Formulaic Sequences and the Implications for Second Language Learning

Qi Xu[1]

[1] Guangdong University of Foreign Studies, China

Correspondence: Qi Xu, Guangdong University of Foreign Studies, Guangzhou 510420, China. E-mail: xqmiracle@gmail.com

Abstract

The present paper is a review of literature in relation to formulaic sequences and the implications for second language learning. The formulaic sequence is a significant part of our language, and plays an essential role in both first and second language learning. The paper first introduces the definition, classifications, and major features of formulaic sequences. Then relevant studies on second language learning are reviewed, and pedagogical implications will be drawn from previous research. It is suggested that more emphasis should be put on prefabs in foreign language teaching, but at the same time, there is also danger of overemphasizing the role of prefabs in SLA research, given limited exposure to the target language in a foreign language learning environment.

Keywords: formulaic sequences, second language learning, foreign language teaching

1. Introduction

Multi-word speech (e.g. formulaic sequences) is a significant part of our language, and it plays an essential role in both first and second language learning (e.g. Wray, 2002, 2008, 2009, 2013). It has long been acknowledged that a number of linguistic strings in our languages are treated like single "big words" (Ellis 1996: 111), for instance, *strong tea*, *give up*, *kick the bucket*, *in front of*, *What's the matter*, etc. Sinclair (1991: 110) regarded them as "single choices, even though they might appear to be analyzable into segments".

Various terminologies related to these strings have been brought up by different scholars in different research fields, such as *chunks*, *collocations*, *fixed expressions*, *multi-word expressions*, *prefabs*, *recurring utterances*, to name but a few. For example, from a probabilistic viewpoint, *multi-word expressions* are defined as "combinations of words that co-occur more often than would be expected by chance alone" (Manning & Schütze, 1999, from Siyanova-Chanturia and Martinez 2014: 1). A *prefab*, as in Erman and Warren's (2000: 31) words, "is a combination 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".

Although there is some overlapping among these terms, it does not mean that they share exactly the same thing in all instances, as demonstrated in the two definitions above. Among all, a clearly defined term, according to Wray (2002: 9), is *formulaic sequence*, i.e. "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored, retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar".

The word *formulaic* is associated with 'unity', 'custom' and 'habit', while *sequence* indicates that more than one internal unit can be detected, and they do not necessarily have to be words. This definition covers all the possibilities of formulaic linguistic units, thus making reference easier.

This paper will introduce the classifications and major features of formulaic sequences. Then relevant studies in L1 and L2 acquisition will be reviewed, and pedagogical implications will be drawn from previous research.

2. Formulaic Sequences

2.1 Classifications of formulaic sequences

Formulaic sequences can be fully fixed in form, e.g. *Have a nice day*, *It's my pleasure*, or semi-preconstructed phrases which require inserting morphological details and/or open class elements, normally referential ones

(Wray, 2002: 7) , like *NP₁ leave + TENSE a/an ADJECTIVE impression on NP₂* (e.g. *The interviewee left a good impression on the manager*).

There is still no consensus on the classifications of formulaic sequences, for the reason that "the categories are neither discrete nor comprehensive" (Hudson, 1998: 13), and "many fixed expressions cannot be accommodated in any of the categories that have hitherto been defined" (Hudson, 1998: 34). Generally speaking, taxonomies are based on one or more of the four features of formulaic sequences, i.e. *form*, *function*, *meaning*, and *provenance*. As shown in Table 1, the classification provided by Becker (1975), despite its cross-associations of different features, gives some basic guidelines to identify formulaic sequences.

Table 1. A six-way division of formulaic sequences (Adapted from Becker, 1975: 6-7)

Category	Example
polywords	*(the) oldest profession; to blow up; for good*
phrasal constraints	*by sheer coincidence*
sentence builders	*(person A) gave (person B) a (long) song and dance about (a topic)*
meta-messages	*for that matter…* (message: 'I just thought of a better way of making my point');
	... that's all (message: 'don't get flustered')
situational utterances	*how can I ever repay you?*
verbatim texts	*better late than never;*
	How ya gonna keep 'em down on the farm?

The first three categories, namely, polywords, phrasal constraints, and sentence builders are form-based, meta-messages are primarily related to meaning, situational utterances are mainly functional, and verbatim texts indicate provenance.

Howarth (1998) amalgamated several widely accepted categorizations (e.g. Cowie, 1988; Gläser, 1988) into one model, illustrated in Figure 1.
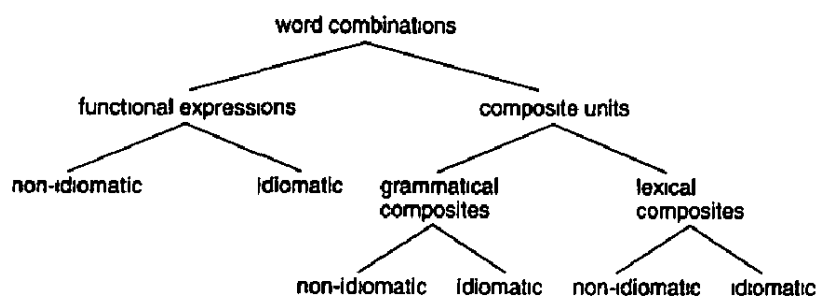
Figure 1. Phrasal categories (Adopted from Howarth, 1998: 27)

This model is significant in that it distinguishes between idiomatic and non-idiomatic expressions, on the basis of both semantic specifications and grammatical restrictions. In this case, composite units bear a syntactic function at the clause or sentence level, and they are further divided into grammatical and lexical composites. Meanwhile, Howarth (1998) argues against a simple two-way division, claiming that categories should be seen as "forming a continuum from the most free combinations to the most fixed idioms, rather than discrete classes" (*ibid*.: 35).

Table 2. Howarth's collocational continuum (Adopted from Howarth, 1998: 28)

	free combinations	restricted collocations	figurative idioms	pure idioms
lexical composites verb + noun	*blow a trumpet*	*blow a fuse*	*blow your own trumpet*	*blow the gaff*
grammatical composites preposition + noun	*under the table*	*under attack*	*under the microscope*	*under the weather*

Table 2 shows the continuum from free combinations to pure idioms, in terms of lexical composites and grammatical composites. On the basis of the categorizations and the collocational continuum mentioned above, I propose that ditransitive constructions can also be treated as formulaic sequences from both grammatical and lexical measures. Take the verb *give* for instance, the two ditransitive structures where it can appear are illustrated with the semi-preconstructed patterns which allow for such open-class elements as morphological tense marking, referential noun phrases or pronouns, etc. (e.g. *Peter's mom gave him some money*; *The professor is giving a talk to the students*).

DOC: NP1   *give* + TENSE   NP2   something
DAT: NP1   *give* + TENSE   something   *to* NP2

From a continuum perspective, it is possible to use ditransitive verbs in different situations, as demonstrated in Table 3. Take the ditransitive verb *give* for example, it may be used in a free double object construction, like *give him a book*, or in a restricted collocational structure, such as *give her a call*, or in a figurative idiom, e.g. *give them a hand*, or even in a pure idiom, like *I give you the chairman*.

Table 3. Collocational continuum for the ditransitive verb *give*

Continuum	Examples
Free combinations	*give someone a book / give a book to someone*
Restricted collocations	*give someone a call / give birth to someone*
Figurative idioms	*give them a hand / give rise to something*
Pure idioms	*I give you the chairman* (used at the end of a formal speech to invite people to welcome a special guest)

2.2 Features of formulaic sequences

One major characteristic of formulaic sequences is their relatively high frequency. This type of frequency is related to phrasal frequency, rather than the frequency of a single word. Each linguistic item that language users encounter will be stored in memory, and will be processed with a large number of exemplars (Bybee, 2006; Ellis 2012), including not only words, but also phrases and grammatical constructions. Only with these various exemplars can frequencies accumulate, thus leading to the success of language learning. Bod (2006: 318) therefore views knowledge of language as "a statistical ensemble of language experiences that slightly changes every time a new utterance is perceived or produced".

It is obvious to see from the above claims the importance of frequency in learning formulaic sequences. According to Bybee (2006: 720), "as a particular string grows more frequent, it comes to be processed as a unit rather than through its individual parts. As it is accessed more and more as a unit, it grows autonomous from the construction that originally gave rise to it."

Another characteristic is predictability, which refers to the possibility to predict the other parts of the formulaic sequences, for instance, *head over…*, *bread and…*, *Don't cry over…* etc. Probabilistic linguists contend that speakers have statistical information about the co-occurrence of words in their minds (Jurafsky, 1996; McDonald & Shillcock, 2003; Seidenberg & MacDonald, 1999). McDonald and Shillcock (2003) also suggest that the surrounding context of a word should be combined into people's lexicon.

3. Formulaic Sequences in Second Language Learning Research

There is unanimous consensus on the importance of formulaic sequences in first and second/foreign language

learning. The appropriate use of formulaic sequences "has been recognized as a prerequisite for any second/foreign language learner who wants to achieve high proficiency and be accepted in an L2 community" (Siyanova-Chanturia & Martinez, 2015: 12). A series of studies have been done in L1 and L2 acquisition.

In psycholinguistic research, focus has been put on language users' comprehension and production of formulaic sequences, through different types of experimental methods, e.g. self-paced reading, elicitation tasks, eye tracking, and so on. Among all the formulaic sequences, idioms have received the most attention from researchers, especially with respect to the order of activating their figurative and literal meanings, and the processing of idiomatic phrases versus novel strings of language. L2 studies have also been done on idiom on-line processing (e.g. Cieślicka, 2006; Conklin & Schmitt, 2008) and idiom comprehension (e.g. Siyanova-Chanturia, Conklin, & Schmitt, 2011).

Apart from idioms, other kinds of formulaic sequences have also been investigated, such as phrases, lexical bundles, grammatical structures, etc. (e.g. Arnon & Priva, 2013; Arnon & Snider, 2010; Tremblay, Derwing, Libben, & Westbury, 2011). Frequency effect on the comprehension and production of formulaic sequences has been detected in L1 research, but little research has explored online processing of formulaic sequences except for idioms in L2.

In corpus linguistics research, a number of corpus-based studies have been carried out to explore the use of formulaic sequences in second language learning. Granger (1998) examined learner phraseology data from the French component of the *International Corpus of Learner English*, and compared with the native speaker performance. Two types of prefabs under investigation consisted of collocations (e.g. *closely linked*) and lexical phrases (e.g. *it is said that…*). She found that learners significantly underuse both restricted collocations and creative combinations. In addition, she also discovered L1 transfer effect in the use of a few stereotyped combinations by learners (such as *deeply rooted*).

4. Implications for Second Language Learning

With regard to pedagogical implications, Granger called for more emphasis on prefabs in foreign language teaching, but at the same time, she also cautioned against the danger of overemphasizing the role of prefabs in SLA research, given limited exposure to the target language in a foreign language learning environment (Granger, 1998: 11). Three types of data were suggested for future exploration, namely "detailed descriptions of English prefabricated language, prefabricated language in the learners' mother tongues, and learner use of prefabs".

With a corpus-driven approach, Biber (2009) compared multi-word patterns used in speech and writing. In his paper, Biber distinguished between two kinds of multi-word patterns, one is 'multi-word lexical collocations' (combinations of content words), and the other is 'multi-word formulaic sequences' (combinations of both content and function words). The findings showed fundamentally different patterns in conversation (preference for continuous fixed sequences) and academic writing (preference for formulaic frames with internal variable slots). Speech normally prefers continuous fixed sequences, e.g. … *I don't know*, *you want to*…, while academic writing is in favor of formulaic frames with internal open slots including both function and content words, e.g. *in the … of, to the … of, at the … of.*

Recent years have seen a trend of combining psychological experimental techniques and corpus-based methodologies, in which way these two methods will complement each other from different perspectives of research (see Gilquin & Gries, 2009).

Ellis et al. (2008) investigated formulaic language in native and second language learners, by triangulating perspectives from corpus linguistics, psycholinguistic experiments and teaching English as a second language research. Corpus-based descriptions were made in terms of pedagogically useful formulaic sequences for academic speech and writing. The experiments' results showed the psycholinguistic validity of corpus-derived formulas. The authors thus gave implications for teaching English for academic purposes (EAP) as to which formulaic sequences should be given priority in teaching. Such a measure of utility is called *formula teaching worth* (FTW), which balances the value of formula frequency and mutual information (Ellis 2008: 392).

Millar (2011) combined corpus research with experimental methods to explore native speakers' processing of learner collocations that are non-native like. The results supported the claim that there are processing advantages for formulaic sequences, and provided empirical evidence for the importance of formulaic sequences in language learning. Lexical Priming theory (Hoey, 2005) was used to explain the processes underlying the increased processing demands regarding learner collocations.

5. Discussions and Conclusion

The present paper has given an overview of formulaic sequences and the implications for second language

learning. It has been found that previous research of formulaic sequences has their major focus on fixed multi-word units, and little attention has been paid to grammatical structures with internal open slots. Xu's (2014) research shows Chinese EFL learners' misuse of word combinations which are grammatically correct but semantically inappropriate, e.g. *bring convenience to someone*. The author proposes that not only L1 transfer, but also the traditional method of teaching words in isolation should be responsible for learners' misuse. The use of non-target-like structures thus calls for more attention of textbooks compliers to formulaic sequences. The fundamental role of formulaic sequences in foreign language teaching and learning has been widely acknowledged (Ellis 2008; e.g. Wray, 2002), and the capability to appropriately use formulaic sequences is regarded as a prerequisite for foreign language learners to achieve higher proficiencies (Siyanova-Chanturia & Martinez, 2015: 12). It is therefore suggested that in English textbooks, the introduction of a verb or any other types of vocabulary should not be restricted to its pronunciation and one-to-one Chinese meaning, but also need to systematically include 'semi-preconstructed' phrases or constructions. When introducing ditransitive verbs, much focus has been put on the two constructions that they can appear; however, what is neglected is formulaic sequences that these verbs can co-occur with.

Take the most frequently used ditransitive verb *give* for example. Mukherjee (2005: 104) found two clusters of lexical items used together with *give* in DAT, i.e. *give + DO + to + PC*. Such words as *access*, *answer*, and *reply*, etc. are "habitually associated with the preposition *to* according to the pattern information in the corpus-based *Macmillan English Dictionary*" (Rundell, 2002, from Mukherjee 2005: 103). They account for 14.6% of all the *to*-dative constructions for *give* in ICE-GB (*The British Component of the International Corpus of English*). The first group of formulaic sequences is listed below.

Formulaic sequence	Example sentence
give access to…	A disc audio prospectus is being prepared to *give access to* our publicity for people with visual impairment. <ICE-GB-S2b-044 #86: 2: A>
give attention to…	So I think we need to *give attention to* this. <ICE-GB_S1b-037 #101:1:B>
give consideration to…	Please could the college *give consideration to* offering me financial support in my study for a Master's degree. <ICE-GB_W1b-022 #167:14>

The other group of formulaic sequences is treated as idioms, e.g. *give birth to someone/something*, which cannot allow for DOC alternation. They take up 34.9% of all instances of *give* in DOC in ICE-GB. Such information may be used in English textbooks compilation on the basis of target learners' proficiency levels.

Formulaic sequences	Example sentences
give birth to…	Part of the reason for this was revealed when she admitted to having lost two children before *giving birth to* the three she brought with her on her visit to Sharptor. <ICE-GB_W2f-007 #94:1>
give rise to…	Fluvial, and marine environment also *give rise to* characteristic bedding such as cross bedding and current bedding FIG 2. <ICE-GB_W1a-020 #10:1>
give way to…	Understandably, this unlikely arrangement soon *gave way to* a more convenient and practical way of signaling, the Telephone Ringer. <ICE-GB_W2b-032 #17:1>

Teaching words in isolation is not effective enough for students to apply them into practical use. However, with the help of the contextual information as shown in the formulaic sequences, it will be an easier job to help students understand a word's meaning and its usages in native English. For future research, it is necessary to investigate grammatical structures from the perspective of formulaic sequences, and provide more implications for second language learning and teaching.

Acknowledgements

References

Arnon, I., & Priva, U. C. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Lang Speech, 56*(3), 349-371. http://dx.doi.org/10.1177/0023830913484891

Arnon, I., & Snider, N. (2010). More than words: Frequency effect for multi-word phrases. *Journal of Memory and Language, 62*, 67-82. http://dx.doi.org/10.1016/j.jml.2009.09.005

Becker, J. (1975). The phrasal lexicon. In R. Shank and B. L. Nash-Webber (Eds.), *Theoretical Issues in Natural Language Processing* (pp. 60-63). Cambridge, MA.: Bolt Beranek & Newman.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics, 14*(3), 275-311. http://dx.doi.org/10.1075/ijcl.14.3.08bib

Bod, R. (2006). Exemplar-based syntax: How to get productivity from exemplars. *The Linguistics Review, 23*, 291-320. http://dx.doi.org/10.1515/TLR.2006.012

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82*(4), 711-733. http://dx.doi.org/10.1353/lan.2006.0186

Cieślicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research, 22*(2), 115-144. http://dx.doi.org/10.1191/0267658306sr263oa

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29*(1), 72-89. http://dx.doi.org/10.1093/applin/amm022

Cowie, A. P. (1988). Stable and creative aspects of vocabulary use. In R. Carter and M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 126-139). London: Longman.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking and points of order. *Studies in Second Language Acquisition, 18*, 91-126. http://dx.doi.org/10.1017/S0272263100014698

Ellis, N. C. (2012). What can we count in language, and what counts in language acquisition, cognition, and use? In S. T. Gries and D. S. Divjak (Eds.), *Frequency Effects in Language Learning and Processing* (pp. 7-34). (Vol. 1): Mouton de Gruyter.

Ellis, N. C., Simpson-Vlach, R., & Carson, M. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 42*(3), 375-396. http://dx.doi.org/10.1515/cllt.2009.003

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text - Interdisciplinary Journal for the Study of Discourse, 20*(1), 29-62. http://dx.doi.org/10.1515/text.1.2000.20.1.29

Gilquin, G., & Gries, S. T. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus linguistics and Linguistic Theory, 5*(1), 1-26. http://dx.doi.org/10.1515/CLLT.2009.001

Gläser, R. (1988). The grading of idiomaticity as a presupposition for a taxonomy of idioms. In W. Hüllen and R. Schulze (Eds.), *Understanding the Lexicon: Meaning, Sense and World Knowledge in Lexical Semantics* (pp. 264-279). Tübingen: Niemeyer.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp. 145-160). Oxford: Oxford University Press.

Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19*(1), 24-44. http://dx.doi.org/10.1093/applin/19.1.24

Hudson, J. (1998). *Perspectives on Fixedness: Applied and Theoretical*. Lund: Lund University Press.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20*, 137-194. http://dx.doi.org/10.1207/s15516709cog2002_1

Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Mass.: MIT Press.

McDonald, S., & Shillcock, R. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science, 14*(6), 648-652. http://dx.doi.org/10.1046/j.0956-7976.2003.psci_1480.x

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics, 32*(2), 129-148. http://dx.doi.org/10.1093/applin/amq035

Mukherjee, J. (2005). *English ditransitive verbs: Aspects of theory, description and a usage-based model*. Amsterdam/New York: Rodopi.

Rundell, M. (Ed.). (2002). *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan.

Seidenberg, M., & MacDonald, M. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23*, 569-588. http://dx.doi.org/10.1207/s15516709cog2304_8

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research, 27*(2), 251-272. http://dx.doi.org/10.1177/0267658310382068

Siyanova-Chanturia, A., & Martinez, R. (2015). The Idiom Principle Revisited. *Applied Linguistics, 36*(5), 549-569. http://dx.doi.org/10.1093/applin/amt054

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning, 61*(2), 569-613. http://dx.doi.org/10.1111/j.1467-9922.2010.00622.x

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*. Oxford; New York: Oxford University Press.

Wray, A. (2009). Future directions in formulaic language research. *Journal of Foreign Languages, 32*(6), 2-17

Wray, A. (2013). Formulaic language. *Language Teaching, 46*(03), 316-334. http://dx.doi.org/10.1017/S0261444813000013

Xu, Q. (2014). *A Corpus-based Study of Alternating Ditransitive Constructions in Chinese Learner English*. Unpublished PhD Thesis, The Chinese University of Hong Kong, Hong Kong.