

# *SLA: A Formulaic Language Perspective*

Bing-yan ZHU

Department of foreign languages & tourism, Hainan College  
of Software Technology,  
Qionghai 571400, China

Shi-xiang LIU\*

Department of foreign languages & tourism, Hainan College  
of Software Technology,  
Qionghai 571400, China  
\*xiangzi0223@126.com

**Abstract**—Formulaic languages are prefabricated multi-word patterns, continuous or discontinuous, which are stored and retrieved whole from memory at the time of use. Both EFL and native speakers' natural language makes considerable use of formulaic language. The formulaic languages, which were noticed 50 years ago, are featured with high frequency, structural cohesion, semantic unity, and variation in form. The past 20 years has witnessed remarkable achievements made in formulaic language-related research home and abroad. Based on related studies, this paper aims to explore how English is acquired and bettered with the help of formulaic language from the perspective of SLA, which may be of implication for English teaching and learning.

**Keywords**—Formulaic Language; SLA; English learners; EFL

## I. INTRODUCTION

The ready-made chunks or multi-word units, i.e. Formulaic language, first observed by John Hughlings Jackson whose research was on the utterance by aphasic patients in the mid-nineteenth century, have been paid successive attention to since the last century by many well-known scholars abroad. For instance, Saussure proposes that “synthesizing the elements of [a] syntagm into a new unit...the mind gives up analysis-it takes a short-cut and applies the concept to the whole cluster of signs, which then becomes a simple unit” [1]. Hymes argues that “a vast portion of verbal behavior...consists of recurrent patterns, of linguistic routines” [2]. Fillmore claims that “an enormously large amount of natural language is formulaic, automatic, and rehearsed, rather than propositional, creative, or freely generated” [3]. Bolinger insists that “speakers do at least as much remembering as they do putting together” [4]. Therefore, Verstraten holds the frequent use of formulaic language is a fairly good indication of the command of L2 [5]. If formulaic language plays such an important role in communication, it will be reasonable to assume that idiomatic selection and fluency of proficient English speakers rest to a considerable extent on knowledge of a body of “institutionalized” or “lexicalized” formulaic language. Since the 1990s, theories on formulaic language have been gradually introduced into China, but few SLA-related studies have been carried out from the perspective of formulaic language.

## II. LITERATURE REVIEW

Wray concludes that 57 terms are documented in literature concerning the same phenomenon, for which proposing a

comprehensive definition remains one of the foremost problems. Wray argues “formulaic language is too commonly used in the literature to be free of such association” [6]. Therefore, she adopts formulaic sequence, for she holds that “formulaic carries with it some association of ‘unity’ and of ‘custom’ and ‘habit’, while sequence indicates that there is more than one discernible internal unit, of whatever kind”. Therefore, Wray defines formulaic language as [6]:

*...a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.*

Formulaic language should be prefabricated multi-word units, used more frequently, structurally coherent and semantically complete. In other words, formulaic language is a prefabricated multi-word unit, retrieved more frequent than chance from memory as a whole, which is coherent structurally, fixed or semi-fixed and complete semantically, compositional or non-compositional, rather than being subject to generation or analysis by the language grammar. Therefore, the following aspects must be taken into consideration: (1) Frequency and collocability. The fact of “probabilities” refers to whether the combination is considered to be formulaic language, though the frequency of co-occurrence can't qualify the formulaic status necessarily. (2) Structural cohesion. Formulaic languages are the recurring sequences or complete syntactic units with complete grammatical structure and relatively well-structured. (3) Semantic unity. On the one hand, most formulaic languages are idiomatic, which suggests that most formulaic languages are native speakers' conventional ways of speaking. On the other hand, all of the formulaic languages are not idioms, but semantic unity is required though not necessarily non-compositional. (4) Flexibility and variation. The dichotomy of fixed and semi-fixed is not suitable. More often, formulaic languages are widely described as “fixed”. However, they are not all “fixed”, but “the continuum model” with varying degrees, such as fixed FS “in a word”, semi-fixed FS “kick the bucket”, which allows variation of tense. Therefore, it is universally recognized formulaic languages, as form-meaning pairings, not only represent the semantic content but also comprise its conventionalized functions, for which formulaic

languages are of great importance for EFL learners to notice as many as possible.

Formulaic language is classified in accordance with its function, from, combination of grammatical structure and function. Yorio conducts a function-based classification: (1) Situational formulas for conversational parameters, e.g. *How do you do*; (2) Stylistic formulas in particular registers of language, e.g. *in conclusion*; (3) Ceremonial formulas, required by certain formal settings, e.g. Ladies and gentlemen; (4) Gambits to organize interactions or some activities, e.g. *It's your turn*[8]. Nattinger & DeCarrico identify three functional categories: (1) Social interaction markers describing social relations, including: conversational maintenance such as summoning, e.g. *what's up*, and conversational purpose, such as offering, e.g. *can I help you*. (2) Necessary topic markers used frequently at the beginning of conversation, such as *My name is\_\_*. (3) Discourse devices which connect the meanings and structures in the discourse, such as *as a result of\_\_*. But this taxonomy is only conversation-oriented, so it is not perfectly suitable for the present study on monologue[8]. As for the from-centered taxonomy, Nattinger & DeCarrico identify four structural categories of formulaic languages, different somewhat in length and grammatical status, canonical or non-canonical, fixed or semi-fixed, continuous or discontinuous: (1) polywords (e.g. *by and large*); (2) institutionalized expressions (e.g. *how are you*); (3) phrasal constraints (e.g. *as far as\_\_*); (4) sentence builders (e.g. *our goal is that\_\_*)[8]. However, Nattinger & DeCarrico realize that their taxonomy of structural categories is a little fuzzy. Combining form and function, Altenberg distinguishes three broad categories of formulaic languages: full clauses, clause constituents and incomplete phrases. In view of the degree of grammatical structures' completeness, full clauses are further classified into dependent clauses and independent clauses while clause constituents are further classified into multiple clause constituents and single clause constituents[9]. With regard to the function formulaic languages perform in discourse, Altenberg further classifies each structural sub-category of formulaic languages in accordance with the embedded function in structural sub-categories[9]. Biber *et al.* classify "lexical bundles" or formulaic languages into 14 major structural categories, though most of them do not represent complete structural units[10]. Based on the most frequent bundles and their functions, a classification scheme includes: referential bundles, text organizers, stance bundles and interactional bundles, each with several sub-categories, closely connected to the functions. Erman & Warren divide prefabs into four categories: lexical, grammatical, pragmatic, and reducible[11]. Lexical and pragmatic prefabs are further classified according to syntactic and functional features respectively. According to the structural properties, lexical prefabs are divided into phrase and clause type. And pragmatic prefabs or the highly conventionalized formulaic language, correspond to specific interactional situations, such as *enjoy your meal* to describe the process of eating between guests. Erman & Warren's scheme broadens the scope of formulaic language research, especially the formulaic language with meta-linguistic functions, such as *and everything* and *sort of*.

It is universally recognized formulaic languages, as form-meaning pairings, not only represent the semantic content but also comprise their conventionalized functions. Nattinger & DeCarrico who treat them as form-function composites, claim that formulaic languages are also endowed with functional meanings[8]. Therefore, these formulaic languages "not only have syntactic shapes, but are capable as well of performing pragmatic acts" [8], for which they can help to lower the difficulty when the speakers retrieve them as a whole from their memory to enrich their written and spoken English. With regard to the semantic feature of formulaic languages from the perspective of corpus linguistics, Sinclair defines them as "extended units of meaning"[12-14]. The term "a progressive de-lexicalization" also refers to the phenomenon that individual words are not always units of meaning, such as *"have"* in the phrase *"have a bath"*, with much of its meaning lost[12]. Instead, the whole meaning spreads across the whole phrase rather than be restricted to one word or another. Hunston & Francis claim that "most of words have no meaning in isolation, or at least are very ambiguous, but have meaning when they occur in a particular phraseology"[15]. Thanks to the description of formulaic languages as "extended units of meaning", we can focus on formulaic languages other than isolated words, and more pedagogical implications have been found in SLA to better EFL teaching and learning.

### III. SLA AND FORMULAIC LANGUAGE

Formulaic languages or formulaic frames with analyzed slots are the basis for creative speech, which serve as "formulaic frames with open slots or slot-and-frame patterns" for creative discourse[16]. Similar to the children's second language acquisition, English learners would easier acquire a foreign language in the form of formulaic languages, imitate again and again, internalize its rules, master the syntactic and grammatical rules, and evolve onto the stage for creative use.

#### A. To memorize formulaic languages consciously or unconsciously with acquisition of their semantic and pragmatic functions for whole retrieval in similar context

Altenberg conclude that formulaic language accounts for 70 percent in native speaker's daily communication[9]. English learners in particular for those in preliminary stage or with poor English depend heavier on the formulaic languages, for English learners are lacking in native-like communication context. Therefore, a wide repertoire of formulaic language is the basis for selective imitation to internalize the rules to chain the formulaic patterns. For instance, the sentence *He has to look up the words in the dictionary* can be segmented into "He || has to || look up the words || in the dictionary" in accordance with the categorization and frequency in discourse. Theoretically, the formulaic chunks *has to*, *look up*, *look up the words*, *in the dictionary*, *look up the words in the dictionary* could "evolve directly into creative language" on condition that the English learners could combine the formulaic languages with certain contexts[17]. For instance, frequently adopting in imitation of native speakers *has to* may contribute to evolving into *have to*, *look up* into *look for*, *look around*, *look at*, *look back*, *look forward to*, *look into*, *look out*, *look through*, *look like*, and so on. If they can frequently applying them to the daily discourse, they should be on the phase of "whole memorized and whole

retrieved” on phrase level without knowing its components separately. For instance, “大雨” can be translated into “big rain” according to semantic memory of some rules. But supposing the phrase “heavy rain” has met several times before (emphasizing frequency) and been considered whole (focusing on noticing), “大雨” may be translated into “heavy rain” based on episodic memory, for which we hold formulaic language like “heavy rain” can be memorized and retrieved whole rather than generated based on rules. Sinclair argues that countless combinational possibilities allowed by grammar are always replaced by frequently used lexical elements in practice[12]. He proposes two principles to interpret the mechanism of language operation: the *open-choice principle* and the *idiom principle*. The open-choice principle, often called “slot-and-filler” model, treats texts “as a series of slots which have to be filled from a lexicon which satisfies local restraints. At each slot, virtually any word can occur” [12]. However, it is obvious in practice that words don’t occur at random, but are arranged in a certain way, so “the open-choice principle does not provide for substantial enough restraints on consecutive choices”[12]. To account for the restraints missed by the open-choice principle, the idiom principle is put forward, which describes that “a language user has available a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments” [12]. The idiom principle, incompatible with the open-choice principle, makes the selection of two or more words together possible, based on their previous and regular co-occurrences. Therefore, it can be concluded that formulaic languages are memorized more easily, and frequently used formulaic languages which are context-dependent would be acquired during the application, lowering the difficulty in English learning.

However, Pawley and Syder argue that formulaic languages uttered by native speakers are not easy to be identified and mastered by non-native speakers, due to the fact that their English doesn’t sound idiomatic due to the absence of formulaic languages[18]. And they find that the native speakers would pause when moving on to the next clause, while the pauses seldom occur in the middle of utterance. What’s more, an even more important aspect concerning formulaic languages is their context-dependent features. Pawley argues that formulaic languages are not random piecing-together words, but “combinations of familiarity and flexibility”, usually used “in close association with the situation in which their use is appropriate”[19]. Myles *et al.* hold FSs “embrace the notion of a multi-morphemic unit memorized and recalled as a whole, rather than generated from individual items based on linguistic rules” [20], which also suggests the processing system is context dependent.

To some extent, the stylistic, syntactic and grammatical changes must be adjusted to a certain situation, as Hakuta hold that formulaic languages occur when learners are forced to make creative speeches to meet communicative needs before they are capable of that[21]. Therefore, they turn to ready-made routines, a shortcut to realize social interactions with insufficient linguistic competence, due to which formulaic languages are considered as social interaction stimuli.

B. To internalize the rules inherent in a wide range of formulaic languages for the phase of “whole memorized and part retrieved” to string together the formulaic chunks with “clause-chaining” style

Pawley and Syder come up with two puzzles, concerning the ability to routinely convey meanings and the ability to produce fluent stretches of spontaneous connected discourse that exceeds human limits in encoding novel speech[18]. At length, they conclude that formulaic languages are indispensable for native-like fluency, for native speakers prefer a “clause-chaining” style to grammatical rules, i.e. the native speakers “string together a sequence of relatively independent clauses, clauses which show little structural integration with earlier or later constructions”[18]. After investigating 16 adults who learn the future tense “will” and “going to” varying from 7 months to 17 and half months, Bardovi\_Harlig finds that the formulaic chunk *I am going to write (about)* is taken as a unit and frequently used at the beginning of the composition[22]. And “going to” serves as a relatively independent construction followed by *write a report* or *read a passage*, which makes formulaic language creative. Based on the data collected from COLSEC, Wei Naixing claims that Chinese English learners who depend on “open-choice principle” other than “idiom principle” for their speech, are poor in chaining strategy like native speakers[23]. And the variety, the total number, and the complexity of formulaic languages they use reflect that the ability to use chaining strategy by EFL students is far from enough.

In practice, we ask our students to build their own formulaic banks in accordance with various subjects, such as offer, order, counter offer, and acceptance in business. Taking business correspondence format for an example, the students are asked to list the formulaic chunks concerning reply at the beginning of the inquiry letter, e.g. *Thank you for your letter of (date) asking if/requiring about...We have received your letter of (date) enclosing/concerning...* Similarly, the formulaic chunks concerning the finding of business report can be imitated again and again: *It was found that... It was discovered that... It was generally the case that...Several staff members expressed the view that...Many staff members suggest that...It was felt that...* Whether in class or not, students are required to write letters in the role sellers or buyers for a whole semester. Practice with frequent retrieval of formulaic language makes students write the business correspondence quicker and more native-like. Skehan argues that using formulaic languages “buy[s] the native speaker time while using language, by reducing the density of information packed into speech”, as “they would ease the operation...and make it more feasible to operate” [24] Therefore, to have a repertoire of knowledge on formulaic languages and the ability to use them in a “clause-chaining” style with less encoding work can be a trigger for native-like use of the target language by English learners, for formulaic languages’ usage “bears the authority of regular and accepted use by members of the speech community” and their conjunction of the meaning and the lexicalized formal expression “forms the social institution”[18]. To some extent, English learners’ fluency or accuracy depends heavily on how many formulaic languages other than the total grammar, for the formulaic language is “grammaticalised lexis” [25] and “a

single lexicogramatical force”[26] to help English learners to use discourse creatively.

On the other hand, the ability to use a certain language is not taught but acquired according to Krashen’s Input Hypothesis on condition that enough input is comprehensible, interesting and relevant, not grammatically sequenced, which is slightly higher than that of the input level (using  $i+1$  to indicate the higher difficulty than the original level of linguistic competence) in order to ensure the multiplier effect of input[27]. From the viewpoint of the traditional English teachers, it is difficult to implement personalized teaching to meet the needs of each learner, for the students’ language basis, learning methods, and learning strategy are various. Therefore, guiding students while storing and retrieving formulaic languages with clause-chaining style in accordance with their interest contributes to internalizing the rules inherent in formulaic languages.

#### IV. CONCLUSION

Formulaic languages, as recurrent combinations of words with various syntactic-semantic structures, combine the grammar and the lexicon. On the other hand, most formulaic languages are context-dependent. Pawley argues that formulaic languages are not random piecing-together words, but “combinations of familiarity and flexibility”, usually used “in close association with the situation in which their use is appropriate” [19]. Therefore, it can be concluded that formulaic language plays an important role in native speakers’ and English learners’ acquisition in particular. They are ready-made routines, serving as a shortcut to realize social communications with insufficient linguistic competence, for the fact that it is idiom-based holistic processing which can better English learners’ oral or written English. A wide repertoire of formulaic languages appropriate in certain contexts can help EFL learners evolve into creative discourse with flexible change in the frame and its components, for which it is of great value for English learners to input as many formulaic chunks as possible to acquire a foreign language in a more natural way.

#### REFERENCES

- [1] Saussure, F. De.Course in general linguistics. New York: McGraw-Hill,1966.
- [2] Hymes, D.H. The ethnography of speaking. In J. A. Fishman (ed.). Readings in the Sociology of Language.The Hague/Paris: Mouton,1968, pp. 99-138.
- [3] Fillmore, C. J. The Need for a Frame Semantics in Linguistics. Statistical methods in linguistics, 12, 1976, pp.5-29.
- [4] Bolinger, D. Meaning and Memory. Forum Linguisticum, 1, 1976, pp.1-14.
- [5] Verstraten, L. Fixed Phrases in Monolingual Learners’ Dictionaries. In P.J.L. Arnaud & H. Bejoint (eds.). Vocabulary and Applied Linguistics. Basingstoke: Macmillan,1992, pp.28-40.
- [6] Wray, A. .Formulaic language and the lexicon. Cambridge: Cambridge University Press, 2002.
- [7] Yorio, C. A. Idiomaticity as an Indicator of Second Language Proficiency. In K. Hyltenstam & L. K. Obler (eds.). Bilingualism across the Lifespan. Cambridge: Cambridge University Press,1980, pp.55-72.
- [8] Nattinger, J. R. & DeCarrico, J. S. Lexical phrases and language teaching. Oxford: Oxford University Press,1992.

- [9] Altenberg, B. On the phraseology of spoken English: the evidence of recurrent word-combinations. In Cowie, A.P. (ed.), Phraseology: Theory, Analysis and Applications. Oxford: Oxford University Press,1998, pp. 101-122.
- [10] Biber, D., Johansson, S., S., Leech, G., Conrad, S., & Finegan, E. Longman grammar of spoken and written English. London: Longman,1999.
- [11] Erman, B. & B. Warren. The Idiom Principle and the Open Choice Principle. Text, 20/1,2000, pp.29-62.
- [12] Sinclair, J. Corpus, concordance, collocation. Oxford: OUP,1991.
- [13] Sinclair, J. Review of the Longman Grammar of Spoken and Written English. International Journal of Corpus Linguistics, 6/2,2001, pp. 339-359.
- [14] Sinclair, J.Trust the text. London: Routledge, 2004.
- [15] Hunston, S. & Francis, G. Pattern grammar: A corpus-driven approach to the lexical grammar of English. Amsterdam: Benjamins, 2000.
- [16] Ellis,N. Constructions,chunking,and connectionism: The emergence of second language structure. In Doughty,C . and M,Long (eds) . Handbook in SLA. Oxford: Blackwell,2003, pp.63-103.
- [17] Wong-Fillmore,L The Second Time Around: Cognitive and Social Strategies in Second Language Acquisition . Doctoral Dissertation,Stanford University,1976: 640.
- [18] Pawley, A. & Syder, F. H. Two Puzzles for Linguistic Theory: Native-like Selection and Native-like Fluency. In Richards, J. & R. Schmit (eds). Language and Communication. London: Longman,1983, pp. 191-226.
- [19] Pawley, A. On Speech Formulas and Linguistic Competence. Linguas Modernas, 12,1985, pp. 84-104.
- [20] Myles F., Hooper, J. & Mitchell. R. Rote or Rule? Exploring the Role of Formulaic Language in Classroom Foreign Language Learning. Language Learning, 1998, 48(3), pp.323-368.
- [21] Hakuta, K. Prefabricated Patterns and the Emergence of Structure in Second Language Acquisition. Language Learning, 24/2, 1974, pp.287-297.
- [22] Bardovi\_Harlig, K.A New Starting Point? Investigating Formulaic Use and Input in Future Expression. Studies in Second Language Acquisition 24/2, 2002, pp.189-198.
- [23] Wei Naixing. A Preliminary Study of the Characteristics of Chinese Learners’ Spoken English. Modern Foreign Languages, 2004, pp. 140-149.
- [24] Skehan, P. A cognitive approach to language learning. Oxford: OUP,1998.
- [25] Schmitt,N. Vocabulary in Language Teaching. Cambridge: Cambridge University Press,2000.
- [26] Lewis,M . The Lexical Approach. Hove and London,England: Language Teaching Publications,1993.
- [27] Krashen, S. The input hypothesis: issues and implications. London: Longman,1985.