Formulaic Language and Collocation

NORBERT SCHMITT, SUHAD SONBUL, LAURA VILKAITĖ-LOZDIENĖ, AND MARIJANA MACIS

The Functions of Formulaic Language

Occupying such an important role in language, it is not surprising that formulaic sequences have a multifaceted functional and pragmatic value. They 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).

Furthermore, the importance of formulaic sequences in reducing cognitive load has long been recognized (e.g., Kuiper, 2004; Wray, 2017). Formulae ensure the physical and social survival of individuals while communicating, help reduce processing effort, and maintain fluency, especially when the speaker feels pressurized. Kuiper (2004), for example, examined naturally occurring speech at auctions and radio sports commentaries and found that, when under time pressure, auctioneers and sports commentators rely heavily on the use of formulaic phrases.

Moreover, formulaic sequences function as "safety zones" (Boers & Lindstromberg, 2009) and are preferred lexical choices by L2 learners who overuse them to mitigate their language errors and communicate their ideas successfully (see section on "Amount of Use"). Similarly, people who suffer from dementia may resort to the use of formulaic sequences to help overcome general processing problems (Wray, 2017).

More recent studies have explored the functions of formulaic language in literature and informal language. Garley, Slade, and Terkourafi (2010) examined the formulaicity in blogs and Old English verse (*Beowulf*) and identified six important functions of formulaic sequences commonly found across both genres: discourse structuring functions (*okay*), filler functions (*if the truth be told*), epithetic functions (*heard under helme*), gnomic functions (*taking a stand*), tonic functions (*spill my guts*), and acronymic functions (*lol*). Likewise, Dickinson (2013) looked at functions of formulaic language in Twitter interactional discourse. He collected data from 50 random public tweets and found that formulaic language seems to have four major functions: situation manipulators (e.g., requests, complaints, and apologies), conveyors of individual identity (*Speak for yourself—I can go entire minutes without checking my mail or Twitter sometimes . . .*), conveyors of group identity (*The* Daily Mail *certainly has an axe to grind about the net . . . maybe are worried about falling newspaper sales?*), and discourse devices and fillers (*I guess, hmm, yeah*).

Based in part on N. Schmitt (2012). Formulaic language and collocation. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. John Wiley & Sons Inc., with permission.

The Encyclopedia of Applied Linguistics. Edited by Carol A. Chapelle.

DOI: 10.1002/9781405198431.wbeal0433.pub2

Formulaic sequences are also typically used for particular purposes in academic discourse. For example, several studies that have analyzed the functions of lexical bundles in academic writing (e.g., Cortes, 2013) concluded that these phrases perform three main functions: stance, discourse organizing, and referential functions, each one containing further subcategories. In the same context, AlHassan and Wood (2015) found that multiword phrases also serve as sentence builders and frames which help L2 learners produce well-structured and more accurate academic written compositions (e.g., *the line graph presents … , In conclusion, it can be seen that*). Yet another function of formulaic sequences is that they can be used to help the speaker be perceived as belonging to a specific community (Wray, 2002).

Overall, it seems that for any conventional activity or function in any type of discourse, there will be associated phrasal vocabulary.

Acquisition of Formulaic Language

The learning of individual words is incremental, and each word has its own particular learning burden (Schmitt, 2010; Nation, 2013). There is no reason to believe that formulaic language is any different in this respect. However, first language (L1) acquirers seem to learn and use formulaic sequences even before they have mastered individual words that make up those sequences (from whole to parts idea, suggested in early work in the field, e.g., Peters, 1983). For example, a child seems to learn the phrase *I wanna cookie* as a holistic phrase with a meaning attached to it and only later to abstract it into *I wanna + noun*. This is the case because recurrent formulaic sequences are frequent in child-directed language (Theakston & Lieven, 2017). For example, Bannard and Matthews (2008) computed the frequencies of 1 to 5-word strings in child-directed speech and found many multiword sequences that were at least as frequent as individual words. So children seem to receive a lot of exposure to formulaic sequences in their input. Children also seem to take and reuse formulaic sequences from the speech they hear. In the same study, Bannard and Matthews (2008) carried out an experiment with two and three-year-old children who were asked to repeat four-word sequences that were either frequent in the speech directed to them (*sit in your chair*) or not (*sit in your truck*). The children were more accurate and faster when repeating familiar phrases, suggesting that phrasal frequency in the input matters for L1 acquisition.

Initially, the use of formulaic sequences in L1 acquisition was seen only as an early and transitional stage. Recently, though, the perspective has changed, mostly because of the technical advances in corpus research making large corpora of child and child-directed language available (Bannard & Lieven, 2012). Now there is no doubt that formulaic sequences exist in child language and in language directed to children, but their theoretical importance depends on the approach taken by the researcher (Theakston, 2016). Following the usage-based approach (e.g., Tomasello, 2009), formulaic language is at the center of L1 acquisition and needs to be incorporated into models of language production and comprehension (Christiansen & Arnon, 2017).

Acquiring a language is essentially acquiring a new skill, so it is driven by repetition, as well as the recency of the exposure, need, and transparency of the mapping between the form and the function (Theakston, 2016). When acquiring a language, infants have to segment the speech flow, attach conceptual meanings to the lexical units, and generalize their usage to some abstract rules. One potential explanation of how children segment units out of the speech stream is statistical learning, that is, tracking distributional information from the input (Jost & Christiansen, 2016). Infants learning their L1s have no way to expect the unit of language to be a word (Jost & Christiansen, 2016). As longer sequences are frequent in their input, applying statistical learning results in extracting lexical units that are longer than individual words (Bannard & Lieven, 2012). We have now some initial understanding

of how formulaic sequences are useful for early language acquisition, but "there is a long way to go before we can hope to fully understand how these mechanisms change over time, how both multiunit strings and grammatical abstractions can co-exist into adult language" (Theakston & Lieven, 2017, p. 600).

There are attempts to answer these questions, though. For example, Wray (2002, 2008) suggests that a "needs-only analysis" is the mechanism of deciding which holistic utterances to segment and which to keep intact. Rather than segmenting every sequence into the grammar system, children will operate with the largest possible unit, and only segment sequences when this 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. Wray (2002) suggests that the relative ratios between the holistically stored units and analytically constructed language may change according to age. From birth to around 20 months, the child will mainly use memorized vocabulary for communication, largely learned through imitation. Later on, the child's grammatical awareness begins, and the proportion of analytic language compared to holistic language increases. At about 18 years, the analytic grammar is fully in place, but formulaic language again becomes more prominent until the balance reaches adult patterns. Children maintain many formulaic sequences into adulthood, even though the components of those sequences are likely to be stored individually as well. This would suggest that dual storage of lexical items is the norm (see also Van Lancker Sidtis, 2012).

One of the reasons why adult speakers continue to rely on formulaic sequences, even when they have analytic capacities to generate language, is the fact that the automatic use of formulaic sequences allows chunking, freeing up memory and processing resources. As already discussed, one of the functions of formulaic sequences seems to be reducing cognitive load. Indeed, formulaic sequences seem to be processed faster than novel phrases by adult L1 speakers and this was shown to be the case for idioms (e.g., Conklin & Schmitt, 2008), collocations (e.g., Vilkaitė, 2016), and longer transparent formulaic sequences (e.g., Arnon & Snider, 2010). However, it has to be noted that this processing advantage can be driven either by the holistic storage of the sequences or by their automatized retrieval (Myles & Cordier, 2017). We do not have clear evidence that would show they are indeed stored holistically rather than just computed faster due to their mutual frequency or predictability (Siyanova-Chanturia & Martinez, 2015). So the idea of holistic storage might be more metaphorical than cognitively real.

Taking into account our knowledge about cognitive processing and the fact that language primarily fulfills a sociointeractional function, Wray (2017) has recently proposed a communicative impact model in order to explain why we use formulaic sequences and what their number in language depends on. She suggests that the use of formulaic language compensates for various disruptions of cognitive processing possibly due to time pressure, various other pressures (such as tiredness or lack of attention), need to communicate with limited resources (as in the case of L2), or language impairment. According to Wray, formulaic language is one of the tools to deal with cognitive pressure in order to maximize fluency and, at the same time, holding the conversational floor, easing the processing for the hearer, and minimizing the chances of misunderstanding or misinterpretation.

So it seems that we start our L1 acquisition from learning formulaic sequences and we keep a vast repertoire of such sequences available to us throughout our lives. The situation of L2 acquisition seems to be rather different, though. While "[l]exical phrases are as basic to SLA as they are to the L1" (Ellis, 2008, p. 97), formulaic language tends to be challenging to L2 learners (Paquot & Granger, 2012).

Why could this be the case? There are a number of potential explanations. Learning an L2 is usually a very different process than learning the L1. Wray (2002) notes that L2 learners

are usually adults, they receive much less exposure to the target language, and the language they are exposed to is not necessarily tailored to their abilities, as is usually the case with infants learning the L1. Also, most L2 learners are already literate in their L1. They have much higher cognitive abilities and tend to approach language learning from a more analytic perspective. As a consequence, while children seem to start learning a language from chunks that they map meanings onto (Tomasello, 2000; Bannard & Lieven, 2012), adults learning an L2 start from the existing knowledge about words as lexical units. Also, infants are learning words and concepts simultaneously while in L2 we already have the concepts available and we only attach form to them (Arnon & Christiansen, 2017).

So it seems that adults and children rely on formulaic sequences to a different extent when learning a language. This idea has been tested by means of computational modeling. For instance, McCauley and Christiansen (2017) used a model called *Chunk-based learner* to explore how children and adults rely on different linguistic sequences. They used corpus data of L1 children and adult speakers as well as L2 adult learners to train their model and then to test how well it can predict the language produced by the different groups. The results showed that, while both L1 and L2 learning involved learning chunks, adult L2 learners learned less useful chunks and relied on them less. It also seems that L1 and L2 learners have different mechanisms to extract chunks from the input (in line with Ellis, Simpson-Vlach, & Maynard, 2008): L1 speakers seem to rely on transitional probabilities between the words but L2 speakers tend to be more influenced by the raw frequency of the chunks. While the most likely explanation for why children are better at learning languages than adults used to be changes in the neuroplasticity and neural commitment of the brain, McCauley and Christiansen (2017) suggest that there is another potential explanation: Children and adults are using building blocks of different granularity when learning a language and they are using them to a different extent.

Looking at the evidence from L1 and L2 acquisition, it can be concluded that, for now, we do not yet have one best model which 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

As demonstrated earlier, formulaic language is very common in language use overall (e.g., Erman & Warren, 2000); native speakers use and produce multiword phrases frequently and effortlessly. However, this does not appear to be the case with non-native speakers. Research suggests that formulaic language seems to be problematic for L2 learners and that its lack/misuse is a major reason why L2 output tends to be judged as unnatural, inappropriate, and non-native. This is evident in both spoken and written L2 output, but most research has focused on the written academic register. We can look at non-native mastery of formulaic sequences along at least three dimensions: amount of use, accuracy/appropriacy of use, and automaticity/fluency of the underlying formulaic language processing. Each dimension will be discussed in turn.

Amount of Use

A considerable amount of research shows that L2 learners' use of formulaic sequences tends to differ from that of native speakers (e.g., Ädel & Erman, 2012; Granger & Bestgen, 2014; Arnon & Christiansen, 2017). A general tendency that emerges from these studies is that non-natives tend to rely on certain favorite formulaic sequences which they know very well

and feel "safe" using in their output (referred to as "safe bets" or "zones of safety") (e.g., Boers & Lindstromberg, 2009). In contrast, they use fewer of other sequences, presumably because they do not know them well and are not as confident in their use.

For example, in one part of their study, Chen and Baker (2010) compared the use of lexical bundles in the writing of Chinese EFL university students and native expert writers. They found that Chinese students tended to favor and overuse certain lexical bundles (e.g., *all over the world*). In contrast, Ädel and Erman (2012) looked at the number of lexical bundles produced in L1 Swedish speakers' written compositions and comparable native-speaker writing. L2 learners underused the target lexical bundles by a large margin. Similar results were reported by Laufer and Waldman (2011), who found a far lower number of verb–noun collocations produced by their Hebrew EFL learners at three proficiency levels (5.9%) compared with their native-speaker counterparts (10%).

As mentioned earlier, spoken L2 production has not been researched to the same extent, but the few existing studies essentially show the same tendency by L2 learners to overuse and underuse formulaic sequences. Erman, Denke, Fant, and Forsberg Lundell (2015) analyzed the output from two spoken tasks (role play and retelling) of advanced Swedish speakers of L2 English, French, and Spanish and compared it to natives of each language, respectively, in the use of various types of formulaic sequences. Results showed that the Swedish speakers were native-like in the amount of use of most target formulaic sequences. However, there was a significant underuse of collocations in the retelling task. Bardovi-Harlig (2008) found similar results regarding underuse of pragmatic formulas (e.g., *You're welcome*).

One of the reasons why non-natives prefer to overuse some formulaic sequences and underuse others might be because L2 speakers seem to acquire and use those sequences that appear frequently and are identified by measures which give a relatively heavy weighting to frequency, such as the *T* score statistic (*good example, hard work*). Conversely, they produce fewer of those collocations that are less frequent, even though these are strongly linked, as identified by the mutual information (MI) statistic (*densely populated, preconceived notions*) (e.g., Ellis et al., 2008; Durrant & Schmitt, 2009). Another reason for this overuse/underuse may be because non-natives tend to focus on single words rather than multiword units, hence these go unnoticed in the input (e.g., Arnon & Christiansen, 2017) and because L2 learners were never explicitly instructed to use these phrases (e.g., Cortes, 2004). A number of studies suggest that direct instruction and enhancement, awareness-raising techniques, have the potential to increase the amount of use of formulaic sequences.

One of Peters and Pauwels's (2015) aims, for example, was to investigate the effect of direct teaching on the use of academic formulaic sequences in writing by Dutch learners of English. The treatment consisted of a range of activities designed to promote noticing, retrieval, and creative use. The findings revealed that students who received direct instruction used on average 10 formulaic sequences in their writing in comparison to 1.5 formulaic sequences on average produced by students who did not receive instruction (see AlHassan & Wood, 2015, for similar findings).

Taken together, the above studies suggest that overuse and underuse of different types of formulaic sequences are characteristic features of L2 learner written and spoken production. Moreover, explicit instruction has the potential to reduce the problem of overuse and underuse to a great extent.

Accuracy/Appropriacy of Use

Just because L2 learners produce formulaic language, it does not necessarily match what natives would produce. Nesselhauf (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 International Corpus of Learner English written by German university students. Almost one quarter of these collocations were judged to be incorrect; and 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). Similarly, Laufer and Waldman (2011) showed that about a third of the collocations in their learner corpus were atypical, with L1 being the source of about half of the errors across all levels. More interestingly, learners at the advanced level produced more atypical collocations than the other two groups due to their false sense of confidence. So, of the few verb–noun collocations that the EFL learners used in this study (see previous section), most were inappropriate even at the advanced level.

Thus it seems that non-native learners make a lot of L1-based errors in their written production. However, language use is not only about what speakers produce but also how well they comprehend texts. Kremmel, Brunfaut, and Alderson (2015, Study 2) showed that appropriate formulaic language knowledge plays a critical role in reading comprehension. After completing a multiple-choice test on 60 formulaic sequences, 15 EFL Austrian learners had their think-aloud protocols recorded as they answered comprehension questions on a reading passage containing these sequences. It was found that, by paying more attention to the formulaic sequences they know, the participants successfully arrived at the correct answer for the comprehension questions.

These results, taken together, seem to suggest that, 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. Moreover, the L1 effect seems to be the overriding main source of errors in the productive use of collocations. On the comprehension side, it seems that reading comprehension is modulated by the accurate recognition of formulaic sequences in context.

Automaticity/Fluency of Use

As stated earlier, formulaic sequences can be overused, underused, and misused by non-native language users, 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? Can they reach a native-like level of fluent/automatic processing? This area has attracted a lot of recent attention, motivated by usage-based approaches and their implications for L2 processing/acquisition (see "Acquisition of Formulaic Sequences" section above).

Wolter and Gyllstad (2013), for example, compared natives to advanced Swedish non-natives in how they process English collocations varying in corpus-based frequency in a timed acceptability judgment task. Their target items included congruent (L1=L2) collocations, incongruent (L1≠L2) collocations, and non-collocate pairs. Both natives and non-natives showed a clear effect of frequency on processing time, with the non-natives exhibiting an additional advantage for congruent collocations over incongruent ones. Sonbul (2015) explored the role that frequency plays in a more natural online reading task both for English natives and advanced non-natives. In addition to gauging eye movements while processing collocations in sentence contexts, the study also looked at offline acceptability judgments (intuitions). Both groups of participants were found to be sensitive to collocational frequency online (initial reading times) and offline, with a clearer effect of frequency for non-natives offline as their proficiency increased (see Siyanova-Chanturia, Conklin, & Van Heuven, 2011, for similar evidence for binominal expressions). Thus non-natives do show automaticity in processing formulaic sequences during comprehension processes.

Does this L2 frequency effect hold for production? This is an important question given the demanding nature of L2 production. To the best of our knowledge, only one very recent study has looked into this matter. Siyanova-Chanturia and Janssen (2018) compared English

natives to advanced non-natives in a timed elicitation task as they articulated frequent binomials and their less frequent reversed forms. Natives' articulation time was found to be modulated by frequency, but this was not the case for the non-natives.

Thus, in line with usage-based models of language development, it seems that L2 language users develop native-like processing of formulaic language incrementally as they receive sufficient L2 input. Moreover, L1 is evidently an influential factor not only in achieving *accuracy* (see previous section), but also in developing *automaticity* even at the advanced level (Wolter & Gyllstad, 2013). Research, however, does not yet provide convincing evidence as to whether or when this language processing advantage transfers to actual language production (Siyanova-Chanturia & Janssen, 2018), and this is reflected in actual language use. While natives tend to resort to formulaic language to get through time-pressurized communicative situations (e.g., Kuiper, 2004), non-natives do not seem to make greater use of formulaic language in such cases, either in speech or writing (e.g., Nesselhauf, 2005; Bardovi-Harlig, 2008). 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 (Erman et al., 2015). However, in terms of writing, neither amount of use nor accuracy of collocation appears to increase with time spent in an English-speaking country (Nesselhauf, 2005). So, even though a year or more spent in an English-speaking country can lead to better *intuitions* of collocation, it seems difficult to extend this into increased *fluency* in language use.

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. Formulaic language is an important element of language overall, perhaps *the* essential element. Research in this area has flourished over the last two decades, with a clear focus on differences between native and non-native processing and acquisition from a usage-based perspective. But there are still many areas in need of further investigation including computational modeling, fluency development, and L1 effect. Thus, phraseology will still be one of the most important areas of inquiry in the applied linguistic field for the foreseeable future.

SEE ALSO: Assessment of Vocabulary; Corpus Linguistics in Language Teaching; Formulaic Sequences; Learner Corpora; Lexical Priming; Teaching Vocabulary

References

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English. *English for Specific Purposes*, *31*(2), 81–92.

AlHassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' writing skills: A quantitative research study. *Journal of English for Academic Purposes*, *17*, 51–62.

Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science*, *9*(3), 621–36.

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

Bannard, C., & Lieven, E. (2012). Formulaic language in L1 acquisition. *Annual Review of Applied Linguistics*, *32*, 3–16.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*(3), 241–8.

Bardovi-Harlig, K. (2008). Recognition and production of formulas in L2 pragmatics. In Z.-H. Han (Ed.), *Understanding second language process* (pp. 205–222). Clevedon, UK: Multilingual Matters.

Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition*. Basingstoke, England: Palgrave Macmillan.

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, *14*(2), 30–49.

Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science*, *9*(3), 542–51.

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.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*(4), 397–423.

Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, *12*(1), 33–43.

Dickinson, P. (2013). "B/w U & me": The functions of formulaic language in interactional discourse on Twitter. *The Linguistics Journal*, *7*(1), 7–38.

Durrant, P., & Schmitt, N. (2009). To what extent do native and nonnative writers make use of collocations? *International Review of Applied Linguistics*, *47*(2), 157–77.

Ellis, N. C. (2008). Sequencing in SLA. *Studies in Second Language Acquisition*, *18*(1), 91–126.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*(3), 375–96.

Erman, B., Denke, A., Fant, L., & Forsberg Lundell, F. (2015). Nativelike expression in the speech of long-residency L2 users: A study of multiword structures in L2 English, French and Spanish. *International Journal of Applied Linguistics*, *25*(2), 160–82.

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

Garley, M., Slade, B., & Terkourafi, M. (2010). A text in speech's clothing: Discovering specific functions of formulaic expressions in Beowulf and blogs. In D. Wood (Ed.), *Perspectives on formulaic language* (pp. 213–33). London, England: Continuum.

Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native speakers: A biagram-based study. *International Review of Applied Linguistics in Language Teaching (IRAL)*, *52*(3), 229–52.

Jost, E., & Christiansen, M. H. (2016). Statistical learning as a domain-general mechanism of entrenchment. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 227–44). Berlin, Germany: De Gruyter.

Kremmel, B., Brunfaut, T., & Alderson, J. C. (2015). Exploring the role of phraseological knowledge in foreign language reading. *Applied Linguistics*, *38*(6), 848–70.

Kuiper, K. (2004). Formulaic performance in conventionalised varieties of speech. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 37–54). Amsterdam, Netherlands: John Benjamins.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2), 647–72.

McCauley, S. M., & Christiansen, M. H. (2017). Computational investigations of multiword chunks in language learning. *Topics in Cognitive Science*, *9*(3), 637–52.

Myles, F., & Cordier, C. (2017). Formulaic sequence (FS) cannot be an umbrella term in SLA: Focusing on psycholinguistic FSs and their identification. *Studies in Second Language Acquisition*, *39*(1), 3–28.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, England: Cambridge University Press.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam, Netherlands: John Benjamins.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–49.

Peters, A. M. (1983). *The units of language acquisition* (Vol. 198). Cambridge, England: Cambridge University Press.

Peters, E., & Pauwels, P. (2015). Learning academic formulaic sequences. *Journal of English for Academic Purposes*, *20*, 28–39.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, England: Palgrave Macmillan.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 1–22). Amsterdam, Netherlands: John Benjamins.

Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 776–84.

Siyanova-Chanturia, A., & Janssen, N. (2018). Production of familiar phrases: Frequency effects in native speakers and second language learners. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 2009–18.

Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, *36*(5), 549–69.

Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition*, *18*(3), 419–37.

Theakston, A. L. (2016). Entrenchment in first language learning. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 315–42). Berlin, Germany: De Gruyter.

Theakston, A. L., & Lieven, E. (2017). Multiunit sequences in first language acquisition. *Topics in Cognitive Science*, *9*(3), 588–603.

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*(1/2), 61–82.

Tomasello, M. (2009). The usage-based theory of language acquisition. In E. L. Bavin (Ed.), *The Cambridge handbook of child language* (pp. 69–87). Cambridge, England: Cambridge University Press.

Van Lancker Sidtis, D. (2012). Two-track mind: Formulaic and novel language support a dual-process model. In M. Faust (Ed.), *The handbook of the neuropsychology of language* (pp. 342–67). Chichester, England: Wiley-Blackwell.

Vilkaitė, L. (2016). Are non-adjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1632–42.

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition*, *35*(3), 451–82.

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

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford, England: Oxford University Press.

Wray, A. (2017). Formulaic sequences as a regulatory mechanism for cognitive perturbations during the achievement of social goals. *Topics in Cognitive Science*, *9*(3), 569–87.

Suggested Readings

Boers, F. (Ed.). (2017). *Multi-word expressions* (Special issue). *Language Teaching Research*, *21*(3).

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, *10*(3), 245–61.

Christiansen, M. H., & Arnon, I. (Eds.). (2017). *More than words: The role of multiword sequences in language learning and use* (Special issue). *Topics in Cognitive Science*, *9*(3).

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

Crossley, S. A., Salsbury, T., & Mcnamara, D. S. (2014). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, *36*(5), 570–90.

Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations: A meta-analysis. *International Journal of Corpus Linguistics*, *19*(4), 443–77.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, *67*(S1), 155–79.

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.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, *33*(3), 299–320.

Polio, C. (Ed.). (2012). *Formulaic language* (Special issue). *Annual Review of Applied Linguistics*, *32*.

Schmitt, N. (Ed.). (2004). *Formulaic sequences*. Amsterdam, Netherlands: John Benjamins.

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (Eds.). (2018). *Understanding formulaic language: A second language acquisition perspective*. London, England: Routledge.

Wood, D. (Ed.). (2010). *Perspectives on formulaic language: Acquisition and communication*. London, England: Bloomsbury.

Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. London, England: Bloomsbury.

Wulff, S., & Titone, D. (Eds.). (2014). *Bridging the methodological divide: Linguistic and psycholinguistic approaches to formulaic language* (Special issue). *The Mental Lexicon*, *9*(3).