Lexical Bundles and Implicit Lexical Knowledge

GAO Hai-long
China Youth University of Political Studies, Beijing, China

In recent years, there have been emergent interests in L2 learners’ implicit lexical knowledge. Researchers employed psycholinguistic methods to investigate the online processing of formulaic sequences in L2 (idioms and collocations) and find evidence for holistic processing of the target linguistic sequences. This paper reviews and comments on lexical bundles and related concepts, trying to make clear this well focused research area in second language acquisition (SLA). Literature is reviewed and criticized on frequency effect in L2 learning, formulaic sequences, and implicit/explicit knowledge/learning, including their theoretical basis, methodology, and conclusion. In this paper, the author argues that sufficient practice, either receptive or productive, can turn explicit knowledge implicit, making it automatically accessed and retrieved. Suggestions on further study in lexical bundles are given.

Keywords: lexical bundles, implicit knowledge, explicit knowledge

Introduction

In recent years, scholars started to look at the explicit/implicit knowledge at the lexical level, especially L2 learners’ implicit lexical knowledge. A few studies have investigated the processing by L2 learners of formulaic sequences which are idiomatic in meaning and structurally complete (Sonbul & Schmitt, 2013) and which are chosen by their high frequency but are structurally complete (“Formulas”) (Nan & Nekrasova, 2007).

Recent research in corpus linguistics (e.g., Hyland, 2008) has evidenced the importance of lexical bundles as a major component in coherent linguistic production and an essential aspect of the shared knowledge of a professional discourse community. However, lexical bundle is a concept closely related to explicit/implicit lexical knowledge/learning on the distinction of which there has hardly been any consensus. This paper tries to tease out the concepts around lexical bundles.

Frequency Effect in L2 Learning

In first language research, it has been adequately demonstrated that language users are sensitive to frequency effects in language (e.g., Hare, Ford, & Marslen-Wilson, 2001). For example, at single-word level, the more frequent a word is in language, the quicker it is visually recognized as a word by speakers of that language in lexical decision tasks (e.g., Monsell, 1991).

The importance of frequency for language acquisition has been given primacy in a group of related models of language acquisition, all of which place frequency at center stage. In these usage-based models (e.g., Bybee, 2007), language learning is viewed as a predominantly inductive and experience-driven process, and the frequency with which structures occur in use plays a pivotal role in the emergence of the language system. On the basis of repeated use in a speech community, symbolic structures of different sizes, which are associations
of semantic and phonological or orthographic structure—essentially form-meaning mappings with a clear resemblance to the Saussurean sign—eventually become automatically retrievable by speakers of that language.

There are some theoretical positions that suggest that L2 learners may not process frequently occurring structures as skillfully as native speakers, which would call into question the notion that usage-based models can be applied to L2 learning in a straightforward manner. The consensus in the research literature would seem to be that, although it is likely that frequency does play a role in L2 acquisition, there is a host of other factors that, in the words of N. C. Ellis (2006a), “filter and color the perception of the second language” (p. 110).

Many researchers have suggested that frequency might not be as central for L2 acquisition or that it might not be processed in the same way for learners as it is for native speakers. In her work on formulaic language, Wray (2002) suggested that L2 learners are (perhaps necessarily) more analytical than L1 learners and, as a result of this, are less able to attend to frequency data in the L2, particularly across larger formulaic sequences. Similarly, N. C. Ellis (2006b) suggested that the fragile features of L2 acquisition are those which, however available as a result of frequency, recency, or context, fall short of intake because of one of the factors of contingency, cue competition, salience, interference, overshadowing, blocking, or perceptual learning, which are all shaped by the L1.

However, there is emerging empirical evidence (Siyanova-Chanturia, Conklin, & Van Heuven, 2011) that is indicative of the fact that learners (perhaps particularly higher proficiency learners) may still be sensitive to frequency effects in L2 input even if their ultimate state of linguistic knowledge typically bears some fundamental differences from that of most L1 speakers.

**Formulaic Sequence vs. Lexical Bundles**

In existing literature, many terms have been used to refer to the formulaic sequences in language. Most literatures hold that formulaic sequences can be a cover term and include idioms and other prefabricated phrases. The most widely quoted definition of “Formulaic Sequences” is from Wray (2002):

> 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)

According to Schmitt and Carter (2004), some types of formulaic sequence have always been obvious in the form of idioms, proverbs, and sayings. They hold that formulas are different from formulaic sequences. Formula is often used, but usually to mean a string of formulaic language with idiosyncratic conditions of use. Jiang and Nekrasova (2007) equaled formulaic sequences with formulas, but made a distinction between formulas and idioms: “the meanings of formulas, but not those of idioms, can be derived by combining meanings of its component words. In second language (L2) learning and teaching, these formulaic sequences are often treated as unanalyzed phrases” (p. 433).

Three criteria can be found regarding the identification of formulaic sequences as compared with non-formulaic sequences.

1. Formulaic sequences equal one big word. R. Weinert (1995): multi-word strings which are produced or recalled as a whole chunk, much like an individual lexical item, rather than being generated from linguistics rules. N. C. Ellis (1996, p. 111): In effect, this means that the words in a formulaic sequence are “glued together” and stored as a single big word.
(2) Frequency in a corpus. Wray (2002, p. 466): …formulaic sequences can be identified by virtue of their being more frequent than other words strings, indeed that frequency is a central definitional criterion. Schmitt and Carter (2004): Another often cited criterion is frequency of occurrence, on the assumption that if a sequence is frequent in a corpus, this indicates that it is conventionalized by the speech community.

(3) Psycholinguistic criterion. Schmitt and Carter (2004): whether a sequence of words is produced more than once by a participant, and whether it is produced with an intact intonation contour.

With the wide use of computer technology and corpus linguistics, it has become possible to statistically analyze recurrent word combinations in discourse. This corpus-driven framework for the study of frequent longer word sequences was first applied in The Longman Grammar of Spoken and Written English (1999), and the term lexical bundle was created to distinguish such recurrent word sequences from other types of multi-word lexical items.

Biber and Conrad (1999) defined lexical bundles as

the most frequent recurring lexical sequences; …which can be regarded as extended collocations: sequences of three or more words that show a statistical tendency to co-occur (e.g., in the case of the, do you want me to, I said to him).

(p. 183)

Lexical bundles is a concept using frequency-driven approach, different from the other approach on formulaic sequences which identify sequences on “idiomaticity”.

According to Biber and Barbieri (2007), lexical bundles have several features which distinguish them from other kinds of formulaic expressions: (1) They are extremely common. (2) They are not idiomatic in meaning (idioms are not to be considered as lexical bundles as they occur much fewer times in corpus). For example, the meanings of bundles like “do you want to” or “I don’t know what” are transparent from the individual words. (3) They do not represent a complete structural unit. For example, Biber et al. (1999) found that only 15% of the lexical bundles in conversation can be regarded as complete phrases or clauses, while less than 5% of the lexical bundles in academic prose represent complete structural units. Instead, most lexical bundles bridge two structural units: They begin at a clause or phrase boundary, but the last words of the bundle are the beginning elements of a second structural unit. Most of the bundles in speech bridge two clauses (e.g., I want to know, well that’s what I), while bundles in writing usually bridge two phrases (e.g., in the case of, the base of the).

The concept of lexical bundles has been used in several later studies to investigate commonly used multi-word items in discourse, focusing mainly on written and spoken registers in university and academic contexts. However, the frequency cut-off applied in these studies varies, e.g., Biber and Barbieri (2007) used 40 per million words, Hyland (2008) opted for 20 per million words.

**Implicit/Explicit Knowledge vs. Implicit/Explicit Learning**

The phenomenon of implicit learning has aroused heated debate in cognitive psychology and in second language acquisition (SLA) over the past few decades (Leung & Williams, 2011). However, in existing literatures, there have been rather confusing discussions relating to implicit/explicit learning and implicit/explicit knowledge. It seems that most scholars tend to treat the two dichotomies as the same thing and use them interchangeably. That is, implicit learning (if it occurs) will automatically lead to implicit knowledge, and explicit learning (if it occurs) will automatically lead to explicit knowledge (Bowles, 2011; GAI & WEN, 2013; Sonbul & Schmitt, 2013). At the same time, the evidence of the implicit knowledge in L2 learners seems
to be the direct product of short period of training in which implicit learning occurs (Rebuschat, 2013). The idea is demonstrated in Figure 1, showing the process from implicit knowledge to implicit learning.

![Figure 1. From implicit learning to implicit knowledge.](image)

However, when the separate definitions are scrutinized, it is obvious that the above assumptions cannot hold ground. For example, Rebuschat (2013) defined implicit learning as “the learning process that results primarily in unconscious knowledge, that is, knowledge that is tacit and inaccessible to conscious introspection” (p. 597). Implicit learning generally occurs without learners’ intention to learn and without learners’ awareness of what has been learned, that is, learners are often unaware that they have acquired knowledge. Explicit learning is defined as “the learning process that results primarily in conscious knowledge”. The author further holds that the terms implicit knowledge and unconscious knowledge as well as explicit knowledge and conscious knowledge are used interchangeably in their study. This viewpoint of implicit learning and implicit knowledge seems to be shared by most literatures the author has found.

On the other hand, according to the widely quoted definition of implicit/explicit knowledge from N. C. Ellis (2006), implicit knowledge is procedural, is held unconsciously, and can only be verbalized if it is made explicit. It is accessed rapidly and easily and thus is available for use in rapid, fluent communication. On the other hand, explicit knowledge consists of the facts that speakers of a language have learned. Explicit knowledge is held consciously, is learnable and verbalizable, and is typically accessed through controlled processing when learners experience some kind of linguistic difficulty in using the L2.

It can be seen that the very important aspect of implicit knowledge is the process when knowledge is accessed rapidly and easily in rapid, fluent communication. The ability (of accessing related knowledge rapid and easily) can be learned through either explicit learning or implicit learning. Typical examples are highly proficient adult L2 learners who have been taught the target language in an explicit classroom. Through years of hard work and practice, they are able to possess implicit knowledge of the target language and access it rapidly and easily in real use of the language without awareness of use the knowledge. In short, that a learner learns something implicitly does not necessarily mean that they can use or access the knowledge in an easy and rapid fashion, especially without periods of practice.

In some recent studies (Leung & Williams, 2011; GAI & WEN, 2013), learners went through a short period of training/learning, and then they were assessed for evidence of implicit knowledge regarding the possible implicit learning involved in the training process. The author of the present paper wants to ask, without periods of practice in both receptive and productive ways, how could implicit knowledge be formed even though there was implicit learning involved?

In this paper, the author argues that whether it is explicit learning, implicit learning, or a combination of explicit learning and implicit learning, as long as there is enough practice in both receptive and productive ways so that the use of the knowledge can be automatically accessed and retrieved, implicit knowledge can be
formed. This framework is shown in Figure 2.

![Diagram of implicit knowledge formation]

The present study focuses on implicit knowledge (rather than implicit learning), which involves the ability to access and retrieve the related knowledge in an easy, rapid, and unconscious way. Evidence of implicit knowledge from the participants in the present study does not indicate whether they have learned the language explicitly or implicitly previously.

**Implicit Lexical Knowledge**

There are views that lexical knowledge is declarative in nature and can never be implicit (Sonbul & Schmitt, 2013). Ullman (2001), for example, developed his declarative/procedural model of linguistic knowledge assuming, based on neurological evidence, that vocabulary knowledge is declarative while grammatical knowledge is procedural. Similarly, Hulstijn (2007) claimed that vocabulary knowledge is symbolic and, thus, explicit in nature. These views seem to treat the lexicon as a simple dictionary containing only word forms connected to their meanings, which has led to the distinction between explicit and implicit lexical knowledge being largely ignored in research.

N. C. Ellis (1994), one of the first scholars to show an interest in the explicit-implicit contrast for lexis, distinguished two groups of components: (1) components related to the connection between the form of a word and its meaning and the various semantic relations between words which can be best learned explicitly and (2) components related to the form of the word and its receptive/productive aspects of usage which are best learned implicitly (but can also be learned explicitly).

R. Ellis (2004) made a similar claim:

> Whereas knowledge of the form of a word and of word collocations is largely implicit …Word meanings probably constitute the largest single area in a learner’s explicit knowledge. In addition, of course, individuals may develop a conscious awareness of the form of at least. (p. 242)

According to Sonbul and Schmitt (2013), due to researchers’ lack of conventional measures for implicit lexical knowledge, most previous lexical studies are explicit in nature, and they only measured breadth, depth, and organization of vocabulary knowledge. Implicit measures used in psychological research are just beginning to be used in investigating vocabulary acquisition issues (e.g., online serial reaction time measures).

**A Brief Review of Previous Studies**

Two previous studies have selected linguistic sequences based on their frequency which are structurally complete. Bod (2001), using a lexical-decision task, showed that high-frequency three-word sentences such as...
“I like it” were reacted to faster than low-frequency sentences such as “I keep it”. More recently, Jiang and Nekrasova (2007) found in two on-line grammaticality judgment tasks that native and non-native speakers of English responded to three-word formulas faster and more accurately than the control phrases which are non-formulas. Although these two studies employed online reaction times to investigate the processing of the linguistic sequences they chose based on frequency, they are structurally complete, and cannot really represent lexical bundles in its strict sense.

Tremblay, Derwing, Libben, and Westbury’s (2011) paper is the only study so far available that investigated the processing of “real” lexical bundles. The study used three self-paced reading experiments to examine the extent to which lexical bundles, (“frequently recurring strings of words that often span traditional syntactic boundaries” in their definition), are stored and processed holistically in L1 English. Three self-paced reading experiments compared sentences containing lexical bundles (e.g., in the middle of the) and matched control sentence fragments (e.g., in the front of the). Lexical bundles and sentences containing lexical bundles were read faster by the subjects than the control sentence fragments in all three experiments. Two follow-up word and sentence recall experiments demonstrated that more sentences containing lexical bundles were correctly remembered.

Conclusion

To conclude, this paper reviews and comments on lexical bundles and related concepts, namely formulaic sequences and implicit/explicit knowledge/learning.

So far there have been no studies that investigate the processing of lexical bundles for L2 learners using online reaction time measures. Future study should address the following questions:

1. Will L2 learners demonstrate processing advantages for lexical bundles over non-lexical bundles, which indicates that they possess the implicit lexical knowledge?
2. Will L2 learners show the same pattern of processing advantages of lexical bundles compared with non-lexical bundles between the self-paced reading task and the word monitoring task?

References

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*, 263-286.
Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgård & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 181-190). Amsterdam and Atlanta: Rodopi.
Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. London: Longman.
Bod, R. (2001). Sentence memory: Storage vs. computation of frequent sentences. Retrieved from http://staff.science.uva.nl/~rens/cuny2001.pdf
Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge—What can heritage language learners contribute? *Studies in Second Language Acquisition, 33*, 247-271.
Bybee, J. L. (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press.
Ellis, N. C. (1994). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review, 11*, 37-56.
Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18*(1), 91-126.
Ellis, N. C. (2006a). Cognitive perspectives on SLA: The associative-cognitive CREED. *AILA Review, 19*, 100-121.
Ellis, N. C. (2006b). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*, 164-194.
Ellis, N. C. (2008). Implicit and explicit knowledge about language. In J. Cenoz & N. H. Hornberger (Eds.), Encyclopedia of Language and Education. Vol. 6: Knowledge about language (2nd ed., pp. 119-132). New York: Springer.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. Language Learning, 54, 227-275.

GAI, S. H., & WEN, Q. F. (2013). Learning without awareness by Chinese EFL learners. Foreign Language Teaching and Research, 45(4), 557-567.

Hare, M. L., Ford, M., & Marslen-Wilson, W. D. (2001). Ambiguity and frequency effects in regular verb inflection. In J. L. Bybee & P. J. Hopper (Eds.), Frequency and the emergence of linguistic structure (pp. 181-200). Amsterdam: Benjamins.

Hulstijn, J. H. (2007). Psycholinguistic perspective on language and its acquisition. In J. Cummins & C. Davison (Eds), The international handbook of English language teaching (pp. 783-796). Norwell, MA: Springer.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. English for Specific Purposes, 27(1), 4-21.

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

Leung, J. H., & Williams, J. N. (2011). The implicit learning of mappings between forms and contextually derived meanings. Studies in Second Language Acquisition, 33, 33-55.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), Basic processes in reading: Visual word recognition (pp. 148-197). Mahwah, NJ: Erlbaum.

Nan, J., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. The Modern Language Journal, 91, 433-445.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. Language Learning, (3), 595-626.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), Formulaic sequences: Acquisition, processing and use (pp. 1-22). Amsterdam: John Benjamins.

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(3), 776-784.

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. Language Learning, 63, 121-159.

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.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. Bilingualism: Language and Cognition, 4, 105-122.

Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. Applied Linguistics, 16, 180-205.

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

Wray, A., & Perkins, M. (2000). The functions of formulaic language: An integrated model. Language and Communication, 20, 1-28.