

# Formulaic Language and Collocation

NORBERT SCHMITT

It is now clear that vocabulary typically behaves not as single words which are held together by syntax, but rather has a strong tendency to occur in multiword phraseological units. Formulaic language is very common in language overall, in both spoken and written modes. For example, Erman and Warren (2000) calculated that 52–8% of the language they analyzed was formulaic, and Foster (2001) came up with a figure of 32% using different procedures and criteria. Biber, Johansson, Leech, Conrad, and Finegan (1999) found that around 30% of the words in their conversation corpus consisted of lexical bundles, and about 21% of their academic prose corpus. Howarth (1998) looked at frequent verbs in a social science/academic corpus and found that they occurred in either restricted collocations or in idioms in 31–40% of the cases. This research convincingly shows that formulaic language is no mere peripheral feature of language, but rather is a ubiquitous and essential feature of how it is used.

## The Functions of Formulaic Language

Being such a big part of language, it is not surprising that formulaic language as a phenomenon is not homogeneous (although many researchers treat it as if it is). It realizes different purposes in language use, including transacting routinized meanings (*that'll be X dollars* = typical way for American shopkeepers to state the cost of a bill), and smoothing social interaction (*yeah, it is* = a routinized way of agreeing with an interlocutor's assertion) (Schmitt & Carter, 2004). It also provides the building blocks upon which one can create more extended strings of language (e.g., with collocations (*valid point*), and with lexical bundles: *it should be noted that* = a standard academic phrase which highlights a point of interest (Biber et al., 1999).

Moreover, one reason that formulaic language is so widespread is that it lexicalizes a wide number of referential, textual, and communicative functions in discourse. It can be used to express a concept (*get out of Dodge* = get out of town quickly, usually in uncomfortable circumstances), state a commonly believed truth or advice (*too many cooks spoil the soup* = it is difficult to get a number of people to work well together), provide phatic expressions which facilitate social interaction (*nice weather today* is a nonintrusive way to open a conversation), signpost discourse organization (*on the other hand* signals an alternative viewpoint), and provide technical phraseology which can transact information in a precise and efficient manner (*2-mile final* is a specific location in an aircraft landing pattern) (Schmitt & Carter, 2004). Likewise, Nattinger and DeCarrico (1992) argue that formulaic language fulfils the functions of maintaining conversations (*How are you?*, *See you later*), describing the purposes for which the conversations take place (*I'm sorry to hear about X*, *Would you like to X?*), and realizing the topics necessary in daily conversations (*When is X?* (time), *How far is X?* (location)). In fact, one might suppose that for every conventional activity or function in a culture, there will be associated phrasal vocabulary. If that is so, there are bound to be a large number of formulaic expressions, perhaps even a larger number than that of single-word vocabulary.

## 2 FORMULAIC LANGUAGE AND COLLOCATION

Formulaic sequences become particularly important in language use when we consider their pragmatic value. For instance, they are very often used to accomplish recurrent communication needs. These recurrent communicative needs typically have conventionalized language attached to them, such as *I'm (very) sorry to hear about . . .* to express sympathy and *I'd be happy/glad to . . .* to comply with a request (Nattinger & DeCarrico, 1992, pp. 62–3). Because members of a speech community know these expressions, they serve as a quick and reliable way to achieve the desired communicative effect. Formulaic sequences also realize a variety of conversational routines and gambits and discourse objectives (Coulmas, 1979, 1981). They are typically used for particular purposes and are inserted in particular places in discourse. For instance, formulaic sequences regularly occur at places of topic-transition and as summaries of gist (Drew & Holt, 1998). Most (all?) conventional speech acts are realized by families of formulaic language and not normally by original expressions (*I'm very sorry* versus *I am feeling apologetic toward you*). Overall, understanding the pragmatic role of formulaic language can tell us much about the nature of interaction (McCarthy & Carter, 2002).

Moreover, formulaic sequences do more than just carry denotative meaning and realize pragmatic function. They can often have a type of register marking called *semantic/collocational prosody* (Stubbs, 2002; Sinclair, 2004; Hunston, 2007). This is often negative—for example, the verb *cause* frequently has a negative evaluation (*cause pain, cause inflation*). However, semantic prosody can also be positive, as in collocations that form around the word *provide* (*provide information, provide services*). This semantic prosody is one means of showing a speaker/writer's attitude or evaluation. For example, their stance can be indicated concerning the knowledge status of the proposition following the formulaic item (*I don't know if X* indicates uncertainty about X), their attitude toward an action or event (*I want you to X* shows a positive attitude toward this action), and their desire to avoid personal attribution (*it is possible to* avoids a directly attributable suggestion) (Biber, Conrad, & Cortes, 2004). Likewise, the choice of formulaic sequences can reflect an author's style and voice (Gläser, 1998). Formulaic sequences can also be used to encode cultural ideas, as Teliya, Bragina, Oparina, and Sandomirskaya (1998) have demonstrated for Russian.

### Acquisition of Formulaic Language

The learning of individual words is incremental and each word has its own particular learning burden (Nation, 1990; Schmitt, 2000), and there is no reason to believe that formulaic language is any different in this respect. This would suggest that many formulaic sequences are partially known for a number of exposures until the point where they become mastered. Certainly, some L1 acquirers learn and use formulaic sequences before they have mastered the sequences' internal makeup. For example, they sometimes seem to acquire an initial phonological mapping of formulaic sequences proceeding from the whole to the individual parts, but with some elements still incompletely grasped, especially the unstressed phonemic constituents (e.g., Peters, 1977; Plunkett, 1993). In L1 acquisition, children with expressive tendencies (having interactional goals; focusing on the social domain) were more likely to learn formulaic expressions than children who had tendencies to name things with individual words (Nelson, 1973). This points out the importance of formulaic language in promoting the kind of social integration which promotes language acquisition in general. In L2 acquisition, formulaic sequences are also relied on initially as a quick means to be communicative, albeit in a limited way. This can lead to faster integration into a peer group, which can result in increased language input. Wong Fillmore (1976) found this was the case with five young Mexican children trying to integrate into an English-medium school environment. She identified eight strategies the children used, and at least three of them directly involved formulaic language:

- Give the impression, with a few well-chosen words (phrases), that you speak the language.
- Get some expressions you understand, and start talking.
- Look for recurring parts in the formulas you know.

The use of formulaic sequences enabled the realization of these strategies even though the children's language capabilities were quite limited. Furthermore, the use of formulaic sequences to facilitate language production is not restricted to L2 children. Schmidt's (1983) study of Wes is a good example of the phenomenon in L2 adults; Wes's speech is filled with formulaic language as a means of fulfilling his desire to be communicative, but not necessarily accurate. Additionally, Adolphs and Durow (2004) found that the amount of social integration into the L2 community (with presumably a commensurate need to be communicative in the L2) was linked to the amount of formulaic language produced in the speech of L2 postgraduate students.

But formulaic sequences may provide language learners with more than an expedient way to communicate; they might also facilitate further language learning. For L1 learners, it has been proposed that unanalyzed sequences provide the raw material for language development, as they are segmented into smaller components and grammar (Peters, 1983). For example, when a child realizes that the phrase *I wanna cookie* (previously used as a holistic unit) is actually *I wanna + noun*, he or she gains information about the way syntax works in the language, as well as the independent new word *cookie*. A number of studies suggest that some children segment formulaic sequences into smaller units, and in doing so, advance their grammatical and lexical knowledge (see Wray, 2000, for one review). It seems that formulaic sequences serve the same purpose for L2 learners (e.g., Myles, Hooper, & Mitchell, 1998; Bardovi-Harlig, 2002). Moreover, there is little doubt that the automatic use of acquired formulaic sequences allows chunking, freeing up memory and processing resources (see Kuiper, 1996; and Ellis, 1996, who explores the interaction between short-term and long-term phonological memory systems). These can then be utilized to deal with conceptualizing and meaning, which must surely aid language learning.

L1 children are exposed to many formulaic sequences in their input, but how do they decide what to analyze and what to keep at the holistic level? Wray (2002) suggests that a "needs-only analysis" is the mechanism. Rather than segmenting every sequence into the grammar system, children will operate with the largest possible unit, and only segment sequences when it is useful for social communication. Thus the segmentation process is driven by pragmatic concerns (communication), rather than an instinctive urge to segment in order to push grammatical and lexical acquisition. The default would be to not analyze, and to retain holistic forms. Thus children maintain many formulaic sequences into adulthood, even though the components of those sequences are likely to be stored individually as well (perhaps being acquired from the segmentation analyses of other formulaic sequences). This suggests that dual storage is the norm. (It is useful to note that the notion of "holistic storage" is actually a metaphor. There is yet no conclusive evidence as to whether formulaic sequences are actually stored in memory as wholes, or whether they are simply computed very efficiently in a manner which gives them the characteristics of holistic units.)

Of course, relying on holistic versus analytical approaches to language acquisition and use is not an either/or proposition, and children will use both approaches in varying degrees. However, Wray and Perkins (2000) and Wray (2002) suggest that the relative ratios between the approaches may change according to age. During phase 1 (birth to around 20 months), the child will mainly use memorized vocabulary for communication, largely learned through imitation. Some of this vocabulary will be single words, and some will consist of sequences. At the start of phase 2 (until about age 8), the child's grammatical awareness begins, and the proportion of analytic language compared to holistic language increases, although with overall language developing quickly in this phase, the

amount of holistically processed language is still increasing in real terms. During phase 3 (until about age 18), the analytic grammar is fully in place, but formulaic language again becomes more prominent. "During this phase, language production increasingly becomes a top-down process of formula blending as opposed to a bottom-up process of combining single lexical items in accordance with the specification of the grammar" (Wray & Perkins, 2000, p. 21). By phase 4 (age 18 and above), the balance of holistic to analytic language has developed into adult patterns.

The course of formulaic sequence development is more difficult to chart in L2 learners. Typically there is early use of formulaic sequences, often after a silent period. As learners' proficiency improves, there is a reasonable expectation of language which is more accurate and appropriate. In natives, this is achieved to a large extent through the use of formulaic sequences (e.g., Nattinger & DeCarrico, 1992). Unfortunately, the formulaic language of L2 learners tends to lag behind other linguistic aspects (Irujo, 1993), but this is not so much a case of the amount of formulaic language use, but rather a lack of native-like diversity. This is probably largely due to a lack of sufficient input. Irujo (1986) suggested that one specific class of formulaic language (idioms) is often left out of speech addressed to L2 learners, leading to a lack of idioms in learner output. More recently, Durrant and Schmitt (2009) show that a more general type of formulaic language (collocations) seems to be tuned to frequency, with L2 learners producing frequent, but not infrequent, collocation pairs. Furthermore, Siyanova and Schmitt (2008) showed that spending a year in an English-speaking country (with presumably a great increase in the amount of L2 input) led to better intuitions of collocation.

However, it may not be just the *amount* of input that is crucial, but also the *quality*. Siyanova and Schmitt (2007) found that the amount of exposure to native-speaking environments did not have an effect on the likelihood of using the multiword verbs. This, however, might be explained by Adolphs and Durow's (2004) findings that sociocultural integration was the key to their case-study learner's acquisition. This suggests that it may not be exposure per se that is important, but the kind of personalized and engaging exposure that presumably occurs in a socially integrated environment.

The nature of formulaic language and its acquisition is likely to become of ever-greater interest as the field turns to more pattern-based models of language acquisition—e.g., pattern grammar (Hunston & Francis, 2000) and usage-based theory (Tomasello, 2003)—which posit that the human facility for language learning is based on the ability to extract patterns from input, rather than being under the guidance of innate principles and parameters which determine what aspects of grammar can and cannot be acquired (see Langacker, 1987; Bybee, 1998; Ellis, 1996, 2002). For example, this line of thinking suggests that we learn the letter sequences which are acceptable in a language (the consonant cluster *sp* can be word-initial in English, but *hg* cannot) simply by repeatedly seeing *sp* at the beginning of words, but not *hg*. This learning is implicit, and may not be amenable to conscious metalinguistic explanation. Of course, learners may eventually reach the point where they can declare a "rule" for this consonant clustering, but the rule is an artifact of the pattern-based learning, rather than the underlying source of learning (e.g., Bod, 2006).

This pattern-based learning also works for larger linguistic units, such as the way in which sequences of morphemes combine to form words (*un-question-able*, *un-reli-able*, *un-fathom-able*). Moving to words, we gain intuitions about which words collocate together and which do not (*blonde hair*, *\*blonde paint*; *auburn hair*, but only for women, not men). Many of these collocations must be based solely on associative pairing, because there is often no semantic reasoning behind acceptable/nonacceptable combinations (*\*blonde paint* makes perfectly logical sense). Neither are most collocations likely to be learned explicitly, because they are not normally taught, and even if they are, only possible cases are illustrated, not inappropriate combinations. Longer formulaic strings, which are also based on

patterns rather than rules, seem to fit very nicely with such sequence-based models of acquisition as well. Time will tell whether this kind of model best captures the mechanics of formulaic sequence acquisition (and that of language in general). However, one thing seems certain; given the increasingly evident importance of formulaic sequences in language use, convincing explanations of the mechanics of their acquisition must become an essential feature of any model of language acquisition.

### Non-Native Use of Formulaic Language

Formulaic language is very widespread in L1 language use (e.g., Biber et al., 1999; Erman & Warren, 2000; Foster, 2001). But what about non-native speakers? There is a widespread feeling that formulaic language is especially problematic for L2 learners, and its lack/misuse is a major reason why L2 output can feel unnatural and non-native-like, at least in their compositions (most research on formulaic language has focused on written discourse). Research has only partially supported this impression. We can look at non-native mastery of formulaic language along at least three dimensions: amount of use, accuracy/appropriacy of use, and quality/speed of the underlying formulaic intuitions. There is a growing literature about the first two dimensions, but only embryonic research on the last. Let us look at each dimension in turn.

#### Amount of Use

It is easy to assume that that problem with non-natives is that they simply do not use as much formulaic language as natives. This is largely incorrect, although there can be an element of avoidance (Laufer & Eliasson, 1993; Laufer, 2000). A series of studies have found that L2 usage depends on *which* formulaic sequences one is focusing on. It is now clear that non-natives actually use *more* of certain favorite formulaic sequences which they know well and tend to overuse as “safe bets,” compared to natives (Granger, 1998; de Cock, 2000; Foster, 2001). Conversely, they use *fewer* of other sequences, presumably because they do not know them as well and are not as confident in their use (e.g., Granger, 1998; Howarth, 1998; Foster, 2001). One type of formulaic sequence which seems to be particularly underused is multiword verbs. For example, Altenberg and Granger (2001) found that their EFL learners had great difficulty with the verb *make*, especially the delexicalized uses, such as *make a decision* and *make a claim*. This is particularly problematic as high-frequency verbs like *make*, *look*, and *do* are used in numerous formulaic sequences. Interestingly though, Granger, Paquot, and Rayson (2006) compared formulaic sequences in a one-million-word native-speaker academic corpus and one million words from the ICLE learner corpus (International Corpus of Learner English) and found more cases of overused formulaic sequences than underused ones.

Durrant and Schmitt (2009) go some way in explaining which formulaic sequences are overused and which underused. Using a corpus composed of written academic output from Turkish and Bulgarian university EFL students and a mixed group of international university students studying in the UK, they found that these students tended to use frequent premodifier-noun collocations at a rate similar to native students. (Congruently, Siyanova and Schmitt (2008) found that their non-natives used adjective-noun collocations in frequencies similar to natives.) These are the kind of collocations which are identified by measures which give a relatively heavy weighting to frequency, such as the *t* score statistic (*good example*, *long way*, *hard work*). However, the non-natives produced many fewer low-frequency collocations (*densely populated*, *bated breath*, *preconceived notions*), even though these combinations are very strongly linked. (This kind of lower-frequency, but tightly bound, collocation is better identified by the mutual information [*MI*] collocation statistic).

Because of their strong ties, and relative infrequency, they are likely to be especially salient for natives, and so their absence in non-native output is particularly noticeable. The authors conclude that the lack of these “MI” collocations is one key feature which distinguishes native from non-native production. In terms of acquisition, L2 learners seem to be able to acquire and use the collocations which appear frequently, but do not seem to pick up as many nonfrequent collocations, whose individual component words may also be infrequent in themselves. This is highly suggestive of the role of frequency in the acquisition process. This is supported by Ellis, Simpson-Vlach, and Maynard (2008), who found that for natives, it is predominately the *MI* of a formulaic sequence which determines processability (as measured by reaction-time experiments), while for non-natives, it is predominately the *frequency*.

#### Accuracy/Appropriacy of Use

Oppenheim (2000) found that much of the language which her non-native subjects produced in consecutive speeches on the same topic consisted of the same recurrent word strings, but that most of these were idiosyncratic in comparison to native-speaker norms. Thus, just because L2 learners produce formulaic language, it is not necessary formulaic in the sense of matching what natives would produce.

Nesselhauf (2003, 2005) gives us some idea of how formulaic language can be “non-native.” She extracted 1,072 English verb–noun combinations from 32 essays in the ICLE written by German university students. Almost one-quarter of these collocations were judged to be incorrect; moreover, the L1 was deemed to be an influence in 45% of the errors. However, the incorrect usage was often the result not of combining words in an unconventional way, but of using conventional word pairs in ways which are not appropriate (Nesselhauf, 2005). This suggests that the difficulty learners have is not only that of learning which words go together, but also learning how to employ the chunks they know. Therefore, at least for the more frequent collocations, the problem may not be so much in the *amount* of formulaic language learners use, but in using the formulaic sequences they know *appropriately* in the right contexts.

#### Quality/Automaticity of Intuitions of Formulaic Language

So formulaic sequences can be overused, underused, and misused by non-native writers (most of this research has been based on analysis of written text), but they are definitely used; there is no question that L2 output is devoid of formulaic language. But how good are the non-native intuitions of this language? There is little research which addresses this; however, three studies found that non-native intuitions were not as well developed as native intuitions.

Siyanova and Schmitt (2008) directly compared native and non-native judgments of the frequency of high-frequency, mid-frequency, and low-frequency adjective-noun collocations on a six-point Likert scale. They also measured how long it took to make these judgments. They found that the natives had fairly good intuitions of the collocation frequency, and that they made their frequency judgments relatively quickly. Compared to these native norms, the non-natives judged the high-frequency collocations as being lower frequency, and judged the low-frequency collocations as being much higher. Furthermore, natives were able to distinguish the frequency difference between mid- and high-frequency collocations, but the non-native as a whole were not. Interestingly though, the non-natives who spent a year or more in an English-speaking country were able to make this distinction. Also, the non-natives took much longer to make their frequency judgments. Taken together, Siyanova and Schmitt conclude that the non-natives’ intuitions were not as developed as the natives’, nor were they as automatized.

Hoffman and Lehmann (2000) elicited native and non-native speakers' intuitions regarding 55 collocations from the BNC with high log-likelihood scores (mainly adjective–noun and noun–noun pairs). Respondents were presented with each node in a questionnaire, and were asked to supply the collocates. On average, the native speakers supplied the “correct” collocate in 70% of cases, which, like the results in Siyanova and Schmitt, indicates relatively good intuitions by the natives. The non-natives did far less well, achieving an average accuracy of only 34%. This shows a major gap between the native and non-native intuitions, although in absolute terms, the non-native results (producing about one third of the collocates) still indicate considerable knowledge.

This poorer intuitive mastery is reflected in learners' production. While natives tend to resort to formulaic language to get through time-pressurized communicative situations, non-natives do not seem to make greater use of formulaic language in such cases, either in speech or writing (Foster, 2001; Nesselhauf, 2005). In terms of speech, non-natives tend to use many recurrent dysfluency markers (such as filled pauses and hesitation markers), although it seems that extensive interaction with native speakers enables them to overcome this (de Cock, Granger, Leech, & McEnery, 1998; Adolphs & Durow, 2004). However, in terms of writing, neither amount of use nor accuracy of collocation appears to increase with time spent in an English-speaking country (Yorio, 1980; Nesselhauf, 2005). So, although a year or more spent in an English-speaking country can lead to better *intuitions* of collocation (Siyanova & Schmitt, 2008), it seems difficult to extend this into increased *production* of formulaic language.

## Conclusion

It seems that mastery of formulaic language takes a long time to acquire and is a hallmark of the highest stages of language mastery. Language testers have picked up on this and often include items which focus on phraseology in their highest-level examinations. Formulaic language is an important element of language overall, perhaps *the* essential element. Research into it is only now gaining momentum, but given its ubiquitousness and demonstrated processing advantages, it looks to be one of the most important areas of inquiry in the applied linguistics field for the foreseeable future.

**SEE ALSO:** Corpus Linguistics in Language Teaching; Teaching Vocabulary

## References

- Adolphs, S., & Durow, V. (2004). Social-cultural integration and the development of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences* (pp. 107–26). Amsterdam, Netherlands: John Benjamins.
- Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of make in native and non-native student writing. *Applied Linguistics*, 22(2), 173–94.
- Bardovi-Harlig, K. (2002). A new starting point? Investigating formulaic use and input in future expression. *Studies in Second Language Acquisition*, 24, 189–98.
- Biber, D., Conrad, S., & Cortes, V. (2004). *If you look at . . .*: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Longman.
- Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review*, 23(3), 291–320.
- Bybee, J. (1998). The emergent lexicon. *Chicago Linguistics Society*, 34, 421–35.

- Coulmas, F. (1979). On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics*, 3, 239–66.
- Coulmas, F. (1981). *Conversational Routine*. The Hague, Netherlands: Mouton.
- de Cock, S. (2000). Repetitive phrasal chunkiness and advanced EFL speech and writing. In C. Mair & M. Hundt (Eds.), *Corpus linguistics and linguistic theory* (pp. 51–68). Amsterdam, Netherlands: Rodopi.
- de Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon on EFL learners. In S. Granger (Ed.), *Learner English on computer* (pp. 67–79). London, England: Longman.
- Drew, P., & Holt, E. (1998). Figures of speech: Figurative expressions and the management of topic transition in conversation. *Language in Society*, 27, 495–522.
- Durrant, P., & Schmitt, N. (2009). To what extent do native and nonnative writers make use of collocations? *International Review of Applied Linguistics*, 47, 157–77.
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18, 91–126.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143–88.
- Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 41(3), 375–96.
- 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. 75–94). Harlow, England: Longman.
- Gläser, R. (1998). The stylistic potential of phraseological units in the light of genre analysis. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 125–43). Oxford, England: Oxford University Press.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145–60). Oxford, England: Oxford University Press.
- Granger, S., Paquot, M., & Rayson, P. (2006). Extraction of multi-word units from EFL and native English corpora: The phraseology of the verb ‘make’. In A. Häcki Buhofer & H. Burger (Eds.), *Phraseology in motion I: Methoden und Kritik* (pp. 57–68). Akten der Internationalen Tagung zur Phraseologie (Basel, 2004). Baltmannsweiler, Germany: Schneider Verlag Hohengehren.
- Howarth, P. (1998). The phraseology of learners’ academic writing. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 161–86). Oxford, England: Oxford University Press.
- Hoffman, S., & Lehmann, H. M. (2000). Collocational evidence from the British National Corpus. In J. M. Kirk (Ed.), *Corpora galore: Analyses and techniques in describing English. Papers from the Nineteenth International Conference on English Language Research on Computerised Corpora (ICAME 1998)* (pp. 17–32). Amsterdam, Netherlands: Rodopi.
- Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics*, 12(2), 249–68.
- Hunston, S., & Francis, G. (2000). *Pattern grammar*. Amsterdam, Netherlands: John Benjamins.
- Irujo, S. (1986). A piece of cake: Learning and teaching idioms. *ELT Journal*, 40, 236–42.
- Irujo, S. (1993). Steering clear: Avoidance in the production of idioms. *International Review of Applied Linguistics in Language Teaching*, 31, 205–19.
- Kuiper, K. (1996). *Smooth talkers*. Hillsdale, NJ: Erlbaum.
- Langacker, R. W. (1987). *Foundations of cognitive grammar. Vol. I: Theoretical prerequisites*. Stanford, CA: Stanford University Press.

- Laufer, B. (2000). Avoidance of idioms in a second language: The effect of L1-L2 degree of similarity. *Studia Linguistica*, 54, 186-96.
- Laufer, B., & Eliasson, S. (1993). What causes avoidance in L2 learning: L1-L2 difference, L1-L2 similarity, or L2 complexity? *Studies in Second Language Acquisition*, 15, 35-48.
- McCarthy, M., & Carter, R. (2002). This that and the other: Multi-word clusters in spoken English as visible patterns of interaction. *Teanga (Yearbook of the Irish Association for Applied Linguistics)*, 21, 30-52.
- Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning*, 48(3), 323-63.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York, NY: Newbury House.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford, England: Oxford University Press.
- Nelson, K. (1973). *Structure and strategy in learning to talk: Monographs of the Society for Research in Child Development* (Serial no. 149, nos 1-2).
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-42.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam, Netherlands: John Benjamins.
- Oppenheim, N. (2000). The importance of recurrent sequences for nonnative speaker fluency and cognition. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 220-40). Ann Arbor: University of Michigan Press.
- Peters, A. (1977). Language learning strategies: Does the whole equal the sum of the parts? *Language*, 53(3), 560-73.
- Peters, A. (1983). *The units of language acquisition*. Cambridge, England: Cambridge University Press.
- Plunkett, K. (1993). Lexical segmentation and vocabulary growth in early language acquisition. *Journal of Child Language*, 20, 43-60.
- Schmidt, R. W. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 137-74). Rowley, MA: Newbury House.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge, England: Cambridge University Press.
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 1-22). Amsterdam, Netherlands: John Benjamins.
- Sinclair, J. (2004). *Trust the text: Lexis, corpus, discourse*. London, England: Routledge.
- Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multiword vs. one-word verbs. *International Review of Applied Linguistics*, 45, 119-39.
- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, 64(3), 429-58.
- Stubbs, M. (2002). *Words and phrases: Corpus studies of lexical semantics*. Oxford, England: Blackwell.
- Teliya, V., Bragina, N., Oparina, E., & Sandomirskaya, I. (1998). Phraseology as a language of culture: Its role in the representation of a collective mentality. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 55-75). Oxford, England: Oxford University Press.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Wong Fillmore, L. (1976). *The second time around* (Unpublished doctoral dissertation). Stanford University, CA.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, 21(4), 463-89.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, England: Cambridge University Press.
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language and Communication*, 20, 1-28.
- Yorio, C. A. (1980). Conventionalized language forms and the development of communicative competence. *TESOL Quarterly*, 14(4), 433-42.

### Suggested Readings

- Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21, 101–14.
- Cheng, W., Greaves, C., & Warren, M. (2006). From n-gram to skipgram to concgram. *International Journal of Corpus Linguistics*, 11, 4, 411–33.
- 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.
- Cowie, A. (Ed.). (1998). *Phraseology: Theory, analysis, and applications*. Oxford, England: Oxford University Press.
- Dechert, H. (1983). How a story is done in a second language. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 175–95). London, England: Longman.
- Granger, S., & Meunier, F. (Eds.). (2008). *Phraseology: An interdisciplinary perspective*. Amsterdam, Netherlands: John Benjamins.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London, England: Routledge.
- Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18, 85–102.
- Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford, England: Oxford University Press.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.
- Schmitt, N. (Ed.). (2004). *Formulaic sequences*. Amsterdam, Netherlands: John Benjamins.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.
- 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*. Amsterdam, Netherlands: John Benjamins.
- Wood, D. (2002). Formulaic language in acquisition and production: Implications for teaching. *TESL Canada Journal*, 20(1), 1–15.