

Use of Formulaic Sequences in Undergraduate and Postgraduate Students' Expository Writings

Servat Shirkhani^{*1}  & Samaneh Omid²

Abstract

Formulaic sequences (FSs) have a significant role in second language learners' fluency. Thus, this study compared the distributions of FS types and FS codes in undergraduate and postgraduate students' writings. A hundred expository compositions written by 50 undergraduate and 50 postgraduate students majoring in teaching English as a foreign language were coded for the FS types and FS codes and the coded data were analyzed. Regarding the FS types used by undergraduate students, the results indicated that the highest frequency belonged to collocations and the lowest to idioms. With regard to FS codes, the results showed that undergraduates' use of grammatical collocations was significantly higher than that of lexical collocations while there was no significant difference in their use of the three lexical bundle codes nor in their use of compound codes. Concerning the distribution of FS types in postgraduate students' writings, the most commonly used FSs were collocations while idioms were not used at all. As with FS codes, no significant difference was found in the postgraduates' use of collocation codes and lexical bundle codes. However, they used open forms significantly less than closed and hyphenated forms. Regarding the comparison of undergraduate and postgraduate students, it was revealed that within the FS types, both groups used collocations with similar frequencies, but they differed in their use of other FS types. Idioms were more used by undergraduates whereas lexical bundles and compounds were more employed by postgraduate than undergraduate students. The study presents implications for language teachers and learners.

Keywords: collocations, compounds, expository writing, formulaic sequences, idioms, lexical bundles, postgraduate students, undergraduate students

Received: 15 March 2025
Received in revised form: 11 July 2025
Accepted: 22 July 2025

¹ *Corresponding Author: Department of English Language Teaching, Khor.C., Islamic Azad University, Khorramabad, Iran, Email: servat.shirkhani@iau.ac.ir,
ORCID ID: <https://orcid.org/0000-0003-2114-5391>

² English Department, CT.C., Islamic Azad University, Tehran, Iran,
Email: samaneh.omidi@iau.ac.ir, ORCID ID: <https://orcid.org/0009-0007-2125-6098>

1. Introduction

Formulaic sequences (FSs) are word sequences or expressions that function as a unit and have the tendency to frequently occur together in language (Zimmerer et al., 2016). They are “combination[s] of at least two words favored by native speakers in preference to an alternative combination which could have been equivalent had there been no conventionalization” (Erman & Warren, 2000, p. 30). FSs realize meanings, themes (Birjandi et al., 2021), and functions and can help individuals understand and produce language more effectively. According to Staples et al. (2013), second language acquisition (SLA) theory states that FSs have a central role in the early stages of second language (L2) learning. Specifically, FSs can be beneficial in facilitating the enhancement of learners’ writing proficiency. It is significant for educators to encourage learners to acquire FSs to enhance and nurture their writing fluency and linguistic competence. If L2 learners integrate these sequences seamlessly into their writing, they may show higher language fluency and coherence, as shown by Lewis (2008) and Ohlrogge (2009). Pellicer-Sánchez and Boers (2018) suggested that mastering formulaic language can significantly enhance language proficiency. Because FSs are widespread in language use, native speakers (NSs) exhibit a rich repertoire of FSs, indicative of heightened language proficiency levels. Consequently, L2 learners must master FSs to attain proficiency commensurate with NSs. This shows that L2 education, in general, and L2 writing instruction, in particular, should give more prominence to the role of FSs in enhancing L2 learners’ production and their writing accuracy and fluency.

Despite the importance of FSs in developing L2 learners’ production (Le-Thi et al., 2017), some studies (e.g., Erman & Warren, 2000) have indicated that the frequency of using FSs in L2 learners’ language is low. On the other hand, as found by Duan and Shi (2021), several factors collectively demonstrate the dynamic and multifaceted nature of FSs in L2 writing, showing that the development of these structures in the L2 is influenced by various factors, including semantic transparency, structure, learners’ proficiency, and their interactions. Although some studies (e.g., Derakhshan, 2019; Duan & Shi, 2021; Esfandiari & Barbary, 2017; Kaya & Yilmaz, 2018; Ma, 2020; Umair et al., 2023; Vilkaitė & Schmitt, 2019; Wang, 2018; Zinkgraf & Verdú, 2021) have examined the use of FSs by L2 learners, there is a scarcity of research on the use of FS codes which can delve into the variety in the use of FSs and better show the functions for which FSs are used. A further issue is that despite the role of factors such as learners’ level of L2

proficiency, scant research (e.g., Esfandiari & Barbary, 2017; Qin, 2014; Vilkaite & Schmitt, 2019) has addressed the relationship between proficiency level and FS use. Because of the significance of FSs in developing L2 learners' writing skills and the role of factors such as language proficiency in the development of FSs in the L2, and due to the rarity of research in this specific area, the present study set out to examine the distributions of the FSs in undergraduate and postgraduate students' writing by attempting to answer the following questions:

1. What are the distributions of the FS types and codes in L2 expository writings by undergraduate students?
2. What are the distributions of the FS types and codes in L2 expository writings by postgraduate students?
3. Is there a significant interaction between students' academic level and their use of FS types and codes in L2 expository writing?

2. Literature Review

FSs are ready-made structures in a language which have fixed grammatical structures, stable meanings, and specific contexts of use. They enable language users to manage sentence creation during communication and writing. As Wray (2002) put it, an FS is "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: in other words, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (p. 9).

2.1. Types of FSs

The prime matter in the categorization of FS types is FS identification in discourse. In fact, identification of FSs is the basis of applied research on formulaic language. However, there is little research about recognizing formulaic language and, therefore, identifying FSs in oral or written texts can be demanding. Erman and Warren (2000) stressed that FSs are sequences of not less than two words which are favored in preference by NSs over sequences that could have been the same if there had been no conventionalization.

Based on previous studies (Fan & Wang, 2024; Wood, 2019), there are different

types of FSs that serve various purposes in communication and writing. According to Wood (2019), the major types of FSs include compounds, collocations, metaphors, idioms, phrasal verbs, lexical phrases, proverbs, n-grams, and bundles. Furthermore, Fan and Wang (2024) highlighted two types of FSs, that is, internal and purely external FSs. They defined internal FSs as sequences of words that have been acquired previously by the learner and, instead of being produced word-by-word during their use, are considered as a whole. On the other hand, purely external FSs were defined as sequences of words with a semantically and grammatically neat structure used as usual ways of conveying certain ideas (Fan & Wang, 2024).

The categorization used in this study is based on Gholami (2022), who classified FSs into four types consisting of idioms, collocations, lexical bundles, and compounds. Based on this classification, any of these FS types might consist of some subtypes based on their structure or function. Following Hyland (2005) and Benson et al. (2010), these subtypes are referred to as FS codes in the present study. Based on this categorization, idioms are word combinations that are semantically noncompositional and nonproductive (Wood, 2019). Collocations are combinations of words which co-occur because they have certain syntactic or semantic relationships with each other (Wood, 2019). There are two FS codes for collocations: lexical collocations and grammatical collocations. Lexical collocations consist of collocated words that have some semantic relationship with each other (e.g., single parent), while in grammatical collocations the words are in a syntactic relationship (e.g., verb + object like make a decision). Lexical bundles are multiword units that have certain functions in discourse and, based on these functions, are categorized into referential bundles, stance bundles, and discourse organizers. According to Wood (2019), referential bundles refer to something real or abstract inside or outside the text (e.g., one of the things), stance bundles are used by the author to express attitudes or assess levels of certainty (e.g., are likely to be), and discourse organizers are used to show the link between discourse and its subsequent parts (e.g., on the other hand). Finally, compounds refer to combinations of two words in which the second is the head of the compound and the first is used to classify the head (Wood, 2019). The three codes of compounds include closed forms, hyphenated forms, and open forms. In closed forms, the constituent words are written as one word (e.g., laptop). In hyphenated forms, the words are separated by hyphens (e.g., brother-in-law), and in open forms, the words are written as separate words (e.g., real estate). Table 1 presents a brief description of the categorization described here, which has been the basis of data coding in this study.

2.2. The Significance of FSs in L2 Writing

There exists a lot of research on the role of FSs in language development (Ellis, 2003; O'Donnell et al., 2013) and language fluency (Russell, 2017; Wood, 2007, 2010). FSs are significant constituents of first language (L1) and L2 discourse in spoken and written forms. They enhance fluency, comprehension, communication, cultural understanding, vocabulary, and pragmatic competence in both writing and speech (Ellis, 2012). FSs convey specific meanings and functions in language, and using them can help language learners sound more fluent and natural in speech (Zinkgraf & Verdú, 2021). Wray (2002) argued that L1 learners' production is fluent and communicative because they tend to save and restore chunks of language in their mental lexicon and also use FSs as standalone units of meaning, but L2 learners tend to divide FSs into parts and store them word-by-word. In fact, processing FSs is faster than creatively generated language (Le-Thi et al., 2017). According to Ellis (2012), psycholinguistic findings have shown that, concerning cognitive processing, using holistic phrases is more advantageous than using separate words and syntactic rules. However, it should be noted that different types of FSs have different processing requirements. For example, regarding the differences between idioms and collocations, Bruening (2020) argued that idioms have a nonliteral meaning that can't be understood by analyzing the individual words, whereas collocations have a literal meaning, with the individual words keeping their usual meanings.

Using FSs in language learning leads to more natural and effective communication that is culturally appropriate (Ellis, 2012). Kaya and Yilmaz (2018) argued that idioms reflect the cultural nuances and heritage of the language. They help students to find a more profound appreciation of the society to which the language belongs. Additionally, FSs are significant in building a community and fostering a sense of belonging within academic circles (Le-Thi et al., 2017). As such, Nattinger and DeCarrico (1992) emphasize that mastering FSs is essential in learning an L2 or a foreign language.

The relevance of FSs for writing practices and their potential to improve L2 writing skills has received considerable attention in the scholarly literature (e.g., Duan & Shi, 2021; Ebrahimi et al., 2021; Esfandiari & Barbary, 2017; Fan & Wang, 2024; Kaya & Yilmaz, 2018; Liou & Chen, 2018; Ma, 2020; Umair et al., 2023; Zhang et al., 2023). Eidian et al. (2013) reported that lexical collocation learning

had a significant effect on improving the writing skills of Iranian university students. Moreover, Esfandiari and Barbary (2017) emphasized the significance of FSs in structuring various pieces of information and presenting them clearly and coherently, especially when dealing with complex concepts. Casal and Yoon (2023) highlighted that FSs tend to have complex impacts on L2 writing in terms of fluency, accuracy, rhetorical effectiveness, metacognitive awareness, and integration of linguistic repertoires. In addition, some studies (e.g., Ma, 2020; Wang, 2018) have stressed the role of FSs in developing academic writing. Ma (2020) found that the use of FSs had a significant effect on fluency, accuracy, and syntactic complexity in writing academic theses in the L2. In addition to research on FSs, in general, some research has addressed the role of FS types, such as idioms (e.g., Kaya & Yilmaz, 2018), collocations (e.g., Umair et al., 2023), and compounds (e.g., Dwivedi & Ghosh, 2023), in fluent language use. Umair et al. (2023) emphasized that proficiency in collocations is often associated with higher language proficiency levels and that mastering collocations shows advanced language skills and can contribute to overall language competence. Moreover, Dwivedi and Ghosh (2023) argued that compounds play an essential role in maintaining the semantic and content integrity of communication, leading to emotionally sensitive and vibrant interactions among participants.

Much research has focused on the relationship between FS use and language proficiency. Many of these studies have stressed the role of FSs in enhancing L2 learners' proficiency (e.g., Le-Thi et al., 2017; Qin, 2014; Rafieyan, 2018; Umair et al., 2023) while others have highlighted the effect of proficiency on the processing of FSs (e.g., Fan & Wang, 2024). Le-Thi et al. (2017) stressed that recognizing and appropriately using FSs can enhance language proficiency. Qin (2014) compared target formulaic bundles between nonnative graduate students studying at different academic levels and expert writers and showed that the number and range of addressed formulaic bundles used by nonnative graduate students differed mostly at higher academic levels. Esfandiari and Barbary (2017) analyzed the most common 4-, 5-, and 6-word lexical bundles employed by native and nonnative English writers in psychology research papers. The study revealed that the use of lexical bundles by Persian writers was lower than that by English writers and that the two groups employed lexical bundles differently with regard to their structure and function. Vilkaitė and Schmitt (2019) observed that students with better control of collocations are at higher levels of language proficiency. On the other hand, Fan and Wang (2024) referred to the impact of L2 learners' writing

proficiency on their processing of FSs. They focused on factors such as retrieval frequency, syntactic modification, and the effects of topic familiarity, suggesting that higher-proficiency students tend to retrieve internal FSs more frequently and make more modifications to them during writing tasks.

Despite the significance of FSs, a review of the existing research shows that even advanced English as a foreign language (EFL) learners may struggle to achieve a native-like level of fluency in their writing. Therefore, focusing on acquiring and using FSs is crucial in teaching L2 writing. Thus, the current study addressed the frequency of use of FS types and FS codes in the expository writings of English major university students at undergraduate and postgraduate levels and the role of academic level in the use of these sequences.

3. Methodology

3.1. Corpus and Setting

The corpus of the study consisted of 100 expository compositions written in writing courses by undergraduate and postgraduate students majoring in teaching English as a foreign language (TEFL) at Islamic Azad University (IAU) in Iran. The participants in the two groups were asked to write on the same topic, which was “The uses of technology in L2 learning.” Fifty of the texts were written by undergraduate students and 50 by postgraduate students. The participants were male and female students who had been selected through convenience sampling. The undergraduates included 3 male and 47 female students, and the postgraduates were 6 male and 44 female students. The average age of the undergraduates was 24, ranging from 21 to 27, and the postgraduates’ age ranged from 23 to 28, with an average of 26. The official language of the students was Farsi, and none of them had been living in an English-speaking context.

3.2. Instruments

The instruments used for data collection in this study included a corpus of student expository writings and a coding table. The corpus comprised 100 writings by undergraduate and postgraduate students as described in the previous section. The coding table designed by the researchers for the purpose of this study consisted of categorizations of FS types and FS codes according to Gholami’s (2022) data

codification. Table 1 presents a schematic representation of the categorization of FS types and FS codes, with a description and an example provided for each.

Table 1
Linguistic Categories

FS types	Description	FS codes	Example
Idioms	Opaque invariant word combinations	-	barking up the wrong tree
Collocations	A sequence of words that are usually used together	A: lexical collocations B: grammatical collocations	Take interest susceptibility to
Lexical bundles	Regular sequences of words without any predefined linguistic structure used for carrying out discourse functions	A: Referential expressions B: Stance expressions C: Discourse organizing	by virtue of take into account in other words
Compounds	A sequence of at least two words combined to make new meaning or a new word	A. Closed forms B. Hyphenated forms C. Open forms	Fireworks to ice skate real estate

3.3. Data Collection and Data Analysis

This descriptive and comparative study was carried out during winter 2024. The descriptive phase examined the distributions of FS types and the codes for each FS type in both types of corpora, that is, the writings by the undergraduates and those by the postgraduate students. The comparative phase compared the frequencies of FS types and codes across the academic level of the students. To gather the required data, the researchers asked writing professors in two branches of IAU, that is, Khorram Abad Branch and West Tehran Branch, to provide the researchers with expository compositions of their undergraduate and postgraduate students in their writing courses. Each student wrote an expository composition of at least 150 words on the given topic. The students didn't have any information about the purpose of the study while writing their compositions. After completing the writing, the students were told that their writing was going to be analyzed for research purposes, with their identities kept confidential. They were given the right to withdraw their writings, but none of the students reported any problem with the issue. On the whole, 100 compositions, consisting of 50 by undergraduates and 50 by postgraduate students, were gathered. The average number of words in the compositions by the undergraduates was 368, ranging from 189 to 591 words. For the postgraduates, the average was 377, with a minimum of 204 and a maximum of

642 words. After receiving the written compositions, the coding table was used as the framework for coding the FS types and codes in either of the two corpora. The coding was done manually by the second researcher based on Table 1 after reading the descriptions and examples of all the FS types and FS codes and discussing them with the first researcher. Next, five from each corpus were selected randomly to be checked by the first researcher to ensure the coding correctness. This decision was made based on the suggestion by O'Connor and Joffe (2020) that 10-25% of data should be coded by at least two coders to establish intercoder reliability. Cohen's kappa coefficients for the undergraduate and postgraduate data were .87 and .84, respectively. Then, the frequency and percentage of use of FS types and codes were calculated to answer the first two questions related to the distributions of FS types and codes in each corpus. Finally, chi-square tests and Phi Cramer's V were used to compare the FS types and codes used by undergraduates with those used by postgraduate students.

4. Results

Analysis of data was done to answer the three questions of the study. In addition to descriptive statistics, Chi-Square Tests and Phi Cramer's V were run, and the resulting crosstabs were used to compare the use of FSs by undergraduate and postgraduate students.

4.1. Types of FSs Used by Undergraduate Students in their L2 Writings

Descriptive statistics and crosstabs within student level (i.e., the undergraduate group) were used to see which FS types and which codes were more frequently used in writings by the participating undergraduate students. The results for FS types are presented in Table 2. In this and the following tables, each subscript letter denotes a subset of the categories whose column proportions do not differ significantly from each other at the .05 level.

Table 2

Crosstabs for Types of FSs Used by Undergraduate Students

Types of FS	Idioms	collocations	lexical bundles	compounds	total
Count	3 _{a, b}	290 _b	26 _{a, c}	9 _c	328
Expected Count	1.4	264.1	39.6	22	328.0
% within Level	0.9%	88.4%	7.9%	2.7%	100%
% of Total	0.4%	42.2%	3.8%	1.3%	47.7%

Based on Table 2, considering the frequency with which any of the four FS types were used by undergraduate students, the highest frequency belonged to collocations with 290 cases (88.4%), which was 7.6 times more than the sum of the frequencies for all the other three types. The next two FS types in frequency were lexical bundles with a frequency of 26 (7.9%) and compounds with a frequency of 9 (27%). The lowest frequency belonged to idioms, with 3 cases, which constituted 0.9 percent of all the FSs types. Comparison of the frequencies with expected frequencies in Table 2 indicates that undergraduate students used idioms (with a frequency of 3 and expected count of 1.4) and collocations (with a frequency of 290 and expected count of 264) more than expected frequencies, while they used lexical bundles and compounds (for lexical bundles the frequency was 26 and the expected 39.6, and for compounds the frequency was 9 and the expected 22.9) less than expected. Regarding the significance of the differences in the frequencies of the four FS types, different scripts in Table 2 show the significant differences based on the frequencies in proportion to expected frequencies. Based on the table, there was a significant difference between the proportionate use of idioms and compounds. The results also showed that the use of collocations was significantly different from lexical bundles and compounds.

The results for codes of collocations, lexical bundles, and compounds are presented in Table 3, Table 4, and Table 5, respectively. With regard to collocations, based on Table 3, grammatical collocations constituted 61.5 percent of the collocations, and lexical collocations accounted for the other 38.5 percent. The table also shows that there was a significant difference between the use of grammatical and lexical collocations. It can be seen from the table that grammatical collocations occurred less than expected (frequency: 179; expected: 198.7), while the frequency for lexical collocations was more than the expected frequency (frequency: 111; expected: 91.8).

Table 3*Crosstabs for Codes of Collocations Used by Undergraduate Students*

	Collocations Grammatical collocation	Lexical collocation	Total
Count	179 _a	112 _b	291
Expected Count	198.7	92.3	291.0
% within Level	61.5%	38.5%	100%

The results of the comparison of the three lexical bundles codes used by the undergraduate students, as presented in Table 4, showed no significant difference among the three FS codes. It was observed that of 26 lexical bundles being used in the writings of all the 50 participating undergraduate learners, 24 cases (92.3%) were discourse organizing functions and 2 cases (7.7%) were stance expressions. Both stance expressions (frequency: 2; expected: 1.6) and discourse organizing functions (frequency: 24; expected: 22.1) were used higher than the expected frequencies. The findings indicated that referential expressions with an expected frequency of 1.9 were not used at all.

Table 4

Crosstabs for Codes of Lexical Bundles Used by Undergraduate Students

	Lexical bundles referential expressions	stance expressions	discourse organizing	Total
Count	0 _a	2 _a	24 _a	26
Expected count	1.9	1.6	22.5	26.6
% within level	0%	7.7%	92.3%	100%

Regarding the codes of compounds, no significant difference was observed among the three FS codes, that is, referential expressions, stance expressions, and discourse organizing functions. According to Table 5, open forms with a frequency of 6 constituted 75.0 percent of the compounds, and hyphenated forms had a frequency of 2, forming 25.0 percent of all the compounds used by the undergraduate students. Based on the table, closed forms with an expected frequency of 9 were employed by none of the undergraduate students.

Table 5

Crosstabs for Codes of Compounds Used by Undergraduate Students

	Compounds			Total
	Closed form	Hyphenated form	Open form	
Count	0 _a	2 _a	6 _b	8
Expected count	9	5.6	1.5	8.0
% within level	0%	25.0%	75.0%	100%

4.2. Types of FSs Used by Postgraduate Students in their L2 Writings

As another concern of the study, the types and codes of FSs used by postgraduate students were examined. Table 6 presents descriptive statistics and crosstabs for FS types within student level (i.e., the postgraduate group).

Table 6*Crosstabs for Types of FSs Used by Postgraduate Students*

Types of FS	Idiom	collocation	lexical bundles	compound	total
Count	0 _{a, b}	264 _b	57 _{a, c}	39 _c	360
Expected count	1.6	289	43.4	25.1	360.0
% within level	0.0%	73.3%	15.8%	10.8%	100%
% of Total	0.0%	38.4%	8.3%	5.7%	52.3%

As the findings of Table 6 indicate, of the 360 FSs used in postgraduate students' writing, the highest frequency belonged to collocations, with 264 (73.3%). Lexical bundles with a frequency of 57 (15.8%) and compounds with a frequency of 39 (10.8%) were the next two categories. Idioms were not used by postgraduate students at all. Based on Table 3, compounds (with a frequency of 39 and expected count of 25.1) and lexical bundles (with 57 frequency and expected count of 43.4) had a higher frequency than the expected count, while the frequencies of collocations (frequency: 264; expected: 289) and idioms (frequency: 0; expected: 1.6) were less than expected. Regarding the significance of the differences in the use of FS types in relation to their expected frequencies, as shown by the subscripts in Table 6, collocations were used significantly more than both lexical bundles and compounds.

As another part of investigating postgraduate students' use of FSs, distributions of the codes of collocations, lexical bundles, and compounds were examined. As seen in Table 7, of the two types of collocations used by postgraduate students, 75.8 percent were grammatical collocations, and only 24.2 percent were lexical collocations. The result indicated that the frequency of grammatical collocations (with a frequency of 200 and an expected count of 180.3) was more than the expected frequency, but the frequency of lexical collocations (with a frequency of 64 and an expected count of 83.7) was less than expected. As the table indicates, this difference between the two codes of collocations was statistically significant.

Table 7*Crosstabs for Codes of Collocations Used by Postgraduate Students*

	Grammatical collocations	Lexical collocations	Total
Count	200 _a	64 _b	264
Expected count	180.	83.7%	264.0
% within Level	75.8%	24.2%	100%

The results for codes of lexical bundles can be seen in Table 8. As shown in the table, among the three codes of lexical bundles, discourse organizing functions, with a frequency of 46 (83.6%), had a noticeably higher frequency than the other two. Referential expressions, with a frequency of 6 (10.9 %), and stance expressions, with a frequency of 3 (5.5%), constituted only 16.4 of all the lexical bundles used by postgraduate students. The results showed that the frequencies of stance expressions and discourse organizing functions (for stance expression, the frequency was 3 and the expected frequency was 3.4, and for discourse markers, the frequency was 45 and the expected frequency was 46.9) were slightly lower than the expected frequencies, while referential expressions (with a frequency of 6 and expected 4.1) were used more than expected. Based on the table, no significant difference was found in postgraduate students' use of the three codes of lexical bundles.

Table 8

Crosstabs for Codes of Lexical Bundles Used by Postgraduate Students

	Lexical bundles referential expressions	stance expressions	Discourse Organizing	Total
Count	6 _a	3 _a	46 _a	55
Expected Count	4.1	3.4	47.5	55.0
% within level	10.9%	5.5%	83.6%	100%

Table 9 includes the results for the distribution of compounds by postgraduate students. Considering the frequency with which any of the three compound types were used, the highest frequency (31) belonged to hyphenated forms, which constituted 79.5 percent of all the compounds used by postgraduate students. The other two categories (i.e., closed forms: 12.8% and open forms: 7.7%) together constituted 20.5 percent of the compounds by this group. While hyphenated forms and closed forms were used more than their expected counts, open forms occurred noticeably lower than expected. For hyphenated forms, the frequency was 31 (79.5%) and the expected count was 27.4, and for closed forms the frequency was 5, and the expected count was 4.1; however, the frequency of open forms was 3, which was quite lower than the expected frequency of 7.5. According to Table 9, considering the occurrence of compound codes in relation to their expected counts, open forms were used significantly less than closed and hyphenated forms.

Table 9*Crosstabs for Codes of Compounds Used by Postgraduate Students*

	Compounds			
	closed forms	hyphenated forms	open forms	Total
Count	5 _a	31 _a	3 _b	39
Expected Count	4.1	27.4	7.5	39.0
within Level	12.8%	79.5%	7.7%	100%

4.3. Comparison of the Types of FSs Used by Undergraduate and Postgraduate Students

To compare the use of FS types by undergraduate and postgraduate students, crosstabs within FS types were provided. Table 10 includes cross-tabulation for types of FSs used by undergraduate and postgraduate students. Since the comparison of frequencies without considering the significance of differences could lead to misinterpretation, the analysis included comparison of column proportions. Moreover, to decrease the probability of occurrence of type 1 error, the cross tabulation was provided with adjusted alpha. The different subscript letters in each row of Table 10 showed that the differences were significant with regard to expected counts.

Examination of cell frequencies in Table 10 indicated some similarities and some differences between undergraduate and postgraduate students in their use of FS types. The results showed that the two student groups' use of FSs was at similar levels; 47.7 percent of all the FSs used by the two groups were by undergraduate students and 52.3 percent were by the postgraduates. In addition, in both of the groups, collocations were used the most and idioms the least. Of all the FSs used by the undergraduates, 88.4 percent were collocations and 0.9 percent were idioms. For postgraduates, 73.3 percent of the FSs were collocations, while idioms were not employed at all. Moreover, in both undergraduate and postgraduate FS employments, only collocations had statistically significant differences with lexical bundles and compounds, and the other categories did not differ significantly within each level.

Comparison of the two student groups across FS types showed that the total frequency of collocations was almost evenly distributed between the two groups; 52.3 percent of the collocations belonged to undergraduate students, and 47.7 percent were by postgraduate students. However, noticeable differences were observed between the two groups in their use of idioms, lexical bundles, and

compounds. Undergraduate students used more idioms than postgraduates, while postgraduate students had higher employment of lexical bundles and compounds than the undergraduates. In fact, all the idioms used were by undergraduates; 68.7 percent of the lexical bundles were by postgraduates and 31.3 percent were by undergraduates; and 81.2 percent of the compounds were used by postgraduates and 18.8 percent by undergraduates.

Table 10

Crosstabs for the Interaction between Students' Level and their Use of FS types in L2 Writing

		FS types				Total	
		idioms	collocations	lexical bundles	compounds		
Level	under graduate	Count	3 _{a, b}	290 _b	26 _{a, c}	9 _c	328
		Expected Count	1.4	264.1	39.6	22.9	328.0
		% within Level	0.9%	88.4%	7.9%	2.7%	100%
		% within FS types	100%	52.3%	31.3%	18.8%	47.7%
	post graduate	Count	0 _{a, b}	264 _b	57 _{a, c}	39 _c	360
		Expected Count	1.6	289.9	43.4	25.1	360.0
		% within level	0.0%	73.3%	15.8%	10.8%	100%
		% within FS types	0.0%	47.7%	68.7%	81.2%	52.3%
	Total	Count	3	554	83	48	688
		% within FS types	100%	100%	100%	100%	100%
% of Total		0.4%	80.5%	12.1%	7.0%	100%	

In addition to the use of FS types, the study was interested in the use of codes of FSs across students' levels. For this purpose, cross-tabulation of data for the interaction between student level and FS codes was run. The results are presented in Table 11. Comparison of the two codes of collocations revealed that both undergraduate and postgraduate students used more grammatical than lexical collocations. However, comparison of the student groups within each FS code showed that postgraduates used more grammatical collocations (52.8%) than the undergraduates (47.2%), while the use of lexical collocations was higher among the undergraduates (63.6%) than the postgraduates (36.4%). With regard to lexical bundles, both groups of students used discourse organizing more than the other two lexical bundle codes. In addition, across-group comparisons indicated that postgraduate students used all three lexical bundle codes, that is, referential expressions (100%), stance expressions (60%), and discourse organizing (65.7%), more than the undergraduates. Finally, the results for compounds showed that while undergraduates used open forms the most and did not use closed forms at all, the

highest frequency of compounds in postgraduates belonged to hyphenated forms and the lowest to open forms. All the closed forms used were by postgraduates. Additionally, most of the hyphenated forms (93.9%) were used by postgraduates, and only 6.1 percent were used by undergraduates. Nonetheless, the open forms were more used by undergraduates (66.7%) than postgraduates (33.3%).

Table 11

Crosstabs for the Interaction between Students' Level and their Use of FS codes in L2 Writing

		FS codes									Total
		collocations		lexical bundles				compounds			
		grammatical collocation	lexical collocation	referential expression	stance expressions	discourse markers	closed form	hyphenated form	open form		
Level	postgradua	Count	179 _a	112 _b	0 _{a, b, c}	2 _{a, b, c}	24 _{a, c}	0 _{a, b, c}	2 _c	6 _{a, b}	328
	undergradua	Expected Count	198.7	92.3	1.9	1.6	22.5	9	5.6	1.5	333/1
		% within level	55.5%	33.8%	0.0%	0.6%	7.3%	0.0%	0.9%	1.8%	100%
		% within FS codes	47.2%	63.6%	.0%	40.0%	34.3%	.0%	6.1	66.7%	47.7%
	postgradua	Count	200 _a	64 _b	6 _{a, b, c}	3 _{a, b, c}	46 _{a, c}	5 _{a, b, c}	31 _c	33 _{a, b}	388
	undergradua	Expected Count	180.3	83.7	4.1	3.4	47.5	4.1	27.4	7.5	350/5
		% within level	56.1%	18.1%	1.7%	0.8%	12.5%	1.4%	8.6%	0.8%	100%
		% within FS codes	52.8%	36.4%	100%	60.0%	65.7%	%100	93.9%	33.3%	52.3%
	Total	Count	384	176	6	5	69	5	34	9	688
		% within FS codes	100 %	100%	100%	100%	100%	100%	100%	100%	100%
		% of Total	55.8%	25.6%	0.9%	0.7%	10.0%	0.7%	4.9%	1.3%	100%

5. Discussion

This study aimed to investigate the distributions of types of FSs that undergraduate and postgraduate students used in their L2 writings. Regarding the FS types used by undergraduate students, the results indicated that collocations occurred around 8 times more than the sum of the frequencies of the other FS types, and idioms had the lowest frequency. The findings further showed that these students used collocations and idioms more than expected, with no significant difference between the two types, and lexical bundles and compounds less than expected, with no significant difference between them. The results revealed significant differences between collocations and lexical bundles, collocations and compounds, and idioms and compounds.

Undergraduate students' noticeably close attention to collocations seems promising in light of studies emphasizing the positive role of knowing collocations

in L2 learners' production enhancement (e.g., Le-Thi et al., 2017; Nattinger & DeCarrico, 1992; Umair et al., 2023; Vilkaitė & Schimtt, 2019). Umair et al. (2023) suggested that knowing collocations assists learners in avoiding common errors in language use, such as awkward word combinations or incorrect word pairings. They added that L2 learners' correct use of collocations can enhance the accuracy of their language production. Moreover, some scholars (e.g., Vilkaitė & Schimtt, 2019) believe that since fluency depends greatly on collocations, learners learn to explore and absorb the collocations more in their speech and writing. On the other hand, the undergraduates' use of lexical bundles is less than expected, which is in line with Esfandiari and Barbary's (2017) finding that native and non-native English writers use lexical bundles in different ways in terms of quantity and quality, highlighting the need for more attention to their role in academic writing. Both idioms and collocations were used more than their expected counts by the undergraduate students; however, the frequency of collocations was much higher than that of idioms. This might be related to Bruening's (2020) distinction between idioms and collocations in terms of their difficulty level. As Bruening (2020) stated, idioms often have meanings that are not easily understood from the individual words, whereas collocations have a more predictable meaning based on the words used together. The finding that undergraduate students used idioms more than the expected count is encouraging because, as many research studies on idioms (e.g., Bagheri & Fazel, 2010; Kaya & Yilmaz, 2018) have suggested, learning idioms can enhance nonnative speakers' (NNSs) language efficiency.

Concerning the distribution of FS types in postgraduate students' writings, the most commonly used FSs were collocations, constituting around 73 percent of the FSs, while idioms were not used at all. Postgraduate students used lexical bundles and compounds more than expected, whereas their use of idioms and collocations was less than expected. The results showed that the use of collocations was significantly different from lexical bundles and compounds. The higher use of lexical bundles and compounds than their expected frequencies among the postgraduate students seems propitious, as some studies have emphasized that knowledge of lexical bundles (e.g., Esfandiari & Barbary, 2017; Qin, 2014) and knowledge of compounds (e.g., Dwivedi & Ghosh, 2023; Peloquin et al., 2020) help develop fluent linguistic production. Qin (2014) emphasized the use of lexical bundles in academic language, stating that L2 learners shouldn't neglect using lexical bundles in attempting to enhance their writing skills. Some authors (e.g., Dwivedi & Ghosh, 2023; Peloquin et al., 2020) have mentioned that compounds play a vital role in

expanding vocabulary and aiding character recognition. They believe that L2 learners with strong compound awareness can more easily break down complex characters into meaningful components and that this awareness helps improve their reading and writing skills and overall literacy development.

With regard to the comparison of undergraduate and postgraduate students' use of FSs, the results showed that the two groups used almost the same number of FSs and that for both groups, collocations had the highest frequencies and idioms the lowest frequencies. Moreover, it was revealed that within the FS types, undergraduate and postgraduate students used collocations with similar frequencies, but they differed in their use of other FS types. Idioms were more used by undergraduates, whereas lexical bundles and compounds were more employed by postgraduate than undergraduate students. Additionally, the results showed that undergraduate students used collocations and idioms more than expected and lexical bundles and compounds less than expected whereas postgraduate students used lexical bundles and compounds more than expected and idioms and collocations less than expected. The finding that undergraduate and postgraduate students used FSs at almost the same level is in contrast with Rafieyan's (2018) finding that FS knowledge can strongly predict learners' proficiency level. Comparing the knowledge of FSs in pre-intermediate, intermediate, and high intermediate students, Rafieyan (2018) found that the knowledge of FSs was the lowest for pre-intermediate and highest for post-intermediate students. The result also contradicts Duan and Shi's (2021) finding that perceived learners' proficiency significantly influenced learners' development of FSs.

The general finding of the study was that the various types of FS types and codes were used with different frequencies and were not compatible with their expected counts in both groups. This is despite the high importance that numerous studies in the literature (e.g., Le-Thi et al., 2017; Rafieyan, 2018; Esfandiari and Barbary, 2017; Ellis, 2012) have given to all FS types in L2 learner production. This finding suggests the need for more effort on the part of teachers to help develop L2 learners' use of all types of FSs. Teachers need to give learners more awareness of the contribution of these fixed structures to their writing fluency and to indicate the necessity of using all types of FSs rather than limiting themselves to just one or two categories of these sequences. This recommendation conforms to Wary's (2002) emphasis on the significance of restoring chunks of language and using all types of FSs to make learners' production more fluent and communicative.

6. Conclusion

The present study investigated L2 learners' use of FSs by undergraduates as compared with postgraduate students. It was revealed that despite differences in the distributions of FS types and codes by undergraduate and postgraduate students, the two groups were similar in terms of the frequencies of total FSs used in their writings and in their noticeably higher use of collocations than the other types of FSs. The results suggest that L2 teachers can direct their learners' attention to the usefulness of all types of FSs and include instruction of FSs in their writing instruction. Teachers' more emphasis on the significant role of FSs in L2 writing can help learners develop a better understanding of the essential role of different FS types in their writing enhancement and work on developing their knowledge and skill of using these sequences in their writing. In that case, we can observe more interaction between our L2 learners' academic level and their use of FSs so that higher level students use more formulaic language and use FSs of more variety than students at lower levels of education.

Similar to other studies, this study has a few limitations which should be acknowledge. The major constraint of the study was that it was delimited to only two branches of the IAU which limits the generalizability of the findings. A further drawback of the study which could influence the generalizability of the results, was that the participants were selected through convenience sampling. Another limitation was that for feasibility concerns only one writing by each student was analyzed while having at least two writings by each participant could have increased the reliability of the corpora. Finally, for practical reasons the recoding procedure of the data was limited to ten percent of the corpora which might increase the risk of coding bias.

In addition, the study has some suggestions for future studies. This study examined the use of FSs across academic levels. Other studies are suggested to investigate the use of FSs across various disciplines, proficiency levels, and learner gender. In addition, future studies are recommended to survey the awareness of learners and their perceptions about the role of FSs in L2 production progress. Moreover, studies on the effectiveness of instruction of FSs on learners' use of these sequences are recommended. Related studies comparing the effectiveness of teaching different FS types would be worth consideration. Additionally, teachers' knowledge of FSs, their awareness of the efficacy of FS instruction, and their perceptions about best practical ways of enhancing learners' implementation of

these expressions in their production are valuable related research issues.

Acknowledgments

We express gratitude to all the teachers and students who contributed to this research.

The authors have no conflict of interest to disclose.

This research received no external funding.

References

- Bagheri, M. S., & Fazel, I. (2010). Effects of etymological elaboration on the EFL learners' comprehension and retention of idioms. *Journal of Pan-Pacific Association of Applied Linguistics*, 14(1), 45–55.
- Benson, M., Benson, E., & Ilson, R. (2010). *The BBI combinatory dictionary of English: Word combinations and phrases*. John Benjamins Publishing.
- Birjandi, S., Shajari, R., Fouladi, A., & Rasulnia, A. H. (2021). The typology of formulaic speech in Charand o'Parand of Dehkhoda. *Language Related Research*, 12(6), 271–298. <https://doi.org/10.52547/LRR.12.6.9>
- Bruening, B. (2020). Idioms, collocations, and structure: Syntactic constraints on conventionalized expressions. *Natural Language & Linguistic Theory*, 38, 365–424.
- Casal, J. E., & Yoon, J. (2023). Frame-based formulaic features in L2 writing pedagogy: Variants, functions, and student writer perceptions in academic writing. *English for Specific Purposes*, 71, 102–114. <http://doi.org/10.1016/j.esp.2023.03.004>
- Derakhshan, A. (2019). The relationship between Iranian EFL learners' proficiency level and their knowledge of idiosyncratic and formulaic implicatures. *Language Related Research*, 10(5), 1–27. <https://dor.isc.ac/dor/20.1001.1.23223081.1398.10.5.12.2>
- Duan, S., & Shi, Z. (2021). A longitudinal study of formulaic sequence use in second language writing: Complex dynamic systems perspective. *Language Teaching Research*, 13621688211002942. <https://doi.org/10.1177/13621688211002942>

- Dwivedi, V., & Ghosh, S. (2023). Semantic relations classification in Hindi compound nouns using embeddings. *International Journal of Information Technology*, 15(6), 2969–2974.
- Ebrahimi, F., Namaziandost, E., Ziafar, M., & Ibna Seraj, P. M. (2021). The effect of teaching formulaic expressions through contrastive lexical approach on Iranian pre-intermediate EFL learners' writing skill. *Journal of Psycholinguistic Research*, 50, 1087–1105. <https://doi.org/10.1007/s10936-021-09778-z>
- Eidian, F., Gorjian, B., & Aghvami, F. (2013). The impact of lexical collocation instruction on developing writing skills among Iranian EFL learners. *International Journal of Language Learning and Applied Linguistics World*, 4(3), 273–283.
- Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. *The handbook of second language acquisition*, 63–103. <https://doi.org/10.1002/9780470756492.ch4>.
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17–44. <https://doi.org/10.1017/s0267190512000025>
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20, 29–62. <https://doi.org/10.1515/text.1.2000.20.1.29>
- Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes*, 29, 21–42. <https://doi.org/10.1016/j.jeap.2017.09.002>
- Fan, K., & Wang, H. (2024). The productive processing of formulaic sequences by second language learners in writing. *Frontiers in Psychology*, 15, 1281926. <https://doi.org/10.3389/fpsyg.2024.1281926>
- Gholami, L. (2022). Incidental corrective feedback provision for formulaic vs. non-formulaic errors: EFL teachers' beliefs and practices. *Language Awareness*, 31(1), 21–52. <https://doi.org/10.1080/09658416.2021.1943421>
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–192. <https://doi.org/10.1177/1461445605050365>
- Kaya, F. B., & Yilmaz, M. Y. (2018). The Frequency of using idioms in writing for

- the students learning Turkish as a foreign language. *Educational Research and Reviews*, 13(16), 602–608. <https://doi.org/10.5897/ERR2018.3585>.
- Le-Thi, D., Rodgers, M. P., & Pellicer-Sánchez, A. (2017). Teaching formulaic sequences in an English-language class: The effects of explicit instruction versus coursebook instruction. *TESL Canada Journal*, 34(3), 111–139. <https://doi.org/10.18806/tesl.v34i3.1276>
- Lewis, M. (2008). *The idiom principle in L2 English: Assessing elusive formulaic sequences as indicators of idiomaticity, fluency, and proficiency* [Unpublished Doctoral dissertation]. Engelska institutionen.
- Liou, H. C., & Chen, W. F. (2018). Effects of explicit instruction on learning academic formulaic sequences for EFL college learners' writing. *Taiwan Journal of TESOL*, 15(1), 61–100. <http://www.tjtesol.org/index.php/>
- Ma, R. (2020). Frequency effect of formulaic sequences on CAF in academic writing: Examples from L2 master's theses. *Chinese Journal of Applied Linguistics*, 43(4), 489–505. <https://doi.org/10.1515/cjal0032-2020->
- Nattinger, J. R., & De Carrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press.
- O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19. <https://doi.org/10.1177/1609406919899220>
- O'Donnell, M. B., Römer, U., & Ellis, N. C. (2013). The development of formulaic sequences in first and second language writing: Investigating effects of frequency, association, and native norm. *International Journal of Corpus Linguistics*, 18(1), 83–108. <https://doi.org/10.1075/ijcl.18.1.07odo>
- Ohlrogge, A. (2009). Formulaic expressions in intermediate EFL writing assessment. *Formulaic language*, 2, 375–386. <https://doi.org/10.1075/tsl.83.07ohl> 6
- Pellicer-Sánchez, A., & Boers, F. (2018). Pedagogical approaches to the teaching and learning of formulaic language. *Understanding formulaic language*, 153–173. <https://doi.org/10.4324/9781315206615-9>
- Peloquin, B. N., Goodman, N. D., & Frank, M. C. (2020). The interactions of rational, pragmatic agents lead to efficient language structure and use. *Topics in Cognitive Science*, 12(1), 433–445. <https://doi.org/10.1111/tops.12489>

- Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System*, 42, 220–231. <http://doi.org/10.1016/j.system.2013.12.003>
- Rafieyan, V. (2018). Role of knowledge of formulaic sequences in language proficiency: Significance and ideal method of instruction. *Asian-Pacific Journal of Second and Foreign Language Education*, 3(1), 1–23. <https://doi.org/10.1186/s40862-018-0050-6>
- Russell, T. S. (2017). *Finding the formula: Formulaic language use in Hong Kong primary school English textbooks* [Unpublished Doctoral dissertation]. University of Birmingham.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL IBT writing section. *Journal of English for Academic Purposes*, 12(3), 214–225. <https://doi.org/10.1016/j.jeap.2013.05.002>
- Umair, H. M., Imran, M., & Sarwat, S. (2023). An analysis of the use of collocations in English essay writing at undergraduate level. *Pakistan Languages and Humanities Review*, 7(1), 43–54. [http://doi.org/10.47205/plhr.2023\(7-I\)05](http://doi.org/10.47205/plhr.2023(7-I)05)
- Vilkaitė, L., & Schmitt, N. (2019). Reading collocations in an L2: Do collocation processing benefits extend to non-adjacent collocations? *Applied Linguistics*, 40(2), 329–354. <https://doi.org/10.1093/applin/amx030>
- Wang, Y. (2018). As Hill seems to suggest: Variability in formulaic sequences with interpersonal functions in L1 novice and expert academic writing. *Journal of English for Academic Purposes*, 33, 12–23. <https://doi.org/10.1016/j.jeap.2018.01.003>
- Wood, D. (2007). Mastering the English formula: Fluency development of Japanese learners in a study abroad context. *JALT journal*, 29(2), 209. <https://doi.org/10.37546/jaltjj29.2-3>
- Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. A&C Black. <https://doi.org/10.5040/9781474212069>
- Wood, D. (2019). Classifying and identifying formulaic language. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 30–45). Routledge. <https://doi.org/10.4324/9780429291586>

- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. <http://doi.org/10.1017/CBO9780511519772>
- Zhang, L., Zhang, L. J., & Xu, T. S. (2023). Effects of teacher explicit instruction in stance-taking on students' perceptions of stance and on their academic writing beliefs. *Frontiers in Psychology, 14*, 1140050. <https://doi.org/10.3389/fpsyg.2023.1140050>
- Zinkgraf, M., & Verdú, M. A. (2021). Dictogloss steals the show? Productive use of formulaic sequences by advanced learners. *Lexis. Journal in English Lexicology*, (18). <https://doi.org/10.4000/lexis.6079>
- Zimmerer, V. C., Wibrow, M., & Varley, R. A. (2016). Formulaic language in people with probable Alzheimer's disease: A frequency-based approach. *Journal of Alzheimer's Disease, 53*(3), 1145–1160. <https://doi.org/10.3233/jad-160099>

About the Authors

Servat Shirkhani holds a PhD in Applied Linguistics and is assistant professor of Applied Linguistics at Islamic Azad University, Iran. She has published books and articles in the field and has had presentations at national and international conferences. Her research interests include interlanguage pragmatics, teacher education, corrective feedback in language instruction, psychology of language learning, and technology- assisted language learning.

Samaneh Omid is a researcher and PhD candidate in Applied Linguistics. She is an English language teacher with over 14 years of experience in teaching at language institutes and universities. Her research interests include educational technology, AI-assisted language learning, and second language acquisition.