# What It Means to Be Processed Faster:

## The Processing of Formulaic Sequences by Native and L2 Speakers

奥　脇　奈津美

OKUWAKI　Natsumi

**Abstract**

Formulaic sequences have been found to present a processing advantage compared to non-formulaic equivalents by native speakers and sometimes by L2 speakers. This is either because formulaic sequences is processed as unanalysed holistic units or because they are processed in a highly automatized manner. It may also be due to the frequency of formulaic sequences, which plays an important part in language processing. In spite of different views regarding what implications these propositions may have for on-line processing, faster processing itself is advantageous for L2 speakers. Based on the review of the concept of formulaicity, I discuss the studies investigating the processing of formulaic sequences and suggest the importance of the question: What does it mean for formulaic sequences to hold a privileged advantage, especially for L2 learners?

## Introduction

The scope of the research into formulaic language includes the study of a variety of multiword units from idioms (*beat around the bush*), proverbs (*the apple never falls far from the tree*), collocations (*densely populated*), binominals (*bride and groom*), multi-word verbs (*look up*), speech formulae (*how are you?*), discourse markers (*first of all*), lexical bundles (*as soon as*) to grammatical constructions (*the -er, the -er*), and such. Corpus studies have now shown that natural language makes considerable use of recurrent formulaic patterns of words (Altenberg, 1998; Erman & Warren, 2000; Foster, 2001; Howarth, 1998).

The extensive use of formulaic sequences may be reflected in the existence of enormous terms found in the literature to describe the set of related phenomena; chunks, fixed expressions, formulae, idioms, collocations, lexical phrases, multiword items, prefabricated patterns, ready-made expressions, and many more (Wray, 2002). In general, formulaic sequences are defined as consisting of multiword structures with a conventional holistic meaning. They are generally not made by the grammatical component to which each word belongs. Wray (2002) provides the most often cited definitions of formulaic sequences.

…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. (p. 9).

Wray (2002) intentionally made this definition as inclusive as possible so that it conveys any kind of linguistic pattern in any research field that is considered formulaic. Since the classical work of Pawley and Syder (1983) on the role of formulaic language as an ordinary and natural form of expressions in native speech, recent years have witnessed a growing interest in the research of various kinds of formulaic language.

One of the important aspects of formulaic sequences is that they may contribute to the reduction of processing effort (Wray, 2002, p. 18). There is growing agreement that native speakers process formulaic sequences faster than non-formulaic sequences (Jiang & Nekrasova, 2007; Underwood, Schmitt, & Galpin, 2004). It has not been known, however, why formulaic sequences are processed faster than non-formulaic equivalents in the first place. Some claim that the advantage in processing is yielded by the representation of formulaic sequences in the mental lexicon. They are "stored and retrieved whole from memory, […] rather than being subject to generation or analysis by the language grammar" (Wray, 2002, p. 9). Such definitions of holistic storage, retrieval, and processing suggest that formulaic sequences are processed as a unitary whole without access to their constituents. A similar but slightly different account was given by Cordier (2013) in that formulaic sequences are used so frequently by language users, and as a result, these sequences are highly automatized, bringing about processing advantage.

This article first reviews the concept of formulaicity; why some linguistic sequences are better treated as wholes, and discusses through which processes formulaicity can emerge. The studies investigating the processing of formulaic sequences are then reviewed, followed by the discussion on the importance of the question: What does it mean for formulaic sequences to hold a privileged advantage, especially for L2 learners? In spite of different views regarding what implications the propositions have for on-line processing, I argue that faster processing *per se* is advantageous for L2 speakers.

## The formulaicity of language

There is an insight in the study of formulaic language that some linguistic sequences are better treated as wholes than being analysed into smaller unites (Durrant & Mathews-Aydinli, 2011). There are some reasons why they should be treated as such. Firstly, there are cases where it is not possible to predict the meaning or syntactic phenomenon of formulaic sequences from a general knowledge of language. Examples include idioms (e.g. *blow the gaff; under the weather*), opaque collocations (e.g. *fat paycheck; amicable divorce*), and

some grammatical constructions (eg. *the -er, the -er*). Making predictions of their meaning is almost impossible unless the language user knows these idioms and collocations. In other cases, sequences are better treated as wholes because they are usually used as such in the speech community and sound more natural than being used with other combinations. For example, in transparent collocations (e.g. *make tea; get a message*), the combination of one form and another is generally arbitrary, and specific knowledge of sequences is needed to make nativelike production (Pawley & Snyder, 1983). Finally, certain linguistic sequences are better treated as a whole because they are highly frequently used in the daily language practice, and as a consequence, the expressions are well-entrenched in a language user's long term memory and retrieved from memory as a whole without the need to compose the sequences word by word online (Wray, 2002, Conlkin & Schmitt, 2012). According to this view, having this representation in the mind would make processing of sequences easier and faster than non-formulaic equivalents.

It should be noted that there is a question as to whether the expression in language can be construed as either formulaic or non-formulaic. The dual processing explanations, offered by such researchers as Sinclair (1991), Wray (2002), and others, claim a clear difference between formulaic and non-formulaic processing. It may be the case, however, that the expression can be formulaic to some extent, being located at any point along a continuum from thoroughly formulaic to thoroughly non-formulaic. This view is shared by those who claim for a continuum nature of formulaicity.

According to Ellis (2003), formulaicity can emerge in various ways. Firstly, if there is a regular association between particular multi-word forms and particular contexts, the sequences will be entrenched as formulae. Secondly, if words co-occur on a regular basis and their mutual association is strong enough to be regarded as collocations, they have a psychological reality as formulae at least for native speakers. Formulaicity can also emerge through the grammatical process, where syntactic representations arise from a gradual process of abstraction from lexical exemplars. If this is the case, abstract syntax should not be any different from vocabulary; instead, there is a continuum between memorized and rule-based constructions, with many of the forms located somewhere between these extremes (Langacker, 1987). This is supported by studies dealing with frequency data based on psychologically motivated explanations (Conklin & Schmitt, 2008; and others).

This section thus shows the cases when some linguistic sequences are better treated as wholes rather than being analysed into small units, which argues for the presence of formulaicity of language. In addition, various ways and processes were shown as to the emergence of formulaic sequences in language.

## Processing advantage for formulaic sequences

The way to conceive formulaicity matters when it comes to the issue of storage and

retrieval of formulaic sequences in the human mind. Research shows that formulaic sequences hold a processing advantage over non-formulaic sequences, but a clear picture has not yet emerged. Even if this is right, it remains to be known how and why language users perceive formulaic expressions faster than non-formulaic equivalents. Another issue is what it means for formulaic sequences to have a privileged advantage in language processing.

Let us first consider the evidence found for faster processing of formulaic sequences. If formulaic sequences are stored in the human mind as whole units, shorter reading times should be found for formulaic sequences compared to equivalent non-formulaic sequences. This is a common assumption that underlies in the studies. A range of investigations have been carried out to see if formulaic sequences have a privileged processing status; reaction time tasks (e.g., Arnon & Snider, 2009; Durrant & Schmitt, 2010; Jiang & Nekrasova, 2007), self-paced reading (e.g., Conklin & Schmitt, 2008; Tremblay et al., 2011), silent reading as evidenced by eye-tracking (Columbus, 2010; Siyanova-Chanturia, Conklin, and van Heuven 2011), and recall (Tremblay & Baayen, 2010). Experiments with native speakers have shown that formulaic strings of various types are processed significantly faster than non-formulaic strings.

In the eye-tracking study which tracks the eye movements of participants when they read passages with embedded formulaic sequences, Underwood, Schmitt and Galpin (2004) measured how often and for what duration the final words in the sequences were fixed on. The assumption is that once a formulaic sequence is recognised from the first few words, it would need less attention to recognise the final word since the participants would be likely to predict the final word of the strings. In other words, if a participant processed a formulaic sequences holistically, he or she would know the last word more correctly and quickly than if he or she had not processed it holistically. The result showed that the terminal words in formulaic sequences gained fewer fixations than the same words in non-formulaic contexts, implying that there was a processing advantage of formulaic sequences in reading. This was true for both native speakers and L2 learners.

Jiang and Nekrasova (2007) also provided the evidence of the processing advantage by comparing reaction times and error rates of how native and L2 speakers recognized and processed formulaic and non-formulaic sequences. They used corpus-derived recurrent word combinations as materials and compared formulaic phrases (*on the other hand*) against matched non-formulaic phrases (*on the other bed*) in two online grammaticality judgement experiments. The results showed that native and L2 speakers of English responded to three-word lexical bundles faster and more accurately than non-lexical bundles. Based on this result, Jiang and Nekrasova argue that formulaic sequences are holistically recognized by both native speakers and proficient L2 speakers.

These studies argue that formulaic sequences are read faster than non-formulaic ones because they are stored and processed as unalalysed units and this gives rise to the processing advantage in formulaic sequences. Observations in recent studies, however, recognize some important issues remain to be addressed. One is the question of whether a

processing advantage in terms of rate shows holistic storage or simply the faster mapping of linguistic components. Another question is that while many research suggest a privileged processing of formulaic language over non-formulaic language for native speakers, it is not known whether this processing advantage can extend to nonnative speakers.

### What it means to be processed faster

Although many studies assume the holistic storage and retrieval of formulaic sequences from the mental lexicon, and that this would make it possible to read formulaic sequences faster compared to equivalent non-formulaic ones, the status of formulaic sequences in the mind can be the matter for debate. The studies on the processing idioms found that they can be processed word by word, not as a whole unit (Cacciari & Tabossi, 1988; Peterson, Dell, Burgess & Eberhand, 2001). In addition, the studies pertaining to the role of frequency in processing suggest no direct relation between a processing advantage and holistic storage of formulaic sequences (Siyanova-Chanturia, 2015).

Then, the question is what it means for formulaic sequences to have a privileged processing. Cordier (2013) points out that some processing advantage for formulaic language found in the studies cannot simply be taken as evidence that these multi-word units are stored as whole units in the lexicon. Similarly, Siyanova-Chanturia (2015, p. 289) caveats, "it is important to bear in mind that the processing advantage *per se* cannot shed much light on the holistic storage of formulaic language (or lack thereof)", and argues that what it does show is that the sequences are frequent in the input and speakers are familiar with these strings, which would enable them to process the sequences very efficiently. If this is on the right track, the studies that claim for the evidence of holistic storage (Jiang & Nekrasova, 2007; Underwood et al. 2004) simply addressed the question of the rate of processing formulaic sequences versus matched control phrases, not of the status of formulaic sequences *per se* in the mental lexicon. In order to explore the holistic status issue, as Siyanova-Chanturia (2015) proposes, one needs to investigate the activation of the individual components within a formulaic sequences in a direct way rather than the rate of processing of formulaic units.

This is the position often taken by the researchers dealing with frequency data in psycholinguistics (Cacciari & Tabossi, 1988; Peterson, Dell, Burgess & Eberhand, 2001; Siyanova-Chanturia, 2015). They argue against the view that a formulaic sequence is represented and processed independently of the properties of its constituents, as it is stored as a whole in the lexicon. Instead, they argue that the crucial role of frequency in processing individual words should also apply to formulaic sequences, and they should be processed just like an individual word. Therefore, rather than processing the chunk *kick the bucket* as one holistic unit, one activates and accesses the individual components of the phrase in one's mental lexicon.

**Lack of L2 knowledge and processing of formulaic sequences for L2 learners**

Whatever it means to be processed faster, processing advantage of formulaic sequences has made a strong case for native speakers. However, studies have not provided satisfactory and clear-cut results for L2 speakers.

L2 speakers are often found to process formulaic sequence in a word-by-word manner like non-formulaic sequences. Siyanova-Chanturia et al. (2011) investigated the processing of formulaic sequences by L2 speakers, focusing on formulaic sequences imbedded in sentences that differed in phrasal frequency (e.g., *bride and groom vs. groom and bride*). The result shows that L2 speakers across a range of proficiency read more frequent sequences faster than less frequent ones, indicating that phrases will be entrenched in memory as L2 ability progresses. Crucially, more proficient L2 learners were more sensitive to the frequency information in the target phrases compared to less proficient learners. Similarly, Siyanova-Chanturia and Spina (2015) focused on the role played by phrase frequency and investigated how L2 speakers (advanced and intermediate) judge collocation frequency. They used 80 noun-adjective collocations, which differed in frequency. It was found that L2 speakers who were very proficient in L2 were remarkably native-like in their intuitions of collocation frequency, suggesting the role of frequency played in the processing of formulaic language. These studies suggest that L2 ability influences the way L2 speakers process formulaic sequences. Thus, studies show there is processing advantage for native speakers, and maybe for very high-proficiency L2 learners, but the evidence is inconclusive for L2 speakers with less proficiency.

If L2 speakers seem to process formulaic language differently from native speakers, and there is a contrast between native and nonnative speakers in the processing of formulaic sequences, this may be due to lack of knowledge of formulaic sequences *per se* on the part of L2 speakers. For L2 speakers to receive the processing advantage afforded by a formulaic sequence, they have to be familiar with the sequences as a premise. Most of the research findings suggest that idioms did not present a privileged processing with L2 learners. This is not surprising, given that L2 speakers tend to lack L2 knowledge on idioms compared to native speakers. It plays a crucial part whether L2 speakers have sufficient linguistic knowledge on the formulaic sequences. As Boers and Lindstromberg (2012) suggest, a L2 speaker must be exposed to the L2 required to reach such a level of familiarity with formulaic sequences. Similarly, Ellis (2008) suggests that learners require considerable language exposure before they acquire formulaic sequences, but many L2 learners may not be able to encounter low-frequency sequences sufficiently enough for them to be firmly established in memory. This is one of the reasons L2 learners are reported not to master native-like formulaic language.

## Conclusion

For L2 learners, mastering formulaic language contributes to their successful and fluent use of L2. As Martinez and Schmitt (2012) propose, it is worthwhile for L2 learners to master this property because of its ubiquity in language use, its᾽ functions in contexts, and its processing advantage. In addition, formulaic language can improve the overall impression of L2 learners᾽ language production.

This article discussed research findings showing that formulaic sequences present a processing advantage compared to non-formulaic equivalents by native speakers and sometimes by L2 speakers. If formulaic sequences are stored in the human mind as whole units, shorter reading times should be found for formulaic sequences. This is a common assumption that underlies in such studies investigating the holistic status of formulaic sequences in human minds. Alternatively, L2 speakers may just need to be exposed to the L2 required to reach a certain level of familiarity with formulaic sequences. In this case, L2 speakers may just require considerable language exposure to formulaic sequences before they are entrenched in long-term memory to be retrieved online. In either case, it is important to address what implications processing advantages have and empirically investigate why formulaic sequences hold such privileged advantage In spite of different views regarding what implications the propositions have for on-line processing, faster processing *per se* is highly advantageous for L2 speakers.

## Acknowledgement

## References

Altenberg, B. (1998). On the phraseology of spoken English: Evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp.101-122). Oxford: Oxford University Press.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62* (1), 67-82.

Boers, F., & Lindstromberg, S. (2012). Experimental and Intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics, 32*, 83-110.

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, *27*, 668-683.

Columbus, G. (2010). Processing MWUs: Are MWU subtypes psychologically real? In D.

Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 194-210). New York: Continuum.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*, 72-89.

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics, 32*, 45-61.

Cordier, C. (2013). *The presence, nature and role of formulaic sequences in English advanced learners of French: a longitudinal study*. Doctoral Dissertation, Newcastle University.

Durrant, P., & Mahews-Aydinli, J. (2011). A function-first approach to identifying formulaic language in academic writing. *Journal of English for Specific Purposes, 30* (1), 58-72.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26* (2), 163-188.

Ellis, N. (2003). Constructions, chunking, and connectionism: the emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition*, 63-103. Oxford: Blackwell.

Ellis, N. (2008). Optimizing the input: Frequency and sampling in usage-based and form-focused learning. In M. H. Long & C. J. Doughty (Eds.), *Handbook of second and foreign language teaching*. Oxford: Blackwell.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text 20* (1), 29-62.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 75-93). Harlow: Longman.

Howarth, P. (1998). The phraseology of learners' academic writing. In A. Cowie (Ed.): *Phraseology: Theory, Analysis and Applications* (pp. 155-175). Oxford: Oxford University Press.

Jiang, N., & Nekrasova, T. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal, 91* (3), 433-445.

Langacker, R. (1987). *Foundations of Cognitive Grammar, vol. 1: Theoretical prerequisites*. Stanford: Stanford University Press.

Martinez, R., & Schmitt, N. (2012). A Phrasal Expressions List. *Applied Linguistics, 33* (3), 299-320.

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Native-like selection and native-like fluency. In J. Richards & R. Schmidt (Eds.). *Language and communication* (pp.191-226).London: Longman,

Peterson, R., Burgess, C., Dell, G., & Eberhard, K. (2001). Disassociation between syntactic and semantic processing during idiom comprehension. *Journal of Experimental*

*Psychology: Learning, Memory, and Cognition, 27* (5), 1223-1237.

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Siyanova-Chanturia, A. (2015). On the 'holistic' nature of formulaic language. *Corpus Linguistics and Linguistic Theory, 11* (2), 285-301.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. (2011). Seeing a phrase 'time and again' matters: The role of phrasal frequency in the processing of multi-word sequences. *Journal of Experimental Psychology: Language, Memory and Cognition, 37*, 776-784.

Siyanova-Chanturia, A., & Spina, S. (2015). Investigation of native speaker and second language learner intuition of collocation frequency. *Language Learning, 65* (3), 533-562.

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151-173). London: Continuum.

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*, 569-613.

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use*. John Benjamins.

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