3L: The Southeast Asian Journal of English Language Studies – Vol 26(4): 99 – 113
http://doi.org/10.17576/3L-2020-2604-08

A Corpus Analysis of Frequently Occurring Words and their Collocations in High-Impact Research Articles in Education

SHARMINE FARAHIN BAHTIAR
Universiti Putra Malaysia, Malaysia

ARSHAD ABD SAMAD
Universiti Putra Malaysia, Malaysia
arshad@upm.edu.my

ABU BAKAR MOHD RAZALI
Universiti Putra Malaysia, Malaysia

HO YOONG WEI
Universiti Putra Malaysia, Malaysia

ABSTRACT

In research writing, the importance of formulaic sequences (FS) or collocations and the need for writers to adapt and utilise these phrasal constructions in their writing cannot be denied. However, English as a second or a foreign language (ESL/EFL) writers often have difficulty in finding the right words or phrases in writing good academic research papers due to limited vocabulary and the lack of native-like fluency. There is also the concern that they are unaware of the rhetorical structure of academic research papers, which can hamper the organization of ideas and flow of writing and lead to messy and unclear production of language. Formulaic sequences and collocations of words represent useful phrasal constructions that writers use to fluently and efficiently express their intended communicative purposes. In this regard, this study uses corpus analysis to list collocations commonly used in high-impact journals written in the field of education. The study also categorises these collocations according to their communicative purposes, referred to as moves and steps, in the rhetorical structure of the Introduction of research articles. The final list which focussed on ten node words from 40 high-impact journal articles consists of 3 to 12 word phrases found in the Introduction section of these articles. They were then categorised according to their specific functions based on an adapted Introduction Move Framework of the Create a Research Space (CARS) schema (Swales 2004) and common moves in the Introduction section of the Academic Phrasebank (Morley, 2014).

Keywords: Rhetorical structure; ESL and EFL; Corpus studies; Collocations; Academic writing

INTRODUCTION

Academic writing is a particular style used in formal essays, which requires writers to use formal language, a logical structure and is supported by evidence. Academics' use of this writing style defines the intellectual boundaries of their disciplines and the areas of their specific expertise. Cohesiveness is also an important characteristic which helps to establish a logical and organised flow in the written expression as well as a unified and coherent whole. In this respect, the use of formulaic sequences (FS), which are recurrent sequences of words, can have a crucial role in academic discourse. Davis and Morley (2015) claim that recognising such sequences can help by providing a scaffold for students to organise their ideas and improve their writing style. Language is to a large extent formulaic in nature (Meunier, 2012; Sinclair, 1991) and learning and mastering FS often used in academic writing can therefore be a good starting point for L2 learners to convey their ideas and messages coherently.

Smadja and McKeown (1990) argued that collocational knowledge, which can be very difficult to acquire for second language learners, is especially effective in sentence generation. Previous studies have shown that non-native English language users tend to have a rather limited knowledge of phraseological structures that typically characterise academic discourse

(Howarth, 1998). He also claimed that English as a Second Language (ESL) or English as a Foreign Language (EFL) learners face particular difficulty in producing appropriate word combinations because of their lack of collocational knowledge. Gilquin and Paquot (2008) claimed that EFL learners struggle with the appropriate use of FS, making their academic prose come across as inappropriate and too colloquial. These findings imply that there is a pressing need for ESL/EFL learners to be exposed to important and frequently used FS or phrases used in academic writing in order to increase overall collocational knowledge.

According to Erman and Warren (2000), discourse comprises slightly more than 50% phraseological elements, and in other studies even higher estimates have been suggested (e.g., Altenberg, 1998), punctuating their importance in writing. Swales and Feak (2012) urge non-native novice writers of the English language to adopt a strategy of searching for phrasal constructions and adapting them for their own writing. In such a situation, a list of phrases or FS compiled through a thorough review of academic written sources can be utilised to assist students in their academic writing. Various academic corpora have been compiled, such as the *Academic Word List* (Coxhead, 2000), *Academic Phrasebank,* (Morley, 2014)*, the Academic Formula List (*Simpson-Vlach and Ellis, 2010) and Ackermann and Chen's (2013) *Academic Collocation List* which can help learners become familiar with such phrases. However, as Biber et al. (2004) argued, it is important to emphasise the communicative function of these phrases and to present them in greater context so as to make them more pedagogically useful. Similarly, understanding the use of the phrases in specific genres such as in academic research papers and according to the rhetorical structures of these genres can have a positive pedagogical impact. In this respect, models such as the Create a Research Space (CARS) model provide learners with a useful reference on how texts are organised. Hence, the focus of this article is to obtain a better understanding of the phrasal constructions or FS that are more salient in the Education field as well as identify their use in the rhetorical structure of the introduction section of academic research papers in the field in order to assist student writers in academic writing.

BACKGROUND TO THE STUDY

Formulaic sequences are frequent "combinations of at least two words favoured by native speakers in preference to an alternative combination which could have been equivalent had there been no conventionalization" (Erman &Warren, 2000, p. 31). They are multi-word units and combinations, such as, *to give an example* and *the first point is,* and are shown to frequently occur in diverse types of available corpora (Peters & Pauwels, 2015) using either intuitive means (Pawley & Syder, 1983) or corpus-based investigations (Altenberg, 1998, Biber, 2006, Sinclair, 1991). Sinclair (1991) summarises from his work that "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments." (1991: 110).

Given the prevalence of the phraseological and collocational nature of academic language discourse, recent research has examined phraseological language, such as collocations, in academic and learner writing (Ang & Tan 2019, Ackermann & Chen 2013, Hadi Kashiha & Chan 2014) and has shown that phraseological units as well as fixed and semi-fixed phrases are not only frequent and pervasive but also play an important role in current academic discourse (Biber, 2006; Hyland, 2008). Cortes (2013) highlighted this importance, claiming that mastering FS is a prerequisite for foreign language learners' success in academic writing. They should know how a text is organised as well as how units are realised both linguistically and lexically, hence the need to demonstrate a high level of proficiency in FS. At this level of proficiency, Peters (1983) believes that writers are able to draw upon a store of

pre-constructed phrases and quickly retrieve them from memory without having to be assembled at each time.

There is a clear relationship between the formation of FS with fluent and natural production of language (Ellis, 2002; Howarth, 1998; Paquot & Granger, 2012; Pawley & Syder, 1983; Sinclair, 1991). The appropriate use of collocations is crucial in ensuring the naturalness of language use in real life situations (Pawley & Syder 1983, Sinclair 1991, Howarth 1998). Hsu (2007) also argues that collocations increase students' writing proficiency with better writers possessing better collocational knowledge. Pawley and Syder (1983) also hinted at stark differences between adult native speakers and ESL/EFL learners since the former have hundreds of thousands of 'lexicalised sentence stems' at their disposal. It is also suggested from their study that ESL/EFL learners should have at least a similar number of chunked expressions to enable them to reduce cognitive effort, save processing time and have language available for immediate use. In short, sufficient mastery of multi-word units like FS or collocations is able to aid ESL/EFL students not only in terms of their writing proficiency but also overall fluency in writing

In spite of the advantages that FS offer, unfortunately many ESL/EFL writers are not able to utilise the benefits. There is evidence that ESL/EFL writers do not seem to use collocations to the same extent as native speakers (Laufer & Waldman 2011), tend to use a smaller range of FS (Adel & Erman, 2012), possess a limited repertoire of phrasal sequences (Howarth, 1998; Pawley & Syder, 1983), and are found to be underusing "the most academic-like recurrent word combinations" (Paquot & Granger 2012, p. 139). Furthermore, ESL/EFL writers' usage of FS tends to be informal (e.g., *to find out*), contains spoken-like features and are often overused in their academic writing (Gilguin & Paquot, 2008).Their lack of collocational knowledge hinders them from producing 'automation of collocation' (Kjellmer, 1991) which is one of the major determiners in differentiating native users of the language from non-native learners. Cowie (1992), in fact, argues that for second language learners "it is impossible to perform at a level acceptable to native users, in writing or in speech, without controlling an appropriate range of multiword units" (p.10). It is therefore crucial that measures are undertaken to make ESL/EFL novice writers more aware of phrasal constructions and collocations and how they should be used.

Flowerdew and Li (2007) stressed that novice ESL/EFL writers need to learn to develop their use of fixed and semi-fixed phrases effectively. Efforts in identifying patterns or structures in research articles can be especially useful. Numerous studies have been conducted to analyse conventions used in specific genres and parts of a genre. With the implicit knowledge of how a certain section of research articles is written, writers can get a better idea of how to organise their ideas and use appropriate strategies. It is therefore not surprising that scholars have examined the generic structure of the *Introduction* section (Kanoksilapatham, 2011, Nguyen & Pramoolsook, 2014), *Abstracts* (Nguyen, 2018), *Methods* (Chang & Kuo, 2011, Lim, 2006), *Results* (Lim, 2010), *Discussion* (Yang & Allison, 2003 and Dobakhti, 2011) and also others who are interested in rhetorical structures as a whole (Kanoksilapatham, 2007).

The various academic corpora that have been compiled provide ESL/EFL writers with common phrases and FS that can ease their writing concerns. Similarly, models of the rhetorical structure of academic articles such as the CARS model (Swales, 1990, 2004) can help learners understand the communicative functions required to write academic research articles. The Academic Phrasebank (Morley, 2014) is especially useful as it was "designed primarily for academic and scientific writers who are non-native speakers of English" (p. 2). The phrases are also organised according to headings which represent common functions in various sections of an academic research paper. Morley (2014) describes his work as providing the "phraseological 'nuts and bolts' of academic writing organised according to the main sections of a research paper" (p. 2). The Academic Phrasebank lists phrases of between 5 to

16 words under functions such as "establishing the importance of the topic" and "highlighting a problem" in writing introductions. Similarly, Swales' CARS model is an attempt to describe functions used in the introduction section of research articles. His earlier model (Swales, 1990) was revised after suggestions made by several researchers (see Afshar et al. 2018). The present model (Swales, 2004) has been described as seminal, inspiring much research with implications for academic writing pedagogy (Cotos et al., 2015). The model is based on "moves" which Swales (2004) refers to as a "coherent communicative function" (p. 229) in a particular written genre. These rhetorical moves are further elaborated by functional steps. Cotos et al. (2015) believe that Swales' work has "strengthened the relationship between linguistic inquiry and EAP pedagogy" (p. 53).

This study attempts to identify common phrases in the *Introduction* section of research articles in the field of Education based on ten frequently occurring words found in the articles. According to Dobakhti (2011), a large number of studies have been carried out to investigate generic structures of research articles in various disciplines due to their saliency in academic papers. Corpus based studies that focus on specific disciplines, i.e. Education in this study, are in line with the view that a discipline specific list is more appropriate than a 'one-size-fits-all' framework that claims to be representative of all disciplines. This approach is taken by Hong, Hua and Mengyu (2017) who limit their research scope to collocations in International Business Management, and Joseph et al. (2014) to collocations in Forestry. In this regard, Hyland and Tse (2007) assert that learners should master the specialised phraseological language common in their respective academic fields as "all disciplines shape words for their own uses" (p. 238). This study has chosen to focus on the *Introduction* section because of the role of the section in presenting the overview and background of the studies, both of which are important functions in a cohesive academic article.

METHODOLOGY

This corpus analysis study is a linguistic analysis on corpora consisting of texts to contextualise the analyses of language. Leech (1994) stated that one of the most important advantages of using corpora in language teaching or learning is that it provides evidence for the function and usage of words and expressions. He also emphasised that with corpus-based research, open-ended and unrestricted supply of data will encourage exploration and discovery of learning.

Gledhill (2011) asserted that especially in corpus linguistics, the context of a specialised corpus must be explicit and display clear design criteria so as to ensure the accuracy of the analysis. The academic written corpus compiled in this study comprised 40 research articles relevant to the field of education, chosen from high-impact Education journals published from 2015 to 2019 namely *Review of Educational Research, American Educational Research Journal, Journal of Teacher Education and TESOL Quarterly*. All four of these journals were in the first quartile (Q1) in the SJR (Scimago Journal and Country Rank) with Impact Factors ranging from 2.15 to 5.51. Articles were randomly selected with the requirement that their authors were associated with academic institutions from English speaking countries. This requirement was considered necessary to ensure some form of commonality in terms of the users of the language. Forty articles (see Appendix A) were hence chosen from the four journals with 10 articles from each journal. The researchers considered 40 articles to be sufficient for the purpose of this study, taking into consideration that similar studies involving similar analysis compiled 20 or fewer (Hirano, 2009; Lim, 2006; Nwogu, 1997; Rahman et al., 2017; and Samraj, 2002). The academic field, journal, quartile and year of publication formed the sampling strata. The researchers made it a point to limit the randomly chosen articles exclusively to the Education field as it is crucial to have a representative target corpus.

The study intended to identify phraseological combinations extracted from the *Introduction* sections of each research article based on ten commonly occurring words. Each of these ten words acted as nodes from which the common collocations were identified manually as well as by using techniques found in the MonoConc Pro 2.0 concordancing software. These node words were selected from the 100 most frequently occurring content words generated after a stop list was used to exclude all function words such as *the*, *in* and *as*. Sinclair (1991) stated that content words, be it a word or multi-word phrase, are especially useful as node items in order to reveal frequent collocates. Although the ten words were not the most frequent in the list, they were preferred over other words, especially nouns such as *teachers*, *knowledge* and *performance* which were more frequent but were considered more limited to the topics of the papers. Additionally, the node words also had to meet the criterion of occurring in at least three of the academic articles. These node words were then used to extract three to twelve word collocations from the corpus using especially the *collocation window* technique available in the concordance software. Gabsalova, Brezina and McEnery (2017) asserted that the collocation window approach is appropriate to find co-occurrences within a specified window span since it enables scholars to identify loose word associations. A three word collocation window was considered a reasonable minimum whereas the twelve word maximum allowed for sufficient context to understand how the node word was used. Furthermore, the window span was comparable to the length of phrases found in Morley's Academic Phrasebank (2014).

The study also intended to identify where the collocations occurred according to the moves models proposed by Swales (1990, 2004) and Morley (2014). In the model adapted from Swales (2004) and Morley (2014) as seen in Table 1, the researchers adopted the first three principal moves from the CARS model and supplemented them with communicative functions, also referred to here as steps, suggested by Morley (2014).

TABLE 1. Adapted Introduction Move Framework

Moves		Steps	
Move 1	Establishing a territory (obligatory)	Step 1.	Establishing context, background and/or importance of topic
		Step 2	Synopsis of literature
		Step 3.	Highlighting problem, or controversy in field of study
Move 2.	Establishing a niche	Step 1.	Highlighting gap and inadequacies of past studies
		Step 2.	Presenting positive justification
Move 3.	Presenting the present work	Step 1.	Outlining purposes of research
		Step 2.	Presenting research question/ hypotheses
		Step 3.	Summarizing methods
		Step 4.	Outlining structure of paper

Several researchers (e.g. Joseph et al., 2014) adapt Swales' CARS model to focus on specific research goals. Similarly, the adapted Introduction Move Framework in Table 1 is a simplified version of the CARS model that incorporates some functions from Morley's (2014) Academic Phrasebank. Move 1 in the Swales' (2004) CARS model consists of only one step which was Topic generalisations of increasing specificity. In this study, the researchers replaced this step with three functions from Morley (2014) which were *Establishing the context, background and/or importance of the topic, Synopsis of literature* and *Highlighting problem, or controversy in the field of study*. For Move 2, the researchers combined two sub-steps into a single step (*Highlighting the gap and inadequacies of past studies*) for simplicity and clarity. In his identification of common phrases associated to functions, Morley distinguished between the two sub-steps but the researchers, found the distinctions unnecessary as the two steps are somehow interchangeable and can be difficult to distinguish. Finally, for

Move 3, the original framework consisted of 7 steps, but the researchers decided to keep only three and add Morley's suggested step of *Outlining purposes of the research*.

RESULTS AND DISCUSSION

The ten content words selected from the corpus to serve as nodes for further analysis were: *research, study, found, provide, focus, examine, identify, suggest, address* and *understand* and their lemmas. Hence, the selection of *study* entails the inclusion of *studies, studying, studied* and *student*.

The researchers then listed these node words according to the moves and steps found in the *Introduction* Section of the selected articles. Table 2 provides the frequency of occurrence of the node words according to moves and steps in the Introduction section of research articles in the field of Education (henceforth abbreviated as InRAEd).

TABLE 2. Frequency of node words according to moves and steps

Moves (M)	Steps (S)	Node words with frequency occurence
M1. Establishing a territory (obligatory)	S1. Establishing context, background and/or importance of topic	Research 20; Found 11; Study 9; Focus 8; Understand 8; Suggest 5; Provide 4; Examine 3; Address 2; Identify 1 (Total: 71)
	S2. Synopsis of literature	Study 22; Research 17; Found 12; Suggest 7; Focus 7; Examine 3; Provide 2 (Total: 70)
	S3. Highlighting problem, or controversy in field of study	Found 10; Research 7; Study 4; Identify 1; Understand 1 (Total: 23)
M2. Establishing a niche	S1. Highlighting gap and inadequacies of past studies	Research16; Study 6; Examine 6; Address 3; Focus 2; Understand 1; Found 1 (Total: 35)
	S2. Presenting positive justification	Research 10; Study 5; Address 3; Examine 2; Focus 1; Provide 1; Identify 1; Suggest 1; Found 1 (Total: 25)
M3. Presenting the present work	S1. Outlining purposes of research	Study 16; Focus 13; Examine 10; Identify 4; Research 2; Address 1 (Total: 46)
	S2. Presenting research question/ hypotheses	Research 10; Study 7; Address 3; Examine 2; Focus 1; Provide 1 (Total: 24)
	S3. Summarizing methods	Study 11; Research 4; Examine 3; Identify 2; Focus 1; Address 1 (Total: 22)
	S4. Outlining structure of paper	Study 12; Research 10; Provide 7; Found 2 (Total: 31)

Collocations for each node word were then generated. These collocations were manually vetted and cross checked by the researchers and a list of phrases containing the collocations was developed (see Appendix B). Ackermann and Chen (2013) argue that human intervention is especially salient in a data-driven collocation listing because a collocational list generated as a result of computational analysis can only be of pedagogical use when it is combined with human judgement. The major collocations for each node word and how they are used according to moves and steps in InRAEd are highlighted below:

i) *Research*. The word *research* is the most often used node word and is found in all the steps in InRAED. However, it should be noted that collocates of the word differ according to the steps in which they are used. In M1S2 (Synopsis of literature), for example, collocations, such as *research shows*, *research highlights* and *research indicates* are found. Whereas in M2S1 (Research gap), the word is noticeably preceded by qualifiers, such as in *little research, lack of research,* and *no prior research*. The only other form of the word *research* that occurs is *researcher/s* while the verb form does not occur. The word is most frequently found in M1S1 (Establishing context), M1S2 (Literature review) and M2S1 (Highlighting gap) where it occurs 20, 17 and 16 times respectively.

ii) *Study*. The word *study* and its lexemes appear most frequently in M1S2 (Synopsis of literature) and M3S1 (Outlining purposes of the research). In M1S2, it is interesting to note that it is more frequent than the word *research* which carries a similar meaning. The use of the word in the plural form may be more popular among the authors as it is able to emphasise number as compared to the word *research* which has a plural form that can sometimes be unclear. Common collocations with the word *study* include *provide*, *focus, portray, show, indicate* and *explore*. Similarly, the word *study* is preferred over *research* in M3S1 even when both are used as nouns. Nevertheless, the two words seem to be used almost as synonyms with the same collocates of the words, for example *studies focus* and *the focus of the research*.

iii) *Found*. As a node word, *found* occurs most frequently in all three steps in the first move (Establishing territory) especially in M1S3 (Highlighting problem) where it is the most frequently used of the ten words. Phrases that use this word and its lexemes include, *found to show, found to support, finding is contrary to, contradictory findings, gaps are found* and *mixed findings*.

iv) *Focus*. The word *focus* collocates often with two other words in this list – *research* and *study*. The word is more often used as a verb – *the study focuses on* – rather than a noun and is used to describe the purpose of the study such as *in this study we focus on…* and *we focus exclusively on*. It is well distributed across all the steps, appearing mostly in M3S1 (Outlining purpose of research), M1S1 (Establishing context) and M1S2 (Synopsis of literature) while it is absent in M1S3 (Highlighting problem) and M3S4 (Outlining structure of paper).

v) *Examine*. Like the word *focus*, the word *examine* also collocates with the words *research* and *study*. However, perhaps due to its more general meaning compared to *focus*, it occurs most frequently in M3S1 (Outlining purpose of research) to help provide an overall purpose of a study or research. At the same time, it also occurs frequently in M2S1 (Research gap) to express gaps in previous studies as in phrases such as *very little research has examined*, *studies have not examined* and *examination of … has not yet been conducted*.

vi) *Provide*. The node word *provide* occurs 15 times in the corpus, almost half of which appear in M3S4 (Outlining structure of paper). Post word collocates of the word in M3S4 include *framework, evidence, description* and *overview*, while the words *the study* and *the table* are collocates that appear before the word.

vii) *Suggest*. The node word *suggest* occurs predominantly in M1 (Establishing territory) collocating with words, such as *evidence*, *literature*, *research*, *studies* and *authors* to form phrases such as "a growing body of literature suggests…" and "prior research suggests…".

viii) *Address*. Appearing exclusively as a verb, the word occurs sparingly throughout but exerts greater presence in M2 (Establishing niche) with almost half of its occurrences appearing in this move. Collocates that occur with the word include *issues*, *gap*, and *question*.

ix) *Understand*. This word is most apparent in M1S1 (Establishing background), occurring 8 times. It appears in only two other steps, each time occurring only once. Among the more common collocations using *understand* include *to better understand, it is important to understand* and *to gain a deeper understanding*.

x) *Identify*. The word *identify* occurs the least frequently among the ten words. Almost half of its occurrences are in M3S1 (Outlining purpose of research) in the *to* infinitive form, i.e., *to identify*, the past tense, i.e. *the review identified,* and the gerund, i.e. *with the goals of identifying*.

Several observations should be noted from the findings. First, many of the node words and their collocations appear in most of the Moves and Steps in InRAEd. Some, however, are more specific and occur only in a few of the steps. The words *provide*, *understand* and *address*, for example, have the lowest frequencies and only appear occasionally in several moves and steps. For instance, *provide* is mainly used in M3S4 (Outlining the structure) to indicate what the article offers to readers; *understand* in M1S1 (Establishing the context) to provide background understanding and *address* in M2 (Establishing a niche) to show the area given attention. Secondly, node words may take up different collocates when used in different moves. A case in point is the word *research* which collocates with words like *support*, *found*, and *shows* in M1S1 but with *needed*, *lack* and *sparse* in M2S1. Both these observations help to justify placing emphasis on moves and steps in academic writing instruction. Finally, the list of phrases generated and presented in Appendix B differs significantly from those in the Academic Phrasebank (Morley, 2014). While the phrases in the Academic Phrasebank seemed to emphasise variety in the words and phrases used, this list provides different ways how specific words can be used to express a move. This focus on specific words may be more readily accepted among novice ESL/EFL writers learning to improve their writing.

CONCLUSION

The main aims of this study were to identify phrases of three-to-twelve words containing high recurrent collocations based on ten node words in the 40 InRAEd articles. Additionally, these phrases were placed in moves and steps according to a simplified Introduction Moves Framework based on Swales (2004) and Morley (2014). The words and their collocations can help raise awareness among ESL/EFL learners on how to effectively form sentences to express Moves and Steps in academic research articles. Similarly, instructors can utilise the list of phrases and collocations to emphasise important phrasal constructions that are used in academic writing.

The development of a list of collocations based on moves and steps is itself a significant contribution. Aside from giving learners ideas of what types of collocations are commonly used by writers in high-impact journals, the list can also help learners develop implicit knowledge of where these collocations commonly belong in the rhetorical structure of academic texts. Writing instructors can therefore utilise the list to emphasise important and highly recurrent phrasal constructions used in academic writing. Future research should analyse other sections of academic papers to generate a more thorough list to help in writing a good research article. Another recommendation is to compile a corpus based on Moves or Steps rather than according to sections in an academic paper. Doing so would help identify the specific words, collocations and phrases that frequently occur for each Move and Step rather than in a larger section of an academic paper such as its Introduction.

REFERENCES

Ackermann, K., & Chen, Y., H. (2013) Developing the Academic Collocation List – corpus driven and expert-judged approach, *Journal of English for Academic Purposes*, *12*(2013) 235–247.

Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English. *English for Specific Purposes*, *31*(2), 81-92.

Afshar, H. S., Doosti, M., & Movassagh, H. (2018). A Genre Analysis of the Introduction Section of Applied Linguistics and Chemistry Research Articles. *Iranian Journal of Applied Linguistics. 21*(1), 163-214.

Altenberg, B. (1998). On the Phraseology of Spoken English: The Evidence of Recurrent Word Combinations. In A. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp.101-122), Oxford: OUP.

Ang, L. H. & Tan, K. H. (2019). From Lexical Bundles to Lexical Frames Uncovering the Extent of Phraseological Variation in Academic Writing. *3L: The Southeast Asian Journal of English Language Studies*. *25(2)*, 99-112.

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

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

Chang, C. F. & Kuo, C. H. (2011). A corpus-based approach to online materials development for writing research articles, *English for Specific Purposes, 30*(3), 222-234.

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.

Cotos, E., Huffman, S. & Link, S. (2015). Furthering and applying move/step constructs: Technology-driven marshalling of Swalesian genre theory for EAP pedagogy. *Journal of English for Academic Purposes, 19,* 52-72.

Cowie, A. (1992). Multiword lexical units and communicative language teaching. In P. Arnaud, & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics* (pp.1-12). London: MacMillan.

Coxhead, A. (2000). *Coxhead's Academic Word List.* Retrieved from http://www.cal.org/create/conferences/2012/pdfs/handout-4-vaughn-reutebuch- cortez.pdf

Davis, M., & Morley, J. (2015). Phrasal Intertextuality: The Responses of Academics from Different Disciplines to Students' Re-Use of Phrases. *Journal of Second Language Writing*. *28*(2015), 20–3.

Dobakhti, L. (2011). *The Discussion Section of Research Articles in Applied Linguistics: Generic Structure and Stance Features* (Master's thesis, Universiti of Malaya). Retrieved from http://studentsrepo.um.edu.my/5641/1/LEILA_DOBAKHTI.pdf

Ellis, N. C. (2002). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition, 24*, 297-339.

Erman, B., & Warren,B. (2000). The Idiom Principle and the Open Choice Principle. *Text*, *20*, 29–62.

Flowerdew, J., & Li,Y.(2007). Language re-use among Chinese apprentice scientists writing for publication. *Applied Linguistics*, *28*, 440–465.

Gabsalova, D., Brezina, V., & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence. *Language Learning, 67*(S1). 155-179.

Gilquin, & Paquot, M. (2008). Too chatty. Learner academic writing and register variation. *English Text Construction*, *3*(1), 41-61.

Gledhill, C. J. (2011). The 'lexicogrammar' approach to analysing phraseology and collocation in ESP texts. *ASp [Online], 59*, 5-23.

Hadi Kashiha & Chan, S. H. (2014). Discourse functions in formulaic sequences in academic speech across two disciplines. *GEMA Online® Journal of Language Studies. 14*(2), 15-27.

Hirano, E. (2009). Research Article Introductions in English for Specific Purpose: a Comparison between Brazilian Portuguese and English. *English for Specific Purposes,* 7, 113-121.

Hsu, J. (2007) Lexical Collocations and their Relation to the Online Writing of Taiwanese College English Majors and Non-English Majors. *Electronic Journal of Foreign Language Teaching 2007*, *4*(2), 192–209.

Hong, A., L., Hua, T., K., & Mengyu, H. (2017). A Corpus-based Collocational Analysis of Noun Premodification Types in Academic Writing, *3L: The Southeast Asian Journal of English Language Studies, 23*(1), 115 – 131.

Howarth, P. (1998). The Phraseology of Learners' Academic Writing. In A. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp.161– 186). Oxford: OUP.

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

Hyland, K. & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly*. *41*(2), 235-253.

Joseph, R., Lim, J. M. & Mohd Nor, N. A. (2014). Communicative moves in Forestry Research Introductions: Implications for the design of learning materials. *Procedia – Social and Behavioural Sciences, 135*:15, 53-69.

Kanoksilapatham, B. (2011). Civil Engineering Research Article Introductions: Textual Structure and Linguistic Characterization. *The Asian ESP Journal*, *7*(2), 55-84.

Kanoksilapatham, B. (2007). Rhetorical moves in biochemistry research articles. In Biber, D., Connor, U. and Upton, T. A. (Eds.) *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam: John Benjamins.

Kjellmer, G. (1991) 'A mint of phrases', in K. Aijmer & B. Altenberg (Eds.) *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. (pp.111-127). Harlow, Essex:Longman.

Laufer, B. & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*. 61(2), 647-672.

Leech, G. (1994). Students' grammar - teachers' grammar - learners' grammar. In M. Bygate, A. Tonkyn & E. Williams (Eds.), *Grammar and the language teacher* (pp.17-30). New York: Prentice Hall.

Lim, J. M. H. (2006). Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes, 25*(3), 282–309.

Lim, J. M. H. (2010). Commenting on research results in applied linguistics and education: A comparative genre-based investigation. *Journal of English for Academic Purposes*, *9*(4), 280-294.

Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics*, 32, 111-129.

Morley, J. (2014). *Academic Phrasebank.* University of Manchester. Retrieved 20 March 2020 from www.kfs.edu.eg/com/pdf/2082015294739.pdf

Nguyen, T. T. L (2018). Rhetorical Structures and Linguistic Features of English Abstracts in Thai Rajabhat University Journals. *3L: The Southeast Asian Journal of English Language Studies, 24*(4), 71-84.

Nguyen, T.T. L & Pramoolsook, I. (2014). Rhetorical Structure of Introduction Chapters written by Novice Vietnamese TESOL postgraduates. *3L: The Southeast Asian Journal of English Language Studies, 20*(1), 61-74.

Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for Specific Purposes, 16*(2) 119-138.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–49.

Pawley, A., & Syder, F., H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp.191–226). New York: Longman.

Peters, A.M. (1983). *The units of language acquisition*. Cambridge: CUP.

Peters, E., & Pauwels, P. (2015). Learning Academic Formulaic Sequences. *Journal of English for Academic Purpose*, *20*(2015) 28-39.

Rahman, M., Darus, S. & Amir, Z. (2017). Rhetorical structure of introduction in applied linguistics research articles. *International Journal of Interdisciplinary Educational Studies. 9*(2), 69-84.

Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes 21*(1) 1-17

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

Sinclair, J. (1991). *Corpus, concordance, collocation: Describing English language*. Oxford: OUP.

Smadja, F.A. & McKeown, K.R. (1990) 'Automatically extracting and representing collocations for language generation', *Proc. of the 28th Annual Meeting of the Association for Computational Linguistics*, pp. 252.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: CUP.

Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge: CUP.

Swales, J., M., & Feak, C., B. (2012). *Academic Writing for Graduate Students: Essential Tasks and Skills* (3rd ed.). Ann Arbor: University of Michigan Press.

Yang, R., & Allison, D. (2003). Research articles in applied linguistics: moving from results to conclusions. *English for Specific Purposes, 22*(4), 365-385.

APPENDIX

APPENDIX A: LIST OF RESEARCH ARTICLE TITLES

Review of Educational Research

1. Developing, analyzing and using distractors for multiple-choice tests in Education: A comprehensive review.
2. Students' thinking about effort and ability: The role of developmental, contextual, and individual difference factors
3. Effective Dementia Education and training for the Health and Social Care Workforce: A systematic review of the literature.
4. Developing a STEM identity among young women: A social identity perspective.
5. Attention Deficit Hyperactivity Disorders and classroom-based interventions: Evidence-based status, effectiveness, and moderators of effects in single-case design research.
6. The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis.
7. The effect of Teacher Coaching on instruction and achievement: A meta-analysis of the causal evidence.
8. The comprehension problems of children with poor reading comprehension despite adequate decoding: A meta-analysis
9. The relation between Mathematics anxiety and Mathematics performance among school-aged students: A meta-analysis.
10. A meta-analysis of family-school interventions and children's social-emotional functioning: Moderators and components of efficacy.

American Educational Research Journal

1. If we can't do it, our children will do it one day'': A qualitative study of West African immigrant parents' losses and educational aspirations for their children.
2. What does it mean to be ranked a ''high'' or ''low'' value-added teacher? Observing differences in instructional quality across districts.
3. Just let the worst students go'': A critical case analysis of public discourse about race, merit, and worth
4. Transforming educational experiences in low-income communities: A qualitative case study of social capital in a full-service community school.
5. Toward equity in mathematics education for students with dis/abilities: A case study of professional learning.
6. Food instability and academic achievement: A quasi-experiment using snap benefit timing.
7. Connections between teachers' knowledge of students, instruction, and achievement outcomes
8. Comparing the effects of analysis-of-practice and content-based professional development on teacher and student outcomes in science.
9. Impacts of a practice-based professional development program on elementary teachers' facilitation of and student engagement with scientific argumentation.
10. Missing bus, missing school: Establishing the relationship between public transit use and student absenteeism.

Journal of Teacher Education

1. Rethinking student teacher feedback: Using a self-assessment resource with student teachers.
2. Whiteness as a dissonant state: Exploring one white male student teacher's experiences in urban contexts.
3. Do I belong in the profession? The cost of fitting in as a preservice teacher with a passion for social justice.
4. Measuring teaching quality of secondary mathematics and science residents: a classroom observation framework.
5. Investigating the role of social status in teacher collaborative groups.
6. Teachers' approaches toward cultural diversity predict diversity-related burnout and self-efficacy.
7. Effects of a data-based decision-making intervention for teachers on students' mathematical achievement.
8. An empirical study of the dimensionality of the mathematical knowledge for teaching construct.
9. Analyzing student learning gains to evaluate differentiated teacher preparation for fostering English learners' achievement in linguistically diverse classrooms.
10. Preservice teachers' mathematics teaching competence: comparing performance on two measures.

TESOL Quarterly

1. Effects of corpus-aided language learning in the EFL grammar classroom: A case study of students' learning attitudes and teachers' perceptions in Taiwan.
2. Mainstream teacher candidates' perspectives on ESL writing: The effects of writer identity and rater background.
3. Teachers' and students' second language motivational self system in English-medium instruction: A qualitative approach.
4. She needs to be shy!: Gender, culture, and nonparticipation among Saudi Arabian female students.
5. Identifying linguistic markers of collaboration in second language peer interaction: A lexico-grammatical approach.
6. Determiner use in English quantificational expressions: A corpus-based study.
7. The effects of international accents and shared first language on listening comprehension tests.
8. The effect of content retelling on vocabulary uptake from a TED talk.
9. Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text?
10. The effects of administration and response modes on grade 1–2 students' writing performance.

APPENDIX B: PARTIAL LIST OF COLLOCATIONS IN INTRODUCTION RAS

	Collocations
M1S1 Context, Background, Importance	- there is a plethora/growing body of **research** on… - Much/extensive **research** has documented/been conducted on… - **researchers** have pursued/noted/**focused** on… - **research** have included a **focus** on … - **research** shows that… - … has led to a growing body of/great deal of empirical/valuable **research** on… - has generated calls for **research**… - **researchers found** significant improvements in… - most **research** is based on… - A strong body of **research** supporting … - the implications of this **research**… - Decades of **research** support the **finding** that… - This body of **research** led… - The relevant **research** on… - ___ **research** is particularly relevant to… - Such **studies** generally compared … - **studies** reviewed above raise a number of important questions regarding… - Of particular relevance to this **study**, - in a **study focused** on … - **studies** have started to shift their **focus** on… - this/the present **study** can help to **identify**/ was undertaken in light of the fact that … - Our **study** is situated at… - many **studies** reporting a significant relation between… - ___ **found** no significant differences in… - … has been **found** to be significantly related to… - This **finding** indicates/highlights that… - … with one of the most robust **findings** being… - This **finding** was critical as… - ___ generally **finding** that… - it was **found** that… - the **focus** has been on… - … has predominately **focused** on … - … **focuses** predominantly on… - added **focus** on… - A growing body of literature/Preliminary evidence/Empirical evidence **suggests** that… - It is important to **address**… - … **addressing** the call for… - … would **provide** a statistically significant benefit for… - … **provide** a valuable perspective to…/a lens for **understanding**… - It is therefore worth **examining**… - we were seeking to unearth new ways to **examine** (the dynamic nature of)… - … best captured through **examination** of… - Such information can **provide** theoretical **understanding** of… - it is important to **understand**/gain a deeper **understanding**… - … are particularly salient for **understanding**… - According to this **understanding**, - (in order) to (better) **understand**… - Another **suggestion** for…
M1S2 Synopsis of Literature	- Prior **research** (seems to) **suggest**/has shown/highlights/indicates that… - **researchers** question/used/report that… - Other recent **research** on … - Several **research studies** have shown (that)… - Their **research** documents how… - Other/Some **researchers** have emphasised/investigated … - Previous/Preliminary **research** has explored/shown (that)… - Unlike previous **research**, - … that has been reported in previous **research** varies from **study** to **study**. - **Studies** on ___ initially explored… - Many (empirical) **studies** have **examined**/reported… - **studies** on … have (further) **examined** (the effects of)… - Existing **studies** of ____ show that… - previous **studies provided** a foundation for…

	- However, (recent) **studies** indicate that/often portray… - The few **studies** that **focus**/have **focused** (specifically) on ___ (have shown that…) - In a **study** by ___, - Overall, these **studies suggested** that… - Some/existing **studies found**/claim that/discuss/have **focused** on… - One **study** by ___ has looked at… - … has been explored in a number of **studies** before - What most of the **studies** have in common is… - Based on **findings** from some of these **studies**,- the authors **found** that… - These **findings** are encouraging; however, … - Further differences are **found** in… - … are consistent with the **findings** reported by … - recent meta-analysis/some others **found** that… - Similar **findings** were also **found** with… - based on the prior **findings** by… - The authors/Some/the results/recent evidence (have) **suggested** that… - As ___ **suggests**, - Other scholars have **focused** on… - Some/they (generally) **focus** heavily on… - … **provides** strong evidence of…
M1S3 Problems/ Controversy	- … a growing concern within education **research** - Although **researchers** tend to agree on… - …has attracted less attention from **researchers** - **researchers** argued that… - … remain under-**researched**, particularly in… - (Test developers) and **researchers** need to **understand**… - … due to the insufficient number of reviewed **studies**. - With the exception of a few **studies**, - Although one **study found** little evidence on… - This **finding** is contrary to… - … has been **found** to show mixed **findings** - This is an important **finding**, but… - However, these **findings** are not drawn from… - This issue is often **found** interwoven with… - In contrast, other **studies** have **found** little support for… - … several reasons for these seemingly contradictory **findings** - Such variability in **findings** led **researchers** to… - … are challenging factors **identified** by…
M2S1 Research gap	- there is comparatively little **research** on… - where further **research** is needed… - …contributed to the lack of **research**… - more **research** is needed to better **understand** - Yet **research** on this area of ___ remains sparse - (Surprisingly) no prior/(very) little **research** (to date) has evaluated/**addressed**/**examined**… - The **study** reported here **addresses** this **research** gap… - …have not been **researched** - …the scarcity of **research** on this topic. - To date, there is no **research** on… - to fill these **research** gaps by… - there has been a recent lag in **research focused** on… - Comparatively few **studies** exist regarding - no **studies** so far have been able to demonstrate… - there are currently no similar **studies** that… - These **studies** have not **examined**… - Few **studies**, however, have **examined**… - an empirical **examination** of ___ has not yet been conducted - gaps are **found** in… - … was not **addressed** - However, no … has **focused** solely on…
M2S2 Positive Justification	- In addition to reviewing **research** on… - … to provide suggestions for future **research** - further **research** must be conducted - proposal for new direction in **research** - will yield highly relevant information that can inform future **research** and practice

	- Following our review of prior **research**, - future **research** needs to look beyond… -… in which additional **research** is needed - identified as needing additional **research** - as evidenced by the plethora of **research** - This **study** (therefore) aimed to add to the existing literature/to fill this void by (**examining**)… - Drawing/expanding on the (designs of the) aforementioned/previous **studies**, - We argue that our **study focused** on… - … to update prior **findings**. - … may **provide** added insight into… - further clarification is needed to **examine**… - These limitations **suggest** the need for… - It is our responsibility to **identify**… - …we seek to attempt/illuminate ways to **address** (the perceived gap)… - … was conducted to **address** these issues.
M3S1 Purpose	- The **research** project compares… - the **focus** of the **research** reported here … - Drawing from the extant literature, the present **study focused** on… - The aims/purpose of the current **study** were/was to… - In the current **study**, we also **focused** on… - the **study** compares/explores/**focuses** on/investigated… - The **study's** aim is to inform… - In/for this **study**, we **focus** in particular/solely on… - Our purpose is to use this ___ **study** to… - … that were the **focus** of this/the current **study**. - For the purposes of this **study** ,- we build on this literature by **focusing** on… - Our **focus** on … is driven by… - In particular, we **focus** exclusively on…- We **focus** our review on… - We **focus** in on… - In the present work, we (will) **examine** (the impacts of)… - this article **examines**… - Our aim/first goal was to **examine**… - In addition/In the present **study**, we **examined** (the relationship between)… - (Specifically,) we (sought to) **examine**… - … with the goals of **identifying**… - The aims of this review were to **identify**… - This review **identified**… - This article seeks to **address** these issues…
M3S2 Research Questions/ Hypotheses	- Our analyses are driven by ___ primary **research** questions - With respect to the ___ **research** question, - The pressing **research** question at this point is this: - to answer the **research** question more comprehensively - the specific **research** questions we will be **addressing** are the following: - Accordingly, we **examine** the following **research** questions - The **research** questions that guide this **study**/was are as follows: - (To this end), this **study** (specifically) **addressed** the following **research** questions: - In the current **study**, we predicted that… - Finally, the **study** hypotheses are presented and explained. - The primary question **addressed** in this **study** is… - By **examining** these questions, - … **provides** support for this hypothesis… - The primary **focus** of these questions is …
M3S3 Methods	- This **study** draws on quantitative and qualitative **research** methods… - (To this end) the current **study** employed/uses a ____ **research** design - **researchers** have sought to test/proposed… - In the current **study**, we use… - Each of these **studies** used… - Drawing on data from a ____ **study**, - The **study** results showed that… - … is an important variable to investigate in the present **study** - Our **study** was designed to… - (In the present review), we **examine studies** using… - We **identified** (60) **studies** of… - We attempt to find answers to these queries by empirically **examining** (the impacts of)…

	- … have been **identified/** can be **addressed** using… - … with data collection generally **focused** on…
M3S4 **Outline of the Paper**	- we summarise **research** on… - In addition to reviewing **research** on… - This **research** reports on… - **research** has demonstrated that… - We conclude with implications for **research** and practice. - **researchers** looked for/are tasked with/**address** the concept of/have drawn on… - By and large, **research studies** calculate… - In this (current) **study**, we drew upon/further explore… - The current **study** was guided by… - The **study** also included… - This article reports on a/an ____ **study** where… - results of the **study** showed/highlight (that)… - Further, this **study** did not report evidence regarding… - Hence, the **findings** of this **study** will help… - this ___ **study provides** information … - using our **findings** as a foundation, - ___ **provide** a comprehensive framework/potential explanation for… - We **provide** evidence on/a description of/an overall, a wider picture of… - The table __ **provides** an overview of …