ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0071

A Review of Formulaic Language and Second Language Writing

Xueshan Zhang[1, a]

[1]Hongkong Polytechnic University, Hongkong, China

[a]Email: zhxs19872016@163.com

Abstract

As the importance of formulaic language in second language acquisition and production, the number of research on this topic has been increasing. This paper made a thorough review of one of its sub-categories: formulaic language and second language writing. The review unfolded from three aspects: investigating the use of formulaic language in second language writing, formulaic language and the assessment of second language writing, and influence of instruction of formulaic language on second language writing.

Keywords

Formulaic language, Second language writing; Writing assessment; Formulaic language instruction.

1. Introduction

The number of studies focusing on formulaic language is on the rise (Wood, 2015). More and more researchers recognized that formulaic language plays a crucial role in both first and second language learning, processing, and use (Sinclair 1991; Biber and Barbieri, 2007; Wray, 1999; Wray, 2002; Martinez and Schmitt, 2012; Appel and Wood, 2016). Therefore, studies on this topic have been conducted to investigate various aspects, such as formulaic language and first language acquisition and production, formulaic language and second language acquisition and production, and formulaic language and aphasia (Wray, 2002). This paper reviews studies focusing on formulaic language and second language production, to be more precise, formulaic language and second language writing. Based on the comprehensive review, pedagogical implications and further research directions are discussed at the final.

2. Background Information About Formulaic Language

Wray (2002) pointed out that technical terms used in formulaic language studies were confusing, as the same term could refer to different things, and different terms could refer to the same phenomenon in different studies (p. 9). Wray (2002) listed more than fifty terms related to the formulaic language that had been used by researchers, such as unanalyzed chunks of speech, lexical bundles, collocations, chunks, n-grams, idioms, and multiword items. Wray (2002) chose the neutral term formulaic sequence: A sequence of words, which can be continuous or discontinuous, tends to be stored and retrieved as a whole, instead of being generated or analyzed by grammar rules (p. 9). In a similar vein, Wood (2015) pointed out that formulaic sequence referred to specific items while formulaic language means a collective of these items (p. 2).

It is well noted that linguists have studied formulaic language for a long time. Fillmore (1979) suggested that most people's language ability relied on the mastery of formulaic utterances" (p. 92). In 1991, Sinclair proposed two principles of language interpretation: the open-choice and the idiom principle. The open choice principle referred to the process of people in which people made choices according to grammatical rules, while the idiom principle meant that people tended to choose the minimal effort way: treating prefabricated units as single choices (p. 110).

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0071

Conklin and Schmitt (2012) indicated that formulaic sequences could be stored in people's long-term memory without the process of selecting lexicon and applying grammatical rules (p. 45).

Based on previous studies, the importance of formulaic language can be discussed in the following three aspects (Martinez and Schmitt, 2012): First, formulaic language exists commonly in language use. Biber et al. (1999) found that lexical bundles constituted 30 % of tokens in their conversation corpus and about 21% in their academic writing corpus. Later, Erman and Warren (2000) reported that formulaic consequences occupied 58.6% of spoken English and 52.3% of written English. With Wmatrix as the analysis tool, Rayson (2008) suggested that 15 percent of text contained formulaic expressions. Although different researchers used various methodologies to come up with different numbers, they concluded that formulaic language was ubiquitous and critical for second language learning. Second, formulaic language has obvious processing advantages and helps people improve speech accuracy and fluency. When people memorize and retrieve formulaic language as a whole and do not have to analyze them into components, they can sound more native-like with fewer efforts and spare more attention to meaning (Wray, 2002). Ellis and Sinclair (1996) advocated that formulaic language was crucial for native and non-native speakers' language fluency. Third, formulaic language can perform functions in speech and writing to help speakers and writers express ideas effectively. Biber, Conrad, and Cortes (2004) summarized its functions into three types: referential, stance, and discourse organizing functions (p. 386). Referential lexical bundles are used to identify crucial characteristics of one aspect of identity, such as time, place. Stance lexical bundles are adopted to express the position, feelings, and attitudes of language users. Discourse organizers show the logical relations between sequential paragraphs, clauses complexes, and simple clauses.

3. Studies on Formulaic Language and Second Language Writing

As has been mentioned before, researchers have studied formulaic language from various perspectives. One of the sub-categories focuses on formulaic language and second language writing. As many second language teachers and students have attested to, writing is one of the most challenging skills to acquire in second language learning. Thus studies on formulaic language and second language writing are critical and necessary. Research about formulaic language in second language writing could fall into three categories: variations in the use of formulaic language between native and non-native speakers with different language proficiency, formulaic language as a writing assessment criterion, and effects of formulaic language instructions on second language writing. Studies related to each category will be reviewed in the following paragraphs.

3.1. Use of Formulaic Language in Second Language Writing

A large number of studies in this area were cross-sectional, while others were longitudinal. It should be noted that besides this research design distinction, according to Jaworska, Krummes, and Ensslin (2015), studies in this area have usually adopted two different approaches: category-based and lexical bundle approach. The category-based approach detects formulaic language on prescribed standards, which means studies with this approach focus on specific structures with defined parts of speech for their components, such as Adj+noun, Adv+adj. The other approach, lexical bundle or distributional approach, depends on corpus-driven methodology and automatic extraction tools to retrieve high-frequency strings of words. In conclusion, the first approach can present a detailed and in-depth analysis about one definite structure use, while the second way can provide a broader picture about the use of formulaic language in language users' production. Therefore, some researchers choose to combine these two ways.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0071

Besides, four other prominent features about research in this area should be mentioned: First, a more significant proportion of studies in this area focused on English as the second language, while a couple of studies centered on other languages as a second language, such as Dutch, German. Moreover, among studies that paid attention to English as the second language, participants in different studies were with various native language backgrounds, such as Chinese, Hebrew, and Spanish. Second, types of second language writing were diverse, such as academic writing, test-takers writing for different tests. Third, some studies only concentrated on advanced second language learners while others conducted comparisons across learners with different language proficiency. Finally, studies with a distributional approach collected data from a wide range of corpora, employing various software tools and statistical procedures.

3.1.1. Cross-sectional Studies

Cross-sectional studies on this topic provided a static comparative picture of the actual use of formulaic language between different language variables. Moreover, most of the studies adopted the corpus-driven or distributional approach, whereas only several exceptional studies preferred the category-based approach. Laufer and Waldman (2011) adopted the category-based approach, examining the use of verb+noun collocation in English argumentative and descriptive essays written by Hebrew learners with three different language proficiency: basic, intermediate, and advanced. Results showed that learners at all levels tended to use fewer verb+noun collocations compared with native writers at the same age. Moreover, learners at intermediate and advanced levels provided a more significant number of deviant usage of verb+noun collocations. In a nutshell, this study indicated that even advanced L2 learners failed to possess appropriate knowledge of verb+noun collocations and master their use. However, this study was confined to investigate the use of the specific collocation structure in argumentative essays. Thus its results lacked the ability of generalization.

Chen and Baker (2010) designed a study with a combination of the two main approaches. Firstly, they adopted an automatic retrieval tool to obtain four-word lexical bundles to conduct the quantitative analysis. Then, they followed the category-based approach to perform an in-depth qualitative analysis. The purpose of their study was to investigate differences in the use of English four-word lexical bundles among published academic articles by native researchers, native students' writing (L1), and non-native students' writing (L2). The researcher reported that published academic articles used the largest number of lexical bundles while writing by L2 students used the least number of lexical bundles. In terms of lexical bundles' structure, native and non-native students' writing tended to use more VP-based strings while native researchers preferred NP-based bundles. Their findings indicated that the preference for NP-based bundles could be treated as an indicator of high-quality articles. As for functions of lexical bundles, native academics preferred more referential expressions, whereas students' writing included a larger number of discourse organizers. It can be concluded that the findings in this study were comprehensive and complicated because of its thorough qualitative analysis of lexical bundles in terms of their structure and function, which was rare in previous research. However, this study only investigated the performance of l2 learners at higher language levels. Another study, which also focused on advanced English learners, but with Spanish as participants' first language background, was conducted by Pérez-Llantada (2014). The researcher aimed to investigate the extent to which formulaic language in second language learner's published writings was native-like, with a corpus-driven approach. This study compared the use of automatically extracted four-word bundles, which was the same as the above study, for three sets of texts: English articles written by native researchers (L1 English), English articles written by Spanish researchers (L2 English), and Spanish articles written by native scholars (L1 Spanish). The analysis suggested that lexical bundles were crucial in all academic writing variables, and the choice of lexical bundles was confined to registers. In addition, a significantly

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0071

larger number of lexical bundles were found in L2 English and L1 Spanish. Meanwhile, some lexical bundles in L2 English articles were deviant from those of L1 English writing in terms of structure and function. In terms of structure, the most prominent type in L1 English writing was " it-clause fragment," while in L2 English articles, it was prepositional phrases. L1 English preferred a wide range of stance expressions that could not be found in L2 English in terms of function. It can be inferred that although both these two studies focused on learners at higher language proficiency levels, their findings were inconsistent. This inconsistency may be due to their participants' L1 background, data extraction, and statistical procedures. It was important to mention that this study shed some light on pedagogy: a genre-based approach for writing training was essential for it could raise learners' consciousness of particular groups of lexical bundles being appropriate for different registers.

Another study focusing on English as a second language was conducted by Staples, Egbert, Biber, and McClair (2013). Instead of limiting to advanced second language learners, their study compared the usage of formulaic language between native and non-native speakers with three different proficiency levels in terms of frequency, function, and fixedness. The study revealed that learners with higher language proficiency tended to use fewer lexical bundles, while the lowest level learners produced the largest number. However, further analysis showed that if lexical bundles in the prompt were excluded, the second-level learners use the widest range of lexical bundles. In terms of function, no significant differences existed across the different language proficiencies, and learners tended to prefer more stance and discourse organizing ones. This finding was in line with that of Chen and Baker (2010), which revealed that second language learners encountered more difficulties in producing referential lexical bundles. Moreover, as for proportions of fixed and variable lexical bundles, the three groups made similar results. However, the fixedness tested in this study was controversial, as the variable lexical bundles in this concept were similar to three-word bundles.

In a similar vein, Appel and Wood (2016) conducted another study. They investigated the use of four to seven-word lexical bundles in academic articles produced by low and high-proficiency language learners. Results indicated that learners with low language proficiency tended to be more dependent on lexical bundles, which was similar to the finding reported by Staples et al. (2013). Furthermore, they reported that learners with high proficiency tended to use more referential strings while learners with low proficiency preferred stance and textual organization lexical bundles. This finding was partially inconsistent with that of Staples et al. (2013).

One more empirical study focusing on advanced learners of German was undertaken by Jaworska, Krummes, and Ensslin (2015). They compared the use of formulaic language in argumentative essays written by native and advanced British learners of German (L2 German), and the findings in this study were similar to that of Pérez-Llantada's (2014) research. They indicated that L2 German tended to use more three-word bundles than L1 German. Additionally, L2 German preferred stance expressions to textual function expressions.

It should be noted that although these studies adopted various automatic techniques to extract lexical bundles from different corpora containing non-native learners' articles and then compared them with data from parallel corpora with a wide range of statistical procedures, they represented overall pictures: leaners encountered difficulties in using native-like formulaic language. They might under-use and over-use some lexical bundles; they might also creatively use collocations that native writers never used. Given these existed problems, more research is needed to explore factors affecting formulaic language use to help teachers provide effective instructions.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0071

3.1.2. Longitudinal Studies

As some researchers argued that cross-sectional studies were static and failed to explore the dynamic development of formulaic language use in second language writing, they advocated that more longitudinal studies were needed. However, as longitudinal studies required more time and energy, involving uncontrollable variables, the number of longitudinal studies was smaller compared to cross-sectional research.

Li and Schmitt (2010) conducted a study lasting a whole academic year to investigate the development of four Chinese advanced English learners' use of English collocations in their academic writing. With BNC academic written corpus as the parallel corpus, it was found that no real change of collocation use had been detected in this process. However, with further analysis, an individual's use of formulaic sequences varied significantly. Although this study was confined to a limited number of writing samples, it still raised researchers' awareness of individual variations in formulaic language development. Moreover, because of the mismatch of registers between students' assignments and articles in the BNC academic written corpus, the BNC academic written corpus was not parallel. More recently, another longitudinal study lasted two and half years. Duan and Shi (2021) undertook the study analyzing 155 articles of 31 Chinese college students majoring in English. It should be noted that this study aimed at a larger sample size compared with Li and Schmitt (2010). After a thorough and complicated data analysis, researchers pointed out that the whole group failed to show changes in the frequency of formulaic language use. However, individuals' development of formulaic language use varied significantly. This result provided additional evidence to the results of Li and Schmitt (2010). Moreover, they also concluded that students used the VP-based lexical bundles decreasingly with longer language learning time, which was consistent with the results of Chen and Baker (2010).

While these two studies focused on the advanced or upper-medium English learners, another longitudinal study concentrated on beginners of Italian with more positive results. Siyanova-Chanturia (2015) conducted a longitudinal study to investigate the development of the use of the structure of N+Adj by thirty-six Chinese learners of Italian. It was suggested that compared with the articles written by beginners at the early stage, later articles constituted a more significant number of high frequent and strong associated N+Adj structures, which indicated that learners improved their ability to use native-like N+Adj collocations significantly through this five-month study.

Another quasi-longitudinal study was conducted by Huang (2015) in Chinese universities to overcome the disadvantages of true longitudinal studies. The researcher compared the usage of three to five lexical bundles in 5590 argumentative essays written by English major students in junior and senior year in terms of frequency and accuracy. Results indicated that the number of lexical bundles used by seniors was much larger than that of juniors. However, with further analysis, seniors failed to show a higher level of accuracy in lexical bundle usage, which suggested that the quality of lexical bundle usage was not improved accordingly as their numbers increased. These findings indicated that more effective instructions about the accurate usage of lexical bundles were needed for learners.

These longitudinal studies showed a general tendency: the development of formulaic language use was comparatively slow, and individuals within groups showed significant variance in their ability to use lexical bundles. More research was needed to explore the reasons for these distinctions. Moreover, advanced learners failed to improve the accuracy of their formulaic language use, preventing their production from becoming more native-like.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0071

4. Formulaic Language and the Assessment of Second Language Writing

Due to the vital role of formulaic language in second language writing, some researchers tried to incorporate the frequency and accuracy of formulaic language use as one predictor of writing quality. Bestgen and Granger (2014) selected COCA (Corpus of Contemporary American English) as the reference corpus to investigate the relationship between learner's use of collocations and the quality of their writing. They reported that a positive correlation existed between students' proper use of bigrams and their article scores, indicating that the learner's ability to use two-word collocations was effective for writing assessment. Bestgen (2017) conducted another study with reference to the British National Corpus (BNC). In this study, the researcher compared the formulaic measures and lexical richness measures for articles assessments. Based on two other learner corpora with error annotation and scores, the researcher suggested that compared with lexical richness measures, formulaic measures could predicate the quality of second language writing more accurately. However, both of these two studies focused on two-words bundles. Therefore further analysis for other types of formulaic language was needed.

5. Effects of Instruction in Formulaic Language on Second Language Writing

Given the vital role of the use of formulaic language in developing writing skills for second language learners and the fact that second language learners' use of formulaic language was deviant from that of native speakers, some researchers began to explore whether instructions of formulaic language could help learners extend their formulaic language repertoire. El-Dakhs, Prue, and Ijaz (2017) performed a study to compare university students' use of formulaic language in essays of re-writing stories. The experiment design included a pre-instruction writing task, writing tasks immediately after instruction, and a final writing task without instruction. Results indicated that the explicit instruction of formulaic language helped students in the experiment group produce a wider range of formulaic language in their following writing tasks overall. However, no significant variance of formulaic language use existed between the pre-experiment writing and the final independent writing. In a nutshell, students' automatic use of formulaic language failed to change through this ten-week program. In other words, second language learners encountered difficulties in using formulaic language without explicit instruction. Another study conducted by Akkoç, Qin, and Karabacak (2018) provided more positive results with participants at upper-intermediate to advanced language proficiency. They were from a turkey university, including both freshmen and sophomore students. The experiment group received explicit instruction in target formulaic language by teacher's detailed and thorough explanations about the meaning, completing related exercises. On the other side, the control group was trained by the traditional mode, in which students listened to presentations about how to write argumentative essays, completing writing tasks without instructions related to formulaic language. The analysis of pre-test, post-test, and delayed post-test results indicated that the intervention was effective in improving students' use of target formulaic language and student's writing quality. As students produced less target formulaic language in the delayed post-test, it can be concluded that the influence of explicit instruction in formulaic language was limited to the specific task. Instruction of formulaic language in these two studies failed to improve students' independent ability of formulaic language use. More research should be conducted to concentrate on using the instruction of formulaic language to help students improve the ability of automatic formulaic language use.

It can be included that for these three above categories, more studies are needed for the final two, as their limited number of studies and problems waiting to be solved mentioned before. Besides these further research suggestions, several pedagogical implications could be

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0071

summarized from existed studies. First, as formulaic language plays a critical role in second language writing, teachers are advised to include high-frequency and strong collocated formulaic language in their course syllables to raise students' awareness of using formulaic language. Research showed that second language failed to detect the strong collocated formulaic language with low frequency. Therefore instructors could explain n-grams of this type in great detail purposely. Second, it is true that registers make distinctions in formulaic language use. Studies indicated that second language learners lacked the awareness of register distinction. In this case, instructors need to provide a solid knowledge of register distinction to learners. Third, for advanced learners, mastering a wider range of formulaic language and their accurate use are critical to break the fossilization state. Thus, more effective and particular instruction of formulaic language use should be devised for learners at higher proficiency.

References

[1] Akkoç, A. B., Qin, J., & Karabacak, E. (2018). The effects of explicit instruction of formulaic language on EFL argumentative writing quality. Indonesian Journal of Applied Linguistics, 8(2), 358-368.

[2] Appel, R., & Wood, D. (2016). Recurrent word combinations in EAP test-taker writing: Differences between high-and low-proficiency levels. Language Assessment Quarterly,13(1), 55-71.

[3] Bestgen, Y. (2017). Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence. System, 69, 65-78.

[4] Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. Journal of Second Language Writing, 26, 28-41.

[5] Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. English for Specific Purposes, 26(3), 263–286.

[6] Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan. 1999. Longman Grammar of Spoken and Written English. Longman.

[7] Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. Language learning & technology, 14(2), 30-49.

[8] Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. Annual Review of Applied Linguistics, 32, 45.

[9] Duan, S., & Shi, Z. (2021). A longitudinal study of formulaic sequence use in second language writing: Complex dynamic systems perspective. Language Teaching Research.

[10] El-Dakhs, D. A. S., Prue, T. T., & Ijaz, A. (2017). The effect of the explicit instruction of formulaic sequences in pre-writing vocabulary activities on foreign language writing. International Journal of Applied Linguistics and English Literature, 6(4), 21-31.

[11] Ellis, N. C. and S. G. Sinclair. 1996. 'Working memory in the acquisition of vocabulary and syntax: putting language in good order. The Quarterly Journal of Experimental Psychology 49(1), 234–50

[12] Erman, B. and B. Warren. 2000. The idiom principle and the open choice principle. Text, 20(1), 29–62.

[13] Fillmore, C.J. 1979. On fluency. In C.J. Fillmore, D. Kempler & S.-Y.W.Wang (Ed.), Individual differences in language ability & language behavior (pp. 85–101). New York: Academic Press.

[14] Huang, K. (2015). More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. System, 53, 13-23.

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

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0071

[16] Jaworska, S., Krummes, C., & Ensslin, A. (2015). Formulaic sequences in native and non-native argumentative writing in German. International Journal of Corpus Linguistics, 20(4), 500-525.

[17] Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach. In Wood, D. (Ed.), Perspectives on formulaic language: Acquisition and communication (pp. 22-46). Continumm.

[18] Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. Applied linguistics, 33(3), 299-320.

[19] Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. Journal of English for Academic Purposes, 14, 84-94.

[20] Rayson, P. (2008). Software demonstration: Identification of multiword expressions with Wmatrix. Paper presented at the Formulaic Language Research Network (FLaRN) conference, Nottingham, UK.

[21] Saussure, F.De. 1966. Course in general linguistics. New York: McGraw-Hill.

[22] Sinclair, J. (1991). Corpus, concordance, collocation. Oxford University Press.

[23] Siyanova-Chanturia, A. (2015). Collocation in beginner learner writing: A longitudinal study. System, 53, 148-160.

[24] 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.

[25] Wood, D. (2015). Fundamentals of formulaic language: An introduction. Bloomsbury Publishing.

[26] Wray, A. (1999). Formulaic language in learners and native speakers. Language teaching, 32(4), 213-231.

[27] Wray, A. (2002). Formulaic language and the lexicon. Cambridge University Press.