Noticing Formulaic Sequences—A Problem of Measuring the Subjective

Hugh Bishop
University of Wisconsin–Madison

Abstract

An important idea in second language acquisition is Schmitt's noticing hypothesis, which asserts the importance of consciousness in second language learning, and which has recently been applied to the hitherto neglected area of vocabulary acquisition. One problem area of the lexicon for second language learners of English is the acquisition of formulaic sequences, which are holistically deployed lexical strings that contribute to fluent target language performance. It was hypothesized that the problem second language users have with formulaic sequences is that they are not noticed and as a result they are not learned. This paper describes how an experiment was carried out to test whether noticing occurs. The results are consistent with the hypothesis that second language readers do not notice unknown formulaic sequences.

1 Introduction

Schmitt and Celce-Murcia (2002) claim that applied linguistics is concerned not only with language, but also with how it is learned, and how it is used in order to achieve some further purpose (1). They assert that a dominant area of interest in applied linguistics is in the teaching and learning of second languages, i.e., second language acquisition (SLA).

Within the field of SLA, the Lexicon, once thought of as the repository of the idiosyncratic, has attracted increasing research interest (Schmitt and McCarthy 1997: 1). The lexicon is of particular importance to the phenomenon of dual encoding, which Skehan (1998) argues arises from the requirement for rapid processing of language in real time (30). He argues that utterances are often not constructed from scratch by application of grammatical rules to individual lexical items, because this would require considerable computational power which would detract from planning language production; rather, utterances are constructed using 'pre-fabricated chunks' of language which are retrieved lexically. Wray (2002) labels these chunks *formulaic sequences* (4). Dual encoding implies the existence not only of a rule-based grammar which operates on words, but also of a memory based lexical system which generates fast access to pre-fabricated language units. Both systems are employed by fluent target language users. Most second language users never entirely master the second language. It may be that not being able to achieve mastery of the memory-based lexical system is an important contributory factor to this problem.

In English, the memory-based lexical system depends on word strings, i.e., formulaic sequences, which consist of idioms, phrasal verbs, compounds and other phrases with varied discourse functions. It is argued that formulaic sequences are a critical lexical component of dual encoding, and are thus, according to Pawley and Syder (1983), of considerable importance in fluent target language performance (191). Formulaic sequences are not generated grammatically. They are stored and deployed as holistic units. This can be a source of confusion for learners since grammatically generated strings of words can appear identical to formulaic sequences. For example, *come in for* can be a formulaic sequence (meaning 'receive' as in *come in for criticism*),

LSO Working Papers in Linguistics 4: 15-19. Madison, Wisconsin: University of Wisconsin-Madison Department of Linguistics. Proceedings of WIGL 2004

but it can also be no more than three separate lexemes assembled grammatically as in *she came in for a cup of coffee*.

Consequently, for second language learners, the acquisition of formulaic sequences is fraught with difficulty (Wray 2002: 182; Arnaud and Savignon 1997: 168). Formulaic sequences appear to be less readily mastered incidentally than words. This is important, because incidental learning, the learning of vocabulary while doing something else such as reading, is an important source of vocabulary for second language learners, especially with low-frequency lexical items that appear only in print (Wesche and Paribakht 1999: 175). Anything that disrupts the acquisition of formulaic sequences is worthy of further study.

Although no direct comparisons between words and formulaic sequences have been made, de Bot, Paribakht and Wesche's (1997: 309) modification of Levelt's (1989) lexical processing model can be used to provide an explanation for the difficulties that learners have. From the model it can also be predicted that formulaic sequences are likely to cause more problems during incidental learning (i.e. while reading) than their single word equivalents, because formulaic sequences are not noticed. De Bot et al. argue that there needs to be recognition of an unfamiliar lexical form, and also the conscious awareness that the form lacks a meaning, before steps can be taken to establish its meaning, as is necessary for subsequent learning. Clearly if a subject does not know that he or she does not know a word, learning is not likely to take place.

Thus Schmidt's noticing hypothesis is an important theoretical part of this argument. Schmidt (1990) argued that if a learner is not consciously aware of a specific language feature, i.e, is unable to articulate that it is problematic, the learner will not be able to learn that language feature whether grammatical, lexical or pragmatic (129).

It is quite plausible that second language readers do not notice unknown formulaic sequences (Arnaud and Savignon 1997: 168), but the verification of this assertion is fraught with methodological difficulty, because of the essentially subjective nature of noticing.

2 Methodological Problems

It is necessary to operationalize the slippery notion of consciousness, which some commentators such as McLaughlin (1990) argue is simply too imprecise (617), and others argue is essentially epiphenomenal, and so irrelevant for study (Jackendoff 1987).

Schmidt, however, takes the view that a precise definition of consciousness is useful and necessary for the study of noticing (Schmidt 1993: 209). He operationalizes it as the ability to verbalize a specified linguistic feature. To verify the idea of noticing, it is necessary to measure this subjective phenomenon. A number of indirect measures have been used. Hegelheimer and Chapelle (2000), in a survey of research on lexical noticing, describe three approaches (41).

The first is to construct *conditions for noticing* in the experimental material by highlighting particular linguistic features. The second method is to use *retrospective think-aloud protocols*. The third method, which is the method employed in this paper, uses *direct observation of observable interactions (concurrent noticing)* such as negotiation of meaning during task activity, or clicking on words on computer screens for help in real time. Swain (1998) states the advantage that concurrent noticing has over the other methods, "...it seems essential in research to test what learners actually do, not what the research assumes instructions and task demands will lead learners to focus on" (80). The current paper will describe an experiment designed to achieve concurrent noticing using computer technology, with its unique capacity for online performance tracking. The experiment will be described briefly below, but see (Bishop in preparation) for fuller details of the development of the measuring instrument.

3 Procedure

Forty-four linguistically heterogeneous upper intermediate ESL students of both genders at the University of Wisconsin–Madison were pre-tested for knowledge of target items on a computerized version of the Vocabulary Knowledge Scale (VKS). Target items consisted of 20 low frequency words, and 20 formulaic sequences which were synonymous with those words were presented in random order for each subject. Subjects' reading levels were also assessed with a TOEFL reading subtest. No difference in reading ability between groups was found.

A week later, to allow for forgetting of the pre-test, subjects were randomly assigned to a control condition and a treatment condition. All subjects read identical texts which differed from each other only in that formulaic sequences were made typographically salient (color and underlining) in the treatment condition. After reading, subjects were required to answer 20 true/false questions focused on sentences containing the target words. Subjects accessed an online glossary by single clicking with the mouse on target words, and double clicking with the mouse on target formulaic sequences. Whenever a request was made for a gloss of an unknown target item, this was construed by the experimenter to mean that the lexical item had been noticed. All mouse clicks were tracked and stored.

4 Results

Unknown words were clicked on significantly more often than unknown formulaic sequences. However, making formulaic sequences typographically salient significantly increased the number of times formulaic sequences (FSs) were noticed, i.e., clicked on . This also appeared to increase global comprehension of the text. Since the data could not be safely assumed to be normally distributed, the non-parametric equivalent of the t-test (Mann-Whitney), and the paired t-test (Wilcoxon Signed Rank) were used.

The mean number of times the various unknown lexical items were clicked upon for glosses is shown in Table 1. The number of clicks on unknown words in the treatment and control conditions does not vary significantly (Mann-Whitney $p = 0.73$), which is to be expected because the words were not altered typographically in either condition, and so were presented identically. Between groups it can also be seen that salient formulaic sequences were glossed significantly more frequently than non-salient sequences (Mann-Whitney $p = .0005$).

In the control condition, the target words (non-salient) were clicked on significantly more often than the non-salient formulaic sequences (Wilcoxon Signed-Rank $p = .01$). However, in the treatment condition, the power of typographic salience is demonstrated. Unknown salient formulaic sequences are looked up significantly more frequently than unknown words (Wilcoxon Signed-Rank $p = .0022$).

Table 1: Number of unknown lexical items glossed

Condition	Treatment				Control			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Words	2.05	2.01	0	6	2.22	1.68	0	5
FS	5.00	3.33	0	10	1.43	1.44	0	4

Additionally, the treatment group achieved higher total scores on the reading comprehension task than the control group (17.24 vs. 16.26). Since the data was not normally distributed, the non-parametric equivalent of the t-test was used. The results were significant at $P = 0.04$ using a Mann-Whitney test.

5 Discussion

The results are consistent with the hypothesis that the problem with formulaic sequences for second language readers is, at least in part, that they are not noticed and so are less readily learned. This is significant given the major role that formulaic sequences play in the dual systems of fluent target language use (Wray 2002; Abel 2003: 329), and given that second language learners rarely reach native speaker target competence (Marinis 2003: 44).

The major problem for the learner seems to be the need to learn large numbers of holistically stored formulaic sequences which must coexist with a potentially vast number of similar looking, but grammatically generated, sequences. A priori, there is no clear way of distinguishing between the two types of unknown word string. More specifically, using Levelt (1989)'s concept of the lexeme and the lemma, it is plausible to argue that the root of the problem lies in the inability of learners to match multi word lexemes with a single lemma (meaning). The lack of awareness that a single meaning needs to be associated with a group of words precludes the association of lexeme (form) and lemma (meaning) which is a necessary condition for lexical learning. This experiment seems to show that formulaic sequences are not noticed and this is likely to lead to the inevitable consequence that they will not readily be learned.

6 Conclusion

The findings of this study are consistent with the claim that learners don't notice unknown formulaic sequences as readily as unknown words. However, making unknown formulaic sequences typographically salient increases readers' willingness to seek glosses, and this glossing might also lead to some increased comprehension of lexical items, although a replication of the experiment a semester later did not show increased comprehension. Solving the problem may not be as simple as identifying the problem, and it will be necessary to make a clear distinction between comprehending lexical items in context and actually learning them.

It is risky to generalize too strongly from such a small-scale study, but nonetheless a number of interesting questions are raised. Can typographical salience help overcome learners' tendencies to ignore lexical items? Does increased attention to local meanings of formulaic sequences in terms of clicking on glosses detract from construction of larger text-based meanings because of constant interruptions to the reading process? Does increased glossing lead to increased incidental learning of formulaic sequences? What effect does it have on intentional learning? Answers to these questions are of both pedagogical and theoretical interest. Further study of the way formulaic sequences are processed by second language readers is certainly required.

References

Abel, Marina. 2003. "English idioms in the first language and second language lexicon: a dual representation approach," *Second Language Research* 19(4): 329-358.

Arnaud, Pierre and Sandra Savignon. 1997. "Rare words, complex lexical units and the advanced learner," in James Coady and Thomas Huckin, eds., *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press. 157-173.

Bishop, Hugh. In preparation. Doctoral dissertation.

De Bot, Kees, T Sima Paribakht, and Marjorie Wesche. 1997. "Towards a lexical processing model for the study of second language vocabulary acquisition," *SSLA* 19: 309-329.

Hegelheimer, Volker, and Carol Chapelle. 2000. "Methodological issues in research on learner- computer interactions in CALL," *Language Learning and Technology* 4: 41-59.

Jackendoff, Ray. 1987. *Consciousness and the computational mind*. Cambridge, MA, MIT Press.

Levelt, Willem. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Marinis, Theodore. 2003. "Psycholinguistic techniques in second language acquisition research," *Second Language Research* 19(2): 144-161.

McLaughlin, Barry. 1990. "Conscious versus unconscious learning," *TESOL Quarterly* 24: 617-634.
Pawley, Andrew, and Frances Syder. 1983. "Two puzzles for linguistic theory: native-like selection and native-like fluency," in Jack Richards and Richard Schmidt, eds., *Language and Communication*. London: Longman. 191-227.
Schmidt, Richard. 1990. "The role of consciousness in second language learning," *Applied Linguistics* 11: 129-158.
Schmidt, Richard. 1993. "Awareness and second language acquisition," *Annual Review of Applied Linguistics* 13: 206-226.
Schmitt, Norbert, and Marianne Celce-Murcia. 2002. "An Overview of Applied Linguistics," in Norbert Schmitt, ed., *An Introduction to Applied Linguistics*. New York: Oxford University Press. 1-16.
Schmitt, Norbert, and Michael McCarthy. 1997 "Introduction," in Norbert Schmitt and Michael McCarthy, eds., *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press. 1-5.
Skehan, P. 1998. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press. 29-41.
Swain, M. 1998. "Focus on form through conscious reflection," Catherine Doughty and Jessica Williams, eds., *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press. 64-81.
Wesche, Marjorie, and T. Sima Paribakht. 1999. "Introduction," *SSLA* 21: 175-180.
Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Hugh Bishop
Department of English
White Hall Room 7187
University of Wisconsin–Madison
Madison, WI 53706

hbishop@wisc.edu