From the perspective of: Functional Analysis of formulaic sequences in Applied Linguistics Research Articles

ALIREZA JALILIFAR* & SEYED MOHAMMAD GHOREISHI
Shahid Chamran University of Ahvaz (Iran)

ABSTRACT

The current study explored two levels of formulaic sequences, general and discipline specific, in a corpus of 200 applied linguistics research articles and examined the functions they serve. Using Antconc software, in total, 2563 sequences were identified including 593 general and 1370 discipline-specific sequences. These numbers account for 30% and 70% of the full list of formulaic sequences, covering 8.54% in comparison to 7.76% of one-million-word corpus, respectively. Functional analysis revealed that general and discipline-specific formulaic sequences follow the same pattern, with text-oriented sequences making it to the top of the list followed by research-oriented and participant-oriented sequences. Discipline-specific sequences surpassed general sequences, providing evidence for the existence of more diversity of discipline-specific formulaic sequences. Structural and functional correlates also showed that the form of formulaic sequences is closely related to the functional use of the sequences rather than whether they are general or specific domain.

KEYWORDS: Formulaic sequence; Phraseology; Research article; Applied linguistics.

1. INTRODUCTION

Among various academic writing tasks, the research article has become the hub of numerous studies. This is due to the crucial role research articles perform in the dissemination of research findings among members of scientific communities (Flowerdew, 1999) and the higher status given to publications in English (Lillis & Curry, 2006).

───────────────────────

*Address for correspondence: Golestan Blvd, Faculty of Letters & Humanities, Shahid Chamran University of Ahvaz, Iran; e-mail: ar.jalilifar@gmail.com

Research has shown that writers in a particular field are sensitive to certain conventions in manipulating linguistics items (e.g. Biber, 2006; Hyland & Tse, 2005; Jalilifar, 2011). In addition, non-native scholars' writings are found to fall short of academic expectations compared to native-speaking academics (Ferguson, Perez-Llantada & Plo, 2011). This has stimulated experts to gain a better understanding of this genre by deriving its distinctive features with the hope that novice writers become more familiar with its nuances.

Studies in corpus linguistics revealed that language in use is characterized by a large number of pre-fabricated word combinations that function as single units even though they seem to be analyzable into individual segments (Sinclair, 1991). This finding has led to the development of phraseology, i.e. "the study of the structure, meaning and use of word combinations" (Cowie, 1994: 3168). Under the catch-all term of formulaic language, there are various types of phraseological units (Biber & Barbieri, 2007: 264). Studies in phraseology have shown that formulaic language is a common feature of language use (Sinclair, 1991: 111), making up between 21 and 52.3 percent of language in use (Biber, Conrad & Cortes, 2004; Erman & Warren, 2000).

Reviewing the literature shows that in the field of phraseology, some terms are used as more general cover terms to the linguistic phenomenon of phraseology, among them set phrases, multiword sequences, phraseological units, lexical bundles and formulaic sequence*s*. However, formulaic sequence has been widely seen as "the most comprehensive term" thus "intentionally all-encompassing, covering a wide range of phraseology" (Schmitt & Carter, 2004: 4). The term formulaic sequence is defined as "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated, that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (Wray, 1999: 214).

Using a bottom-up approach, referred to as the distributional approach, Biber, Johansson, Leech, Conrad and Finegan identified the most frequent formulaic sequences in academic discourse which they called lexical bundles. They defined lexical bundles as "bundles of words that show a statistical tendency to co-occur" (1999: 989) and as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status" (1999: 990) (e.g. *from the perspective of, as a result of, the extent to which*).

Research has documented that the unnatural and unidiomatic nature of papers written by non-native English students is due to a dearth or misuse of formulaic sequences (e.g. Adel & Erman, 2012; Alipour & Zarea, 2013; Amirian, Ketabi & Eshaghi, 2013; Jalali, 2014; Qin, 2014). Research has also shown that formulaic sequences vary across different registers and genres (e.g. Biber, 2006; Hyland, 2012; Jalali, 2013*,* Jalali. 2015), and characterize specific disciplines (e.g. Alipour, Jalilifar & Zarea, 2013; Hyland, 2008; Jalilifar, Ghoreishi & Emam Roodband, 2017; Kashiha & Heng, 2013).

In terms of their structures, Biber et al. (1999) found that only 15% of formulaic sequences in conversation and 5% in academic prose are structurally complete, and that most formulaic sequences connect two units, that is, the last word of the sequence is often the first element of the next unit. Although formulaic sequences do not often represent complete structural units, Biber et al. (1999) found that they have considerable grammatical correlates that vary considerably depending on the register. For example, sequences in conversation are often clausal (e.g. I want you to, it's going to be), whereas in academic written prose, most formulaic sequences are commonly phrasal (e.g. as a result of, on the other hand). Likewise, Biber and Barbieri (2007) found that formulaic sequences often consist of incomplete nominal chunks (e.g. the nature of the, as a result of) or clausal chunks (e.g. *I don't know how, I thought that was*).

Despite their structural incompleteness, formulaic sequences serve important functions in discourse. Biber et al (2004) put forward a taxonomy that reflects the purposes of formulaic sequences in text and distinguishes among three main functions: 1) stance expressions which are used to express attitudes or assessments of certainty (e.g. you might want to, it is important to), 2) discourse organizers which reflect relationships between prior and coming discourse (e.g. if you look at, on the other hand) and 3) referential expressions which are used to make direct reference to physical or abstract entities, or to the textual context itself (e.g. the nature of the, as shown in figure).

Later, Hyland (2008) created another functional framework by investigating formulaic sequences in a large corpus of research articles. His categories better represented the formulaic sequence functions he found in his corpus. He classified formulaic sequences into three functional categories: research-oriented sequences that are used to structure research activities (*at the beginning of*, *at the same time*, *in the present study)*; text-oriented sequences that help writers organize the text and develop their argument (*on the other hand*, *these results suggest that*, *in the next section)*; participant-oriented sequences that involve writers and readers in the developing text (*may be due to*, *it is possible that, should be noted that*). Hyland (2008) discovered considerable difference in the function of formulaic sequences across disciplines. He found that the writers in physical sciences utilized more research-oriented sequences, whereas text-oriented and participant-oriented sequences were employed with more frequency in social sciences.

In another study, Strunkyte and Jurkunaite (2008) compared humanities and natural sciences in terms of distribution of different structural and functional types of formulaic sequences. The results of this study indicated that there is an overt relationship between structural and functional categories. Thus, stance sequences are composed entirely of verbal phrase fragments, while text-organizing sequences are composed of noun phrase or verb phrase fragments. Referential sequences are the only functional category, which are realized in all four structural types.

Despite significant progress on formulaic sequences, past research has provided little direction regarding the identification and analysis of general and discipline specific sequences in a specific discipline. Developing general formulaic sequences and determining their proportion to discipline-specific formulaic sequences in a specific discipline can help settle the debate over whether general academic formulaic sequences exist, as Simpson-Vlach and Ellis' (2010) results indicate, or if formulaic sequences are strictly discipline-specific as Hyland's (2008) findings suggest.

Additionally, it can help to discover whether the principles of developing academic vocabulary lists can be applied to formulaic sequences as well. That is, it can also contribute to the way we may construe different levels of formulaic sequences according to their contribution to the realization of different registers. Just as the way vocabulary researchers have conceptualized different levels of vocabulary- general vocabulary, academic vocabulary, and specialized vocabulary (Carter, 1998)- we can construe general formulaic sequences as middle level academic formulaic sequences consisting of multi-word expressions occur more frequently in academic texts and consistently across different disciplines without being field-specific. Similarly, we can construe discipline-specific formulaic sequences as technical or field specific expressions.

In addition, exploring structural and functional correlates between general and discipline-specific sequences can determine the relationship between the forms of the formulaic sequences and the functions they serve. Besides, the functional utility of general and discipline-specific formulaic sequences provides enough face validity for EAP material designers to move beyond individual lexical items in developing appropriate teaching materials. Functional utility of formulaic sequences can provide learners with convincing arguments as to why it is important to focus on these sequences and highlight the importance of formulaic sequences in comprehension and production of academic texts.

Thus, the current study aims to use the list of common formulaic sequences across three disciplinary areas obtained in our previous study (Jalilifar et al, 2017) as a baseline to identify and calculate the proportion of general and discipline-specific formulaic sequences in applied linguistic research articles. Additionally, the study aims to compare the general and discipline-specific formulaic sequences in terms of the functions they serve. Thus, to address the above goals, the following main questions stand out:

1. What is the proportion of general and discipline-specific formulaic sequences in a corpus of research articles from the field of applied linguistics?
2. What differences and similarities can be identified between general and discipline-specific formulaic sequences in terms of their functional characteristics?

IJES, vol. 18(2), 2018, pp. 161–186
Print ISSN: 1578-7044; Online ISSN: 1989-6131

2. MATERIALS AND METHOD

2.1. Corpora

For the purpose of the study, we built two general corpora. First, a corpus of over six million words was built from three major disciplinary areas. To this aim, Glanzel and Schubert's (2003) classification scheme was used for the selection of the disciplines and subject areas. From each discipline, five subject areas were randomly selected, and each subject area was represented by four journals as recommended by experts. Thus, altogether, 60 journals were selected. Then volumes and research articles from each journal were randomly selected amounting to 1374 research articles. This corpus was also used in our former study (Jalilifar et al., 2017) in which we identified shared formulaic sequences across three main academic disciplines. Table 1 and Table 2 provide the basic information about the disciplinary areas (adopted from Jalilifar et al., 2017).

Sciences	Social Sciences	Arts & Humanities
Agricultural Science	Education	Arts
General Medicine	Economics	Literature
Chemistry	Sociology	Applied Linguistics
Physics	Management	Philosophy
Computer Science	History	Religion

Table 1. Subject Areas in the Three Disciplines.

Disciplines	No. of Texts	No. of words
Sciences	528	2020551
Social Sciences	380	2027307
Arts and Humanities	466	2008878
Total	1374	6056736

Table 2. The Number of Articles and Word Counts in the Corpora.

We then compiled a second corpus, including slightly over one million words from a sample of 200 written articles extracted randomly from applied linguistics journals. This corpus was used to identify general and discipline-specific formulaic sequences and their proportions in this sample. In the selection of the journals and articles, we followed the same procedure in the compilation of the first corpus.

2.2. Analytical steps

In order to retrieve formulaic sequences, Antconc 3.2.1 (Anthony, 2007) was used. To ensure smooth processing of the texts, the articles were first refined by being cleared from non-text materials such as page numbers, references, figures and tables.

As for the length of the formulaic sequences, for the sake of thoroughness, we set out to query formulaic sequences with varying lengths. Three-word formulaic sequences were included in the data set since their pedagogical importance was shown in Simpson-Vlach and Ellis (2010). The results of their study showed that the majority of the top 50 general formulaic expressions tend to be three-word phrases (e.g. *in terms of, in order to, in other words*). These three-word strings, together with four-word formulaic sequences which are the most frequently researched strings in academic prose (e.g. Biber & Barbieri, 2007; Biber et al., 2004; Cortes, 2002, 2004; Hyland, 2008) as well as rare five-word formulaic sequences, were all considered in the current study.

The actual frequency cut offs are somewhat arbitrary in the literature. In this study, we followed Biber et al. (1999) by setting a minimum frequency of 10 times per million words. The lower cut-off frequency point allowed us to cast a wider net in identifying formulaic sequences that might be overlooked in previous studies. Finally, following Biber et al. (1999), the dispersion criterion (the number of texts in which a sequence has to occur) was set at five or more texts.

To identify general formulaic sequences, we first generated three separate lists of formulaic sequences from the three main disciplines. The lists were then collapsed into a single list to calculate the frequencies. Those strings that occurred three times were considered as general formulaic sequences across the three disciplinary areas. In order to explore the overlap between general formulaic sequences and formulaic sequences in a specific discipline, we first retrieved all the formulaic sequences from the corpus of applied linguistics texts. We then uploaded this list and the original list of general formulaic sequences. Those formulaic sequences that occurred twice were considered as general formulaic sequences in applied linguistics. Finally, to identify discipline-specific formulaic sequences, the whole list of formulaic sequences from applied linguistics and the list of identified general formulaic sequences in applied linguistics were collapsed into one single list. Those formulaic sequences that occurred once were considered as discipline-specific (i.e., specific to applied linguistics).

Finally, general and discipline-specific formulaic sequences in applied linguistics were functionally categorized and compared based on a modified version of Hyland's (2008) classification used by Salazar (2014), as illustrated in Table 3. This classification scheme was found to be particularly useful for the present study, since it is adapted to the specific concerns of research-focused written genres.

Research-oriented	Text-oriented	Participant-oriented
Help writers to structure their activities	Concerned with the organization of the text and its meaning	Focus on the writer or reader of the text
Location Indicate place and direction *at the site, the tip of* **Procedure** Indicate events, actions and methods *the onset of* **Quantification** Indicate measures, quantities, proportions and changes thereof *total volume of* **Description** Indicate quality, degree and existence *the appearance of* **grouping** Indicate groups, categories, parts and order *a wide range of*	**Additive** Establish additive links between elements *in addition to* **Comparative** Compare and contrast different elements *as compared with* **Inferential** Signal inferences and conclusions drawn from data *the results suggest that* **Causative** Mark cause and effect relations between elements *as a result of* **Structuring** Organize stretches of discourse or direct the reader elsewhere in text *as described previously* **Framing** Situate arguments by specifying limiting conditions *in the case of* **Citation** Cite sources and supporting data *it has been proposed that* **Generalization** Signal generally accepted facts or statements *little is known about* **Objective** Introduce the writer's aims *we asked whether*	**Stance** Convey the writer's attitudes and evaluations *is likely to* **Engagement** Address readers directly *it should be noted that* **Acknowledgments** Recognize people or institutions that have participated in or contributed to the study *would like to*

Table 3. Functional Classification of Formulaic Sequences (Salazar, 2014: 52).

3. RESULTS AND DISCUSSION

After the application of retrieval criteria, in total, 5999 formulaic sequences were extracted from the three main disciplines, amounting to 307497 individual cases.

After collapsing the three lists and calculating the frequencies, we identified 661 general formulaic sequences of varying lengths across disciplines (See Table 4, adopted from Jalilifar et al., 2017). In addition, these 661 formulaic sequences are used 152442 times for a total of 475304 words out of the 6 million words in the three sub-corpora (See Table 5).

	3-words	3-words	3-words	Total
General formulaic sequences	566	80	15	661
%	85%	12.5%	2.5%	100%

Table 4. Frequency Information of General Formulaic Sequences in the Three Disciplines.

	Arts and humanities	**Social sciences**	**Sciences**	**Total**	**Word tokens**
3-words	40600	46045	49390	136035	408105
4-words	4791	4843	5202	14836	59344
5-words	440	584	547	1571	7855
Total	45831	51472	55139	152442	475304

Table 5. Frequency Information of General Formulaic sequences Tokens Across Disciplinary Areas.

The identification of 661 general formulaic sequences is not consistent with Hyland and Tse (2007) who questioned the utility of developing a list of generic lexical items for the students of academic writing courses (as mentioned in Jalilifar et al., 2017) and also opposed to Hyland's (2008) argument that "there are not enough formulaic sequences common to multiple disciplines to constitute a core academic phrasal lexicon" (Simpson-Vlach & Ellis, 2010: 509).

This apparent finding can be explained by considering the specifications of the corpora and the extraction criteria used in their studies. In Hyland and Tse's (2007) study, the corpus contained 3.3 million words from a range of academic disciplines and genres. Similarly, Hyland (2008) used a 3.5-million-word corpus of major academic genres (research articles, doctoral dissertations and Master's theses). In addition, they analyzed only 4-word formulaic sequences and set the minimum frequency of occurrence to 20 times per million words. However, in this study, we used a larger corpus including 6 million running words sampled from only published research articles and we started with a lower cut-off frequency. Moreover, our extracted 661 formulaic sequences are limited to those shared among the three disciplines, the criterion that was not considered in Hyland and Tse's (2007) and in Hyland's (2008) study.

3.1. The proportion of general and discipline-specific formulaic sequences in applied linguistics

To answer the first research question, we first extracted formulaic sequences in the applied linguistics corpus. In total, 1963 formulaic sequence types of varying lengths were extracted, representing 51496 individual cases. To identify general formulaic sequences in the applied linguistics corpus, this list was then collapsed with the list of general formulaic sequences from our multidisciplinary corpus to examine the overlap. In total, 593 formulaic sequences

in the applied linguistics corpus were found to overlap with the list of 661 general formulaic sequences. This leaves 68 formulaic sequences occurring only in the shared formulaic sequences list from the three disciplines. In terms of sequence types, this number covers 30% of the full list of formulaic sequences in the applied linguistics corpus. However, when the proportion of overlap was calculated in terms of tokens, the coverage increased. As shown in Table 6, general formulaic sequences in the applied linguistics corpus represent 27417 tokens which account for 53% of all formulaic sequences cases.

Regarding discipline-specific formulaic sequences, in total, 1370 formulaic sequences were identified. This number covers 70% of the full list of formulaic sequences types in the applied linguistics corpus. Nonetheless, when the proportion of overlap was calculated in terms of tokens, the coverage decreased. As shown in Table 6, discipline-specific formulaic sequences account for 47% of all formulaic sequences tokens, representing 24079 cases. Figure 1 and Figure 2 display the proportion of general and discipline-specific formulaic sequences types and tokens in the applied linguistics corpus. With respect to the coverage of general and discipline-specific formulaic sequences over the corpus, general formulaic sequences cover 8.54% and discipline-specific formulaic sequences cover 7.76% of the one-million-word corpus of applied linguistics.

General types	%	General tokens	%	Discipline-specific types	%	Discipline-specific tokens	%
593	30.%	27417	53%	1370	70%	24079	47%

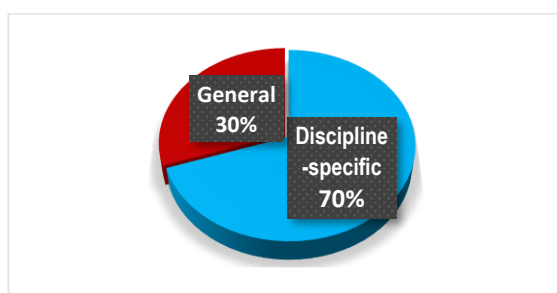**Table 6.** Distribution of General and Discipline-specific Types and Tokens.

Figure 1. Proportion of general and discipline-specific formulaic sequences types.
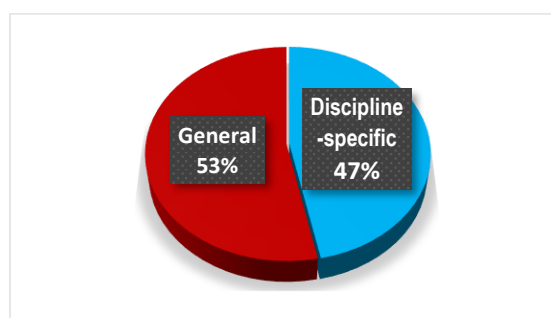

Figure 2. Proportion of general and discipline-specific formulaic sequences tokens.

The proportions of general and discipline-specific formulaic sequences confirm previous research that there is not just "one single pool of formulaic sequences" that speakers or writers draw on, but that ''each register employs a distinct set of formulaic sequences, associated with the typical communicative purposes of that register'' (Biber & Barbieri, 2007: 265). This has a clear pedagogical implication for academic writing students and course designers, suggesting that they should be aware of the importance of both general and

discipline-specific formulaic sequences in order to incorporate them properly in their instructional materials.

Accordingly, this suggests that general lists of formulaic sequences may be suitable for students studying in specific disciplines, as many of these formulaic sequences appear in the sample applied linguistics research articles. This finding is opposed to that of Hyland and Tse (2007: 248) who criticized the argument that "there is a general vocabulary of value to all students preparing for, or engaged in, university study". Similarly, Ward (2009) argued that studies are required to develop word lists from individual disciplines as learners in different domains have different lexical needs. On the contrary, our results may suggest that the development of an inventory of general formulaic sequences might be useful for students who need to read and write in specialized areas.

3.2. Functional analysis of general and discipline-specific formulaic sequences in applied linguistics

To answer the second research question, the identified general and discipline-specific formulaic sequences in the applied linguistics articles were classified according to their functions. Table 7 displays the functional classification of the general and discipline-specific formulaic sequences and their corresponding frequencies. Figure 3 graphically compares the distribution of main functional categories of general and discipline-specific formulaic sequences types and Figure 4 compares the distribution of functional subcategories of general and discipline-specific formulaic sequences.

As shown in Table 7, text-oriented formulaic sequences make it to the top of the list with 58.51% of general formulaic sequences, 347 cases, and 55.18% of discipline-specific formulaic sequences, with 756 cases. Placing second are research-oriented formulaic sequences with 191 types of general formulaic sequences (32.20%) and 494 types of discipline-specific formulaic sequences (36.05%). Participant-oriented formulaic sequences rank a distant third, with 9.27% of general types (n=55) and 8.75% of discipline-specific types (n=120).

The concentration of text-oriented general and discipline-specific formulaic sequences is in line with Hyland's (2008) observation that in soft sciences text-oriented formulaic sequences are employed by writers to make the presentation of their research texts more discursively elaborate. Hyland also showed the domination of research-oriented formulaic sequences in hard sciences. He explained this difference by the methods writers carry out their experiments in hard sciences. "Highlighting research rather than its presentation" (Hyland, 2008: 15) requires writers in hard sciences to use more research-oriented formulaic sequences, which "emphasizes the empirical over the interpretive, minimizing the presence of researchers and contributing to the ''strong'' claims of the sciences" (Hyland, 2008: 15).

Function	General	%	Discipline-specific	%
Research-oriented formulaic sequences	**191**	**32.20**	**494**	**36.05**
Location	2	0.33	6	0.43
Procedure	72	12.14	187	13.64
Quantification	14	2.36	70	5.10
Description	77	12.87	166	12.11
grouping	26	4.36	65	4.74
Text-oriented formulaic sequences	**347**	**58.51**	**756**	**55.18**
Additive	44	7.35	44	3.21
Comparative	21	3.54	38	2.77
Inferential	48	8.02	175	12.77
Causative	38	6.40	60	4.37
Structuring	43	7.25	127	9.27
Framing	97	16.35	139	10.14
Citation	26	4.38	99	7.22
Generalization	0	0	1	0.07
Objective	30	5.05	73	5.32
Participant-oriented formulaic sequences	**55**	**9.27**	**120**	**8.75**
Stance	29	4.49	71	5.18
Engagement	25	4.21	45	3.28
Acknowledgment	1	0.16	4	0.29
Total	**593**	**100%**	**1370**	**100%**

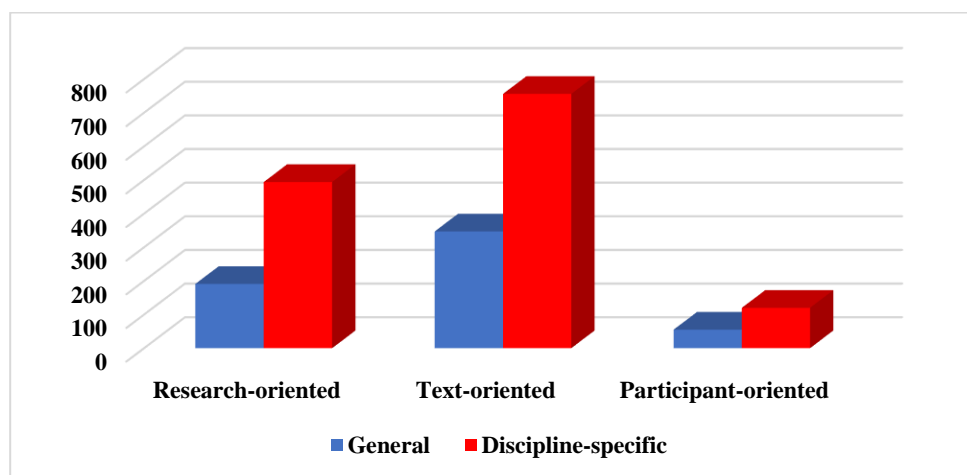**Table 7.** Frequency Distribution of Functions of General Formulaic Sequences.

Figure 3. The distribution of main functional categories of general and discipline-specific formulaic sequences.

Additionally, the frequencies obtained in the current study agree with the results of our previous study (Jalilifar, et al., 2017) in which we found the predominance of text-oriented formulaic sequences in the published articles across three disciplinary areas. In that study, text-oriented, research-oriented, and participant-oriented formulaic sequences accounted for 58.5% and 32.5% and 9% of total general formulaic sequences types, respectively.

With respect to the ranking of the functional subcategories, however, we found differences between general and discipline-specific formulaic sequences. The five most frequent functions of general formulaic sequences are: framing (97 types, 16.35%), description (77 types, 12.87%), procedure (56 types, 10.45%), inferential (48 types, 8.02%) and additive (44 types, 7.35%). By comparison, the five most frequent functions of discipline-specific formulaic sequences are: procedure (174 types, 13.13%), inferential (175 types, 12.77%), description (61662 types, 12.11%), framing (139 types, 1.14%) and structuring (127 types, 9.27%). In the following section, the functional characteristics of general and discipline-specific formulaic sequences with their structural correlates will be discussed at length.
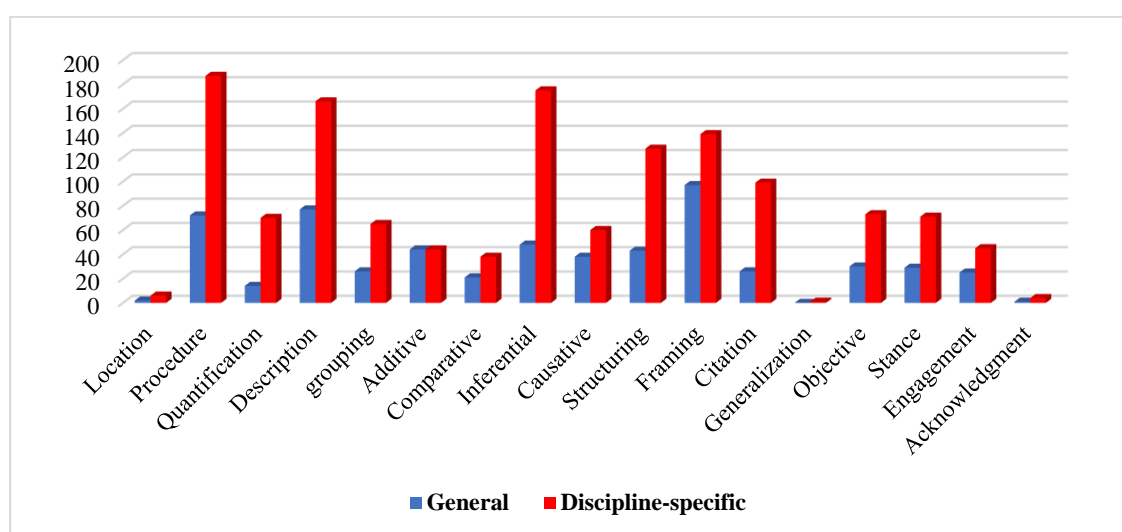

Figure 4. The distribution of subcategories of general and discipline-specific formulaic sequences.

3.2.1. Research-oriented formulaic sequences

Research-oriented formulaic sequences place second on the list of ranking of main functional categories for both general and discipline-specific formulaic sequences. In addition, the frequency distribution of research-oriented general formulaic sequences revealed that description formulaic sequences make it to the top of the list, with 77 types accounting for 40.39% of all research-oriented general formulaic sequences. This is followed by procedure (n=72), grouping (n=26), quantification (n=14) and location (n=2). Discipline-specific formulaic sequences, on the other hand, revealed a different ranking of subcategories. As illustrated in Table 7, procedure formulaic sequences surpass other research-oriented subcategories and make up 13.64% (n=187) of the total discipline-specific formulaic sequences. This finding is in line with Salazar (2014) who finds the predominance of procedure formulaic sequences in her study. This is followed by description (n=166), quantification (n=70), grouping (65) and location (n=6).

As commented previously, discipline-specific procedure formulaic sequences exceed general formulaic sequences. A possible reason for this is the topic-specificity of this particular formulaic sequence function to denote events, actions and methods in describing research processes and activities. Formulaic sequences such as *application of the, assessment of the, creation of the* may refer to certain techniques that were utilized by the applied linguistics authors. The high concentration of procedure formulaic sequences shows that writers in applied linguistics know the importance of reporting research practices with objectivity and precision. Structural analysis of procedure formulaic sequences showed similarities between general and discipline-specific formulaic sequences. They typically take the form of noun phrase + preposition fragments (1), prepositional phrases (2), passive structures (3), and to clause fragments (4).

> *(1) An analysis of* the instances of direct quotes by the IT writers shows why it is so common.
> (2) A redefinition of learning and knowing must be developed, and researchers should play an important role *in the process of* teacher training.
> (3) The data *produced by the* COPS were indeed surprising, most noticeably regarding the lack of student-initiated talk.
> (4) There is a space for Flickr members *to provide a* profile of themselves on their home page.

Another interesting observation in relation to procedure formulaic sequences is that the majority of formulaic sequences in this category incorporate noun phrases. A large number of noun phrases depicting research procedures may suggest the writers' concern in reporting steps they have taken in conducting their research and reporting the results of their study. This may also indicate the writers' attempt to document research activities and report results in an impersonal and objective way. This is consistent with the observation made by Salazar (2014) who finds that noun phrases are commonly used in the Experimental- Materials and Methods- sections of the biomedical research articles suggesting the scientists' concern for objective reporting of the various steps of their studies involved in research and experimentation.

Research-oriented formulaic sequences that contribute to the description of research objects and procedures stood at the top of the list of research-oriented general formulaic sequences, with 77 types. These formulaic sequences place second in the research-oriented discipline-specific list, representing 166 types. With respect to structural characteristics, like procedure formulaic sequences, description formulaic sequences are realized by principally similar structures. They typically take the form of noun phrase + preposition fragments (5), prepositional phrases (6), adjectival phrases (7) and to clause fragments (8).

(5) The predominance of content and listener orientation seems logical.

(6) At the heart of the procedures of discipline, it manifests the subjection of those who are perceived as objects.

(7) Children are observed and interacted with in context-embedded situations where they feel safe and where they are familiar with the interlocutors.

(8) The author first provides information about the effects of trauma and then identifies teaching approaches that are sensitive to the needs of those affected by trauma.

Another frequently used category is that of research-oriented quantification formulaic sequences, with 56 types of general formulaic sequences and 70 types of discipline-specific formulaic sequences. An analysis of quantification formulaic sequences revealed that discipline-specific formulaic sequences display more diversity in keywords denoting quantification. Keywords such as value, couple, overall, percentage, extent, percentage are found to be specific to discipline-specific formulaic sequences. In terms of structures, likewise, discipline-specific formulaic sequences displayed more structural diversity. General formulaic sequences are principally realized by prepositional phrases (*portion of the*). By comparison, discipline-specific formulaic sequences take the form of proportional phrases (*relative frequency of*), noun phrases (*the total number*), to clause fragments (*to the extent that*), that-clause fragments (*the extent that*), adverbial clause fragments (*as measured by*), adjectival fragments (*little or no*) and coordinate conjunctive fragments (*and to what extent*).

Another research-oriented formulaic sequence is that of grouping, ranking fourth in both general and discipline-specific sequence lists. In terms of types, discipline-specific formulaic sequences represent more frequencies (n=65) than general formulaic sequences (n=26). In terms of structural categories, discipline-specific formulaic sequences are realized by similar structures. They are principally realized by noun + of fragments (*rest of the*), and prepositional phrases (*in a variety*) *and* passive structures (*is divided into*).

The least employed category of research-oriented formulaic sequences is that of location. This category comprises 0.33% of general and 0.43% of discipline-specific formulaic sequences and are commonly realized by prepositional phrases (*at the university of*). This is consistent with the observation made by Salazar (2014) that location and grouping formulaic sequences appeared in smaller numbers and that location formulaic sequences were basically realized by prepositional phrases.

In brief, the use of research-oriented formulaic sequences reflects writes' preoccupation in producing an objective and unbiased account of procedures adopted in their experimentations, so that the subsequent data interpretation can be conducted in a verifiable and reproducible manner. This is in line with Hyland's (2008) argument that the use of research-oriented formulaic sequences expresses the empirical over the interpretive, and contributes to minimizing the presence of researchers and highlights research practices and

the methods, procedures and equipment used, and this allows writers to highlight procedures rather than interpretations

3.2.2. Text-oriented formulaic sequences

Text-oriented functions are associated with the largest number of formulaic sequence types in both general and discipline-specific formulaic sequences. As stated earlier, text-oriented formulaic sequences ranked first with 58.51% of general formulaic sequence types (347 cases) and 55.18% of discipline-specific formulaic sequence types (756 cases). This finding agrees with Salazar (2014) in that text-oriented formulaic sequences make the most widely used category among the three main functional categories constituting nearly half of formulaic sequence types and tokens. The concentration of text-oriented formulaic sequences agrees with the argument made by Hyland (2008) that text-oriented formulaic sequences are more a characteristic of soft-knowledge fields such as applied linguistics, and that they play a central role in the discursive practice in the soft sciences.

A comparison of general and discipline-specific formulaic sequences shows that the distribution of their subcategories is not analogous. However, with respect to the structures by which these subcategories are realized, there are similarities that are worth taking note of

As shown in Table 7, one of the most frequently used functional categories across general and discipline-specific formulaic sequences is framing formulaic sequences. Framing formulaic sequences "are used to focus readers on a particular instance or to specify the conditions under which a statement can be accepted" (Hyland, 2008: 16). Framing formulaic sequences constitute the predominant text-oriented general formulaic sequences, accounting for 16% (97 cases) of general formulaic sequences. By comparison, framing formulaic sequences rank second in the text-oriented discipline-specific list, with 139 cases accounting for 10% of discipline-specific formulaic sequences.

The dominance of framing devices in the text-oriented category agrees with Hyland in that framing formulaic sequences comprised a high proportion of text-oriented formulaic sequences in the applied linguistics texts. Hyland argues that writers employ framing devices in soft sciences far more than hard sciences since "the research often has to be contextualized far more carefully and the connections between components explained in greater detail for readers unfamiliar with the thread of prior research" (2008: 6). The widespread use of framing formulaic sequences is also in line with Salazar (2014) in that framing signals are the most frequent function in the text-oriented category.

Structural analysis of framing formulaic sequences revealed that similar structures are used to form this functional category across general and discipline-specific formulaic sequences. Framing formulaic sequences are typically realized by prepositional phrase structures (9), noun phrases (10), *that clause fragments* (11), and passive structures (12):

(9) There is no research that addresses this issue <u>from the perspective of</u> goal setting for vocabulary teaching.
(10) It is likely to be stronger precisely in the use of <u>features associated with</u> orality and the manifestation of speaker stance.
(11) In the course of the unit, students participate in an activity in which they change the rules of their school and also explore <u>the notion that</u> the U.S Constitution is not set in stone.
(12) The word-family principle <u>is based on the</u> conviction that the lexicon can be usefully divided into larger morphologically related units.

Another most frequently used functional category across general and discipline-specific formulaic sequences is inferential associated with formulaic sequences that are used to indicate conclusions of the study and inferences the writers want readers to draw from the arguments (Hyland, 2008). Inferential formulaic sequences constitute the predominant text-oriented discipline-specific formulaic sequences, accounting for 12% (175 cases) of discipline-specific formulaic sequences. By comparison, inferential formulaic sequences rank second in the text-oriented general list, with 48 cases accounting for 8% of general formulaic sequences. As these numbers show, discipline-specific formulaic sequences surpass general formulaic sequences in this category.

Structural analysis of inferential formulaic sequences revealed that similar structures are used to realize this functional category across general and discipline-specific formulaic sequences. Inferential formulaic sequences are typically realized by prepositional phrase structures (13), noun phrases (14), that-clause fragments (15), to clause fragments (16) and passive structures (17):

(13) The analysis of <u>the relationship between</u> involvement and class size was limited to the 55 MICASE class sessions.
(14) <u>The findings of this</u> study provide significant insights into the relationships between learners' profiles, awareness, and learning in L2 pragmatics in two respects.
(15) It could therefore <u>be argued that</u> non-addressing of such errors are, in-and-of itself, a management practice worthy of future examination.
(16) The difference could <u>be attributed to</u> the observation that IT professionals often include substantial quotes from prior texts in the form of computer code
(17) The difference might <u>be associated with</u> the teaching practices at the EMI university.

Additive signals are the third most frequently used formulaic sequences among general formulaic sequences, accounting for 7.35% of the general formulaic sequence list. By comparison, this category ranks seventh in the discipline-specific list, comprising 2.77% of discipline-specific formulaic sequences. Additive signals are used to establish additive (18),

contrastive links (19) between elements. They are also used for exemplification (20) and clarification (21). A comparison of general and discipline-specific formulaic sequences shows that these signals typically take the form of prepositional phrases in both groups of formulaic sequences.

(18) <u>In addition to</u> investigating language play in other classes, researchers would benefit from teacher interviews and language play.

(19) <u>On the other hand,</u> using the French ordination instead of computer may not be the most effective means of disputing English linguistic hegemony.

(20) <u>For example, the</u> paradigm might be shifted from teacher-centered to student centered learning, from rote learning to discovery learning.

(21) <u>In other words</u>, there is not enough capacity available for long-term learning.

The next frequent signals in the text-oriented category are structuring formulaic sequences. These formulaic sequences organize the text by providing signals that guide readers through the text (Salazar, 2014). These signals make it to the third place of discipline-specific formulaic sequences and second of the general formulaic sequence list. They make up 7.25% of general and 9.27% of discipline-specific formulaic sequences.

A structural analysis revealed that these signals are realized by similar structures across general and discipline-specific formulaic sequences. They typically take the form of noun phrase + prepositions (22), noun phrases (23), prepositional phrases (24), passive structure combined with prepositions (25) and adverbial clause fragments (26):

(22) The major <u>sections of the</u> article review research on aptitude and pedagogy, and aptitude as relevant to understanding sensitive period effects.

(23) Through investigating and describing the intertextuality of two sets of discourse flows collected from two professionals, this study aims to address the <u>following research questions:</u>

(24) It seems likely that this technique indicates an epistemically motivated hedge, as distinct from the dialogically expansive hedges discussed <u>in the previous section</u>.

(25) As a result, the articles adopted in the corpus were selected from the 21 subject areas, which are <u>shown in Table</u> 1 below.

(26) <u>As mentioned above</u>, part of this problem grew out of the isolation of applied linguistics from its parent university department.

This finding agrees with Salazar (2014) in that structuring formulaic sequences, which take the form of adverbial-clause fragments and passive structures combined with prepositions, are used to facilitate comprehension by guiding readers through the text and referring the readers to tables and figures.

IJES, vol. 18(2), 2018, pp. 161–186
Print ISSN: 1578-7044; Online ISSN: 1989-6131

The next most frequent group of general formulaic sequences in the text-oriented category includes causative signals. Causative markers are employed to facilitate comprehension by providing readers with signals which highlight cause-and-effect relationships (Salazar, 2014). Frequency counts revealed that causative formulaic sequences place fifth in the general formulaic sequence list, accounting for 6.40 of general formulaic sequences (38). On the other hand, in the discipline-specific list, they rank sixth, with 60 cases (4.37%). Causative signals, in both general and discipline-specific formulaic sequences, are typically realized by   noun phrase + of (*27*), prepositional phrases (*28*), noun phrases +post-modifier fragments (*29*), verb phrases (*30*), passive constructions (*31*) and verb +to clause fragments (*32*):

> *(27) Test score validation also requires, according to Messick, consideration of <u>the consequences of</u> the uses of tests, otherwise known as test impact.*
>
> *(28) Bilingual discourse features should also be interpreted <u>in the light of</u> findings on complexity in written academic language*
>
> *(29) The results show that linguistic competence has a statistically <u>significant positive effect on</u> achievement tests scores.*
>
> *(30) Researchers should <u>play an important role</u> in the process of teacher training.*
>
> *(31) Future research may include questions concerning whether verbal reports <u>are influenced by the</u> practice effect in subsequent phases of data collection.*
>
> *(32) In contrast, neither the main effect of input modality and that of WM capacity nor the combined effect on mental effort <u>turned out to be</u> significant.*

Another functional category across general and discipline-specific formulaic sequences is objective used to introduce writers' aims. Objective signals make up 5.05% of general formulaic sequences and 5.32% of discipline-specific formulaic sequences. Similar to other functional categories, discipline-specific objective formulaic sequences surpass their general counterparts.

Objective signals, in both general and discipline-specific formulaic sequences, are principally realized by verb +to clause fragments (33). However, there are instances of noun phrase + of (34), prepositional phrases (35), and noun phrases +post modifier fragments (36):

> *(33) The overarching aim of meta-analysis is to consolidate data collected across a number of studies <u>in order to determine</u> the extent of the relationships in question.*
>
> *(34) The <u>purpose of this study</u> was to investigate the potential of using multimedia to facilitate teaching and recalling technical vocabulary.*
>
> *(35) Lastly, <u>in an effort to </u>decrease the formality of the participant observer relationship, I mingled informally with the participants.*
>
> *(36) In this study, I will detail my experience using creative writing assignments in <u>an attempt to </u>raise the critical consciousness of nine ethnic Japanese students.*

Citation signals, which are used to cite research resources and supporting data, account for 4.38% of general and 7.22% of discipline-specific formulaic sequences. Similar to objective signals, discipline-specific citation formulaic sequences exceed their general counterparts. This suggests that applied linguists prefer to overly rely on more specialized expressions in citing resources. Structural analysis revealed that there are similarities in the way they are realized in general and discipline-specific formulaic sequences. They typically take the form of noun phrase +of fragments (*a number of studies*), noun phrase +other post modifiers (*research in this area*), prepositional phrases (in accordance with), passive structure (*have been shown to*), verb phrases (*research suggests that*) and adverbial fragments (*as noted by*). This structural analysis is in line with Salazar (2014) in that citations are frequently realized by adverbial-clause phrases, as well as by a variety of passive structures, including anticipatory-*it* and *that*-clause constructions controlled by passive verbs.

The remaining text-oriented functions are comparative and acknowledgement signals. Comparative signals that are used to compare and contrast elements account for 3.54% of the general and 2.77% of the discipline-specific formulaic sequences. This functional category ranks eighth in both lists. Like other categories, structural analysis indicated similarities in the way comparative signals are realized in general and discipline-specific formulaic sequences. They typically take the form of noun phrase +of fragments (*same level of*), noun phrase +other post modifiers (*significant differences in*), prepositional phrases (*in comparison with*), adjective phrases (*a statistically significant*) and adverbial clause fragments (*as opposed to*):

Generalization signals (*little is known about*), which are used to signal generally accepted facts or statements appear in smaller quantities in comparison to other functional categories in discipline-specific formulaic sequences (n=1). No instance of this functional category was found among general formulaic sequences.

3.2.3. Participant oriented formulaic sequences

This last main functional category corresponds to the dialogic interaction between the participants in the text: the writer and the reader. Participant-oriented formulaic sequences occur less frequently than the two other functional categories, representing only 9.27% of general and 8.75% of all discipline-specific formulaic sequences. This finding agrees with previous research (e.g. Cortes, 2004; Hyland, 2008; Salazar, 2014) in that participant-oriented formulaic sequences are associated with the lowest proportion of the three main functional categories.

A large proportion of participant-oriented formulaic sequences are the expressions that are used to express stance, in line with the findings of Salazar (2014), Cortes (2006) and Hyland (2008). Stance markers help writers communicate their assessments of certain claims

and convey meanings such as confidence, possibility and importance (Salazar, 2014). Discipline-specific formulaic sequences exceed general formulaic sequences, suggesting a preference on the part of applied linguistics writers to use more discipline specific expressions to convey their attitudes towards their assertions and establish the appropriate relationship with their reader. Stance markers make up 4.49% of all general formulaic sequences (29 cases) and 5.18% of discipline-specific formulaic sequences (71 cases). As for structural features, general and discipline-specific formulaic sequences use similar structures to realize this function. Stance markers typically take the form of anticipatory it + adjective phrase (37), be+ adjective phrase (38), adjective clause + that clause fragment (39) and verb/adjective + to clause fragment (40):

> *(37) It is possible that these linking adverbials require higher level writing skills and/or more sophisticated writing tasks or research topics.*
> *(38) It is clear from the data that students were given, and took advantage of, opportunities to be creative even in simple ways.*
> *(39) It is perhaps not surprising that vocabulary composed most of the playful LREs.*
> *(40) There seems to be little attempt to make links between the papers other than the obvious one of the influence of Biber's work.*

Structural realization of stance phrases in our study agrees with Hyland (2008) and Salazar (2014) in that most stance expressions are realized by impersonal structures such as adjective phrases and anticipatory *it* constructions. The use of impersonal structures "indicate the scientific writers' efforts to soften the expression of their attitudes and opinions by means of indirect forms (Salazar, 2014: 105) and also signal objectivity, as it "reduces the writer's role as agent and interpreter and allows research to be presented as independent of any particular scientist" (Hyland, 2008: 19).

Engagement markers refer to another participant-oriented category which occurs much less frequently. These formulaic sequences involve readers in the argument presented by the writer by helping them focus on certain things and see them in a particular way (Salazar, 2014). Writers utilize evaluative adjectives of necessity and importance to perform engagement functions (Salazar, 2014). Engagement markers make up 4.21% of all general formulaic sequences (25 cases) and 3.28% of discipline-specific formulaic sequences (45 cases). As for structural features, general and discipline-specific formulaic sequences use similar structures to realize this function. Stance markers typically take the form of anticipatory it + adjective phrase (41), be+ adjective phrase (42), adjective/verb clause + that clause fragment (43), verb/adjective + to clause fragment (44), adverbial phrases (45) and noun phrase + of fragment (46):

(41) It is worth noting that the risk-taking variable correlated only with the anxiety measure and exhibited no significant correlations with motivation and self-confidence.

(42) It is essential to clarify at the outset the status of the claims that they warrant.

(43) It should be noted that the vocabulary analysis in this study was strictly quantitative.

(44) It is important to note that these randomly generated lists included many general high-frequency words of English.

(45) Overall, as can be seen from the examples above, there is a strong tension between the quantitative and the qualitative principles of the wordlist compilation.

(46) These findings imply the necessity of addressing how ELF speakers adopt pragmatic strategies to facilitate communication.

The final category of participant-oriented formulaic sequences is acknowledgements. Formulaic sequences with this classification are "used to thank individuals or entities for financial assistance or the provision of experimental materials" (Salazar, 2014: 106). This last category accounts for 0.16 percent of general formulaic sequences and 0.29 percent of discipline-specific formulaic sequences and is typically realized by passive structures (is supported by) and verb+ to clause fragments (would like to thank).

4. CONCLUSION

The current study was an extension of our previously published work (Jalilifar et al., 2017) in which we identified 661 shared formulaic sequences across three disciplinary areas. In the present study, we used this list of shared formulaic sequences as a baseline to identify and calculate the proportion of general and discipline-specific formulaic sequences in the field of applied linguistics and to compare the functions they served.

In the first stage of the study, in total, 593 general formulaic sequences and 1370 discipline-specific formulaic sequences were identified. The pervasiveness of discipline-specific formulaic sequences indicates that disciplinary variation is important in the choice of the formulaic sequences and testifies that specialized formulaic sequences constitute an important element of academic discourse competence. This is in line with the observation made by Jalali and Zarei (2016) that formulaic sequences vary more considerably within genres of a specific discipline than those in the same genre of different disciplines. This suggests that "gaining control of academic discourse requires a sensitivity to expert users' preferences for certain sequences of words over others" (Hyland, 2012: 166). In other words, novice writers can benefit from identifying formulaic sequences used by experts in a discipline. The use of these expressions can help novice writers gain communicative competence in presenting their arguments in a proper way (Hyland, 2012).

Frequency distribution of the main functional categories revealed that general and discipline-specific formulaic sequences follow the same patterns, with text-oriented formulaic sequences at the top of the list followed by research-oriented and participant-oriented formulaic sequences. This similarity in distribution shows that general and discipline-specific formulaic sequences have the same contribution to effective communication in applied linguistics research articles including discursive organization of the text and its meaning, engaging with the readers, and structuring research activities and experiences.

A comparison of general and discipline-specific formulaic sequences in respect to their functional subcategories also showed that discipline-specific formulaic sequences surpass general formulaic sequences in all functional subcategories. This shows that in presenting their research, writers in applied linguistics tend to use more discipline specific formulaic sequences in organizing their texts. This can be seen as a confirmation of previous research proving that writers of academic texts do not draw on a single repertoire of multi-word expressions and that the use of these expressions is conventionalized in specific registers and writers employ these expressions in order to obtain the desired communicative purposes (Biber & Barbieri, 2007). This has a clear pedagogical implication for academic writing course designers and practitioners, and calls for an awareness of disciplinary similarities and differences in order to incorporate them properly in the syllabus and teaching materials.

The findings also revealed the relationship between the form and function of these recurrent expressions. We observed that general and discipline-specific formulaic sequences use similar structures to realize functional categories. The structural and functional correlates indicate that the forms of formulaic sequences are related to the functional use of the formulaic sequences rather than whether they belong to the shared disciplinary formulaic sequences or to domain specific ones. This has clear pedagogical implications for EAP practitioners. It is essential to raise the students' awareness towards the relationship between formulaic sequences and their corresponding functions.

It should be noted that studies in phraseology often produce lists of multi-word expressions (e.g. Ackermann & Chen, 2013; Martinez & Schmitt, 2012; Shin & Nation, 2008; Simpson-Vlach & Ellis, 2010) and encourage EAP material designers to use them for material design and course development. However, it should be noted that the need of the students should be taken into account before a list of formulaic sequences is used for a specific purpose. For example, if the purpose of the course is to prepare students to read general academic texts, then, the teachers may decide to have students work with general formulaic sequences that are shared among a range of disciplines. On the other hand, if the course intends to prepare students to understand and produce texts in a technical area, focusing on discipline-specific formulaic sequences derived from domain specific texts may be useful. Nonetheless, it should be noted that any list of general or specialized formulaic

sequences should include only strings with identifiable discourse functions, in order to be of maximum utility.

Our list of identified general and discipline-specific formulaic sequences is the result of a corpus-based study. As Coxhead (2000) notes, such studies create lists of decontextualized multi-word expressions and concordance lines. The use of corpus approaches does not mean that language teachers and learners should rely on decontextualized formulaic sequences. Researchers should provide teachers and learners with the information about the contexts in which the formulaic sequences are used. As we observed in our study, most formulaic sequences are multifunctional and have different pragmatic meanings in different contexts.

The present research is among the few studies on formulaic sequences that maintains three-word sequences in the analysis. The high occurrence rate of three-word sequences motivated many researchers to exclude them in their studies and to work exclusively on four-word sequences, which are far more manageable (e.g. Biber et al 2004; Cortes 2004; Hyland 2008a). In this study, the decision to include three-word formulaic sequences has contributed to a more complete picture of formulaic sequences. However, it gave rise to the problem of overlap. Here, we observe that a number of four-word sequences incorporate shorter sequences, which appear to inflate the frequency rates. For example, the sequence "*as opposed to*" is extended to form "as *opposed to the*". We decided to preserve overlap cases for the sake of thoroughness and a more comprehensive list of formulaic sequences.

Nevertheless, the question of whether shorter strings or longer ones should take priority in teaching and learning was not addressed in this study. For pedagogical purposes, these overlapping cases should be eliminated to avoid unnecessary repetition and make the final list as brief as possible. It would be a great step forward if future research addresses this problem by developing clear criteria to decide which sequences of overlapping items should be preserved and which should be eliminated. Future research may consider the frequency rates of three and four-word sequences. In cases of overlap, where shorter sequences occur with similar frequency, the shorter ones may be eliminated from the list. On the other hand, in cases where there are differences in occurrence rates between the overlapping sequences, and the shorter sequence can function as an independent sequence, the overlapping sequences can be preserved. Moreover, longer sequences ending in articles such as *a, an and the* can be disregarded since they do not provide further functional information that can justify their inclusion.

In addition, studies can balance frequency with intuition to develop a more manageable list of pedagogically useful formulaic sequences for classroom application. Future research may also consider some measurement of statistical cohesiveness, for example, mutual information or the log-likelihood statistics in conjunction with frequency methods to provide insights into which formulaic sequences are perceived to be the important ones for teaching.

In closing, we should note that formulaic sequences are just one aspect of phraseological tendencies of language use (Salazar, 2014) and it is necessary for those interested in this area to determine how formulaic sequences fit into other formulaic expressions, with the hope that the identified formulaic language helps students and non-native academics to communicate efficiently in academic settings (Byrd & Coxhead, 2010). As Salazar (2014) notes, the use of these functional recurrent expressions combined with other features of academic writing results in well-organized texts. It is hoped that this study represents a significant contribution towards reaching a better understanding of the crucial role played by formulaic sequences in written academic discourse.

REFERENCES

Ackermann, K. & Chen, Y.H. (2013). Developing the academic collocation list: A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*(4), 235–47.

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

Alipour, M. & Zarea, M. (2013). A disciplinary study of lexical bundles: The case of native versus non-native corpora. *Taiwan International ESP Journal, 5*(2), 1-20.

Alipour, M., Jalilifar, A. R. & Zarea, M. (2013). A corpus study of lexical bundles across different disciplines. *Iranian EFL Journal, 9*(6), 11-35.

Amirian, Z., Ketabi, S. & Eshaghi. H. (2013). The use of lexical bundles in native and non-native post-graduate writing: The case of applied linguistics MA theses. *Journal of English Language Teaching and Learning, 11*, 1-29.

Anthony, L. (2007). *Antconc 3.2.1*. Available online at http://www.antlab.sci.waseda.ac.jp/

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

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

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

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

Byrd, P. & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, *5*, 31-64.

Carter, R. (1998). *Vocabulary: Applied linguistics perspectives*. London: Routledge.

Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds.), *Using corpora to explore linguistic variation* (pp. 131-45). Amsterdam: John Benjamins.

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

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

Cowie, A. P. (1994). Phraseology. In R. E. Asher (Ed.), *The encyclopedia of language and linguistics* (pp. 3168-3171). Oxford: Oxford University Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213-38.

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

Ferguson, G. R., Perez-Llantada, C. & Plo, R. (2011). English as an international language of scientific publication: a study of attitudes. *World Englishes, 29*(3), 41-59.

Flowerdew, J. (1999). Problems in writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing, 8*, 243–264.

Glanzel, W. & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics, 56* (3), 357-367.

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

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics, 3*, 150-169.

Hyland, K. & Tse, P. (2005). Hooking the reader: A corpus study of evaluative that in abstracts. *English for Specific Purposes, 24*(2), 123–139.

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

Jalali, H. (2013). Lexical bundles in applied linguistics: Variations across postgraduate genres. *Journal of Foreign Language Teaching and Translation Studies, 2*(2), 1-29.

Jalali, H. (2014). Examining novices' selection of lexical bundles: The case of EFL postgraduate students in applied linguistics. *Journal of Applied Linguistics and Language Research, 1*(2), 1-11.

Jalali, H. (2015). A Corpus-driven study of *it* lexical bundles in applied linguistics postgraduate genres. *Journal of Applied Linguistics and Language Research, 2*(1), 36-49.

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, 7*(4), 87-114.

Jalilifar, A. R. (2011). World of attitudes in research article discussion sections: A cross-linguistic perspective. *Journal of Technology and Education, 5*(3), 177-186.

Jalilifar, A.R., Ghoreishi, S.M. & Emam Roodband, S. A. (2017). Developing an inventory of core lexical bundles in English research articles: A cross-disciplinary corpus-based study. *Journal of World Languages, 6*(3), 184-203.

Lillis. T. & Curry. M., (2006). Professional academic writing by multilingual scholars. Interactions with literacy brokers in the production of English-medium texts. *Written Communication, 23*(1), 3-35.

Kashiha, H. & Heng, C.S. (2013). An exploration of lexical bundles in academic lectures: Examples from Hard and Soft Sciences. *The Journal of Asia TEFL, 10*(4), 133-161.

Martinez, R & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299-320.

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

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. Amsterdam: John Benjamins.

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

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

Sinclair, J. (1991). *Corpus, concordance, collocation: Describing English Language*. Oxford: Oxford University Press.

Shin, D. & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal, 62*(4), 339–348.

Strunkyte, G. & Jurkūnaite, E. (2008). *Written academic discourse: Lexical bundles in humanities and natural sciences*. Unpublished master's thesis, Vilnius University, Lithuania.

Ward, J.W. (2009). A basic engineering word list for less proficient foundation engineering undergraduates. *English for specific purposes, 28*, 170–182.

Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching, 32*, 213-231.