The Processing on Different Types of English Formulaic Sequences

Li Qian[1]

[1] Center of Lexicographic Studies, Guangdong University of Foreign Studies, Guangzhou, China

Correspondence: Li Qian, Center of Lexicographic Studies, Guangdong University of Foreign Studies, Guangzhou, 510420, China. Tel: 86-20-36207232. E-mail: lqchristina@gdufs.edu.cn

Abstract

Formulaic sequences are found to be processed faster than their matched novel phrases in previous studies. Given the variety of formulaic types, few studies have compared processing on different types of formulaic sequences. The present study explored the processing among idioms, speech formulae and written formulae. It has been found that in addition to the processing advantage of formulaic sequences as compared to the nonformulaic phrases, frequent written formulaic sequences were processed faster than the infrequent idioms. The results suggested that when processing advantage was concerned, both the holistic storage view and the frequency effect need to be considered.

Keywords: formulaic sequences; processing; formulaic sequence types

1. Introduction

Formulaic sequences, such as idioms, speech formulae, collocations and binominal expressions, are widely used in native-like communication (Oppenheim, 2000; Foster, 2001; Erman & Warren, 2000). Since Pawley and Syder (1983)'s classic work concerning the role of formulaic sequences in native speakers' speech, the recent thirty-two years have witnessed rapid progress in the field. Annual Review of Applied Linguistics (ARAL) has contributed its 2012 edition to the research on formulaic language, which covers different aspects of the field. Among them, the processing of formulaic sequences has become one of the increasingly interesting topics.

2. Literature Review

Conklin and Schmitt (2012) have categorized research on the processing of formulaic sequences mainly into two aspects: the processing of idioms and of nonidiomatic formulaic sequences. Idioms, as a type of non-transparent language expression, have been considered as the prototype of formulaic language (Nekrasova, 2009). Specifically, researchers focused on two respects of idiom processing: the literal vs. figurative interpretations of idioms and the processing of idioms vs. novel phrases.

As for the competition between literal and figurative meanings of idioms, researchers proposed different models to explain the phenomenon. For example, Lexical Representation Hypothesis proposed by Swinney and Cutler (1979) argues that speakers initiated the compilation of the literal meaning and the activation of the figurative meaning almost at the same time. Because idioms are stored like morphologically complex words, the figurative meaning is first activated. On the other hand, the Idiom Decomposition Hypothesis (Gibbs, Nayak, & Cutting, 1989) suggests that whether an idiom is decomposable or not will decide the way of idiom processing. The decomposable idioms are analyzed linguistically, and the idiom meaning is consistent with the analysis result. Processing time can be saved as a result of such consistency. Hence, decomposable idioms enjoy a processing advantage. Differently, Configuration Hypothesis (Cacciari & Tabossi, 1988) proposes that at the very beginning, for an idiom, both the component words and their literal meaning are activated. As the discourse information accumulates, the idiom will be identified as a fixed item. At this time, the figurative meaning is retrieved.

In addition to the above models, researchers are concerned with the processing of idioms in certain experimental settings. Tabossi, Fanari and Wolf (2009) showed that both decomposable and non-decomposable idioms are reacted more quickly than matched literal phrases by native speakers. Underwood, Schmitt, and Galpin (2004) investigated the processing of idioms embedded in a reading text. It was revealed that native speakers fixed their eyes less (and with a shorter duration) on the terminal words of idioms than on the nonformulaic words. In a recent study, Siyanova-Chanturia, Conklin, and Schmitt (2011) revealed that the idioms are processed significantly faster than the nonformulaic phrases by native speakers.

In brief, the previous studies consistently showed the speed advantage for idioms as compared to novel phrases. However, there are some problems for choosing idioms as the research focus in the processing research (Conklin & Schmitt, 2012): firstly, some of the idioms may be unfamiliar to nonnative speakers and L1 children, which may be an intervening factor in certain studies; secondly, the figurative and literal meanings of idioms may bring about certain ambiguity in processing; thirdly, idioms with different degrees of transparency will influence the processing of idioms. Hence, Conklin and Schmitt (2012) suggested non-idiomatic formulaic sequences be a better test case.

Some researchers turned the focus to the comparison between the non-idiomatic formulaic sequences and the novel matched ones. A recent eye-tracking study by Siyanova-Chanturia, Conklin and Van Heuven (2011) investigated processing of formulaic sequences with different phrasal frequency. It has been revealed that first, frequent formulaic sequences are processed faster than less frequent ones; second, regardless of frequency, native speakers processed the entrenched binomials significantly faster than reversed forms. Tremblay and Baayen (2010) used behavioral and electorphysiological measures to explore the processing of the phrase *in the middle of*. They found a frequency effect for this four-word expression. Although the evidence is somewhat incomplete, the above findings suggest that frequent formulaic sequences may be processed differently from less frequent ones by native speakers.

To summarize, as far as the processing of formulaic sequences is concerned, idioms and non-idiomatic formulaic sequences are separately researched. It is not clear that whether the processing advantages are similarly enjoyed by both idioms and non-idiomatic formulaic sequences, and whether the processing advantages for both types of formulaic sequences arise from the same effect. Few studies have ever compared the processing differences among varied types of formulaic sequences. In the present study, first, we tend to confirm the previous finding that formulaic sequences are processed faster than the matched novel phrases; second, we are going to explore whether different types of formulaic sequences (i.e. idioms, speech formulas and written formulas) are processed differently by native speakers.

3. The Present Study

3.1 Research Questions

a) Do native speakers process formulaic sequences and the matched novel phrases in different ways?

b) Do native speakers process different types of formula sequences in different ways?

3.2 Method

3.2.1 Participants

There were 20 English native speakers who participated in the study. All participants were students at a British University (8 graduate and 12 undergraduate students), among them 12 females and 8 males. Their ages ranged from 18 to 29.

3.2.2 Research Design

The research material (see Appendix) can be divided into three parts: a) three types of English formulaic sequences, namely idioms, speech formulaic sequences and formulaic sequences in academic writing (i.e. written formulaic sequences) (10 for each type). They were chosen from some corpus-based studies (Biber et al 1999; Nattinger & DeCarrico, 1992); b) for each formulaic sequence, a matched novel phrase was constructed. We replaced one or two words in a formulaic sequence with another one with similar length (in terms of number of syllables) and word frequency. For example, for the idiom *bear in mind*, the first word *bear* was replaced by another word *hold* to form a non-formulaic sequence *hold in mind*. We have made an effort to make sure that the number of syllables of the replacement items was equal to or smaller than that of the formulaic sequence, and the words used as the replacement words were matched in frequency on the basis of BNC frequency list (Leech et al, 2001). There are 30 matched novel phrases in the study; c) 15 ungrammatical sequences were constructed which were used as the distracters in the study.

From the 75 items, we have created two counterbalanced material lists, each of which included 15 formulas (5 for each formula type), 15 controlled novel phrases, and 15 ungrammatical sequences (see Appendix).

The study had a 3*2*2 design, with the formula types (idiom vs. speech formulaic sequences vs. written formulaic sequences), and the item grammaticality (grammatical vs. ungrammatical) and formulaicity (formulaic vs. nonformulaic). All the variables are the within-subject variables.

3.2.3 Procedures

In this study, the items were presented on the computer screen one by one (with 5s interval) in a random order. The participants were required to judge whether or not the items are grammatical. To respond, they pressed the key "q" for grammatical ones, and "p" for ungrammatical ones (Note 1).

Participants' reaction time and error rate were collected for data analysis. For item presentation and data collection, we used the computer program "Psychopy" developed by Peirce (2007) at Nottingham University. Each participant was randomly assigned to take either of the two test sets individually. Prior to the test, they read the written instructions and were given a training session for 20 practice items.

3.2.4 Results and Analysis

The reaction time and the error rate for all types of sequences were calculated for analysis. The descriptive data of the reaction time and error rate are showed in table 1. For the within-subject analysis, GLM-Repeated Measures procedures of SPSS were used.

Table 1. Native speakers' mean reaction time (in milliseconds) and error rate (in percentage) on different types of sequences

	Reaction time Mean (SD)	Error rate (percentage) Mean (SD)
Formulaic sequences		
Idiom	946 (257)	2 (6.32)
Speech	897 (207)	1 (3.16)
Written	865 (243)	1 (3.16)
Non-formulaic		
Idiom replacement	1209 (434)	18 (17.5)
Speech replacement	1103 (368)	12 (14.76)
Written replacement	1140 (447)	7 (10.29)
Ungrammatical	1179 (400)	15.3 (20.31)

GLM-Repeated Measures analysis on reaction time showed that there was significantly different reaction time concerning different types of formulaic sequences, $F_1(6, 2.26)=23.83$, p=.000 (p<.05), partial eta squared (η^2)=.412. Specifically, the reaction time on idioms (946ms) was significantly longer than that on the written formulaic sequences (865ms), p=.014, and there was no significant difference between the reaction time on idioms (946ms) and on speech formulaic sequences (897ms) (p=.078), and between the reaction time on speech formulaic sequences (897ms) and on written formulaic sequences (865ms) (p=.226). When formulaic sequences and novel phrases were compared, the reaction time on each type of formulaic sequences was significantly shorter than that on each type of the matched novel phrases (p=.000). As for the comparison between the ungrammatical sequences and the other types of sequences, the reaction time on the ungrammatical sequences (1179ms) was significantly longer than those on the three types of formulaic sequences (946, 897, 865 for idioms, speech formulaic sequences and written formulaic sequences) (p=.000), and significantly longer than the reaction time on novel speech phrases (1103ms) (p=.011).

GLM-Repeated Measures analysis on error rate revealed that there were significantly different error rates among different types of sequences, $F_2(6, 4.59)=2.58$, p=.028 (p<.05) partial eta squared (η^2)=.223. Specifically, we can find some patterns when the error rate of the idiom replacement was compared with that of the other types of sequences. The error rate of the idiom replacement (n=18) was significantly higher than that of the idiom judgment (n=2) (p=.022), of the speech formulaic sequences judgment (n=1) (p=.019), of the written formula judgment (n=1) (p=.019), and of the written formula replacement judgment (n=7) (p=. 04). It can be suggested that native speakers produced more errors when they were judging the sequences which were constructed based on the idioms.

4. Discussion

The present study aimed to explore whether native speakers process different types of formulaic sequences—idioms, speech and written formulaic sequences in varied ways. The results showed that

i)   Native speakers processed the formulaic sequences significantly faster than they processed the matched novel phrases.

ii)   Native speakers produced significantly more errors when they were judging the matched novel phrases as compared to the formulaic sequences.

iii)   When processing different types of formulaic sequences, native speakers processed idioms by using the longest time, and processed written formulaic sequences fastest among the three types.

As for Research Question 1, the first finding supports the previous argument that formulaic sequences enjoy processing advantage as compared to their matched novel phrases. In this study, the frequency and length of formulaic sequences and their matched novel phrases have been controlled to match each other. As such, the reaction time difference cannot be significantly influenced by these factors. The processing advantage can be explained by Heteromorphic Distributed Lexicon (HDL) proposed by Wray (2002). According to HDL, mental lexicon is made up of five lexicons serving different functions. Specifically, they are a) lexicons serving a grammatical role in the production of novel sentences; b) referential expressions, including mono- and polymorphic words, and word strings, such as idioms; c) context-dependent words and expressions that show little creativity and are used mainly for communication; d) memorized texts, and e) expressions served as automatic responses to different types of stimuli (i.e. external or psychological). Three forms of units are stored in the lexicon: morphemes, words, and word strings. In HDL, we can see that lexical units can be represented in different forms. For example, *bear in mind* is stored holistically in the lexicon as a word string, whereas *bear*, *in* and *mind* can be stored in the referential lexicon separately as three individual words. As an idiom, for most cases, *bear in mind* are retrieved and processed from the mental lexicon holistically, and only when in a certain biased context which need structural analysis is it fully analyzed.

In the grammaticality judgment test of this study, for the novel items, participants were supposed to analyze the syntactic structure since they do not have these representations as word strings in the mental lexicon. Therefore, the syntactic analyzing process needs a relatively longer time. However, for the formulaic sequences, they are assumed to be stored as a word string (e.g. idioms), or some context-dependent expressions in communication (e.g. written or speech formulaic sequences) or some automatic responses to stimuli (e.g. speech formulaic sequences). They can be retrieved holistically from the mental lexicon, which brings about a shorter reaction time in the judgment.

Similarly, the second finding of the study, that is formulaic sequences were judged at a higher accuracy rate than the non-formulaic sequences, can be explained by HDL as well. For formulaic sequences, they are stored in the mental lexicon holistically, although their elements are stored separately as well. When the grammaticality judgment is on the formulaic sequences, participants just need to match the presented items with the items in the mental lexicon, and no syntactic analysis has been involved. The existing lexicon and the direct matching guarantee the accuracy rate of the judgment. However, for the non-formulaic novel phrases, participants failed in matching them with the holistically-stored lexicon. They had to turn to the analysis of the grammaticality of the structure, which does not only take additional time of processing, but risk making some judgment errors in a limited time length.

For research question 2, the comparison is among the different types of formulaic sequences. In the literature, few studies have been carried out to compare the idioms with other types of formulaic sequences. In this study, idioms are found to be processed with significantly longer reaction time than written formulaic sequences. This finding cannot be explained by HDL. According to HDL, formulaic sequences, although in different types, are assumed to be stored holistically in the mental lexicon. As such, participants are supposed to process them in a similar way which results to similar reaction time for different types of items. This is incompatible with the third finding of the study.

As for the thrid finding, usage-based models can account for the differences. Usage-based models view language as a statistical accumulation of experience that develops and changes when more and more utterances are encountered (Goldberg, 2006; Tomasello, 2003). This view predicts that it is highly probable that all frequently exposed or frequently used units, words and phrases will be processed faster than less frequent ones. In the present study, the frequencies of the formulaic sequences and of their matched novel phrases have been controlled. However, we did not control the relative frequencies among the three types of formulaic sequences.

Since in this study, the reaction time for written formulaic sequences is significantly shorter than that of idioms, we did a post hoc analysis on their frequency differences. We compared the frequency lists of both types, and found that the average frequency of written formulaic sequences (average frequency in BNC= 4667) are significantly higher than that of the idioms (average frequency in BNC=296) (p=.000). Hence, it may be suggested that in addition to the holistic storage view, frequency effect functions as well. Those formulaic sequences which are frequently exposed or used may enjoy a processing advantage than the infrequent formulaic sequences, although both of them are stored holistically in the mental lexicon.

To conclude, this study supports the argument that formulaic sequences are processed faster than the matched novel phrases. In addition, infrequent idioms are processed with longer time than the frequent written formulaic sequences, although both of them are under the category of formulaic sequences. Heteromorphic Distributed Lexicon Model can account for the first finding; while usage-based frequency effect explains the second one. In summary, it may be suggested that when the processing advantage of formulaic sequences is concerned, the holistic storage view and the frequency effect need to be considered for such a combined effect.

Acknowledgements

References

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language, 27*, 343-346. http://dx.doi.org/10.1016/0749-596X(88)90014-9

Conklin, K, & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics, 32*, 45-61. http://dx.doi.org/10.1017/S0267190512000074

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

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, UK: Longman.

Gibbs, R., Nayak, N., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language, 28*, 576-593. http://dx.doi.org/10.1016/0749-596X(89)90014-4

Goldberg, A. (2006). *Constructions at work: The nature of generation in language*. Oxford: Oxford University Press.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: based on the British National Corpus*. Longman, London.

Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford Unviersity Press.

Nekrasova, T.M. (2009). English L1 and L2 Speakers' Knowledge of Lexical Bundles. *Language Learning, 59*, 647-686. http://dx.doi.org/10.1111/j.1467-9922.2009.00520.x

Oppenheim, N. (2000). The importance of recurrent sequences for non-native speaker fluency and cognition. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp.220-240). Ann Arbor: University of Michigan Press.

Pawley, A., & Syder, F. H. (1983). Two puzzles of recurrent sequences for non-native speaker fluency and cognition. In. J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-225), London, UK: Longman.

Peirce, J. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8-13. http://dx.doi.org/10.1016/j.jneumeth.2006.11.017

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*, 1-22. http://dx.doi.org/10.1177/0267658310382068

Siyanova-Chanturia, A., Conklin, K., & van Heuven, J. B. (2011). Seeing a pharse "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology*:

Learning, Memory and Cognition, 37, 776-784. http://dx.doi.org/10.1037/a0022531

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expression. *Journal of Verbal Learning and Verbal Behavior, 54*, 161-184. http://dx.doi.org/10.1016/S0022-5371(79)90284-6

Tabbossi, P., Fanary, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory and Cognition, 37*, 529-540. http://dx.doi.org/10.3758/MC.37.4.529

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.

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, UK: Continuum International.

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* (pp. 153-172). Philadelphia: John Benjamins. http://dx.doi.org/10.1075/lllt.9.09und

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511519772

Note 1. "q" and "p" are the two keys which are located far enough from each other on the keyboard to avoid any operation confusion.

Appendix: Test Materials

	Formulaic sequences	**Matched nonformulaic items**	**Ungrammatical items**
Idiomatic	a piece of cake	a plate of cake	a come surprise as
	as a matter of fact	as a place	boy in went
	for the time being	for the well-being	computer red that
	in a nutshell	in the shell	bus the on
	up to date	up to ten	late usual as
	bear in mind	hold in mind	chance a stand
	come as a surprise	come to a party	at paper the look
	keep an eye on	keep him awake	way the same in
	miss the boat	take a boat	test the in
	stand a chance	stand a test	on conclusion
			than bigger far
Spoken	for the most part	for the best time	as longer as
	as you know	as you sleep	context in of
	as I was saying	as he was doing	car to goes
	guess what	guess it	time over a long
	something like that	a thing like that	
	what happens is	what helps when	
	by and large	big and large	
	you know	you tell	
	by the way	by the route	
	and how about	and how might	
Written	as a result	as a father	
	in addition	in tradition	
	in other words	in other plays	
	for example	for a few	
	the point is	the pay is	
	in conclusion	in a feeling	
	in the case of	in the tree	
	to tell the truth	to tell the price	
	in the same way	in the same file	
	it depends on	it sits on	

Copyrights