



A phraseological examination of research articles in the field of environment using key phrase frame

Elyas Barabadi ^a , Mohammad Ali Robotjazi ^{a*} , Mokarrameh Bayata ^a 

^a *University of Bojnord, Bojnord 9453155111, Iran*

Received 6 July 2019 | Received in revised form 25 November 2019 | Accepted 26 November 2019

APA Citation:

Barabadi, E., Robotjazi, M. A., & Bayat, M. (2020). T A phraseological examination of research articles in the field of environment using key phrase frame. *Eurasian Journal of Applied Linguistics*, 6(1), 81-100.

Doi: [10.32601/ejal.710217](https://doi.org/10.32601/ejal.710217)

Abstract

Given that formulaic sequences play an important role in enhancing L2 learners' academic writing skills and that phraseological sequences constitute a large proportion of written discourse, the current study was an attempt to draw up a list of key phrase frames— phraseological sequences with a variable slot— in a corpus of research articles (RAs) in the field of environment. A complete list of such frames, extracted by *kfNgram* software, was subjected to a series of corpus statistics such as frequency, range, and keyness, as well as manual filtering. This resulted in the generation of 65 four-word and 20 five-word key phrase frames specific to the field of environment. Structural analysis revealed that the vast majority of these key phrase frames were content word sequences, while functionally most of them belonged to referential expressions. Pedagogically speaking, L2 learners are recommended to develop a rich repertoire of formulaic sequences such as key phrase frames which are considered the cornerstone of academic discourse.

© 2020 EJAL & the Authors. Published by *Eurasian Journal of Applied Linguistics (EJAL)*. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Phrase frame; phraseology; discipline-specific; L2 writing pedagogy; environmental science

1. Introduction

Fixed or semi-fixed multi-word sequences compose many utterances in English language and have been investigated in corpus-based studies of phraseology concerning both ESP and EAP as well as General English for EFL learners to employ for communicative purposes. As in case, Altenberg (1998, in Garner, 2016: 31–68) studied the spoken data from the *London-Lund Corpus* (Svartvik, 1990) and estimated that “80% of the words in spoken language help form these sequences.” This has also been supported by psycholinguistic research works showing that multiword sequences are represented in the mental lexicons of native speakers as well as non-native speakers of English (Ellis, Simpson-Vlach, & Maynard, 2008; Jiang & Nekrasova, 2007). This fact signifies the acquisition of a productive knowledge of linguistic phraseology by L2 learners at different proficiency levels and the provision

* Corresponding author. Tel.: +98-915-184-2572

E-mail address: ma.robotjazzy@ub.ac.ir

<http://dx.doi.org/10.32601/ejal.710217>

of lists of academic formulaic expressions for them proves central. Based on this assumption, in the past, research has tried to identify continuous and discontinuous sequences in which frames surrounding a variable slot are formed, and this notion has been investigated under different combinations as Firth's (1957) notion of Collocation, Stubbs' (2002) category of Chains, Hunston's (2008) grouping of Semantic Sequences, Eeg-Olofsson and Altenberg's (1994) classification of Prefabricated Expressions, Ellis et al.'s (2008) premise of Formulaic Language, Simpson-Vlach and Ellis's (2010) notion of Academic Formulas, Cheng, Greaves, and Warren (2006) and Forchini and Murphy's (2008) N-Grams, and Biber's (2009), Lexical Bundles among which one key difference has been the methodology applied in identifying the lexical sequences.

Nevertheless, Biber argues that not all these approaches can be considered accurate; rather, differences in investigatory endeavors in phraseology in turn take different kinds of multi-word sequences (Biber, 2009). From all the methodologies adopted for investigating the sequences, one common concept, which is increasingly attaining its locus proper in L2 pedagogy and constitutes the formulaic expressions, is phrase-frames (PFs) that contain one or more variable slots. PFs, which were originally referred to as 'collocational frameworks' (Renouf & Sinclair, 1991) and 'lexical frames' (Gray & Biber, 2013) as the recurrent word sequences and fundamental units of meaning, are considered as one type of linguistic construction that can become a L2 learner's linguistic system and be represented in his mental lexicon. Thus, that PFs constitute the formulaic frames in an L2 learner's linguistic knowledge enhances the necessity that L2 teachers equip their students with a repertoire of multiple conventionalized modes of stringing words together through established sequence frames. Having recognized the significance of real linguistic and communicative use of PFs and the importance of providing, categorizing, and teaching of these frames, researchers in corpus linguistics have conducted research on phraseology and its valuable contribution to L2 acquisition and pedagogy and have tried to derive different types of lists from various academic corpora such as the '*Academic Formulas List*' (AFL, Simpson-Vlach & Ellis, 2010) and the '*Phrasal Expressions List*' (Martinez & Schmitt, 2012).

To add to the significance of PFs, Fletcher (2006, 2011) holds that the inclusion of these frames in the academic lists enriches academic formulaic expression appreciation both in writing and conversations acquisitively and pedagogically. These variants in a frame usually establish similar semantic or functional clusters such as in '*the * of the study* which can be filled by *purpose* and *goal* as well as *motivation* or *rationale*' (Lu et al. 2018) and they are more variable in writing than in conversation (Gray & Biber, 2013). These explanations provide us with the insight that teaching variable frames helps teachers present abundant language to learners while reducing the cognitive reliance on memory and helps learners understand the degree of formulaicity of varied linguistic constructions. Moreover, as Biber et al. (1999) and Hyland (2008) maintain, registers, genres, and disciplines vary linguistically, and determining how meaning is created in the subsets of language is suggestive of a specialized grammar and vocabulary (Römer, 2010). Thus, these valuable intuitions

help researchers in applied linguistics come to a better understanding of formulaic language in particular academic corpora and convince them that providing lists of formulaic expressions for specific genres assists teachers in satisfying the undertaking of genre pedagogy and help learners enjoy the real benefits of pulling together language, content, and contexts (Hyland, 2007). Therefore, provoked by these insights, we aimed to explore PFs in an EAP corpus of environmental RAs and derive a pedagogically beneficial list of PFs in texts related to that field. To this end, first a set of PFs was identified and filtered through methodological procedures involving *kwNgram* software and manual examination. Then, all PFs were structurally and functionally analyzed. Given the above explanations, the present study aimed at answering the following research questions:

1. What are the key PFs (both four and five-word sequences) of RAs in the field of environment?
2. What is the structural profile of these four and five-word PFs?
3. What is the functional profile of these four and five-word PFs?

2. Literature review

Results of many corpus-based researches have emphasized the significance of L2 learners' acquiring productive knowledge of language phraseology. Many researchers exploring this aspect of language learning have examined the fixed multi-word sequences practiced by different learners in academic settings and compared and contrasted their use with and against that employed by native-speakers.

Sinclair (2008a) asserts that "the normal primary carrier of meaning is the phrase and not the word; the word is the limiting case of the phrase and has no other status in the description of meaning." (p. 409) and continues that phrase is central and pivotal in linguistic analysis emphasizing the superiority of an approach that focuses on phrases rather than words in a vacuum (Sinclair, 2004). Sinclair (2008b) stresses that "one of the great strengths of a phraseological approach is the preservation of the integrity of text for much longer than alternative approaches to description, and in turn this entails the preservation of meaning" (p. xvii). In the same vein, Hunston (2002) draws on a theory of "language as phraseology" and emphasizes that an integrative and interdisciplinary approach be adopted for the phraseological studies (p. 137).

Phraseological research in different corpora, especially in English textual bodies, has shown that most utterances in English are comprised of fixed or semi-fixed multi-word sequences. Altenberg (1998), for example, analyzed the spoken data from the *London-Lund Corpus* (Svartvik, 1990) and came to this conclusion that most of the words in spoken language include phrasal combinations. Several other studies conducted on phrase-frames (Ellis et al., 2008; Jiang and Nekrasova, 2007; Siyanova-Chantura, and Martinez, 2015) have indicated that languages largely include fixed or semi-fixed multi-word sequences that are used in communication through spoken or written modalities.

In recent years, research methodology in corpus linguistics has given a way to different perspectives on how to investigate frequent multi-word units in a variety of text types or genres. Based on the frequency and distribution of PFs in various texts, the most frequently used sequences of words have been investigated by researchers in different ways and, therefore, touched upon in specialist literatures, e.g. ‘chunks’ (O’Keefe, McCarthy, & Carter, 2007), ‘PFs’ (Fletcher, 2002-2007), ‘formulaic frames’ (Biber, 2009), etc., of which Fletcher’s (2002-2007) PFs— as sets of variants of n-grams of varied length identical except for one word, phrase, or clause— have been acting as a theoretical framework for the description of phraseological patterns in different texts.

Gray and Biber (2013), in a paper entitled *Lexical Frames in Academic Prose and Conversation*, worked on lexical bundles comprised of discontinuous sequences where words form a frame surrounding a variable slot such as ‘*I don’t * to, it is * to*’ (p. 109). By using corpus-driven approach to identify the set of recurrent four-word continuous and discontinuous patterns in conversational and academic writing and direct computational analysis of the corpora, they revealed a more complete set of frames than alternative approaches, resulting in the documentation of highly frequent frames rarely identified in previous research. They also found that phraseology in English conversation is connected to lexical patterning, while phraseology in academic writing is more inherently linked to grammatical constructions, which testifies the bigger pervasiveness of patterned language in spoken language than in written language. However, as for the frames in academic writing, they are typically more subject to change than frames in conversation and are mainly comprised of function words.

Concerning the genre-bundle and field-dependency analysis, Cunningham (2017) examined a nearly 2,300,000-word corpus of 128 mathematics RAs collected from eight scholarly mathematical journals through ‘key phrase frames framework’ (pp. 71-2) and found out that while there are plenty of resources accessible for teaching research writing of outdated research papers, teachers enjoy little to refer in teaching graduate students in mathematics. Adopting the PFs theory, recurring multiword units with a variable slot or blank, Cunningham tried to establish the concept of key PFs mainly consisting of nouns, while coherence and cohesion of referential, logical and discourse types being emphasized as supporting the PFs assumption.

Continuing with the approach orientation trend, Schmitt (2013) drew on multiword phraseological units in a research study entitled ‘*Formulaic Language and Collocation*’. Herein, the old-run assumption that vocabulary typically does not take its meaning in a vacuum but in a context, held together by syntax and frames, with an emphasis on formulaic language in both spoken and written modes, was reestablished. In this study, it was emphasized that formulaic language is not merely a peripheral feature of language; rather, it is a permeating and indispensable feature of how that language is and shall be used and mastery and teaching of formulaic language prove vital in academic settings.

In a research work ‘*Is there a Formula for Formulaic Language?*’, to highlight the significance of formulaic language concept, Forsyth and Grabowski (2015) tried to

detect and measure the traces of formulaic language pattern variability by adopting the PF theoretical framework proposed by Fletcher (2002-2007). The degree to which a text type can incorporate inflexible sequences of words was the focus of the study. By focusing on modes of identifying the traces of formulaic language and enumerating its repetition in texts, they showed that formulaicity is an unformulated concept due to the disagreement among linguistics researchers and that one piece of corpora can be more formulaic than another.

To better appreciate the conceptualization of formulaicity and attribute it to phraseology and to signify the phraseological methodology in investigating the corpuses, Garner (2016), by adopting a PF approach, investigated phraseology in learner writing across proficiency levels. The quantitative and qualitative characteristics of PFs employed by L1 German learners of English at five different proficiency levels in the *EF-Cambridge Open Language Database* were investigated. The findings showed that PFs in texts grew less predictable, more functionally complex and inconsistent as proficiency level grew.

To add to the literature of phraseological studies, Grabowski (2015) conducted a study on the intra-disciplinary register variation of PFs in English pharmaceutical corpuses. This corpus-driven investigation tried to explain the use, structure and discourse roles of contiguous sequences of four word PFs as in '*be * with caution, to take * medicine*'. The analyses indicated 3 significant results: 1. the use and discourse functions of PFs showing variations in pharmaceutical texts, 2. the correlation between the frequency of PFs and their pattern variability depending on a register or genre, and 3. the discourse functions of PFs as distinct from those of their textual variants. The results indicated that intra-disciplinary register varies and the patterns of use, composition and discourse functions of the 50 most frequent PFs are not consistent across "patient information leaflets, summaries of product characteristics, clinical trial protocols and chapters/sections from academic textbooks on pharmacology" (p. 285).

Phraseology as a newly established discipline has also inclusively dealt with different corpuses with various non-academic genres and registers. In this turning point, Fuster-Marquez (2014) has taken a further step by empirically investigating the lexical bundles and PFs in the language of hotel websites. This analysis proved that hotels and similar systems enjoy highly specialized discourses which in turn require specific investigative attempts. In this study, Marquez investigated a British hotel website corpus containing 242,000 words from a database and found that bundles and PFs follow comparable codes but vary in tractability. In addition, PFs indicate a close association with Sinclair's (1991) 'idiom and open-choice principle' (pp. 84-5, 105-6) which does not limit the bundles but allows different insertions. In the same vein, Biber et al. (2004) assert that registers differ both in quality and quantity with incomparable lexical bundles. Hotel websites, for example, have highly distinguished phraseological references different from other registers or text types. Since these websites follow a motivational register in linguistic choice to influence customers, the word sequential choice enjoys both a referential and evaluative

function with a focus on motivational choice of phrases, words and sequences like *quite convenient, perfectly suited for*, etc. (pp. 105-6).

‘*Academic Writing Revisited: A Phraseological Analysis of Applied Linguistics High-Stake Genres from the Perspective of Lexical Bundles*’ is a study conducted by Jalali and Zarei (2016). In their work, the academic genre of Applied Linguistics RAs used by two master-level and PhD EFL scholars as beginner discourse members in the same discipline were investigated. The findings indicated that, in many cases, postgraduate students were able to use target bundles as the writers of published articles did with little difference.

Römer (2009) in ‘*The inseparability of lexis and grammar*’ studied the crossing point of lexis and grammar and offered corpus proofs for the inseparability of these two areas conventionally separated in language pedagogy and linguistic studies. The study touched upon the influential research strands and model-building strategies in “pattern grammar and collostructural analysis, lexical-grammatical pattern continuum, the way meaning is generated, and the expletive subject *it* pattern in expert and apprentice academic writings” (pp. 1, 158-9). The findings of the study indicated that research strands and theories applied for phraseological studies are not diverse enough but the interconnectivity of lexis and grammar and the close relationship between constructions and meanings are emphatically held up.

In his work, ‘*Establishing the phraseological profile of a text type*’, Römer (2010) reestablished that meaning does not primarily reside in individual words but in a phrase. He examined recurring phrases in language trying to convince corpus researchers that a profile of the phraseological sequences in a text or text collection can be developed. By applying his model (the Phraseological Profile Model–PPM) to a 3.5-million-word corpus of online academic book reviews, he found that the PPM can substantially facilitate the analysis of the occurrence and distribution of the phraseological strands and sequential boundaries “in linguistic book reviews, and how it helps to determine the extent of the phraseological tendency of language” (pp. 95-8).

Working on text- and genre-specific concept of phrases and sequential wordings, Win and Masada (2015) explored technical PFs from research paper titles of graduate students and proposed a method for the researchers’ exploring the technical frames through word n-grams extraction. They investigated the trigrams in the titles through applying mathematical procedures like PageRank algorithm to the co-occurrence graph and came up with weighted and unweighted trigrams as “technical keyphrases” at the higher ranks. They concluded that “while the top ranked trigrams obtained by unweighted PageRank have just a self-contained meaning, those obtained by their method are technical PFs, i.e., a word sequence that forms a complete technical phrase only after putting a technical word (or words) before or/and after it” (pp. 558-9, 563).

3. Method

3.1. Corpus

Having consulted the Environment Category of the Journal Citation Reports of 2018, we developed the corpus of the study from 125 RAs consisting of all sections except for references and appendixes from five 2014–2018 high-ranking Environmental Science journals with impact factors between 28.92 and 8.3. This corpus included 862343 words from five articles. Moreover, to come up with a fairly homogenous corpus, one additional criterion was taken into account: only RAs having Introduction, Methods, Results, Discussion, and Conclusion were selected.

Table 1 The Journals Used in the Corpus

Journal	No. of texts	Word count
Energy and Environmental Science	25	219887
Nature climate change	25	158193
Global change biology	25	139908
Environmental health perspectives	25	242076
Frontiers in ecology and environment	25	94156

3.2. Procedure

A five-step procedure was employed in order to identify PFs specific to Environmental Science. To identify them, a complete list of all PFs was prepared inductively (bottom-up approach compared to bundle-to-frame approach) using the kfNgram software (Fletcher, 2011). Next, in order to reduce the data and identify the most relevant key PFs, we used frequency and range cut-offs, with the data being manually filtered by two experts in Environmental Science and Applied Linguistics. Then, by using the software, the measure of keyness, which is based on the comparison of the frequency of PFs in this corpus to a reference group corpus, was applied to identify key PFs only from the initial candidates. Finally, they were examined in terms of structure and function.

3.2.1. Identification of PFs

Having compiled the corpus, all possible PFs with different lengths were extracted. The list included n-grams of size four and five, and following the previous research (Grabowski, 2015; Fuster-Markuez & Pennock-Speck, 2015; Römer, 2010), we initially extracted n-grams of size four. Additionally, n-grams of size five were included for two reasons: first, longer n-grams make it possible to extract more semantically complete PFs (Lu et al., 2018); second, n-grams of size five allow us to identify those PFs concerning discipline, genre, and register-specific (Cunningham, 2017; Römer, 2009). With the size of n-grams decided, all the variants of each p-frame were obtained through the software. The generated n-grams were used to identify PFs by combining those identical n-grams except for the one variable slot. For example, the n-grams *at the watershed scale* and *at the global scale* are used to create the frame *at the * scale* with a frequency of 71.

3.2.2. Manual filtering

The initial extracted candidate PFs went under careful manual examination. First, all PFs containing proper names, punctuation marks, digits, and other symbols were discarded. Following Lu et al. (2018), we excluded all linguistically incomplete PFs (**surface area of the* with fillers such as *the*, *specific*, and *protected*) or did not enjoy clausal boundaries (e.g. *the adsorbent * the* with fillers such as *and*, *however*, and *from*). To achieve these aims, we carried out a significant amount of concordance analysis by examining each p-frame in its original textual context. Then, in the case of four-word PFs, only sequences with internal variable slot were retained, and those with a variable slot at the beginning or end of the frame were excluded because such frames could be subsumed under five-word PFs. As for the five-word PFs, sequences with a variable slot in any position were considered as potential candidates. However, it should be noted that many of the five-word PFs with a variable slot at the beginning or end of the frame were removed because either they were linguistically incomplete, or they crossed clausal boundaries. Another filtering criterion for retaining a p-frame was having at least three variants. For example, the p-frame *food security * biodiversity* with only two variants (e.g. *and* and *or*) was removed because such frames are too fixed to be considered a true p-frame. Finally, two experts in Environmental Science and Applied Linguistics were asked to decide whether a particular p-frame was specifically related to the field of environment. However, this phase was conducted after applying the keyness measure. For example, although the five-word sequence *to enhance the * of* (fillers: *performance*, *development*, *removal*, and *rate*) could successfully be included as a key p-frame, it was rejected as a p-frame irrelevant to environment by both experts.

3.2.3. Frequency and range criteria

Two additional methodological considerations for reducing the data and hence identifying key PFs were the frequency cut-off points of p-frame and the number of texts (e.g. journals) in which a p-frame should appear (e.g. range) to be determined. Regarding frequency, the raw count of each p-frame was normed to one million words. Then, only PFs that met the cut-off point of 20 counts per million words (i.e. 17 in our corpus) were retained and the rest were removed. As for the range, previous researchers (Cunningham, 2017; Golparvar & Barabadi, 2020) held that a p-frame should appear at least in 75% of the texts constituting a corpus, but as Lu et al. (2018) rightly argued determining range threshold is an arbitrary decision which is contingently made by the researchers. Given all these considerations, any p-frame had to appear in three texts out of five texts (e.g. five journals) that comprised the corpus in this study. We used Antconc 3.2.4m (Anthony, 2014) to search the five files related to the five environmental journals separately and to examine the concordance plots.

3.2.4. Key PFs identification

Identification of key PFs distinctive of environment was the last step taken to reduce the data. To achieve this aim, the normalized frequency of each p-frame in our

study was compared to its counterpart in the British Academic Written English Corpus (BAWE) as the reference. Following Cunningham (2017), we excluded those more frequent PFs in the reference corpus. To carry out the comparison, we applied sMAPE formulation involving “the difference of the two values over the average of the two values” (Cunningham, 2017: 73). Simply put, this formula involves $(\text{Environment} - \text{BAWE}) / (\text{Environment} + \text{BAWE})/2$. sMAPE has a maximum value of 2, given that the frequency of a particular frame in BAWE is not equal to zero. In the present study, sMAPE score of 1.95 and above was used to identify those frames ‘having occurred 100 times more frequently in the corpus of the current study than in BAWE. Finally, the resulting frames meeting the previous criteria were subjected to Fisher’s exact test to gain clear assurance that a particular frame presence in the two corpora is significantly different. The results of Fisher’s test indicated that all key PFs meeting the sMAPE threshold score of 1.95 or above could do well on Fisher’s exact test.

3.2.5. Structural and functional analysis of key PFs

Those PFs selected as ultimate candidates were further structurally and functionally analyzed based on Gray and Biber’s (2013) structural classification and Simpson-Vlach and Ellis’ (2010) functional classification respectively (referential, stance, and discourse organizing expressions). Gray and Biber’s (2013) structural taxonomy classifies all entries (e.g. p-frames) into three main categories: (a) verb-based frames with a verb as part of the frame like *evaluated the * of*; (b) other-content-word frames with one or more content words other than verb like *the surface * of the*; and (c) function-word frames with only function words like *and * of hgs*. It is worth noting that a specific p-frame can have different functions in a text depending on the type of words appearing in the variable slot. Nonetheless, almost all the realizations of PFs in this study performed the same function. The discourse functions of PFs were determined by examining the concordance line for each frame, enabling us to understand the use of each p-frame in context.

4. Results

4.1. The high frequency PFs

To be considered as a key p-frame and hence be included in the analysis, a frame has to meet several criteria: 1. it should occur 20 times per million in the corpus, 2. it should appear in at least 3 of the 5 journals which comprised the corpus in this study, 3. the frame should obtain an sMAPE value of 1.95 or above, and 4. $p < 0.0001$ on Fisher’s exact test. Overall, 85 PFs fitted these criteria. To be more specific, 65 were four-word PFs, and the rest were five-word PFs. In most cases, the variable slots were in the medial, while in few cases PFs with variable slots at the initial and final positions in the case of five-word PFs were included in contrast to previous studies (see Golparvar & Barabadi, 2020; Cunningham, 2017). The reason for the inclusion of PFs with variable slot in the initial and final position was that some key PFs specific

to the field of environment would have been missed out if we had included PFs with only medial variable slot. Table 2 presents 10 four-word PFs:

Table 2 List of 10 Four-word PFs Along with Their Selected Fillers

P-frame	Frequency	Fillers and their frequencies
The * of nitrate	84	Reduction (24), removal (18), conversion (6), use (6), risk (4)
For * removal of	81	The (79), effective (1), cost-efficient (1)
The * of ZNO	75	Thickness (8), surface (7), morphology (5), doping (5), transmittance (4)
At the * scale	71	Watershed (22), global (8), landscape (4), lab (3), pixel (3)
The * surface area	56	Specific (13), large (7), biochar (6), high (3), total (3)
To * climate change	18	Modest (4), mitigate (3), near-term (2), combat (2), predict (1)
Desorption of * from	17	Contaminants (6), hms (2), phenanthrene (2), pahs (2), adsorbate (1)
The * of deforestation	17	Location (3), prevalence (2), intensity (1), quantity (1), perspective (1)
The * of pathogens	17	Number (4), concentration (3), mobilization (2), quantity (1), transport (1)
Of * hydropower plants	17	Small (10), multiple (3), large (2), the (1), diversion (1)

As shown in Table 2, *the * of nitrate* is the most frequent four-word p-frame in the corpus with 84 occurrences. The word *reduction* is the most frequent filler of this p-frame with 24 occurrences. That is, the sequence *the reduction of nitrate* marks the most frequent appearance of this p-frame. Other realizations of this p-frame include *the removal of nitrate*, *the conversion of nitrate*, *the risk of nitrate*, etc. *The * of deforestation* is another four-word phraseological sequence which is filled by words such as *location*, *prevalence*, *intensity*, *quantity*, and *perspective*. As can be observed in Table 2, these four-word PFs deal with a range of issues in the field of environment such as chemicals (*the reduction/removal/conversion/use of nitrate – the presence/addition/mineralization/ effect of citric- the conversion/ modification/ solubility/ identification of chitosan*), chemical processes (*desorption of contaminants/ hms/ phenanthrene/ pahs/ adsorbate from*), environmental phenomena (*the location/ prevalence/ intensity/ quantity/ perspective of deforestation – to mitigate/ combat/predict/absorb climate change*), and bacteria and viruses (*the number/ concentration/ mobilization/ quantity/ transport of pathogens*). Similar issues were also raised in five-word PFs. Table 3 presents 10 five-word PFs:

Table 3. List of 10 Five-word PFs Along with Their Selected Fillers

P-frame	Frequency	Fillers and their frequencies
*For the removal of	79	Absorbents (6), chitosan (5), disinfection (4), method (4), pathway (2)
*on the surface of	51	Absorbed (5), coating (4), deposited (2), obtained (1) loaded (1)
The * capacity of the	37	Absorption (17), nitrate (2), discharge (2), buffering (2), sink (1)
The effect of * on	35	Temperature (9), rhamnolipids (4), OAs (2), warming (2), biogas (2)
The surface * of the	32	Area (21), properties (2), functionality (1), energy (1), morphology (1)
*of the hyporheic zone	32	Degradation (4), impact (4), size (2), location (2), volume (2)
The * of the absorbent	32	Cost (4), preparation (4), pore (3), activation (3), characteristics (2), basicity (2)
The * properties of the	30	Electrical (4), mechanical (3), physicochemical (3), sensory (2), surface (2)
The * of the sludge	29	Activation (9), carbonization (4), conversion (2), disposal (2), preparation (1)
*in the great lakes	21	Production (3), periphyton (2), concentration (2), scale (1), value (1)

As can be observed in Table 3, the p-frame *absorbents for the removal of* occurs the most among the five-word sequences. Other realizations of this p-frame are *chitosan for the removal of*, *disinfection for the removal of*, *method for the removal of*, and *pathway for the removal of*. The sequence *the * capacity of the* is another frequent five-word p-frame whose fillers include words such as *absorption*, *nitrate*, *discharge*, *buffering*, and *sink*. A good example of five-word p-frame with variable slot at the final position is the sequence *the concentration of the ** which is filled by nouns such as *stimulus*, *precursor*, *surfactant*, and *biomass*. A detailed examination of five-word PFs indicated that a range of topics such as natural places (*Degradation/ impact/ size/ location of the hyporheic zone*), causal and effect relationship especially related to environmental phenomena (*the effect of temperature/ rhamnolipids/ OAs/ warming/ biogas on*), and chemicals (*the cost/ preparation/ pore/ activation/ characteristics/ basicity of the absorbent*). A close examination of the range of topics dealt with in all PFs in this study indicates that a very specialized domain of knowledge which is highly discipline-specific is emphasized. This discipline-specific phraseological sequence can be helpful for academics involved in this particular community of practice.

4.2. Structural analysis

Establishing the structure of PFs identified in the previous sections involved analyzing the type of words constituting each p-frame. As demonstrated in Table 4, the most dominant and frequent type of p-frame in terms of structure is related to content words in both four-word and five-word PFs with 91% and 80% respectively. The next dominant category is related to verb-based phraseological items comprising

7.7% of four-word PFs and 20% of five-word PFs. Finally, the least frequent type of phraseological items was related to function word frames.

Table 4 The Structural Categories of PFs

Length	Verb-based frames		Other content word frames		Function word frames	
4-word PFs	5	7.7%	59	91%	1	1.53%
5-word PFs	4	20%	16	80%	-	-

The following section presents one example of 4 and 5-word PFs for each structural category. However, in the case of five-word sequences, there is no instance of function-word p-frame.

A. Verb-based PFs

Evaluated the **performance/effect/feasibility** of (4-word p-frame)

Attributed to the **formation/production/presence** of (5-word p-frame)

B. Other content word PFs

The **maximum/highest/increased/areal** current density (4-word p-frame)

The surface **area/properties/functionality** of the

C. Function word PFs

And **dissolution/quantification/methylation** of hgs (4-word p-frame)

4.3. Functional analysis

Table 5 reports the distribution of four-word and five-word PFs by prominent function as proposed by Simpson-Vlach and Ellis (2010). For four-word PFs, referential PFs constitute the largest category (89.23%, n = 58) while discourse organizing frames come next (10.76%, n = 7). Interestingly, there was no instance of stance frames. For five-word PFs, similarly, referential frames constitute the largest category (80%, n=16), followed by an equal number of stance frames (10%, n = 2) and discourse organizing frames (10%, n = 2).

Table 5 Distribution of PFs by Their Prominent Function

Length	Referential	Stance	Discourse	Total
Four-word	58 (89.23%)	0	7 (10.76%)	65
Five-word	16 (80%)	2 (10%)	2 (10%)	20
All	74 (87%)	2 (2.35%)	9 (10.58%)	85

In the following section, we provide some examples of PFs with different prominent functions. Moreover, in order to provide more adequate explanation of functional aspect of the PFs, we tried to present specific examples for the subcategories of the three primary functions. As mentioned earlier, referential frames account for the largest proportion of PFs of articles in Environmental Science. For example, the phraseological sequence *the * of deforestation* is a referential frame whose fillers (e.g.

location, prevalence, intensity, and quantity) deal with different attributes or aspects of this environmental phenomenon. Example 1 specifies one prominent attribute of *deforestation*. As such, this frame belongs to the specification of attributes which is a subcategory of referential expression. Example 2 also refers to the same functional subcategory among five-word sequences. In this example, the frame specifies one major attribute of *absorbent*.

1. ...distinction between the definition of a spatial deforestation potential and the estimation of the quantity of deforestation...
2. The chemical treatment involves the activation of the absorbent in the presence

Deictive and locative frames were the second largest sub-category of referential PFs among both four-word and five-word sequences. These frames refer to “physical locations in the environment (e.g. the real world) or to temporal or spatial reference points in the discourse (e.g. a and b, at this point)” (Simpson-Vlach and Ellis, 2010: 505). Given that our corpus is comprised of articles in the field of environment, deictives and locatives are quite expected. For instance, the phraseological sequence *in the * soil* is a four-word PFs whose different fillers (e.g. rhizosphere, subsurface, bulk, and forest) refer to very specific locations of the soil in the environment. Example 3 demonstrates how this referential sequence makes it possible to know about the exact location of “organic rich material” which is in subsurface soil.

3. ...hot spots of N removal/retention are found especially at the interface between permeable and fine organic rich material *in the subsurface soil*...

The sequence ** in the great lakes* is five-word p-frame with a deictive and locative function. The slot is filled by words such as *production, periphyton, concentration, health and values*.

4. These maps incorporate foundational research addressing the relationships between water quality and fish *health in the Great Lakes*

Identification and focus was another subcategory of referential expressions appearing in our corpus. According to Simpson-Vlach and Ellis (2010), this functional category is prevalent in academic discourse because of the key role that exemplification and identification play in academic speech and writing. Yet, there were only three instances of this functional category in our corpus. Example 5 reflects a typical expository phrase (e.g. as a robust electrode) whose function is to identify the focus of a previous noun (e.g. The PDMS-CNT).

5. The PDMS-CNT was used to make the device flexible and also serve *as a robust electrode* on the base of the device.

Stance expressions are the second major functional category examined in this study. This functional category is concerned with one’s attitude and perspective toward an idea or proposition in a text. There were only two instances of this functional category among five-word PFs. The p-frame *attributed to the * of* is a stance expression with an epistemic stance sub-function which is concerned with knowledge claims and demonstration. Example 6 demonstrates how this p-frame mitigates the

knowledge claim concerning the cause of deviation between predicted and experimental data:

6. The small deviation between predicted and experimental data was **attributed to the differences of** the blade width and chord angle between the model and the actual structure of the impeller.

The third primary functional category investigated in this study is related to *discourse organizing expressions* with topic elaboration as the only subcategory emerged in the data. This subcategory involves phrases that aim at explaining and elaborating an already introduced topic either in cause and effect relationship or non-causal relationship. For instance, the p-frame *evaluated the * of*, having fillers such as *performance, effect, feasibility, and capabilities*, is a discourse organizing expression with a topic elaboration sub-function. Example 7 is a non-causal elaboration p-frame which elaborates a topic previously introduced:

7. In this sense, Zhu and Feng (2003) **evaluated the capabilities of** mixed anionic/nonionic surfactants in enhancing the water solubility of PAHs.

*The effect of * on* is another *discourse organizing expression* with a topic elaboration sub-function which explicates the cause and effect relationship. This five-word p-frame has fillers such as *temperature, rhamnolipids, warming* and *biogass*. Overall, nine PFs in the corpus had a discourse organizing function all having a topic elaboration sub-function.

8. **The effect of temperature on** foliar emissions has been extensively explored, but the effect on floral emissions has not, so...

5. Discussions and conclusions

This study aimed at extracting key PFs from a corpus of RAs in Environment Science. The impetus for conducting this research came from recent studies (Cunningham, 2017; Golparvar & Barabadi, 2020; Lu et al., 2018) which demonstrated the variation in p-frame use across a variety of disciplines, genres, and registers. Following Lu et al. (2018), we opted for a fully inductive approach to identify PFs by drawing on all continuous lexical sequences, instead of focusing only on lexical bundles. The use of a set of statistical techniques especially the keyness formula as well as manual analysis yielded 65 four-word key PFs and 20 five-word key PFs. These PFs were then analyzed structurally based on Gray and Biber's (2013) structural taxonomy and functionally based on Simpson-Vlach and Ellis' (2010) functional classification.

The results of this study indicated that the vast majority of both four-word and five-word PFs were related to content word frames especially nouns in their core such as *the * of the sludge* with fillers like *activation, preparation, pore* and *cost*. This facet of results is supported by findings of Lu et al. (2018) and Golparvar and Barabadi (2020) in which content word phraseological frames were more dominant in their corpora than verb-based and functional based frames. However, this finding is not supported by Cunningham's (2017) study in which majority of PFs identified in his corpus of

mathematics RAs were verb-based. This disparity in findings of Cunningham's (2017) and other studies (Golparvar and Barabadi, 2020; Lu et al., 2018) including the present study can be attributed to the fact that "...even within the highly controlled genre of published research articles, mathematics seems different" (Cunningham, 2017: 71). The extensive use of nouns in PFs can be an indication that academic writing is more compressed structurally (Biber & Gray, 2010). In fact, the heavy use of nouns in PFs is likely to make the academic texts more challenging and hence more difficult to comprehend (Fang, Schleppegren, & Cox, 2006).

Besides, the prevalence of content word PFs as found in this study does not support Gray and Biber's (2013) findings where there was a heavy reliance on function-word frames in the corpus of academic books and RAs. This inconsistency can be attributed to the use of keyness measure for identifying key PFs in the current study. The initial candidate PFs extracted through kfNgram resulted in many phraseological sequences that were function-word or verb-based frames. However, after employing the keyness measure, many of these function-word and verb-based frames were left out since such frames did not meet the keyness criterion, and hence they were not identified as key PFs distinctive of environment texts. It seems that when the keyness criterion is taken into account, nominal sequences have a higher chance of being included as a key p-frame at least in the corpus of this study.

As for the functions of PFs, the results of the study indicated that referential frames were the most prevalent and stance frames were the least among both four and five-word phraseological sequences. Particularly, specification of attributes was the most common sub-category of referential frames. Such phraseological sequences aim at framing both concrete and abstract entities. For example, the p-frame *the * of deforestation* with fillers such as *location, prevalence, intensity, and quantity* deal with different attributes of deforestation. Simpson-Vlach and Ellis (2010) state that specification of attributes is the largest pragmatic sub-category for all academic phrases both in written and spoken genres. This finding is consistent with those of previous research on the discourse functions of phraseological sequences in academic writing (Chen & Baker, 2010; Golparvar and Barabadi, 2020; Le et al., 2018). Another dominant sub-category of referential expression in our corpus was related to deictives and locatives. These frames which refer to physical locations in the environment were the second most frequent among both four and five-word sequences. The prevalence of PFs with this sub-function can be attributed to the specialized content and nature of our corpus which is about environmental issues. It is not surprising that in such a corpus there are references to different physical locations in the environment (e.g. the p-frame ** in the great lakes* with fillers such as *production, periphyton, and concentration*).

Another primary function of PFs is related to stance expressions that make up only 2.35% of phraseological sequences in the current study. While stance PFs make up the smallest category in this study, they make up the second largest category in both Lu et al.'s (2018) corpus (i.e. 17.8%) and Golparvar and Barabadi's (2020) corpus (i.e. 18.1%). This discrepancy between our results and those of Lu et al. (2018) and

Golparvar and Barabadi (2020) can be linked to the different content that makes up these corpora. While our corpus consisted of RAs in Environmental Science (e.g. natural and hard sciences), Lu et al.'s (2018) and Golparvar and Barabadi's (2020) corpora were made up of RAs from social sciences. This discrepancy in the use of stance expressions suggests that stance choices are motivated by disciplinary practices (Hyland & Jiang, 2016). It seems that natural and hard sciences (e.g. environment, biology, electrical engineering) tend to make limited use of stance expressions to create an "author-evacuated" and objective prose. This kind of prose is based on studies which are more empirically grounded and quantitative. In contrast, humanity and social sciences tend to make more extensive use of stance expressions to "...signal their attitudes toward the content of their talk" (Hyland & Jiang, 2016: 3). Thus, it can be argued that writers of RAs in social sciences tend to use stance expressions more extensively to construct a dialogue with their readers and to "...balance claims for significance, originality, and plausibility of their work" (p. 4). On the other hand, writers of natural sciences like Environmental Science make very limited use of stance expressions because they see their work as objective and empirically-grounded and hence feeling no need to persuade their readers of their claims by using stance expressions.

The phraseological patterns demonstrated in this study along with the disciplinary insights obtained from these patterns can help EAP instructors and material developers to make more informed decisions when trying to incorporate the phraseology of this particular discipline into ESP/EAP materials (Cunningham, 2017; Gilmore & Miller, 2018). Specifically, a list of discipline-specific frames can be ideal for teaching ESP courses since such a list can help students arrive at a more accurate understanding of linguistic choices within a specific discipline (Cunningham, 2017). Similarly, Lu et al. (2018) have suggested that such phraseological lists can be employed to enhance EAP students' genre competence by assisting them to notice and analyze the linguistic features (e.g. PFs) that characterize RAs. Given these suggestions, we expect that the list of key PFs in this study will help novice researchers in Environmental Science write academic papers in line with the norms and expectations of their particular community of practice by using this list as a useful reference.

Having derived a list of PFs from a corpus of RAs introductions in six social sciences, Lu et al. (2018) argued that such discipline-specific list of PFs can be pedagogically useful not only to researchers and instructors but also to student writers. Previous research (Alhassan & Wood, 2015; Ding, 2007; Jones & Haywood, 2004) has demonstrated that formulaic sequences can play an important role in enhancing L2 learners' academic writing skills since such phraseological sequences constitute a large proportion of written discourse. Corpus-based evidence suggests that academic writing abounds with formulaic sequences (Bestgen, 2017; Boers & Lindstromberg, 2012; Ellis, Simpson-Vlach & Maynard, 2008). Such a view of academic discourse suggests that mastery of academic writing goes beyond learning only lexicon and syntax. L2 learners need to develop a rich repertoire of formulaic

sequences such as key PFs which are considered the cornerstone of academic discourse (Biber, 2006; Lewis, 2000; Martinez & Schmitt, 2012). In case of PFs, it can be argued that they can enhance learners' writing proficiency because they function like frames on which L2 learners may draw when embarking on a writing task. Li and Schmitt (2009) have suggested that one major reason why written texts produced by L2 writers seem unnatural and weird is that such texts lack formulaicity. Indeed, for L2 learners, having a small collection of formulaic sequences poses a serious obstacle to developing a high level of writing proficiency because they are likely to overuse this limited repertoire, and hence their writing seems unnatural and strange (Bestgen, 2017). However, it should be noted that mere exposure to formulaic sequences such as PFs may not guarantee their use in writing; L2 learners need explicit instruction to acquire such phraseological sequences, and hence integrate them in their writing (Alhassan & Wood, 2015; Bestgen, 2017). As such, EAP instructors should be encouraged to help their students learn and integrate a large number of formulaic sequences such as key PFs because successful communication in a discipline is to a large extent dependent on mastery of formulaic sequences of that particular discourse (Jones & Haywood, 2004). Thus, our final suggestion is that facilitating L2 learners' development of writing competence requires knowledge of formulaic sequences such as key PFs as it needs knowledge of single-words (Bestgen, 2017).

In the present study, the use of keyness criterion led to a significant drop in PFs selected in the final list. Future studies can drop this criterion to allow for the inclusion of more PFs. Our concern is particularly that the use of a mathematical formula like keyness might lead to the exclusion of some useful PFs. Other researchers can simply use the output of *keyness* software and then a panel of experts can decide whether a specific p-frame belongs to a particular discipline or not. Furthermore, future researchers can investigate specific PFs in different sections of RAs separately. Given that each section of RAs has its own moves and steps, identifying PFs of specific sections of RAs can be insightful both for EAP instructors and students. Finally, future research can examine the pedagogical usefulness of phraseological list in this study in actual writing classes.

References

- AlHassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes*, 17, 51-62.
- Anthony, L. (2014). *Antconc 3.2.4m [computer software]*. Tokyo, Japan: Waseda University.
- Bestgen, Y. (2017). Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence. *System*, 69, 65-78.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied linguistics*, 25(3), 371-405.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14 (3), 275-311.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2–20.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, 32, 83-110.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30–49.
- Cheng, W., Greaves, C., & Warren, M. (2006). From n-gram to skipgram to conigram. *International Journal of Corpus Linguistics*, 11(4), 411-43.
- Cunningham, J. K. (2017). A phraseological exploration of recent mathematics research articles through key phrase frames. *Journal of English for Academic Purposes*, 25, 71-83.
- Ding, Y. (2007). Text memorization and imitation: the practices of successful Chinese learners of English. *System*, 35(2), 271-280.
- Eeg-Olofsson, M., & Altenberg, B. (1994). Discontinuous recurrent word combinations in the London-Lund Corpus. In U. Fries, G. Tottie and P. Schneider (Eds.). *Creating and using English language corpora: Papers from the fourteenth international conference on English language research on computerized corpora Zurich 1993* (pp. 63-77). Amsterdam: Rodopi.
- Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375-396.
- Fang, Z., Schleppegrell, M., & Cox, B. (2006). Understanding the language demands of schooling: nouns in academic registers. *Journal of Literacy Research*, 38(3), 247–273.
- Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis* (pp. 1-32). Oxford: Blackwell.
- Fletcher, W. H. (2006). *Phrases in English*. Retrieved from <http://phrasesinenglish.org/>.
- Fletcher, W. H. (2002–2007). *KfNgram*. Annapolis: USNA. Retrieved from <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html>.
- Fletcher, W. H. (2011). *KfNgram*. Annapolis, MD: USNA.
- Forchini, P., & Murphy, A. (2008). N-grams in comparable specialized corpora: Perspectives on phraseology, translation, and pedagogy. *International Journal of Corpus Linguistics*, 13(3), 351-367.
- Forsyth, R. S., & Grabowski, L. (2015). Is there a formula for formulaic language? *Poznan Studies in Contemporary Linguistics*. Retrieved from <https://www.researchgate.net/publication/287157710>.
- Fuster-Márquez, M. (2014). Lexical bundles and phrase frames in the language of hotel websites. *English Text Construction*, 7(1), 84-121.
- Fuster-Márquez, M., & Pennock-Speck, B. (2015). Target frames in British hotel websites. *International Journal of English Studies*, 15(1), 51-69.
- Garner, R. J. (2016). A phrase-frame approach to investigating phraseology in learner writing across proficiency levels. *International Journal of Learner Corpus Research* 2(1), 31–68. <https://doi.org/10.1075/ijlcr.2.1.02garissn> 2215–1478/
- Gilmore, A., & Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes*, 51, 1-17.

- Golparvar, S. E., & Barabadi, E. (2020). Key phrase frames in the discussion section of research articles of higher education. *Lingua*, 102804.
- Grabowski, L. (2015). Phrase Frames in English Pharmaceutical Discourse: a corpus-driven study of interdisciplinary register variation. *Research in Language*, 13(3). <https://doi.org/10.1515/rela-2015-0025>
- Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, 18, 109-136.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. (2008). Starting with the small words: Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics*, 13(3), 271-295.
- Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16(3), 148-164.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Hyland, K., & Jiang, F. (2016). Change of attitude? A diachronic study of stance. *Written Communication*, 33(3), 251-274.
- Jalali, H., & Zarei, G. R. (2016). Academic writing revisited: a phraseological analysis of applied linguistics high-stake genres from the perspective of lexical bundles. *The Journal of Teaching Language Skills (JTLS)* 7(4), 87-114. ISSN: 2008-8191.
- Jiang, N., & Nekrasova, T. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91(3), 433-445. <https://doi.org/10.1111/j.1540-4781.2007.00589x>
- Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: an exploratory study in an EAP context. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 269-300). Amsterdam: John Benjamins.
- Lewis, M. (2000). Materials and resources for teaching collocation. In M. Lewis (Ed.), *Teaching collocations: Further developments in the lexical approach* (pp.186-204). Boston, MI: Heinle.
- 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.
- Lu, X., Yoon, J., & Kisselev, O. (2018). A phrase-frame list for social science research article introductions. *Journal of English for Academic Purposes*. <https://doi.org/10.1016/j.jeap.2018.09.004>
- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33, 299-320.
- O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom. Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- Renouf, A., & Sinclair, J. McH. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds.). *English Corpus Linguistics* (pp. 128-143). London: Longman.
- Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, 7, 141-163.
- Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Construction*, 3(1), 95-119.
- Schmitt, N. (2013). Formulaic Language and Collocation. *The Encyclopedia of Applied Linguistics* (pp. 1-7), Edited by Carol, A. Chapelle: Blackwell Publishing Ltd. <https://doi.org/10.1002/9781405198431.wbeal0433>
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied linguistics*, 31, 487-512.
- Sinclair, J. McH. (1991). *Corpus concordance collocation*. Oxford: Oxford University Press.

- Sinclair, J. McH. (2004). *Trust the Text. Language, corpus and discourse*. London: Routledge.
- Sinclair, J. McH. (2008a). The phrase, the whole phrase, and nothing but the phrase. In S. Granger and F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 407–410). Amsterdam: John Benjamins.
- Sinclair, J. McH. (2008b). Preface. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. xv–xviii). Amsterdam: John Benjamins.
- Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics* 36(5), 549–569.
- Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics*, 7(2), 215-24.
- Stubbs, M. (2007). An example of frequent English phraseology: Distributions, structures and functions. In R. Facchinetti (Ed.), *Corpus Linguistics 25 Years On* (pp. 89-105). Amsterdam: Rodopi.
- Svartvik, J. (1990). *The London Corpus of Spoken English*. Lund: Lund University Press.
- Win, Y., & Masada, T. (2015). Exploring Technical Phrase Frames from Research Paper Titles. *29th International Conference on Advanced Information Networking and Applications Workshops*. IEEE. <https://doi.org/10.1109/WAINA.2015.37>
-

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the Journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).