Journal of English Language Teaching and Applied Linguistics
ISSN: 2707-756X
DOI: 10.32996/jeltal
Journal Homepage: www.al-kindipublisher.com/index.php/jeltal

JELTAL
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Phrase-frames in Low- and Intermediate-level EFL Learners' Essays: Variability, Structures, and Functions

Qidi Li
School of International Studies, Zhejiang University, Hangzhou, China
Corresponding Author: Qidi Li, **E-mail**: li_qidi@zju.edu.cn

| ABSTRACT
A recent trend in formulaic language research has been investigating phrase-frames, which are discontinuous formulaic sequences with a variable slot. The current study aims to investigate the phrase-frames used by low- and intermediate-level learners of English as a foreign language (EFL). The phrase-frames are extracted from a self-compiled learner writing corpus across nine grades, and they are analyzed in terms of their variability, structures, and functions. The results show that as learners studied for more years, they would use phrase-frames that are more variable. Besides, they would use phrase-frames of different structures more flexibly, especially function word frames, and use phrase-frames for more varied functions. Through the research into the characteristics of phrase-frames and discussion about the reasons behind the differences between groups, this study contributes to a more comprehensive understanding of formulaic language development in low- and intermediate-level EFL learners and provides some insights into formulaic language teaching.

| KEYWORDS
Phrase-frames; Formulaic Language; EFL Writing; Writing Development; Chinese EFL Learners

| ARTICLE INFORMATION
ACCEPTED: 19 July 2023 **PUBLISHED:** 13 August 2023 **DOI:** 10.32996/jeltal.2023.5.3.7

1. Introduction

In recent decades, formulaic language has attracted great attention in the field of L2 writing. Formulaic language has been defined in different ways for different research purposes, and a relatively general definition of formulaic language might be "any multiword string that is perceived by the agent (i.e., learner, researcher, etc.) to have an identity or usefulness as a single lexical unit" (Wray, 2019, p. 267). It is widely acknowledged that formulaic language plays a vital role in acquisition, proficiency, and idiomaticity in both L1 and L2. (Ellis, 2006).

One of the issues that concern researchers in this field is the acquisition and usage of formulaic sequences by L2 learners of different proficiency levels or different L1 backgrounds. Studies have shown that L2 learners use less varied formulaic sequences than native writers and that they "tend to rely too heavily on a limited repertoire of phrases" (J. Li & Schmitt, 2009, p. 99). In previous studies, continuous formulaic language, like lexical bundles and collocation, is often regarded as a window into the characteristics of L2 learners' writing development. Most L2 writing studies, especially corpus-based studies, have focused on lexical bundles and collocation because they are "particularly well-suited to corpus analysis" (Durrant, 2022, p.141). Relevant research generally reveals that factors like the use of varied lexical bundles or highly associated collocations are positively associated with a variety of L2 proficiency metrics (Kim & Kessler, 2022; Eguchi & Kyle, 2023).

In recent years, phrase-frames (hereafter referred to as p-frames), as a special type of formulaic language, have received increasing attention in formulaic language research, especially in L2 writing settings. Different from continuous formulaic sequences, p-frames are discontinuous and are defined as "multi-word units with a variable slot in the internal positions" (Liu & Chen, 2022, p. 2), and

the word that can fit into the slot is regarded as a filler (Ren, 2021). For example, the sequence *it is * to* is a typical 4-word p-frame, and the filler in the slot can be an adjective like *important*, *meaningful*, *impossible*, *hard*, etc. There have been studies on p-frame usage by English as a foreign language (EFL) learners, mostly across proficiency levels or of different L1 backgrounds (Garner, 2016; Geluso, 2022; Ren, 2022; Tan & Römer, 2022). However, few studies have focused on the characteristics of p-frames used by beginners or low- and intermediate-level EFL learners. Based on a self-compiled corpus of Chinese low- and intermediate-level EFL learners' essays, the current study intends to fill the research gap by exploring the variability, structures, and functions of p-frames used by these learners across three grade groups. By doing so, this study hopes to add to our understanding of p-frames in EFL learners' essays and provide some pedagogical help for formulaic language teaching and learning.

2. Literature review

The earliest studies on p-frames mainly focused on pre-selected discontinuous sequences. For example, Renouf & Sinclair (1991) selected 7 different frameworks with an intermediate word as their research target, like *a * of* and *be * to*. After some corpus analysis tools became more widely used, more studies expanded their research targets to a wider range of p-frames that appeared in texts.

In recent years, studies on p-frames can be largely divided into four types. First, some scholars focused on differences across L2 proficiency levels (Garner, 2016; Tan & Römer, 2022). Tan and Römer (2022) traced the language development of Chinese EFL learners across proficiency levels through 3-and 4-p-frames. They carried out the analysis in terms of the variability, predictability, and functions of p-frames, which is also a commonly adopted research pattern in p-frame studies. According to their analysis, in L2 learners' writing, there are usually a higher number of relatively fixed p-frames. Besides, as learners' proficiency increases, p-frames would become less predictable, and learners would move closer to L1 writing. In terms of functional characteristics of p-frames, the most frequently used 3-word p-frames across all proficiency levels were referential expressions, while the main functions of 4-word proficiency levels differed among proficiency levels, perhaps due to the influence of different writing tasks. Second, many scholars investigated how L2 learners used p-frames differently from native English students. Ren (2022) compared the 5-word p-frames used in the essays of native and non-native English students. No statistically significant inter-corporal differences were found in terms of p-frame types and tokens, but native English students used a larger number of fixed and predictable p-frames. However, the non-native English students that participated in this study had spent years learning in UK-based universities, which might distinguish them from L2 learners studying in non-English environments. Geluso (2022) picked out a more specific type of p-frames, i.e., preposition-based phrase frames, and focused on their grammatical complexity and functional characteristics. The third type of p-frame analysis focuses on L2 learners of different native languages, in which native language transfer of L1 is often regarded as a crucial factor accounting for the differences in using p-frames (O'Donnell et al., 2013; Juknevičienė & Grabowski, 2018). The last type focuses on characteristics of p-frames in a specific type of register, like academic prose (Gray & Biber, 2013), pharmacy (Grabowski, 2015), research article introductions (Lu et al., 2018), business emails (Xia et al., 2023), etc. Some studies were not restricted to only one type of research goal mentioned above and included various factors that might influence the use of p-frames.

When analyzing p-frames, most scholars focused on the following dimensions: variability, predictability, structures, and functions. Variability and predictability are usually categorized according to the type-token ratio (TTR) and normalized entropy (H_{norm}), respectively, which can both be automatically calculated in corpus analysis tools like AntConc (Anthony, 2022). To analyze the structures of p-frames, many studies have drawn on the three-way classification system put forward by Gray & Biber (2013), namely verb-based frames, content word frames, and function word frames (see Section 3.2.3 for more detailed descriptions). Functional analysis is also vital for capturing the characteristics of p-frames since "it is those functions which may help one distinguish between less and more advanced writers" (Juknevičienė & Grabowski, 2018, p. 305). Until now, there has been no functional taxonomy developed, especially for p-frames, and most previous studies adopted the taxonomy established by Biber et al. (2004), which was originally designed for analyzing lexical bundles. In most cases, p-frames realize the same function despite the variable word (Garner, 2016). However, it is still not uncommon to see p-frames that fulfill different functions with different fillers. When determining the specific function of p-frames, there are also different opinions. Some adopt the variant-based approach (Römer, 2010; Lu et al., 2018; Tan & Römer, 2022), taking both the variant and the context in which it occurs into consideration. In contrast, some studies assign functional labels to p-frames based on the function of the fixed components only, without considering the variant or the context (Grabowski, 2015). It is to avoid cases where a p-frame is assigned multiple functions.

Though the measures mentioned above have formed a generally accepted research pattern in p-frames studies, there were also some scholars who questioned the effectiveness of certain dimensions. For example, some studies simply compared the predictability of p-frames across proficiency levels. However, Tan and Römer (2022) pointed out that certain p-frames (e.g., *as * as*) were supposed to be much less predictable than other p-frames in which a range of nouns occur (e.g., *the * is*). Therefore, it should be more careful when choosing the characteristics to be investigated and deciding how to measure the characteristics.

The characteristics of p-frames used by L2 learners have been extensively explored. So far, however, there has been little discussion about how Chinese EFL learners of low and intermediate level use this special type of formulaic sequence. Besides, though the dimensions like variability, structures, and functions of p-frames have been repeatedly discussed, most studies stopped at describing the differences only, and the reasons behind the differences as well as their pedagogical implications, have received little attention. Nevertheless, this deserves our profound investigation since the differences may reveal the issues that exist in our teaching methodology.

In response to this scarcity, the current study compares the variability, structures, and functions of the p-frames extracted from our self-built corpus and discusses the reasons behind the differences (if any). Specifically, three research questions will be addressed:

1. How does the variability of p-frames vary across studying stages for low- and intermediate-level Chinese EFL learners?
2. How do the structures of p-frames vary across studying stages for low- and intermediate-level Chinese EFL learners?
3. How do the functions of p-frames vary across studying stages for low- and intermediate-level Chinese EFL learners?

3. Methods

3.1 Corpus

The corpus we used for this study is a self-compiled corpus of 896 student essays, which were written by students from an elementary school and a middle school in Zhejiang Province in eastern China. The students ranged from the fourth grade of elementary school to the twelfth grade of middle school, and they can be regarded as beginner and intermediate EFL learners (Jiang et al., 2019). The primary school learners were required to write essays based on the picture they saw, and the topics included "my teacher", "my favorite season", "my hobbies", "my classroom", etc. The essays were all narrative essays. Junior and senior school learners wrote essays according to the essay prompt, and they were required to write narrative essays like "One Interesting/ Unforgettable Weekend/ Festival" or argumentative essays like "Should Everyone Learn to Play a Musical Instrument". Table 1 presents the distribution of essays in each grade. In the current study, all the essays were divided into three groups according to the studying stage, i.e., primary group (G4, G5, G6), junior group (G7, G8, G9), and senior group (G10, G11, G12).

Table 1: Distribution of essays in each grade

Grade Level	Number of compositions	Total words	The mean length of each text
G4	144	7,439	52
G5	145	8,480	58
G6	167	12,271	73
Total (primary group)	456	28,170	62
G7	86	8,167	95
G8	110	13,983	127
G9	70	7,413	106
Total (junior group)	266	29,560	111
G10	40	5,224	130
G11	45	5,483	122
G12	89	12,274	138
Total (senior group)	174	22,981	132

3.2 Procedure

3.2.1 Automatic p-frame extraction

To identify the p-frames in the corpus, the automated text analysis tool AntConc was used. Before generating the p-frame list, the length of the p-frames should be taken into consideration. To make sure that the generated p-frame list is within a manageable size for manual categorization and concordance checks (Chen & Baker, 2010), sequences of less than 4 words were excluded from this study. Meanwhile, 5-word p-frames often cross phrase boundaries and therefore do not form complete and meaningful units, especially for L2 learners at lower proficiency levels (Tan & Römer, 2022). Therefore, the current study focused on p-frames that were four words long. Besides, only p-frames with an inner slot were considered in this study because 4-word p-frames with an initial or final slot could be regarded as 3-word lexical bundles rather than a p-frame. Based on these considerations, we generated a list of 4-word p-frames with an inner slot for each group using AntConc.

3.2.2 Manual revision

After automatic p-frame extraction, we manually checked the lists and excluded the p-frames that were not suitable for this study. First, p-frames broken by punctuation were all removed (e.g., *very * I like* in *Summer is very beautiful. I like summer.*). Second, variants were checked for spelling and grammatical errors. For example, the filler *like* does not form a grammatical sentence in the p-frame *I * play football*. Considering that the current study does not consider errors in the use of p-frames, all the hits of *I like play football* were corrected into *I like playing football* so that the variability of p-frames would be more authentic. After manual revision, we finally got the list of the most frequently used 100 p-frames for each group. All these p-frames occurred over 4 times across over 4 different essays, which met the criteria for lexical bundle analysis (Liu & Chen, 2022).

3.2.3 Manual annotation

In order to address the second and third research questions, manual annotation was required for the identification of the structures and functions of the p-frames.

For structure identification, this study adopted the three-way classification system put forward by Gray & Biber (2013). The first type is verb-based frames, which contain one or more modal, auxiliary, or main verbs (e.g., I like * best, we can * a, the main * is, want to * in). The second is content word frames, which refer to frames without verbs but with other content words, including one or more nouns, adjectives, or adverbs (e.g., they * very happy, I * lots of, on the * floor, a good * to). The last type is function word frames, which consist of only function words such as prepositions, determiners, conjunctions, pronouns, complementizers, etc. (e.g., the * of the, if you * to, the * in my, at the * of). This classification system covered all the p-frames that occurred in the corpus.

For the functions of p-frames, most previous studies were carried out in the framework established by Biber et al. (2004), which was primarily used as a functional taxonomy for lexical bundles. Under this taxonomy, functions of lexical bundles were divided into four categories, namely referential expressions, stance expressions, discourse organizers, and special conversational expressions. However, when manually annotating the functions, we found that there were expressions that could not fit any of the four aZAcategories, and special conversational expressions were not found in the current corpus, perhaps due to the genre of the essays. Therefore, we decided to adopt the functional taxonomy used by Tan and Römer (2022), who added a fifth function, i.e., activity expression, to the most widely used system established by Biber et al. (2004). When identifying the main function of a p-frame, we adopted the variant-based approach. In cases where the p-frame might perform multiple functions, we followed Tan and Römer's (2022) pattern and assigned the p-frame the function performed by its most frequent set of variants.

4. Results

4.1 Variability

To assess the variability of all the p-frames extracted, we referred to the type-token ratio (TTR) provided by AntConc. The most frequently used 100 p-frames at each level were placed into three variability categories, i.e., fixed, variable, and highly variable, according to their TTR value. The number of 4-frames in each variability category is presented in Table 2. As the numbers show, the three groups mostly differed in terms of the number of fixed and highly variable frames. As learners studied for more years, there were fewer fixed p-frames and more highly variable p-frames. The distribution of TTR values was statistically significantly different across groups (χ^2=29.106, p= .000 < .05). To further determine pairwise significant differences, we conducted post-hoc Chi-square tests with a Bonferroni correction. Significant pairwise differences were found between every two groups (for the primary and junior groups, χ^2=6.950, p= .032 < .05; for the primary and senior groups, χ^2=21.583, p=.000 < .05; for the junior and senior groups, χ^2=11.009, p= .004 < .05).

Table 2: Number of 4-frames in each variability category

	Primary	Junior	Senior	χ^2	p
Fixed (<=.30)	33	17	10		
Variable (>.30 and <=.70)	50	64	50	29.106	.000
Highly Variable (>.70)	17	19	40		

4.2 Structural aspect

Table 3 shows the distribution of 4-frames by structural categories across the three groups. It is obvious that verb-based frames accounted for the most part among all the three groups, but senior school students, compared with primary and junior school students, were more flexible in applying p-frames of different structures. To be more specific, senior school students used more function word frames. The Chi-square test showed that a significant difference was found among the three groups (χ^2=20.940, p=.000 < .05). Post-hoc Chi-square tests also proved that there were significant differences between each group except between the primary and junior groups (for the primary and junior groups, χ^2=5.564, p= .062 > .05; for the primary and senior groups, χ^2=19.130, p= .000 < .05; for the junior and senior groups, χ^2=7.842, p= .020 < .05). These statistics showed that from primary

group to senior group, students could flexibly use p-frames of more types of structure and that they would use function word frames more often in their essays. Besides, such improvement might be more significant after they entered senior school.

Table 3: Number of 4-frames in each structural category across groups

	Primary	Junior	Senior	χ^2	p
Verb-based frames	80	72	53		
Content word frames	15	13	24	20.940	.000
Function word frames	5	15	23		

To better investigate the structural characteristics of p-frames used by Chinese learners, we referred to a corpus of native English students' writing as our reference corpus. The corpus we chose was the *Growth in Grammar* (GiG) corpus, which is composed of 2,898 texts written by 983 children at schools in England. The children's ages range from 6 to 16, being roughly consistent with the ages of the students who participated in the current study. We randomly selected 1000 texts from the GIG corpus and got the most frequently used 100 p-frames in the same way as discussed in Section 3.2. by AntConc. Then we also manually annotated the structures of the p-frames used by native English students and compared the results with that of the senior group (see Table 4). As Table 4 shows, the English children used function word frames most often, which accounted for 49% of all the frames investigated. Meanwhile, a significant difference was still found between the GIG corpus and the senior group (χ^2=28.608, p=.000 < .05). The results indicated that even the senior school students still underused function word frames compared with native English students.

Table 4: Number of 4-frames in each structural category in GIG and the senior group

	GIG	Senior	χ^2	p
Verb-based frames	27	53		
Content word frames	5	24	28.608	.000
Function word frames	49	23		

4.3 Functional use

As is shown in Table 5, L2 learners of the three groups used p-frames mainly for different functions. For the primary group, learners used referential expressions and stance expressions most often (accounting for 54% and 34% of the total, respectively). Meanwhile, there were only 1 discourse organizing expression and 11 activity expressions. For the junior group, the numbers of stance expressions, discourse organizing expressions, and activity expressions all witnessed a slight increase, but the number of the latter two types of expressions still accounted for only a small part of all the expressions (31%). In contrast, in the senior group, we can see a largely even distribution of the four types of expressions. According to the Chi-square test, a significant difference was found among the three groups (χ^2=36.225, p= .000) and between each group except between the junior and senior groups (for the primary and junior groups, χ^2=14.633, p= .002 < .05; for the primary and senior groups, χ^2=28.337, p= .000 < .05; for the junior and senior groups, χ^2=7.416, p= .060 > .05). The results indicated that as learners studied English for more years, they would use p-frames for more varied functions in their essays. Meanwhile, the junior and senior school learners used p-frames largely in a similar pattern in terms of functions.

Table 5 Number of 4-frames in each functional category across groups

	Primary	Junior	Senior	χ^2	p
Referential expressions	53	31	26		
Stance expressions	35	38	33	36.225	.000
Discourse organizing expressions	1	5	17		
Activity expressions	11	26	24		

5. Discussion

The purpose of the current study was to find out the characteristics of p-frames used by Chinese L2 learners in terms of variability, structures, and functions. The materials came from a self-compiled corpus of essays written by low- and intermediate-level Chinese EFL learners.

The first research question concerns the variability of p-frames. After categorizing the p-frames according to their TTR value, this study found that as learners move on to higher studying stages, the fillers that fill the slots would be more varied. This finding is consistent with previous studies that at higher proficiency levels, the number of variables and highly variable p-frames would

increase (Garner, 2016; Tan & Römer, 2022). This is often attributed to the fact that as L2 writers' proficiency level improves, "their phraseological inventory becomes more varied and productive" (Tan & Römer, 2022, p. 8). It can also explain why the learners of the current corpus could use p-frames in a more variable way as they study English for more years. From the input perspective, as the learners move on to the next studying stage, they will be exposed to more complex materials, which could add to their phraseological inventory. From the output perspective, learners in higher studying stages would be more likely to create essays that consist of complex vocabulary, which may also lead to higher variability of p-frames used in the essays.

The second research question focuses on the structure of p-frames, i.e., verb-based frames, content word frames, and function word frames. This study finds that as L2 learners study English for more years, they would use more varied types of p-frames, and such difference showed most significantly in terms of function word frames. This finding contradicts previous findings that little to no difference is found in the number of verb-based frames, content word frames, and function word frames used by German EFL learners across proficiency levels (Garner, 2016). This may be attributed to the influence of L1 transfer. German, like English, has a huge resource of formulaic sequences of different word types. Meanwhile, in Chinese, we do not stress word types of formulaic sequences, especially function words. However, the current finding is in line with Garner's (2016) conclusion that L2 learners, compared with native English students, tend to underuse small function words. According to De Wilde's (2023) research into young L2 English learners' narrative writing, function words, which are often high-frequency words in students' essays, might be difficult for L2 learners, especially beginners. Therefore, for the learners in our primary group, who have been studying English for only 1 to 3 years, function words might be an obstacle to authentic L2 writing. As the students move up to higher studying stages, they would be in better command of function words and therefore show decreasing reliance on content verbs and nouns, which might lead to more function word frames and fewer verb-based and content word frames in their essays.

The last research question centers on the functions of p-frames. As is shown in Table 5, learners of the primary group used referential expressions and stance expressions most often and underused discourse organizing expressions and activity expressions. As they studied for more years, they showed a greater tendency to use p-frames for more varied functions. For the senior group, the p-frames showed a roughly even distribution among the four types of function, indicating that the p-frames they use are more functionally diverse. The finding is consistent with previous studies that for L2 learners, the overwhelming majority of p-frames were referential (Garner, 2016), and such a tendency is more significant among beginners. Garner (2016) attributed such a tendency to the influence of essay prompts, which can also explain the difference in functions detected in the current study. As introduced in Section 3.1, primary school learners were required to write essays based on the picture they saw, and the topics included "my teacher", "my favorite season", "my hobbies", "my classroom", etc. Therefore, their essays were mostly descriptive, which may result in a large number of referential expressions. Meanwhile, as for learners of junior and senior groups, their essay prompts would be more complex and require them to write under different types of registers. For example, the junior and senior school students were asked to write a letter to their friends or express their opinions on whether children should do housework at home. In this case, they would organize their words to clarify their intentions or express their opinions. Besides, writing strategies might also be an important factor that increased the number of discourse organizing expressions in the senior group. According to Kim and Kessler (2022), there are usually more discourse-organizing bundles that structure the essay in higher-scoring essays. Their finding can also explain the difference in the number of discourse organizing expressions we found among the three groups in the current study.

6. Conclusion

Based on a self-compiled learner corpus, this study has demonstrated how low- and intermediate-level Chinese EFL learners used p-frames in their essays at different studying stages. The analysis was carried out in terms of the variability, structures, and functions of p-frames. An external corpus of native English children's essays was also referred to when necessary to further investigate the characteristics of learners' use of p-frames. The comparative analysis has shown that as learners study for more years, they would be able to use more variable p-frames. The structures and the functions fulfilled by the p-frames would also be more flexible and diversified.

The present study hopes to contribute to a more comprehensive understanding of the ways in which p-frame competence of low- and intermediate-level EFL learners develops. It also reveals some issues in the English teaching process, which are worth our profound reflection. First, EFL learners' ability to use p-frames of varied structures grew slowly at the lower level, i.e., from primary school to junior school stages. This reminds us that we should use materials that contain more diversified p-frame structures in teaching, especially function word structures. Second, though learners at higher grades were better at using function word frames, they were still not skilled enough compared with native English students when using function word frames, or "small function words" as referred to in some studies (Juknevičienė & Grabowski, 2018). This should be paid more attention to since it is vital for authentic and native-like writing output.

However, there are still some limitations in the current study. First, although the materials in this study are authentic and representative of low- and intermediate-level Chinese EFL learners, the corpus size, compared with studies that utilized huge datasets with open access, is still relatively small and, therefore, may not reveal more detailed information about learners' usage of p-frames. Second, this study discussed the general structures and functions of 4-word p-frames only, while some more fine-grained measures may provide new insights into this topic. For example, Ren (2022) found that the position of fillers might be relevant to the function of the whole frame. Specifically, A*CDE p-frames were mainly content word frames, while for AB*DE p-frames, function word p-frames far outnumbered the other two structural categories. Besides, the word type of the filler might also be worth studying to figure out the whole picture of p-frames usage by EFL learners (Juknevičienė & Grabowski, 2018). These dimensions deserve further exploration in future studies on this topic.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Anthony, L. (2022). AntConc (Version 4.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software.

[2] Biber, D., Conrad, S., & Cortes, V. (2004). If you look at . . .: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, *25*(3), 371–405. https://doi.org/10.1093/applin/25.3.371

[3] Chen, Y., & Baker, P. K. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning & Technology*, *14*(2), 30–49. http://llt.msu.edu/vol14num2/chenbaker.pdf

[4] De Wilde, V. (2023). Lexical characteristics of young L2 English learners' narrative writing at the start of formal instruction. *Journal of Second Language Writing*, *59*, 100960. https://doi.org/10.1016/j.jslw.2022.100960

[5] Durrant, P. (2022). *Corpus Linguistics for Writing Development: A Guide for Research*. Taylor & Francis.

[6] Eguchi, M., & Kyle, K. (2023). L2 collocation profiles and their relationship with vocabulary proficiency: A learner corpus approach. *Journal of Second Language Writing*, *60*, 100975. https://doi.org/10.1016/j.jslw.2023.100975

[7] Ellis, N. C. (2006). Cognitive perspectives on SLA: The associative-cognitive CREED. In Kathleen Bardovi-Harlig & Zoltán Dörnyei (eds.), *Themes in SLA research (AILA Review 19)*, 100–121. Amsterdam: John Benjamins. https://doi.org/10.1075/aila.19.08ell

[8] Geluso, J. (2022). Grammatical and functional characteristics of preposition-based phrase frames in English argumentative essays by L1 English and Spanish speakers. *Journal of English for Academic Purposes*, *55*, 101072. https://doi.org/10.1016/j.jeap.2021.101072

[9] Garner, J. W. (2016). A phrase-frame approach to investigating phraseology in learner writing across proficiency levels. *International Journal of Learner Corpus Research*, *2*(1), 31–67. https://doi.org/10.1075/ijlcr.2.1.02gar

[10] Grabowski, Ł. (2015). Phrase frames in English pharmaceutical discourse: a corpus-driven study of intra-disciplinary register variation. *Research in Language*, *13*(3), 266–291. https://doi.org/10.1515/rela-2015-0025

[11] Jiang, J., Bi, P., & Liu, H. (2019). Syntactic complexity development in the writings of EFL learners: Insights from a dependency syntactically-annotated corpus. *Journal of Second Language Writing*, *46*, 100666. https://doi.org/10.1016/j.jslw.2019.100666

[12] Juknevičienė, R., & Grabowski, Ł. (2018). Comparing Formulaicity of Learner Writing through Phrase-Frames: A Corpus-Driven Study of Lithuanian and Polish EFL Student Writing. *Research in Language*, *16*(3), 303–323. https://doi.org/10.2478/rela-2018-0016

[13] Kim, S. S., & Kessler, M. J. (2022). Examining L2 English University students' uses of lexical bundles and their relationship to writing quality. *Assessing Writing*, *51*, 100589. https://doi.org/10.1016/j.asw.2021.100589

[14] Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, *18*(2), 85–102. https://doi.org/10.1016/j.jslw.2009.02.001

[15] Li, X., & Yu, H. (2021). Parataxis or hypotaxis? Choices of taxis in Chinese–English translation. *Lingua*, *251*, 103026. https://doi.org/10.1016/j.lingua.2020.103026

[16] Liu, C., & Chen, H. H. (2022). A phraseological exploration of university lectures through phrase frames. *Journal of English for Academic Purposes*, *58*, 101135. https://doi.org/10.1016/j.jeap.2022.101135

[17] Lu, X., Yoon, J., & Kisselev, O. (2018). A phrase-frame list for social science research article introductions. *Journal of English for Academic Purposes*, *36*, 76–85. https://doi.org/10.1016/j.jeap.2018.09.004

[18] O'Donnell, M., Römer, U., & Ellis, N. C. (2013). The development of formulaic sequences in first and second language writing. *International Journal of Corpus Linguistics*, *18*(1), 83–108. https://doi.org/10.1075/ijcl.18.1.07odo

[19] Ren, J. (2022). A comparative study of the phrase frames used in the essays of native and nonnative English students. *Lingua*, *274*, 103376. https://doi.org/10.1016/j.lingua.2022.103376

[20] Renouf, A. & Sinclair, J. McH. 1991. "Collocational frameworks in English". In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics*. London: Longman, 128–143.

[21] Römer, U. (2010). Establishing the phraseological profile of a text type. *English Text Construction*, *3*(1), 95–119. https://doi.org/10.1075/etc.3.1.06rom

[22] Tan, Y., & Römer, U. (2022). Using phrase-frames to trace the language development of L1 Chinese learners of English. *System*, *108*, 102844. https://doi.org/10.1016/j.system.2022.102844

[23] Wray, A. (2019). Concluding question: Why don't second language learners more proactively target formulaic sequences? In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language*. Routledge. https://doi.org/10.4324/9781315206615-14

[24] Xia, D., Sulzer, M. A., & Pae, H. K. (2023). Phrase-frames in business emails: a contrast between learners of business English and working professionals. *Text & Talk*. https://doi.org/10.1515/text-2022-0027