase	Framing signal
91	in terms of*	18	Introduction	Prepostitional phrase with embedded of-phrase fragment	Framing signal
	In terms of	6	Introduction	Prepostitional phrase with embedded of-phrase fragment	Framing signal
	in terms of	16	Methods	Prepostitional phrase with embedded of-phrase fragment	Framing signal
	in terms of	38	Results	Prepostitional phrase with embedded of-phrase fragment	Framing signal
	In terms of	17	Results	Prepostitional phrase with embedded of-phrase fragment	Framing signal
	in terms of	24	Discussion	Prepostitional phrase with embedded of-phrase fragment	Framing signal
	In terms of	10	Discussion	Prepostitional phrase with embedded of-phrase fragment	Framing signal
92	in the following	12	Results	Other prepositional phrase	Structuring
93	in the future	6	Discussion	Prepositional phrase expressions	Framing signal
94	in the process	5	Results	Prepositional phrase expressions	Framing signal
95	in the study	5	Methods	Prepositional phrase expressions	Location
	in the study	7	Results	Prepositional phrase expressions	Location
96	in the target	6	Introduction	Prepositional phrase expressions	Location
97	in the text	6	Introduction	Prepositional phrase expressions	Location
	in the text	7	Results	Prepositional phrase expressions	Location
98	in the use	6	Results	Prepositional phrase expressions	Framing signal
99	in this group	5	Results	Prepositional phrase expressions	Location

100	in this study*	39	Methods	Prepositional phrase expressions	Location
	In this study	10	Methods	Prepositional phrase expressions	Location
	in this study	19	Results	Prepositional phrase expressions	Location
	In this study	7	Results	Prepositional phrase expressions	Location
	in this study	35	Discussion	Prepositional phrase expressions	Location
	In this study	14	Discussion	Prepositional phrase expressions	Location
101	indicate that the*	6	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
	indicated that the	10	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
	indicates that the	10	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
102	interview questions were	5	Methods	Noun/pronoun phrase+be	Procedure
103	investigation of each*	12	Methods	noun phase with of-phrase fragment	Procedure
	investigation of the	6	Methods	noun phase with of-phrase fragment	Procedure
104	is consistent with	9	Discussion	Copula be + noun/ adjective phrases	Resultative signal
105	is important to	6	Discussion	verb or adjective to-clause fragment	Stance feature
106	is in line	5	Discussion	Other verbal fragment	Other
107	is not only	6	Introduction	Copula be + noun/ adjective phrases	Other
108	is obtained for	5	Results	Passive+prepositional phrase fragment	Procedure
109	is one of	5	Results	Copula be + noun/ adjective phrases	Quantification
110	is possible that	6	Results	Copula be + noun/ adjective phrases	Stance feature
111	is similar to	5	Results	Copula be + noun/ adjective phrases	Resultative signal
112	is suggested that	6	Discussion	verb phrase or noun phrase+that-clause fragment	Stance feature
113	is used as	7	Introduction	Passive+prepositional phrase fragment	Procedure
114	is used to	5	Results	verb or adjective to-clause fragment	Procedure
115	it can be*	13	Results	Anticipatory it + verb/ adjective phrase	Stance feature
	It can be	10	Discussion	Anticipatory it + verb/ adjective phrase	Stance feature
	it could be	5	Discussion	Anticipatory it + verb/ adjective phrase	Stance feature
116	It is also	5	Discussion	Anticipatory it + verb/ adjective phrase	Stance feature
117	it is necessary	5	Discussion	Anticipatory it + verb/ adjective phrase	Stance feature
118	it is possible	5	Results	Anticipatory it + verb/ adjective phrase	Stance feature

119	It must be	5	Results	Anticipatory it + verb/ adjective phrase	Stance feature
120	It should be	7	Results	Anticipatory it + verb/ adjective phrase	Stance feature
121	it was found	7	Results	Anticipatory it + verb/ adjective phrase	Resultative signal
	it was found	7	Discussion	Anticipatory it + verb/ adjective phrase	Resultative signal
122	it would be	5	Results	Anticipatory it + verb/ adjective phrase	Stance feature
123	knowledge of the	5	Introduction	noun phase with of-phrase fragment	Intangible framing
	knowledge of the	13	Discussion	noun phase with of-phrase fragment	Intangible framing
124	large number of	5	Introduction	noun phase with of-phrase fragment	Quantification
125	likely to be	6	Discussion	other expressions	Stance feature
126	majority of the*	5	Results	noun phase with of-phrase fragment	Quantification
	majority of them	6	Results	noun phase with of-phrase fragment	Quantification
127	mean value of	5	Results	noun phase with of-phrase fragment	Intangible framing
128	meaning of the	34	Results	noun phase with of-phrase fragment	Description
	meaning of the	8	Discussion	noun phase with of-phrase fragment	Description
129	might not be	6	Discussion	other expressions	Stance feature
130	more likely to	6	Results	other expressions	Stance feature
131	most of the	5	Introduction	noun phase with of-phrase fragment	Quantification
	most of the	19	Results	noun phase with of-phrase fragment	Quantification
	most of the	7	Discussion	noun phase with of-phrase fragment	Quantification
132	need to be	5	Discussion	verb or adjective to-clause fragment	Stance feature
133	not be able	6	Discussion	Copula be + noun/ adjective phrases	Resultative signal
134	number of the	7	Results	noun phase with of-phrase fragment	Quantification
135	of data collection	11	Methods	Prepositional phrase expressions	Procedure
136	of each interviewee	12	Methods	Prepositional phrase expressions	Procedure
137	of the findings	5	Discussion	Prepositional phrase expressions	Resultative signal
138	of the participants	13	Methods	Prepositional phrase expressions	Resultative signal
	of the participants	12	Results	Prepositional phrase expressions	Resultative signal
	of the participants	20	Discussion	Prepositional phrase expressions	Resultative signal
139	of the questionnaire	11	Methods	Prepositional phrase expressions	Resultative signal
140	of the respondents	16	Results	Prepositional phrase expressions	Resultative signal

141	of the student	11	Discussion	Prepositional phrase expressions	Resultative signal
	of the students	11	Methods	Prepositional phrase expressions	Resultative signal
	of the students	28	Results	Prepositional phrase expressions	Resultative signal
142	of the Study*	8	Introduction	Prepositional phrase expressions	Resultative signal
	of the study	6	Introduction	Prepositional phrase expressions	Resultative signal
	of the study	17	Methods	Prepositional phrase expressions	Resultative signal
	of the study	21	Discussion	Prepositional phrase expressions	Resultative signal
	of this study	14	Introduction	Prepositional phrase expressions	Resultative signal
	of this study	10	Methods	Prepositional phrase expressions	Resultative signal
	of this study	7	Results	Prepositional phrase expressions	Resultative signal
	of this study	37	Discussion	Prepositional phrase expressions	Resultative signal
143	On the other	6	Introduction	Other prepositional phrase	Transition signal
	On the other	15	Discussion	Other prepositional phrase	Transition signal
	On the other	6	Results	Other prepositional phrase	Transition signal
144	On the whole	5	Results	Other prepositional phrase	Transition signal
145	one of the*	22	Introduction	noun phase with of-phrase fragment	Quantification
	One of the	7	Introduction	noun phase with of-phrase fragment	Quantification
	one of the	7	Results	noun phase with of-phrase fragment	Quantification
	one of the	10	Discussion	noun phase with of-phrase fragment	Quantification
146	out of the	7	Results	Other prepositional phrase	Quantification
147	part of the	10	Methods	noun phase with of-phrase fragment	Framing signal
	part of the	14	Results	noun phase with of-phrase fragment	Framing signal
	part of the	7	Discussion	noun phase with of-phrase fragment	Framing signal
148	participants in the*	8	Methods	noun phrase with other post-modifier fragment	Description
	participants in the	11	Discussion	noun phrase with other post-modifier fragment	Description
	participants in this	5	Methods	noun phrase with other post-modifier fragment	Description
149	participants were able	6	Discussion	Noun/pronoun phrase+be	Resultative signal
150	pattern of the	6	Results	noun phase with of-phrase fragment	Intangible framing
151	point out that*	5	Discussion	verb phrase or noun phrase+that-clause fragment	Resultative signal
	points out that	6	Introduction	verb phrase or noun phrase+that-clause fragment	Resultative signal

152	presented in Table	7	Results	Passive+prepositional phrase fragment	Structuring
153	purpose of the*	5	Methods	noun phase with of-phrase fragment	Framing signal
	purpose of this	6	Introduction	noun phase with of-phrase fragment	Framing signal
154	related to a*	9	Results	Passive+prepositional phrase fragment	Stance feature
	related to the	6	Introduction	Passive+prepositional phrase fragment	Stance feature
	related to the	12	Results	Passive+prepositional phrase fragment	Stance feature
155	reliability of the	5	Methods	noun phase with of-phrase fragment	Intangible framing
156	research has been	6	Introduction	Noun/pronoun phrase+be	Other
157	result of this	5	Discussion	noun phase with of-phrase fragment	Resultative signal
158	result shows that	5	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
159	results from the	8	Results	noun phrase with other post-modifier fragment	Resultative signal
160	results of the*	24	Results	noun phase with of-phrase fragment	Resultative signal
	results of the	17	Discussion	noun phase with of-phrase fragment	Resultative signal
	results of this	11	Discussion	noun phase with of-phrase fragment	Resultative signal
161	results show that	7	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
162	revealed that in	6	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
163	revealed that the	10	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
164	schematic knowledge of	11	Discussion	noun phase with of-phrase fragment	Intangible framing
165	seemed to be*	8	Results	verb or adjective to-clause fragment	Stance feature
	seems to be	5	Introduction	verb or adjective to-clause fragment	Stance feature
	seems to be	6	Results	verb or adjective to-clause fragment	Stance feature
166	seen from the	6	Results	Passive+prepositional phrase fragment	Structuring
167	seen in Table	14	Results	Passive+prepositional phrase fragment	Structuring
168	should be conducted	9	Discussion	Other verbal fragment	Stance feature
169	should be noted	9	Results	Other verbal fragment	Stance feature
170	show that the*	8	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
	showed that the	10	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
	showed that the	6	Discussion	verb phrase or noun phrase+that-clause fragment	Resultative signal
171	shown in Table	27	Results	Passive+prepositional phrase fragment	Structuring

172	shown in the	11	Results	Passive+prepositional phrase fragment	Structuring
173	shows that the	17	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
174	similar to that	5	Results	other expressions	Stance feature
175	so that they	5	Results	other expressions	Stance feature
176	some of the	6	Results	noun phase with of-phrase fragment	Quantification
177	stated that the	5	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
178	students should be	5	Discussion	Noun/pronoun phrase+be	Stance feature
179	study aims to	5	Introduction	Other noun phrase	Other
180	study demonstrated that	5	Discussion	verb phrase or noun phrase+that-clause fragment	Resultative signal
181	study found that	6	Discussion	verb phrase or noun phrase+that-clause fragment	Resultative signal
182	such as the	5	Results	other expressions	Transition signal
	such as the	6	Discussion	other expressions	Transition signal
183	suggests that the	6	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
184	terms of the	7	Results	noun phase with of-phrase fragment	Framing signal
	terms of the	7	Discussion	noun phase with of-phrase fragment	Framing signal
185	The above table	5	Results	Other noun phrase	Structuring
186	the acquisition of	13	Introduction	noun phase with of-phrase fragment	Intangible framing
187	the basis of	5	Methods	noun phase with of-phrase fragment	Framing signal
188	the beginning of	6	Methods	noun phase with of-phrase fragment	Location
	the beginning of	7	Results	noun phase with of-phrase fragment	Location
189	the case of	6	Introduction	noun phase with of-phrase fragment	Structuring
190	the characteristics of	12	Introduction	noun phase with of-phrase fragment	Intangible framing
	the characteristics of	6	Methods	noun phase with of-phrase fragment	Intangible framing
	the characteristics of	12	Results	noun phase with of-phrase fragment	Intangible framing
	the characteristics of	5	Discussion	noun phase with of-phrase fragment	Intangible framing
191	the concept of	5	Introduction	noun phase with of-phrase fragment	Intangible framing
192	the context of	5	Introduction	noun phase with of-phrase fragment	Framing signal
193	the current study	5	Results	Other noun phrase	Location
194	the data from	5	Results	noun phrase with other post-modifier fragment	Resultative signal
195	the data were*	7	Methods	Noun/pronoun phrase+be	Procedure
	The data were	7	Methods	Noun/pronoun phrase+be	Procedure

196	the degree of	7	Results	noun phase with of-phrase fragment	Quantification
197	the development of	8	Discussion	noun phase with of-phrase fragment	Intangible framing
	the development of	7	Results	noun phase with of-phrase fragment	Intangible framing
	the development of	8	Discussion	noun phase with of-phrase fragment	Intangible framing
198	the effectiveness of	7	Introduction	noun phase with of-phrase fragment	Intangible framing
199	the effects of	8	Discussion	noun phase with of-phrase fragment	Intangible framing
200	the fact that	6	Introduction	verb phrase or noun phrase+that-clause fragment	Other
	the fact that	19	Results	verb phrase or noun phrase+that-clause fragment	Other
	the fact that	9	Discussion	verb phrase or noun phrase+that-clause fragment	Other
201	the findings from*	5	Results	noun phrase with other post-modifier fragment	Resultative signal
	the findings of	6	Results	noun phase with of-phrase fragment	Resultative signal
	the findings of	15	Discussion	noun phase with of-phrase fragment	Resultative signal
	The findings of	13	Discussion	noun phase with of-phrase fragment	Resultative signal
202	the form of	6	Methods	noun phase with of-phrase fragment	Intangible framing
203	the frequency of*	26	Results	noun phase with of-phrase fragment	Quantification
	The frequency of	5	Results	noun phase with of-phrase fragment	Quantification
	the frequency of	10	Discussion	noun phase with of-phrase fragment	Quantification
204	the importance of	16	Introduction	noun phase with of-phrase fragment	Intangible framing
	the importance of	10	Discussion	noun phase with of-phrase fragment	Intangible framing
205	the influence of	2	Results	noun phase with of-phrase fragment	Intangible framing
206	the kind of	10	Results	noun phase with of-phrase fragment	Intangible framing
207	the level of	6	Results	noun phase with of-phrase fragment	Quantification
	the level of	6	Discussion	noun phase with of-phrase fragment	Quantification
	the levels of	5	Results	noun phase with of-phrase fragment	Quantification
208	the majority of*	12	Results	noun phase with of-phrase fragment	Quantification
	The majority of	5	Results	noun phase with of-phrase fragment	Quantification
209	the meaning of*	9	Methods	noun phase with of-phrase fragment	Description
	the meaning of	47	Results	noun phase with of-phrase fragment	Description
	the meaning of	6	Discussion	noun phase with of-phrase fragment	Description

	the meanings of	7	Results	noun phase with of-phrase fragment	Description
	the meanings of	5	Discussion	noun phase with of-phrase fragment	Description
210	the medium of	5	Introduction	noun phase with of-phrase fragment	Intangible framing
211	the nature of	7	Results	noun phase with of-phrase fragment	Intangible framing
212	the number of*	8	Methods	noun phase with of-phrase fragment	Quantification
	the number of	30	Results	noun phase with of-phrase fragment	Quantification
	The number of	7	Results	noun phase with of-phrase fragment	Quantification
213	the other hand	17	Results	other expressions	Framing signal
214	the part of	6	Results	noun phase with of-phrase fragment	Framing signal
	the part of	5	Discussion	noun phase with of-phrase fragment	Framing signal
215	the participants had	6	Discussion	verb phrase with noun/pronoun	Resultative signal
216	the participants in	7	Methods	noun phrase with other post-modifier fragment	Procedure
	the participants in	15	Discussion	noun phrase with other post-modifier fragment	Procedure
217	the participants were	18	Methods	Noun/pronoun phrase+be	Procedure
	The participants were	6	Methods	Noun/pronoun phrase+be	Procedure
	the participants were	6	Results	Noun/pronoun phrase+be	Procedure
	the participants were	8	Discussion	Noun/pronoun phrase+be	Procedure
218	the pattern of*	10	Results	noun phase with of-phrase fragment	Description
	the patterns of	6	Results	noun phase with of-phrase fragment	Description
219	the percentage of	9	Results	noun phase with of-phrase fragment	Quantification
220	the present study	13	Methods	Other noun phrase	Location
	the present study	17	Results	Other noun phrase	Location
	the present study	22	Discussion	Other noun phrase	Location
221	the process of	6	Introduction	noun phase with of-phrase fragment	Procedure
	the process of	7	Results	noun phase with of-phrase fragment	Procedure
222	the proportion of	5	Results	noun phase with of-phrase fragment	Quantification
223	the purpose of	5	Introduction	noun phase with of-phrase fragment	Framing signal
	the purpose of	9	Methods	noun phase with of-phrase fragment	Framing signal
224	the questionnaire was*	7	Methods	Noun/pronoun phrase+be	Procedure

	The questionnaire was	6	Methods	Noun/pronoun phrase+be	Procedure
225	the respondents had	6	Results	verb phrase with noun/pronoun	Resultative signal
226	the result of*	5	Results	noun phase with of-phrase fragment	Resultative signal
	The result of	5	Results	noun phase with of-phrase fragment	Resultative signal
	the result of	5	Discussion	noun phase with of-phrase fragment	Resultative signal
	The result of	5	Discussion	noun phase with of-phrase fragment	Resultative signal
	The results from	6	Results	noun phrase with other post-modifier fragment	Resultative signal
	the results of	16	Results	noun phase with of-phrase fragment	Resultative signal
	The results of	9	Results	noun phase with of-phrase fragment	Resultative signal
	the results of	22	Discussion	noun phase with of-phrase fragment	Resultative signal
	The results of	8	Discussion	noun phase with of-phrase fragment	Resultative signal
227	The results show	5	Results	verb phrase with noun/pronoun	Resultative signal
228	the role of	11	Introduction	noun phase with of-phrase fragment	Intangible framing
	the role of	13	Discussion	noun phase with of-phrase fragment	Intangible framing
229	the scores of	5	Results	noun phase with of-phrase fragment	Quantification
230	the sense that	5	Results	verb phrase or noun phrase+that-clause fragment	other
231	the strategy of	5	Results	noun phase with of-phrase fragment	Intangible framing
232	the students were*	7	Methods	Noun/pronoun phrase+be	Procedure
	The students were	5	Methods	Noun/pronoun phrase+be	Procedure
	the students were	8	Results	Noun/pronoun phrase+be	Procedure
233	the study of	6	Results	noun phase with of-phrase fragment	Description
	the study of	7	Discussion	noun phase with of-phrase fragment	Description
234	The subjects were	5	Methods	Noun/pronoun phrase+be	Procedure
235	the success of	5	Discussion	noun phase with of-phrase fragment	Intangible framing
236	the type of*	6	Results	noun phase with of-phrase fragment	Intangible framing
	the types of	8	Results	noun phase with of-phrase fragment	Intangible framing
237	the use of*	33	Introduction	noun phase with of-phrase fragment	Procedure
	the use of	9	Methods	noun phase with of-phrase fragment	Procedure

	the use of	56	Results	noun phase with of-phrase fragment	Procedure
	the use of	43	Discussion	noun phase with of-phrase fragment	Procedure
	The use of	6	Discussion	noun phase with of-phrase fragment	Procedure
	their use of	10	Results	noun phase with of-phrase fragment	Procedure
238	there is a	17	Results	Noun/pronoun phrase+be	Stance feature
	there was a	15	Results	Noun/pronoun phrase+be	Stance feature
	there was a	15	Results	Noun/pronoun phrase+be	Stance feature
239	There should be	7	Discussion	Noun/pronoun phrase+be	Stance feature
240	there is no*	5	Results	Noun/pronoun phrase+be	Stance feature
	there was no	5	Results	Noun/pronoun phrase+be	Stance feature
241	This can be	8	Results	Noun/pronoun phrase+be	Stance feature
242	This finding is	9	Discussion	Noun/pronoun phrase+be	Resultative signal
243	This indicates that	7	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
244	This is because	5	Introduction	other expressions	Stance feature
245	This is in	5	Discussion	Noun/pronoun phrase+be	other
246	This means that	5	Discussion	verb phrase or noun phrase+that-clause fragment	Resultative signal
247	this statement is	7	Results	Noun/pronoun phrase+be	Intangible framing
248	this study is*	5	Discussion	Noun/pronoun phrase+be	Location
	this study was	11	Methods	Noun/pronoun phrase+be	Location
	this study was	11	Discussion	Noun/pronoun phrase+be	Location
249	This suggests that	6	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
250	to answer the	6	Introduction	other expressions	Procedure
251	to be a*	8	Introduction	other expressions	Procedure
	to be a	9	Results	other expressions	Procedure
	to be an	5	Introduction	other expressions	Procedure
	to be the	7	Results	other expressions	Procedure
	to be the	8	Discussion	other expressions	Procedure
	to be the	8	Discussion	other expressions	Procedure
252	to be able	7	Discussion	other expressions	Stance feature
253	to be aware	7	Discussion	other expressions	Stance feature
254	to complete the	7	Methods	other expressions	Procedure
255	to determine the	8	Methods	other expressions	Procedure
256	to find out	6	Methods	other expressions	Procedure
	to find out	7	Discussion	other expressions	Procedure
257	to identify the	7	Results	other expressions	Procedure
	to identify the	6	Results	other expressions	Procedure
258	to investigate the	7	Introduction	other expressions	Procedure
259	to participate in	5	Methods	other expressions	Procedure
260	to retain the	7	Results	other expressions	Procedure

261	to the participants	7	Results	Prepositional phrase expressions	Resultative signal
262	to use a*	8	Discussion	other expressions	Procedure
	to use a	8	Discussion	other expressions	Procedure
	to use the	6	Introduction	other expressions	Procedure
	to use the	6	Results	other expressions	Procedure
	to use the	7	Discussion	other expressions	Procedure
263	understanding of the	7	Results	noun phase with of-phrase fragment	Intangible framing
264	use of a*	8	Discussion	noun phase with of-phrase fragment	Procedure
	use of the	10	Introduction	noun phase with of-phrase fragment	Procedure
	use of the	13	Results	noun phase with of-phrase fragment	Procedure
	use of the	7	Discussion	noun phase with of-phrase fragment	Procedure
	use of the	7	Discussion	noun phase with of-phrase fragment	Procedure
265	used as a	10	Introduction	Passive+prepositional phrase fragment	Procedure
266	used by a*	12	Results	Passive+prepositional phrase fragment	Resultative signal
	used by children	6	Results	Passive+prepositional phrase fragment	Resultative signal
267	used in the*	17	Methods	Passive+prepositional phrase fragment	Procedure
	used in the	15	Results	Passive+prepositional phrase fragment	Procedure
	used in the	8	Introduction	Passive+prepositional phrase fragment	Procedure
	used in this	10	Methods	Passive+prepositional phrase fragment	Procedure
	used in this	5	Results	Passive+prepositional phrase fragment	Procedure
268	used to analyze	8	Methods	verb or adjective to-clause fragment	Procedure
269	used to determine	5	Methods	verb or adjective to-clause fragment	Procedure
270	validity of the	6	Methods	noun phase with of-phrase fragment	Intangible framing
271	was able to*	10	Discussion	verb or adjective to-clause fragment	Stance feature
	were able to	8	Results	verb or adjective to-clause fragment	Stance feature
	were able to	16	Discussion	verb or adjective to-clause fragment	Stance feature
272	was based on	6	Methods	Passive+prepositional phrase fragment	Structuring
273	was carried out	7	Methods	Passive+prepositional phrase fragment	Procedure
274	was divided into	8	Methods	Passive+prepositional phrase fragment	Procedure
275	was employed to	5	Methods	verb or adjective to-clause fragment	Procedure

276	was found that	5	Introduction	verb phrase or noun phrase+that-clause fragment	Resultative signal
	was found that	9	Results	verb phrase or noun phrase+that-clause fragment	Resultative signal
	was found that	11	Discussion	verb phrase or noun phrase+that-clause fragment	Resultative signal
277	was found to	13	Results	verb or adjective to-clause fragment	Resultative signal
	was found to	6	Discussion	verb or adjective to-clause fragment	Resultative signal
278	was identified according	7	Results	Passive+prepositional phrase fragment	Procedure
279	was used to	6	Methods	verb or adjective to-clause fragment	Procedure
280	were administered to	7	Methods	verb or adjective to-clause fragment	Procedure
281	were asked to	17	Methods	verb or adjective to-clause fragment	Procedure
	were asked to	10	Results	verb or adjective to-clause fragment	Procedure
282	were consistent with	5	Discussion	Copula be + noun/ adjective phrases	Resultative signal
283	were divided into	8	Methods	Passive+prepositional phrase fragment	Procedure
284	were found to	6	Discussion	verb or adjective to-clause fragment	Resultative signal
285	were reported to	6	Discussion	verb or adjective to-clause fragment	Resultative signal
286	were required to	5	Methods	verb or adjective to-clause fragment	Procedure
287	were used in	6	Methods	Passive+prepositional phrase fragment	Procedure
	were used in	7	Results	Passive+prepositional phrase fragment	Procedure
288	were used to	11	Methods	verb or adjective to-clause fragment	Procedure
289	with regard to	5	Results	Other prepositional phrase	Framing signal
	with regard to	5	Discussion	Other prepositional phrase	Framing signal
	With regard to	5	Discussion	Other prepositional phrase	Framing signal