Annual Review of Applied Linguistics (2012), **32**, 17–44.

doi: 10.1017/S0267190512000025

Formulaic Language and Second Language Acquisition: Zipf and the Phrasal Teddy Bear

Nick C. Ellis

This article revisits earlier proposals that language learning is, in essence, the learning of formulaic sequences and their interpretations; that this occurs at all levels of granularity from large to small; and that the language system emerges from the statistical abstraction of patterns latent within and across form and function in language usage. It considers recent research in individual differences, the psycholinguistics of language processing, and longitudinal studies of first (L1) and second (L2) language acquisition. The first section reviews studies of individual differences in phonological short-term memory (PSTM) and working memory (WM) and their correlations with vocabulary and grammar acquisition in L2. The second section summarizes evidence that language processing is sensitive to the statistical properties of formulaic language in terms of frequency and transitional probability. The third section examines the definition of formulas and formulaicity using different statistical metrics. The fourth section evaluates longitudinal research in L1 and L2 into the putative developmental sequence commonly proposed in usage-based approaches, from formula to low-scope pattern to creative construction. The final section weighs the implications of the statistical distributions of formulaicity in usage for developmental sequences of language acquisition. Zipf's law and the "phrasal teddy bear" explain the paradox whereby formulas seed language acquisition and yet learner language is formula-light in comparison to native norms.

Fifteen years ago, a review of second language acquisition, "Sequencing in SLA" (Ellis, 1996) contended that language acquisition is essentially sequence learning." One strand of this argument was that language learners with better sequencing ability in phonological short-term memory (PSTM) were more successful in acquiring vocabulary and grammar. The article argued that learners' long-term knowledge of lexical sequences in formulaic phrases serves as the database for the acquisition of language grammar, and it proposed what was called *chunking* as a general process of SLA. This was at odds with earlier views claiming that formulaic language was outside the creative language process (e.g., Krashen & Scarcella, 1978).

A subsequent article (Ellis, 2002a) reviewed evidence that language processing was sensitive to the sequential probabilities of linguistic elements, at all levels from phonemes to phrases, in comprehension as well as in fluency and idiomaticity of speech production. The article claimed that this sensitivity to sequence information in language processing was evidence of learners' implicit knowledge of formulas, and that this knowledge in itself served as the basis for linguistic systematicity and creativity. Therefore, it proposed that a common pattern of developmental sequence in both first (L1) and second (L2) language acquisition was from formulaic phrase, to limited-scope slot-and-frame pattern, to fully productive schematic pattern. At the time, although there were various sources of convergent evidence for this proposal, direct evidence of this putative acquisition sequence in analyses of dense longitudinal learner corpora of language acquisition was scant. Ellis (2003) phrased the argument in terms of linguistic constructions rather than formulas.[1]

None of these ideas were particularly innovative. Psycholinguistic, functional-linguistic, corpus-linguistic, cognitive-linguistic, and usage-based approaches to first language acquisition (L1A) were coming to the fore at that time. Now, a decade or more later, as this *ARAL* volume attests, there has never been more interest in formulaic language in applied linguistics. This review article therefore looks at subsequent research progress on each of these fronts.[2]

The first section reviews studies of individual differences (IDs) in PSTM and L2 vocabulary and grammar acquisition. The second section summarizes evidence that language processing is sensitive to the statistical properties of formulaic language in terms of frequency and transitional probability, and the third section consequently considers different statistical operationalizations of formulas and formulaicity. The fourth section evaluates research into the putative developmental sequence, from formula to low-scope pattern to creative construction, in L1 and L2, and the final section considers the implications of the statistics of formulaicity in usage for developmental sequences of language acquisition. Zipf's (1935) law and the "phrasal teddy bear" explain the paradox whereby formulas seed language acquisition and yet learners typically do not achieve native-like formulaicity.

INDIVIDUAL DIFFERENCES IN PSTM, VOCABULARY, AND GRAMMAR

Ellis (1996, p. 92) reviewed how language learners differ in their ability to repeat sequences in PSTM. The basic idea was that PSTM provides a window on the surface structure of spoken language, one that is limited in capacity and duration and allows the learning and analysis of the sequential forms of language and their relations. IDs in this capacity therefore determine learners' facility to acquire vocabulary and grammatical constructions. Much of the work considered in this review, as in the more major proposal by Baddeley, Gathercole, and Papagno (1998) of PSTM as a language acquisition mechanism, concerned L1 acquisition, although one or two studies (e.g., Ellis & Sinclair, 1996; Service, 1992) involved L2 acquisition. Since then, various recent articles have extended these demonstrations to L2 speech fluency, vocabulary, and grammar:

O'Brien, Segalowitz, Collentine, and Freed (2006) investigated the relationship between PSTM and L2 fluency gains in native English-speaking adults learning

Spanish in two learning contexts. PSTM (operationalized as serial nonword recognition), Spanish oral fluency (temporal/hesitation phenomena), and L2 lexical, narrative, and grammatical abilities (use of free grammatical morphemes and subordinate clauses) from speech samples were assessed at two times, 13 weeks apart, while the students were learning either at their home university or abroad in an immersion context. After the variance attributable to learning context was partialled out, initial PSTM was significantly associated with L2 oral fluency development, explaining 4.5–9.7 percent of unique variance. O'Brien, Segalowitz, Collentine, and Freed (2007) further reported that PSTM significantly contributed to the development of L2 narrative skills for less proficient participants (18 percent of variance explained) and to gains in correct use of function words for more proficient participants (16%). Together, these findings suggest that PSTM plays roles in oral fluency development, in narrative development at earlier stages of L2 learning, and in the acquisition of grammatical competence at later stages.

Kormos and Sáfár (2008) studied the relationship between PSTM and performance in end-of-year reading, writing, listening, speaking, and use of English tests in 121 15- to 16-year-old secondary school students in their first year of intensive language training in a bilingual education program in Hungary. Although no significant relationships involving PSTM were found within the beginning learners in this study, in the preintermediate learners, PSTM was moderately correlated with students' performance in writing and use of English and with the fluency ($r = .46$) and vocabulary ($r = .57$) components of the oral test.

Hummel (2009) explored the relationship among aptitude, PSTM, and L2 proficiency in 77 nonnovice francophone adult ESL (English as a second language) learners at a university. Exploratory factor analysis revealed three main factors corresponding to the variables examined: L2 proficiency, aptitude, and PSTM. Aptitude and PSTM together predicted 29 percent of the variance in L2 proficiency. Alone, PSTM correlated 0.35 with L2 proficiency, 0.36 with L2 vocabulary, and 0.33 with L2 grammar. In this study, PSTM was a stronger predictor of success in the lower than in higher proficiency subgroup.

Wen (2011) investigated the effects of PSTM (nonword repetition) and working memory (WM; speaking span) on eight L2 speech performance measures (speech rate, reformulations, false starts, replacements, global accuracy, length of accurately produced clauses, syntactic complexity, and lexical diversity) in 40 EFL (English as a foreign language) participants doing L2 task-based speech planning and performance. The results suggested that both PSTM and WM were related to syntactic complexity and lexical diversity measures of L2 learners' oral speech. Wen argued that higher PSTM scores were related to greater fluency (speech rate and replacement), global accuracy, lexical density, and syntactic complexity of the L2 learners' oral speech. Higher WM was related to repair fluency and the length of clauses that could be produced accurately.

These findings demonstrate a connection between PSTM and language learning, suggesting a role for PSTM in consolidating, entrenching, and automatizing activation of stable, long-term mental representations of novel phonological material such as individual words, morphemes, and lexical sequences. But is there a direct involvement of PSTM in the acquisition of the grammatical relations between the sequenced items? Or is the relationship between PSTM and grammar abilities entirely mediated by vocabulary knowledge, as some authors (e.g., Andrade & Baddeley, 2011; Service & Kohonen, 1995) have suggested? Three recent studies suggest direct effects of PSTM on grammar acquisition.

Williams and Lovatt (2003) demonstrated that individual differences in PSTM were related to adults' ability to learn determiner–noun agreement rules in semiartificial microlanguages but that there were additional statistically independent effects of knowledge of other gender languages. This suggests that while PSTM might provide working storage of linguistic form that might serve as the evidence for induction of patterns or rules, rule induction itself is modulated by prior knowledge and learned attentional biases (Ellis, 2006b).

French and O'Brien (2008) examined PSTM in L2 grammar learning in a group of native French-speaking children undergoing a five-month intensive English program. PSTM (as referenced by Arabic [ANWR] and English [ENWR] nonword repetition tasks), L2 receptive and productive vocabulary, and L2 grammar (knowledge of morphosyntactic structures) were assessed during the first and last months of the program. Time 1 PSTM predicted L2 grammar scores at Time two (*r* values between .79 and .82). After controlling for initial grammar ability, PSTM significantly predicted grammar development (28 percent of variance explained) in addition to the contribution made by vocabulary knowledge (9.5%). Although PSTM as measured by ENWR increased between Time 1 and Time 2, ANWR did not improve. The findings show that phonological memory plays an important role in L2 grammar development that is unmediated by lexical knowledge. They also provide evidence that language-specific phonological memory improves with language development, as claimed in the PSTM-LTM (long-term memory) bootstrapping model as described in Ellis (1996, 2001).[3]

Martin and Ellis (2012) analyzed PSTM and WM and their relationship with vocabulary and grammar learning in an artificial mini-language based on the characteristics of a number of natural languages. Nonword repetition and nonword recognition were used as measures of PSTM, listening span indexed WM. Participants learned the singular forms of vocabulary in an artificial foreign language before being exposed to plural forms in sentence contexts. They were later tested on their ability to induce the grammatical constructions and to generalize them to novel utterances, the plural system as a whole being quasi-regular. Individual differences in final abilities in vocabulary and grammar correlated between 0.44 and 0.76, depending on the measure. Despite these strong associations, hierarchical regression and structural equation modeling (SEM) demonstrated significant independent effects of PSTM and WM upon L2 vocabulary learning and upon L2 grammar learning. There were substantial independent effects of PSTM ($\beta = 0.41$) and WM ($\beta = 0.42$) upon trained vocabulary, an effect of trained vocabulary upon generalization test grammar ($\beta = 0.33$) allowing mediated indirect effects of PSTM (0.14) and WM (0.14) upon grammar, and direct effects of PSTM ($\beta = 0.25$) and WM ($\beta = 0.30$) upon generalization test grammar. Thus these memory systems are indeed involved in vocabulary learning, but they are also involved in grammar induction from language usage over and above that.

In sum, PSTM affects the efficiency of learning novel word forms and the retention of sequences of forms, with evidence that the latter contributes to grammatical development through processes of analysis. What is the nature of these processes? They must involve some form of distributional learning and analogy: "Usage-based theories hold that the acquisition of language is exemplar based. It is the piecemeal learning of many thousands of constructions and the frequency-biased abstraction of regularities within them" (Ellis, 2002a, p. 143).

A recent study by Misyak and Christiansen (2012) highlights the relationships between statistical learning ability, verbal working memory (vWM), and language comprehension using an individual-differences framework. Participants were administered separate statistical learning tasks involving adjacent and nonadjacent dependencies, along with language comprehension tasks and a battery of other measures assessing vWM, STM (digit span), vocabulary, reading experience, cognitive motivation, and fluid intelligence. The comprehension tasks entailed the tracking of adjacent and/or nonadjacent natural language dependencies. Strong interrelationships were found among statistical learning, vWM, STM, and comprehension. The specific pattern of intercorrelations between statistical learning and vWM/STM indicated that adjacent statistical learning is relatively strongly associated with both vWM and STM performance ($r = .46$ and .40, respectively), whereas nonadjacent statistical learning was more associated with vWM than with STM ($r = .53$ and .13, respectively). Thus, the kind of learning and memory skills involved in vWM tasks may be more closely related to the learning of nonadjacencies than adjacencies, whereas STM may be more closely associated with mechanisms subserving the learning of adjacent dependencies. Martin and Ellis (2012) also found that the contributions of PSTM and WM to grammar learning were separable.

This recent research supports that PSTM is involved in vocabulary and grammar acquisition, that the short-term representation of lexical items and lexical sequences promotes their long-term consolidation, and that this evidence allows statistical learning of grammatical patterns and dependencies. See Williams (2011) for a more thorough review of PSTM and WM and language learning, Acheson and MacDonald (2009) for review of verbal working memory and language production, and Rebuschat and Williams (forthcoming) for an edited collection of current work on statistical learning of language.

Priorities for the future include scrutiny of interrelations between PSTM and WM, implicit and explicit learning, and statistical learning of adjacent and discontinuous dependencies as individual differences as they affect variously the acquisition of lexis, formulaic phrases, and creative grammatical constructions. These processes and involvements might differ dependent upon stage of learner proficiency and language typology, especially in terms of reliance upon word order as a grammatical cue.

SENSITIVITY TO FORMULAIC SEQUENCES IN COMPREHENSION AND PRODUCTION PROCESSING

There has been considerable recent activity investigating learners' sensitivity to formulaic sequences, both in comprehension and production, and in fluency therein. Phonetic processing and lexical perception are affected by formulaic knowledge. Hilpert (2008) demonstrated that syntactic context in the form of constructions and collocations affects both phonemic categorization and low-level phonetic processing in native speakers. One experiment used the English *make*-causative construction which has a strong bias toward verbs of emotion and psycho-physiological reaction: the verb *cry* occurs 73 times in the *make*-causative construction, the verb *try* just 11; in discourse as a whole, *try* is 10 times more frequent than *cry*; thus *make me cry* is more formulaic than *make me try*. The carrier phrase was *They made me*, followed by a signal that ranged on an eight-step continuum from /trai/ to /krai/. Over many trials, participants had to say whether they heard /t/ or /k/. The resulting categorization curve was half a step toward the right side of the continuum, that is, more instances of ambiguous sounds were identified as *cry*, when they were presented in the *make*-causative constructional carrier phrase rather than alone.

Kapatsinski and Radicke (2008) provided data that similarly point to a competition between larger units and their parts when the whole-form is of sufficient frequency. Participants had to respond whenever they detected the particle *up* in a verb–particle combination (e.g., *give up*). Reaction times were faster the more frequent the collocation up to a point; but for collocations in the highest frequency bin, there was a slowdown in reaction times.

Reading time is affected by formulaic knowledge. 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* by native speakers. Ellis, Frey, and Jalkanen (2008) used lexical decision to demonstrate that native speakers preferentially process frequent verb-argument and booster/maximizer-adjective two-word collocations. Durrant and Doherty (2010) used lexical decision to assess the degree to which the first word of low- (e.g., *famous saying*), middle- (*recent figures*), high-frequency (*foreign debt*), and high-frequency and psychologically associated (*estate agent*) collocations primed the processing of the second word in native speakers. The highly frequent and high-frequency associated collocations evidenced significant priming. Arnon and Snider (2010) used a phrasal decision task, in which participants had to decide whether a phrase was possible in English or not, to show that comprehenders were also sensitive to the frequencies of compositional four-word phrases: More frequent phrases (e.g., *don't have to worry*) were processed faster than less frequent phrases (*don't have to wait*) even though these were matched for the frequency of the individual words or substrings. Tremblay, Derwing, Libben, and Westbury (2011) examined the extent to which lexical bundles (LBs, defined as frequently recurring strings of words that often span traditional syntactic boundaries) are stored and processed holistically. Three self-paced reading experiments compared sentences containing LBs (e.g., *in the middle of the*) and matched control sentence fragments (*in the front of the*) such as *I sat in the middle/front of the bullet train*. LBs and sentences containing LBs were read faster than the control sentence fragments in all three experiments.

There is a substantial literature demonstrating sensitivity to such sequential information in sentence processing (for a review, see MacDonald & Seidenberg, 2006). Eye-movement research shows that the fixation time on each word in reading is a function of the frequency of that word (frequent words have shorter fixations) and of the forward transitional probability (the conditional probability of a word given the previous word $P(w_k|w_{k-1})$: for example, the probability of the word *in* given that the previous word was *interested* is higher than the probability of *in* if the last word was *dog*) (McDonald & Shillcock, 2003, 2004). Parsing time reflects the more frequent uses of a word (e.g., the garden-path effect caused by *The old man the bridge*, in which *man* is used as a verb). Phrase frequency affects parsing in a similar way. For example, ambiguity resolution is driven not only by how often a verb appears as a past participle and how likely a noun is to be an agent, but also by the exact frequencies of the noun–verb combination. Reali and Christiansen (2009) demonstrate such effects of chunk frequency in the processing of object relative clauses. Sentences such as *The person who I met distrusted the lawyer*, are easier to process when the embedded clause is formed by frequent pronoun–verb combinations (*I liked* or *I met*) than when it is formed by less frequent combinations (*I distrusted* or *I phoned*).

Generally, analyses of large corpora of eye movements recorded when people read text demonstrate that measures of *surprisal* account for the costs in reading time that result when the current word is not predicted by the preceding context. Measuring surprisal requires a probabilistic notion of linguistic structure (utilizing transitional probabilities or probabilistic grammars). The surprisal of a word in a sentential context corresponds to the probability mass of the analyses that are not consistent with the new word (Demberg & Keller, 2008).

Maintenance of material in short-term memory and accurate subsequent production is also affected by knowledge of formulaic sequences. Bannard and Matthews (2008) identified frequently occurring chunks in child-directed speech (e.g., *sit in your chair*) and matched them to infrequent sequences (e.g., *sit in your truck*). They tested young children's ability to produce these sequences in a sentence-repetition test and found that two- and three-year-olds were significantly more likely to repeat frequent sequences correctly than to repeat infrequent sequences correctly. Moreover, the three-year-olds were significantly faster to repeat the first three words of an item if they formed part of a chunk (e.g., they were quicker to say *sit in your* when the following word was *chair* than when it was *truck*). Tremblay et al. (2011) similarly used word and sentence recall experiments to demonstrate that more sentences containing LBs (the same ones as in their earlier mentioned comprehension experiments) were correctly remembered by adults in STM experiments.

Language processing generally shows recency effects where a construction recently experienced in discourse is picked up and reused productively in the "dance of dialogue" (Pickering, 2006). This phenomenon, known as priming, is observed in phonology, conceptual representations, lexical choice, and syntax. Syntactic priming refers to the phenomenon of using a particular syntactic structure given prior exposure to the same structure. This behavior has been observed when speakers hear, speak, read, or write sentences (Bock, 1986; McDonough & Trofimovich, 2008; Pickering & Ferreira, 2008; Pickering & Garrod, 2006).

People have LTM as well for the particular wording used to express something (as any parent who misreads a favorite bedtime story can readily attest). Some learning takes place after just one incidental exposure. Gurevich, Johnson, and

Goldberg (2010) showed that adult native speakers recognize at above-chance rates full sentences that they have been exposed to only once in texts of 300 words that were presented noninteractively with no advanced warning of a memory test. Verbatim memory occurred even when lexical content and memory for gist was controlled for. Even after a six-day delay, participants reliably reproduced sentences they had heard before when asked to describe scenes, even though they were not asked to recall what they had heard.

These effects cumulate: "All lexical items are primed for grammatical and collocational use, that is, every time we encounter a lexical item it becomes loaded with the cumulative effects of these encounters, such that it is part of our knowledge of the word that it regularly co-occurs with particular other words or with particular grammatical functions" (Hoey, 2004, p. 21; 2005). These experiments concern native speakers. What about L2 learners?

Jiang and Nekrasova (2007) examined the representation and processing of formulaic sequences using online grammaticality judgment tasks. English as a second language speakers and native English speakers were tested with formulaic and nonformulaic phrases matched for word length and frequency (e.g., *to tell the truth* vs. *to tell the price*). Both native and nonnative speakers responded to the formulaic sequences significantly faster and with fewer errors than they did to nonformulaic sequences.

Conklin and Schmitt (2007) measured reading times for formulaic sequences versus matched nonformulaic phrases in native and nonnative speakers. The formulaic sequences were read more quickly than the nonformulaic phrases by both groups of participants.

Ellis and Simpson-Vlach (2009) and Ellis, Simpson-Vlach, and Maynard (2008) used four experimental procedures to determine how the corpus-linguistic metrics of frequency and mutual information (MI, a statistical measure of the coherence of strings) are represented implicitly in native and nonnative speakers, thus to affect their accuracy and fluency of processing of the formulas of the Academic Formulas List (AFL, Simpson-Vlach & Ellis, 2010). The language processing tasks in these experiments were selected to sample an ecologically valid range of language processing skills: spoken and written, production and comprehension, form-focused and meaning-focused. They were (a) speed of reading and acceptance in a grammaticality judgment task where half of the items were real phrases in English and half were not, (b) rate of reading and rate of spoken articulation, (c) binding and primed pronunciation—the degree to which reading the beginning of the formula primed recognition of its final word, and (d) speed of comprehension and acceptance of the formula as being appropriate in a meaningful context. Processing in all experiments was affected by various corpus-derived metrics: length, frequency, and MI. Frequency was the major determinant for nonnative speakers, but for native speakers it was predominantly the MI of the formula which determined processability. For review of L2 processing of formulaic language, see also Conklin and Schmitt (this volume).

Priming, the phenomenon in which prior exposure to specific language forms or meanings influences a speaker's subsequent language comprehension or production, has now been extensively researched and observed in L2A (Gries &

Wulff, 2005, 2009; McDonough & De Vleeschauwer, in press; McDonough & Kim, 2009; McDonough & Mackey, 2008; McDonough & Trofimovich, 2008).

Finally, extensive exposure to formulaic sequences increases fluency of speech production. Taguchi (2007) examined the development of spoken discourse among 22 Anglophone college students enrolled in an elementary Japanese course who received extensive practice on grammatical chunks through communicative drills and the memorization of dialogues that contained the target chunks. The development of the students' spoken discourse was examined in spontaneous conversations and narrative tasks administered twice during the semester at five-week intervals. Results showed that the students produced twice as many grammatical chunks in the second data-collection session, with a wider range of chunk types. The participants also showed increasing sensitivity to discourse features over time, suggesting that memorized chunks served as a basis for the creative construction of discourse. For L2 instruction of formulaic language, see also Boers and Lindstromberg (this volume).

Broadly, these findings argue against a clear distinction between linguistic forms that are stored as formulas and ones that are computed or openly constructed. Grammatical and lexical knowledge are not stored or processed in different mental modules, but rather form a continuum from heavily entrenched and conventionalized formulaic units (unique patterns of high token frequency) to loosely connected but collaborative elements (patterns of high type frequency) (Bybee, 2010; Ellis, 2008b, in press; Ellis & Larsen-Freeman, 2009; Robinson & Ellis, 2008). "The linguist's task is in fact to study the whole range of repetition in discourse, and in doing so to seek out those regularities which promise interest as incipient sub-systems. Structure, then, in this view is not an overarching set of abstract principles, but more a question of a spreading of systematicity from individual words, phrases, and small sets" (Hopper, 1987, p. 143).

Language users (both L1 and L2) are sensitive to the sequential statistics of these dependencies, large and small. "Words used together fuse together" (Bybee, 2005, p. 112) (after Hebb's [1949] research, often summarized as showing that "cells that fire together, wire together").[4] These collaborations, conspiracies, and competitions occur at all levels of granularity and points in a sequence—remember, for example, the demonstration of Bannard and Matthews (2008) that the three words beginning a chunk of a four-word sequence is said more quickly when it precedes a more highly related collocation. The phenomenon is entirely graded. If this is correct, we should be operationally defining formulaic language in statistical terms of frequencies and coherence. We will return to this in the next section.

The results encourage an emergentist view whereby all linguistic material is represented and processed in a similar fashion, where learners are sensitive to the frequencies of occurrence of constructions and their transitional probabilities, and hence where they have learned these statistics from usage, tallying them implicitly during each processing episode. What I have given here is but a small sample of a very substantial amount of research in different disciplines which point to these conclusions. The interested reader should see Bybee and Hopper (2001); Bod, Hay, and Jannedy (2003); and Diessel (2007) for reviews of frequency effects in language processing; Corrigan, Moravcsik, Ouali, and Wheatley (2009) and Conklin and Schmitt (this volume) for reviews of processing formulaic language; Trousdale and Hoffman (2012) for reviews of construction grammar; and Robinson and Ellis (2008) for reviews of usage-based theories of SLA.

Having said that, the same caveat must be stated clearly: "To the extent that language processing is based on frequency and probabilistic knowledge, language learning is implicit learning. This does NOT deny the importance of noticing (Schmidt, 1993) in the initial registration of a pattern recognition unit. NOR does it deny a role for explicit instruction" (Ellis, 2002a, p. 145). Frequency is important in tuning the system, but it is by no means the only factor that counts in acquisition (Ellis, in press). The L2 learning literature, rife with demonstrations of how years of input can fail to become intake, shows that implicit tallying does not take place for low salient cues for which pattern recognition units have never been consolidated (Ellis, 2002b, 2005, 2006b).

DEFINING, OPERATIONALIZING, AND IDENTIFYING FORMULAIC LANGUAGE

Earlier in this article I argued against a firm distinction between linguistic forms that are stored as formulas and ones that are openly constructed. Instead it proposed that formulaicity is a dimension to be defined in terms of strength of serial dependencies occurring at all levels of granularity and at each transition in a string of forms. At one extreme are formulaic units that are heavily entrenched (unique patterns of high token frequency), and at the other are creative constructions consisting of strings of slots each potentially filled by many types. Broadly, the more frequent and the more coherent a string, the faster it is processed.

It follows that formulas need to be operationalized in statistical terms that measure frequency and coherence. An admirably open and widely cited definition of a formulaic sequence is that of Wray (2000, p. 465): "A sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is stored and retrieved whole from the memory at the time of use, rather than being subject to generation or analysis by the language grammar." Let us try to apply this to the sentences listed in Figure 1—which of them are formulaic?

You have good reasons, I'm sure, for liking some of these more than others. But should you be allowed a *yes* or *no* response, or a more-or-less scale? You might argue that they all fit the definition. You might ask about the context and how often each has been heard in a particular household. You might ask for a recording so that you can measure fluency of production, hesitations, and runs, and analyze the prosody. You might identify multiple formulaic sequences within some of these. You might try a corpus search. All would be sensible. For purposes of replicable research, however, we need some objective metrics to assess degree of formulaicity. As you were pondering, you probably considered such factors as the frequency of the whole n-gram, the length of utterance, the

i. Put it in.

ii. Put it in the fridge.

iii. Polly put the kettle on.

iv. Put the butter on the table.

v. Put that in your pipe and smoke it.

vi. Put another nickel in the Nickelodeon.

vii. Gabe cleared the music stands from the stage.

viii. Why don't you kids ever clear the dishes from the table?

ix. Boy, you gonna carry that weight, Carry that weight a long time.

x. Dad's spilled Digestive crumbs all over the kitchen floor again, typical!

Fig. 1. Which of these sentences are formulaic, and why?
While you're at it, in the Roman numbering system, is *iv* a better formula than *ix,* or *i* than *ii,* or *iii* than *ii,* or what?

coherence (or what might be called the *gluedness*) of the string, the frequency of each of the words involved, alone and in combination, whether a tune goes with it, and more.

Given psycholinguistic research such as that already summarized, statistical operationalizations can clearly contribute, and they provide the added bonus of allowing triangulation with corpora samples of the usage that serves as the source of our knowledge of formulaicity. Corpus-linguistic techniques can provide a range of methods for the quantification of recurring sequences (as clusters, n-grams, collocations, phrase-frames, etc.) and for gauging the strength of association between the component words. Three broad options for the basis of determination of formulaic sequences come immediately to mind: frequency, association, and native norms. In the next sections, we consider each in turn.

Frequency

Formulas are recurrent sequences. One definition, then, is that we should identify strings that recur often. This is the lexical bundle approach of Biber and colleagues (Biber, Conrad, & Cortes, 2004; Biber, Conrad, & Reppen, 1998) and based solely on frequency. It has the great advantages of being methodologically straightforward and having face validity—we all agree that high-frequency strings like *How are you*, *Nice day today*, and *Good to see you* are formulaic sequences. There is psycholinguistic support for such measures too in frequency effects in language processing and cognition (Bybee, 2010; Ellis, 2002a). But we also know some formulas that are not of particularly high frequency, like *blue moon*, *latitude and longitude*, and *raining cats and dogs*. And other high-frequency strings, like *and of the*, or *but it is*, don't seem very formulaic. Definitions in terms of frequency alone result in long lists of recurrent word sequences that collapse distinctions that intuition would deem relevant. High-frequency n-grams occur often. But this does not imply that they have clearly identifiable or distinctive functions or meanings; many of them occur simply by dint of the high frequency of their component words, often grammatical functors. The fact that a formula is above a certain frequency threshold does not necessarily imply either psycholinguistic salience or coherence.

ASSOCIATION

Psycholinguistically salient sequences, on the other hand, like *once in a blue moon*, *on the other hand*, or *put it on the table* cohere much more than would be expected by chance. They appear to be somehow glued together and thus measures of association, rather than raw frequency, are more relevant. There are numerous statistical measures of association available, each with their own advantages and disadvantages (Evert, 2005; Gries, 2008, 2009, 2012, in press). For example, MI is a statistical measure commonly used in information science to assess the degree to which the words in a phrase occur together more frequently than would be expected by chance (Manning & Schütze, 1999; Oakes, 1998). A higher MI score means a stronger association between the words, while a lower score indicates that their co-occurrence is more likely due to chance. MI is a scale, not a test of significance, so there is no minimum threshold value; the value of MI scores lies in the comparative information they provide. MI privileges coherent strings that are constituted by low-frequency items, like *longitude and latitude*.

In corpus linguistics, Gries and colleagues are making important progress developing measures for collostructional analysis of the degree of association between a word and a construction based on the Fisher-Yates exact probability test (Gries, 2008, 2009, 2012). In psychology, there is huge interest in determining the best measure of association of cue and outcome to predict learning—this field of work did not by chance become known as contingency learning (Ellis, 2006a; Hattori & Oaksford, 2007; Shanks, 1995). In cognitive science, approaches to rational cognition based upon Bayesian probabilistic reasoning (e.g., Oaksford & Chater, 1999) is arguably the fastest growing area. There are important developments triangulating measures of association based upon corpora of usage to psycholinguistic processing data such as eye movements in reading (Demberg & Keller, 2008; Wiechmann, 2008).

Many of these statistics focus on two × two contingency tables measuring the association between two things (such as two words in a collocation, or a word and a construction) and we also need measures for computing sequential dependencies over multiple units (Jurafsky & Martin, 2009) as well as computational models, often based upon serial recurrent networks, for analyzing structure in time (Ellis & Larsen-Freeman, 2009; Elman, 1990; Elman et al., 1996). As you will have gathered, this is an area where much is happening.

Native Norms

Definitions purely in terms of frequency or association might well reflect that language production makes use of sequences that are ready-made by the speaker or writer, but these need not necessarily be native-like. Nonnative academic writing can often be identified by the high-frequency of use of phrases that come from strategies of translation (like *make my homework*, or *make a diet*) or formulas that occur frequently in spoken language but which are frowned upon as informal in academic writing (like *have a nice day!*, or *it is stupid to*. . .). An additional, divergent, criterion for formulaicity is that it reflects native-like selection and native-like fluency (Pawley & Syder, 1983). Thus we can also operationalize the formulaicity of L2 language by how well it uses the formulaic sequences and grammatico-lexical techniques of the norms of its reference genre. For example, O'Donnell, Römer, and Ellis (2012) search for instances of formulaic academic patterns[5] of the AFL (Simpson-Vlach & Ellis, 2010) in learner corpora of native and nonnative English academic writing at different levels of proficiency. They show that these three criteria of frequency, association, and native norm are divergent and complementary.

We are only beginning to explore how these different statistical and corpus-based operationalizations affect acquisition and processing, and this is a research area where much remains to be done. There is strong consensus that research on formulaic language, phraseology, and constructions is in dire need of triangulation across research in L1 acquisition, L2 acquisition, corpus linguistics, usage-based linguistics, and psycholinguistics (Divjak & Gries, in press; Ellis, 2008b; Gries, 2008, 2009).

THE PUTATIVE DEVELOPMENTAL SEQUENCE FROM FORMULAIC PHRASE, TO LIMITED-SCOPE SLOT-AND-FRAME PATTERN, TO FULLY PRODUCTIVE SCHEMATIC PATTERN

Assessing the degree to which formulas feed into the acquisition process entails search for developmental sequences for particular constructions which are seeded by particular memorized formulaic phrases. Note that this developmental sequence does not imply that all apparently formulaic strings are first acquired as wholes. Far from it. Some formulaic sequences are readily learnable by dint of being highly frequent and prototypical in their functionality (e.g., *put it on the table*). Hasselgren (1994) described how in an L2 we "regularly clutch for the words we feel safe with: our 'lexical teddy bears' " (p. 237) and shows how even advanced L2 learners often overuse high-frequency basic words like *very* rather than risking making a word selection error going for a less frequent but more appropriate booster collocate such as *exceedingly* or *gravely*. Highly frequent and prototypically functional phrases like *put it on the table*, *how are you?*, *it's lunch time*, and the like are similarly broad ranging and safe—let us call them *phrasal teddy bears*. These are the likely candidates as construction seeds. Many other formulaic sequences are not readily learnable—they are less frequent, often indeed rare, and many are nontransparent and idiomatic in their interpretation (e.g., *once in a blue moon*). Learners require considerable language experience before they encounter these once, never mind sufficient times to commit them to memory (Ellis, 2008a; Ellis, Simpson-Vlach et al., 2008). This is why learners typically do not achieve native-like idiomaticity (Granger, 2001; Pawley & Syder, 1983). These low-frequency, low-transparency formulas are targets for learning, rather than seeds of learning. Thus the observations that learner language is often light in frequency of formulaic language compared to native norms (see Paquot & Granger, this volume) and acquisition of native-like targets can be challenging (see Bardovi-Harlig, this volume).

Can We See Formulaic Seeds Longitudinally in L1 Acquisition?

Perhaps the most hotly contended question in child language research is whether children's early language (a) makes use of abstract categories and principles for composing sentences by combining those categories in ordered sequences, or (b) consists of a repertoire of more concrete constructions or formulas, many based on particular lexical items (e.g., *jump*, *put*, and *give*) rather than abstract syntactic categories like *verb*. The corresponding theoretical positions are (a) children don't need to learn grammar because the principles and categories of grammar are innate, requiring only minimal exposure to the language to be triggered, or (b) the process of syntactic development consists of acquiring a large repertoire of constructions and formulas and statistically inducing increasingly abstract categories on the basis of experience hearing the types of items that fill the sequential slots of language usage. The last 20 years has seen considerable research that points to the second alternative, or so I see it at least. I have neither space nor inclination here to dispute the case and gladly defer to Bannard and Lieven (this volume) as well as to recent reviews (Ambridge & Lieven, 2011; Behrens, 2009; Dąbrowska, 2004; Diessel, 2012; Lieven, Behrens, Speares, & Tomasello, 2003; Tomasello, 1992, 2003).

One important evidential source has been dense longitudinal corpora of naturalistic language development that capture perhaps 10 percent of the child's speech and the input they are exposed to, collected from two–four years of age when the child is undergoing maximal language development (Behrens, 2008; Maslen, Theakston, Lieven, & Tomasello, 2004). Without such dense sampling, it is difficult if not impossible to clearly identify sequences of development of linguistic items of relatively low frequency as they unfold over time (Tomasello & Stahl, 2004).

Using dense corpora, Lieven and colleagues have used the traceback method (Dabrowska & Lieven, 2005) of analyzing adult–child conversation to show that very often when a child produces what seems to be a novel utterance, the ingredients for that utterance are to be found earlier in the transcript. That is, the novel utterance has not been generated from scratch but rather a previous sentence has been manipulated, replacing one content word. Even when children are more productive than that, the data-dependent nature of children's underlying knowledge is evidenced in the relations between the frequency of structures in input and the frequency of children's production of those structures. Children are initially conservative in their language in that their production is more formulaic than openly combinatorial. These are the essential observations for the developmental sequence from formula to limited-scope pattern to creative construction in L1A (Lieven et al., 2003; Tomasello, 2000, 2003).

There are observations to the contrary, for example, Naigles, Hoff, and Vear (2009) reported findings from a verb diary study in which they had eight mothers record their child's very first 10 uses of 34 different verbs and found more variety in the ways in which children use verbs from their very first uses than one would expect if first use is entirely supported by formulas based on memorized strings of input. On average, the children used 65 percent of their verbs in multiple syntactic configurations before the 10th instance. The first change in syntactic use (e.g., adding a subject or object to a verb first used in its bare form) occurred, on average, 15 days after the first use. Flexibility was also seen at the individual level: One of the faster learners produced *Drop, I drop, I drop something,* and *I dropped it* all within 30 days. Even one of the slower learners produced *I don't like* and *I like cheese* within a single day. What we don't know, of course, from such observations, is how the rest of the discourse unfolded that the day, that is, how the co-adaptive input might have supported these developments, nor indeed in the days before—the two-year-old children who show flexible verb use have been listening to and processing speech for some time (less than two years, nine months) before they produce their first and flexibly used verbs. Thus, it is perhaps not surprising that they show flexibility of verb use and other signs of abstract representations at an early age. It might be that, in effort after objectivity, a goal admirably achieved in the dense-corpora-plus-trackback studies, the empirical analysis overly privileges focus upon learner productions rather than their comprehension of patterns that are latent in the input. A large body of evidence now reveals that infants can acquire considerable knowledge about the patterns of language, including abstractions from the input (for reviews, see Gerken, 2007; Saffran & Thiessen, 2007). These abstract categories and formulaic patterns interact in the comprehension and production of language—this, surely, is what constructing a language is about. We will open this issue up again in the last section.

Can We See Formulaic Seeds Longitudinally in L2 Acquisition?

What about when learners *re*construct an L2? There are longitudinal studies in support of this sequence there too, though the available corpora are far from dense.[6] In an extensive study of secondary school pupils learning French as a foreign language in England, Myles (2004; Myles, Mitchell, & Hooper, 1999) analyzed longitudinal corpora of oral language using 13 tasks in 16 Beginner (years 7, 8, and 9) and 60 Intermediate learners (years 9, 10, and 11) studied cross-sectionally. These data showed that multimorphemic sequences that go well beyond learners' grammatical competence are very common in early L2 production. Notwithstanding that these sequences contain such forms as finite verbs, *wh*-questions and clitics, Myles denied this as evidence for functional projections from the start of L2 acquisition because these properties are not present outside chunks initially. Analyses of inflected verb forms suggested that early productions containing them were formulaic chunks. These structures, sometimes highly complex syntactically (e.g., in the case of interrogatives), cohabited for extended periods of time with very simple sentences, usually verbless, or when a verb was present, normally untensed. Likewise, clitics first appeared in chunks containing tensed verbs, suggesting that it is through these chunks that learners acquire them. Myles characterizes these early grammars as consisting of lexical projections and formulaic sequences, showing no evidence of functional categories. "Chunks do not become discarded; they remain grammatically advanced until the grammar catches up, and it is this process of resolving the tension between these grammatically advanced chunks and the current grammar which drives the learning process forward" (p. 152). The study also investigated the development of chunks within individual learners over time, showing a clear correlation between chunk use and linguistic development:

> In the beginners' corpus, at one extreme, we had learners who failed to memorize chunks after the first round of elicitation; these were also the learners whose interlanguage remained primarily verbless, and who needed extensive help in carrying out the tasks. At the other extreme, we had learners whose linguistic development was most advanced by the end of the study. These were also the learners who, far from discarding chunks, were seen to be actively working on them throughout the data-collection period. These chunks seem to provide these learners with a databank of complex structures beyond their current grammar, which they keep working on until they can make their current generative grammar compatible with them. (Myles, 2004, p. 153)

Eskildsen and Cadierno (2007) investigated the development of *do*-negation by a Mexican learner of English. *Do*-negation learning was found to be initially reliant on one specific instantiation of the pattern, *I don't know*, which thereafter gradually expanded to be used with other verbs and pronouns as the underlying knowledge seemed to become increasingly abstract, as reflected in token and type frequencies. The emerging system was initially exemplar based, and development was based on the gradual abstraction of regularities that link expressions as constructions (see also Eskildsen, in press).

Mellow (2008) described a longitudinal case study of a 12-year-old Spanish learner of English, Ana, who wrote stories describing 15 different wordless picture books during a 201-day period. The findings indicate that Ana began by producing only a few types of complex constructions that were lexically selected by a small set of verbs which gradually then seeded an increasingly large range of constructions.

Sugaya and Shirai (2009) described acquisition of Japanese tense-aspect morphology in L1 Russian learner Alla. In her 10-month longitudinal data, some verbs (e.g., *siru* "come to know," *tuku* "be attached") were produced exclusively with imperfective aspect marker *-te i-(ru)*, while other verbs (e.g., *iku* "go," *tigau* "differ") were rarely used with *-te i-(ru)*. Even though these verbs can be used in any of the four basic forms, Alla demonstrated a very strong verb-specific preference. Sugaya and Shirai followed this up with a larger cross-sectional study of 61 intermediate and advanced learners (based on the ACTFL scale), who were divided into 34 lower and 27 higher proficiency groups using grammaticality judgment tasks. The lower proficiency learners used the individual verbs in verb-specific ways and this tendency was stronger for the verbs denoting resultative state meaning with *-te i-(ru)* (e.g., achievement verbs) than the verbs denoting progressive meaning with *-te i-(ru)* (e.g., activity, accomplishment, and semelfactive verbs). Sugaya and Shirai concluded that the intermediate learners begin with item-based learning and low-scope patterns and that these formulas allow them to gradually gain control over tense-aspect. Nevertheless, they also consider how memory-based and rule-based processes might co-exist for particular linguistic forms, and that linguistic knowledge should be considered a formulaic-creative continuum. We return to this later in this article.

Having said that, there are studies of L2 that have set out to look for this sequence and found less compelling evidence. Bardovi-Harlig (2002) studied the emergence of future expression involving *will* and *going to* in a longitudinal study of 16 adult learners of ESL (mean length of observation 11.5 months; 1,576 written texts, mainly journal entries, and 175 oral texts, either guided conversational interviews or elicited narratives based on silent films). The data showed that future *will* emerges first and greatly outnumbers the use of tokens of *going to*. Bardovi-Harlig described how the rapid spread of *will* to a variety of verbs suggests that, "for most learners, there is either little initial formulaic use of *will* or that it is so brief that it cannot be detected in this corpus" (p. 192). There was some evidence of formulaicity in early use of *going to*:

> For five of the 16 learners, the use of *I am going to write* stands out. Their production over the months of observation show that the formula breaks down into smaller parts, from the full *I am going to write about* to the core *going to* where not only the verb but also person and number vary. This seems to be an example of learner production moving along the formulaic-creative continuum. (Bardovi-Harlig, 2002, p. 197)

But other learners showed greater variety of use of *going to*, with different verbs and different person-number forms, from its earliest appearance in the diary. Bardovi-Harlig concludes that "although the use of formulaic language seems to play a limited role in the expression of future, its influence is noteworthy" (p. 198).

Eskildsen (2009) analyzed longitudinal oral L2 classroom interaction for the use of *can* by one student, Carlo. *Can* first appeared in the data in the formula, *I can write*. But Eskildsen noted how formulas are interactionally and locally contextualized, which means that they may possibly be transitory in nature, their deployment over time being occasioned by specific recurring usage events.

Hall (2010) reported a small-scale study of the oral production of three adult beginner learners of ESL over a nine-week period in a community language program meeting three days per week for two hours each day. A wide variety of tasks was used to elicit the data included picture description and semistructured interviews. Hall reported that formulas were minimally present in the learner output and that constructions and formulas of similar structure coexisted, but that a developmental relationship between formulas and constructions was not clearly evident. He concluded that the amount of elicited data was too limited to substantiate the learning path under investigation, and that more controlled task dimensions were also needed.

What Factors Might Determine the Outcomes of Such Studies?

The outcome of such studies searching for developmental sequences seeded by use of formulaic patterns must rest on methodological factors. The data have to be dense enough to identify repeated uses at the time of emergence. Sparse sampling is going to miss this (Tomasello & Stahl, 2004). The use of formulas and constructions are determined by context, function, genre, and register (Biber, 2003). Eskildsen (2009) identified how they are interactionally and locally contextualized, hence possibly transitory in nature, their deployment being occasioned by specific recurring usage events. If the elicitation tasks vary, so the chance of sampling the same formula and its potential variants diminishes accordingly.

Equally, they may vary as a function of L1 acquisition versus L2 acquisition. L1 acquisition may indeed be more formulaic than L2. When child learners are learning about language from formulaic frames (Ambridge & Lieven, 2011; Mintz, 2003; Tomasello, 2003), and the analysis of sequences of words (Elman, 1990; Kiss, 1973; Redington & Chater, 1998), they are learning from scratch about more abstract categories such as verb, pronoun, preposition, noun, transitive frame, and so on. It is debatable whether the units of early L1 acquisition are words at all (Peters, 1983). Adult L2 learners already know about the existence of these units, categories, and linguistic structures. They expect that there will be words and constructions in the L2 that correspond to such word classes and frames. Once they have identified them, or even once they have searched them out and actively learned such key vocabulary, they are more likely therefore to attempt creative construction, swopping these elements into corresponding slots in frames.

It is also the case, as in all other areas of language processing, that recognition of formulas is easier than production. Ellis and Ferreira-Junior (2009a, 2009b) showed that naturalistic adult L2 learners used the same verbs in frequent verb-argument constructions as are found in their input experience; indeed, the relative ordering of the types in the input predicted uptake with correlations in excess of $r = 0.90$. Nevertheless, while they would accurately produce short, simple, formulaic sequences such as *come in* or *I went to the shop*, structurally more complex constructions were often produced in the simplified form of the so-called *basic variety* (Klein & Perdue, 1992; Perdue, 1993), which involves a pragmatic topic-comment word ordering, where old information goes first and new information follows.

Transfer from L1 is also likely to affect the process (Granger, 2001). So too are types of exposure—children are naturalistic language learners from thousands of hours of interaction and input. While some adults learn naturalistically, others take grammar-rich courses. A dictionary and a grammar book neither provide naturalistic input nor encourage fluent idiomatic expression of formulaic speech. Nevertheless, even for foreign language acquisition, Myles (2004) demonstrated the viability of this sequence of acquisition.

LESSONS FROM THE STATISTICS OF USAGE: ZIPFIAN PATTERNS OF ACQUISITION

It is time, formulaicly speaking, to bring things to a close. In so doing, I must return to the issue of the formulaic-creative continuum (i.e., the interplay of formulas and abstract categories in development) and relate it to what I believe is the most important lesson of many to be learned from defining formulas statistically and from analyzing the statistics of usage.

Let us begin with our formula (i) *put it in,* and put it in its context of usage in a large corpus of English, such as the Corpus of Contemporary American English (COCA; Davies, 2008; http://corpus.byu.edu/coca/). As I search today,[7] *put it in* occurs 3,620 times. Consider it as a formulaic exemplification of the schematic verb-object-locative (VOL) verb-argument construction (VAC), which can describe a routine generic caused-motion function of moving something to a new place or in a new direction. Compare it to other VOL VACs. Search for *put it [i*].* ([i*] indicates a preposition.) This is very common (8,065 token occurrences), from *put it in* (3,620), *put it on* (1,926*), put it onto* (745), . . .(these seem like phrasal teddy bears to me)... dropping rapidly to *put it away* (1). Actually, these frequencies broadly follow a Zipfian distribution[8] (Ninio, 2011; Solé, Murtra, Valverde, & Steels, 2005; Zipf, 1935), as in language overall, but not following the particular ordering found in language as a whole—each slot attracts particular types of occupant (Ellis & O'Donnell, in press). We would get a very good idea of the locative by abstracting over these types and tokens of prepositions. How about the types of verb that work in these constructions? Searching *[v*] it [i*]* produces *put it in* (3,608), *give it to* (2,521), *do it in* (2,059), *put it on* (1,917)—again pretty formulaic. There are many more types here, but the frequencies still follow a Zipfian distribution. Figure 2 shows the results of a parallel analysis of the verbs types in VOL constructions from the native English speakers in the ESF corpus. There's some noise, of course, but abstracting over the verb types, of which *put* takes the lion's share, we get a pretty good idea of the semantics of caused-motion verbs.

Back to COCA, let's get more specific with *put it in the ** which generates *put it in the oven* (53), *put it in the refrigerator* (28), *put it in the back* (27), *put it in the freezer* (26), . . . *put it in the hold* (2). The sorts of everyday places that we put things in are reasonably clear in their semantics too, when averaged thus. And who puts? Searching *[p*]/[n*] put it* generates *you put it* (1067), *he put it* (975), *I put it* (891), . . . *who put it* (72), *official put it* (62), and so on. We get a clear idea of the sorts of entity who do the putting. Yes there are exceptions, but there is semantic coherence over the general exemplar cloud.

In each of these analyses there is a broadly Zipfian-type token-frequency distribution within the slot; the most frequent, path-breaking slot-filler for each VAC is much more frequent than the other members; the most frequent slot-filler is semantically prototypical and generic of the VAC island as a whole.

Our analysis here in the COCA corpus was seeded with a frequent formulaic prototype VOL, *put it in*, with its characteristic form and its generic interpretation. Scrutiny of its component islands and the types they attract in usage generated other VOLs with high-frequency prototypical occupants. Abstracting

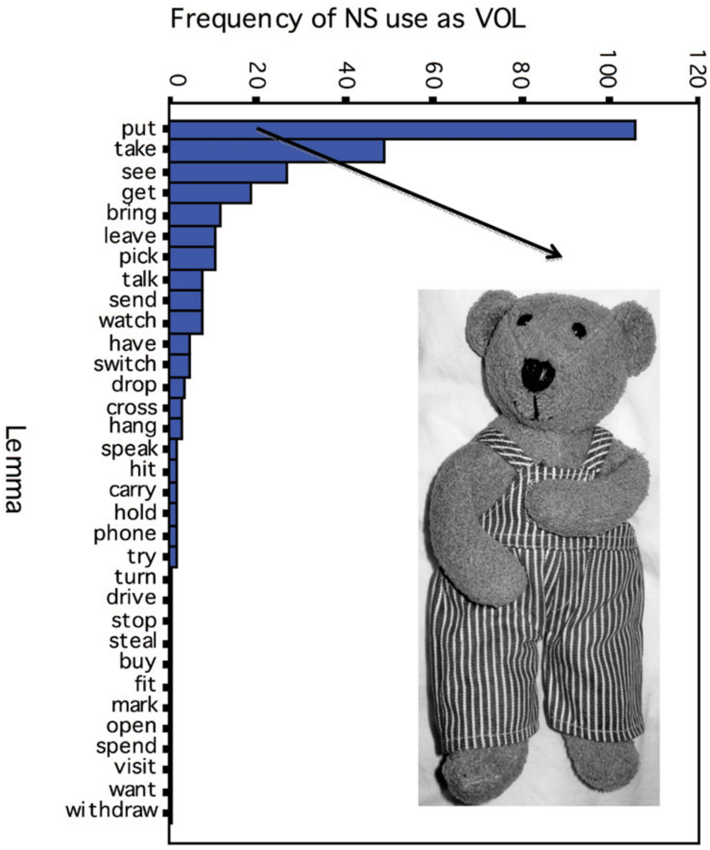

Fig. 2. The Zipfian-type token-frequency distribution of verb lemmas in the VOL VAC in the native English participants of the ESF project (Ellis & Ferreira-Junior, 2009b; Perdue, 1993). *Put* takes the lion's share and seeds the abstract schema with phrasal teddy bears such as *put it in*, *put it on*, *put it onto*.

over the typical types in the various islands results in a generalized schema for the VOL, with the different slots becoming progressively defined as attractors. Each island in each construction archipelago thus makes a significant contribution to its identification and interpretation (Ambridge & Lieven, 2011; Bybee, 2010; Ellis & Ferreira-Junior, 2009a, 2009b; Ellis & Larsen-Freeman, 2009; Ellis & O'Donnell, in press; Goldberg, 2006; Tomasello, 2003).

Usage-based theories of language hold that learners acquire constructions in a similar fashion—from the statistical abstraction of patterns of form-meaning correspondence in their usage experience—and that the acquisition of linguistic constructions can be understood in terms of the cognitive science of concept formation following the general associative principles of the induction of categories from experience of the features of their exemplars. In natural language, the Zipfian-type token-frequency distributions of the occupants of each of these construction islands, their prototypicality and generality of function in these roles and the reliability of mappings between these together conspire to make language learnable. Phrasal teddy bears, formulaic phrases with routine functional purposes, play a large part in this experience, and the analysis of their components gives rise to abstract linguistic structure and creativity.

Is the notion of language acquisition being seeded by formulaic phrases and yet learner language being formula-light having your cake and eating it too? Pawley and Syder (1983) thought not. While much of their classic article concentrated on the difficulty L2 learners had in achieving native-like formulaic selection and native-like fluency, nevertheless they stated, "Indeed, we believe that memorized sentences are the normal building blocks of fluent spoken discourse, and at the same time, that they provide models for the creation of many (partly) new sequences which are memorable and in their turn enter into the stock of familiar uses" (p. 208). Neither does Granger (2001), whose analyses of collocations and formulas in advanced EFL writing shows that "learners use fewer prefabs than their native-speaker counterparts" while at the same time they use some lexical teddy bears as "general-purpose amplifiers" in booster and maximizer phrases—"the analysis showed a highly significant overuse of *very* as the all-round amplifier par excellence ... one could postulate that the learners' underuse of *–ly* amplifiers is compensated for by their overuse of *very*." (p. 151).

This characterization of the developmental sequence as being from formula to low-scope pattern to creative construction is less true to the traditional idea of a formula as categorically defined, and more so to that of formulaicity as a variable reflecting sequential dependencies in usage and degree of entrenchment in the learners' mind. It remains an empirical priority to put rich, quantitative flesh on the core, skeletal claim that "grammar is what results when formulas are re-arranged, or dismantled and re-assembled, in different ways" (Hopper, 1987, p. 145).

ACKNOWLEDGMENTS

I am grateful to the Freiburg Institute for Advanced Studies (FRIAS), where I was based as an External Senior Fellow while writing this article; Heike Behrens, Stefan Pfänder, Peter Auer, Karin Madlener, Lars Konieczny, and the Graduiertenkolleg DFG 1624 Frequenzeffekte, University of Freiburg, for encouraging reflection on chunking and for constructive comments thereon; Charlene Polio for inviting this article; and my friends Matthew Brook O'Donnell, Ute Römer, Rita Simpson-Vlach, Stefanie Wulff, and Stefan Gries for researching formulaicity with me.

NOTES

1 Cognitive-linguistic and usage-based approaches view language learning as the acquisition of linguistic constructions as form-meaning mappings (Robinson & Ellis, 2008; Trousdale & Hoffmann, 2012).

2 Note that it does not reiterate prior quoted evidence, the brief being to concentrate upon the last five years, nor, in the space available, can it approach an exhaustive review of recent developments.

3 "3.1. Repetition of sequences in phonological STM allows their consolidation in phonological LTM.

3.2. The same cognitive system that does the LTM for phonological sequences does the perception of phonological sequences. Thus the tuning of phonological LTM to regular sequences allows more ready perception of input which contains regular sequences. As a result L2-experienced individuals' phonological STM for regular sequences is greater than for irregular ones.

3.3. "The cyclical reciprocal interactions of the processes of 3.1 and 3.2 allow learners to bootstrap their way to knowledge of SL structure" (Ellis, 1996, p. 92).

4 Some readers are probably getting tired of my quoting this quip—if so, well, that's verbatim formulaic recall for you.

5 Of course there is the decision here too whether to define these in terms of frequency or coherence.

6 If I was asked what is the research priority in SLA, I would reply with the need for dense longitudinal acquisition corpora.

7 Your numbers will differ because the corpus is always growing.

8 In natural language, Zipf's law (Zipf, 1935) describes how the highest frequency words account for the most linguistic tokens. The frequency of words decreases as a power function of their rank in the frequency table, with the most frequent word occurring approximately twice as often as the second most frequent word, which occurs twice as often as the fourth most frequent word, and so forth.

ANNOTATED BIBLIOGRAPHY

Collins, L., & Ellis, N. C. (Eds.). (2009). Input and second language construction learning: frequency, form, and function. *The Modern Language Journal, 93*, whole issue.

A special issue of *The Modern Language Journal* exploring how the acquisition of linguistic constructions is driven by the frequency and frequency distribution of exemplars within construction, the matching of their meaning to the construction prototype, and the reliability of their mappings.

Ellis, N. C., & Cadierno, T. (2009). Constructing a second language. *Annual Review of Cognitive Linguistics, 7*(Special section), 111–290.

A special section of *Annual Review of Cognitive Linguistics* exploring cognitive linguistic theories of SLA.

Gries, S. T., & Divjak, D. S. (Eds.). (in press). *Frequency effects in language: learning and processing*. Berlin, Germany: Mouton de Gruyter.

A state-of-the-art collection on frequency effects in language learning and processing.

Robinson, P., & Ellis, N. C. (Eds.). (2008). *A handbook of cognitive linguistics and second language acquisition*. London, UK: Routledge.

Reviews of usage-based approaches to L1 and L2 acquisition.

Rebuschat, J., & Williams, J. N. (Eds.). (in press). *Statistical learning and language acquisition*. Berlin, Germany: Mouton de Gruyter.

A state-of-the-art collection on statistical language learning.

REFERENCES

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin, 135*, 50–68.

Ambridge, B., & Lieven, E. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge, UK: Cambridge University Press.

Andrade, J., & Baddeley, A. D. (2011). The contribution of phonological short-term memory to artificial grammar learning. *Quarterly Journal of Experimental Psychology, 64*, 960–974.

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

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173.

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*, 241–248.

Bardovi-Harlig, K. (2002). A new starting point? *Studies in Second Language Acquisition, 24*, 189–198. doi: 10.1017/S0272263102002036

Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics, 47*, 383–411.

Behrens, H. (Ed.). (2008). *Corpora in language acquisition research: Finding structure in data*. Amsterdam, the Netherlands: John Benjamins.

Biber, D. (2003). Variation among university spoken and written registers: A new multidimensional analysis. In P. Leistyna & C. F. Meyer (Eds.), *Corpus analysis: Language structure and language use* (pp. 47–67). Amsterdam, the Netherlands: Rodopi.

Biber, D., Conrad, S., & Cortes, V. (2004). "If you look at . . .": Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*, 371–405.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. New York, NY: Cambridge University Press.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387.

Bod, R. (2001). *Sentence memory: Storage vs. computation of frequent sentences.* Paper presented at the CUNY-2001, Philadelphia, PA.

Bod, R., Hay, J., & Jannedy, S. (Eds.). (2003). *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Bybee, J. (2005). *From usage to grammar: The mind's response to repetition*. Paper presented at the Linguistic Society of America, Oakland, CA.

Bybee, J. (2010). *Language, usage, and cognition*. Cambridge, UK: Cambridge University Press.

Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam, the Netherlands: John Benjamins.

Corrigan, R., Moravcsik, E. A., Ouali, H., & Wheatley, K. M. (Eds.). (2009). *Formulaic language: Vol. 2. Acquisition, loss, psychological reality, and functional explanations*. Amsterdam, the Netherlands: John Benjamins.

Dąbrowska, E. (2004). *Language, mind and brain: Some psychological and neurological constraints on theories of grammar*. Edinburgh, UK: Edinburgh University Press.

Dabrowska, E., & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics, 16*, 437–474.

Davies, Mark. (2008). *The Corpus of Contemporary American English: 425 million words, 1990–present*. Available online at http://corpus.byu.edu/coca/.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition, 109*, 193–210.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology, 25*, 108–127.

Diessel, H. (2012). Construction grammar and first language acquisition. In G. Trousdale & T. Hoffmann (Eds.), *The Oxford handbook of construction grammar*. Oxford, UK: Oxford University Press.

Divjak, D. S., & Gries, S. T. (Eds.). (in press). *Frequency effects in language: Linguistic representation*. Berlin, Germany: Mouton de Gruyter.

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory, 6*, 125–155.

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. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33–68). New York, NY: Cambridge University Press.

Ellis, N. C. (2002a). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*, 143–188.

Ellis, N. C. (2002b). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition, 24*, 297–339.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 33–68). Oxford, UK: Blackwell.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*, 305–352.

Ellis, N. C. (2006a). Language acquisition as rational contingency learning. *Applied Linguistics, 27*, 1–24.

Ellis, N. C. (2006b). Selective attention and transfer phenomena in SLA: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*, 1–31.

Ellis, N. C. (2008a). Optimizing the input: Frequency and sampling in usage-based and form-focussed learning. In M. H. Long & C. Doughty (Eds.), *Handbook of second and foreign language teaching*. Oxford, UK: Blackwell.

Ellis, N. C. (2008b). Phraseology: The periphery and the heart of language. In F. Meunier & S. Grainger (Eds.), *Phraseology in language learning and teaching* (pp. 1–13). Amsterdam, the Netherlands: John Benjamins.

Ellis, N. C. (in press). What can we count in language, and what counts in language acquisition, cognition, and use? In S. T. Gries & D. S. Divjak (Eds.), *Frequency effects in cognitive linguistics: Vol. 1. Statistical effects in learnability, processing and change*. Berlin, Germany: Mouton de Gruyter.

Ellis, N. C., & Ferreira-Junior, F. (2009a). Construction Learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal, 93*, 370–386.

Ellis, N. C., & Ferreira-Junior, F. (2009b). Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics 7*, 111–139.

Ellis, N. C., Frey, E., & Jalkanen, I. (2008). The psycholinguistic reality of collocation and semantic prosody (1): Lexical access. In U. Römer & R. Schulze (Eds.), *Exploring the lexis-grammar interface*. Amsterdam, the Netherlands: John Benjamins.

Ellis, N. C., & Larsen-Freeman, D. (2009). Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning, 59*, 93–128.

Ellis, N. C., & O'Donnell, M. B. (in press). Statistical construction learning: Does a Zipfian problem space ensure robust language learning? In J. Rebuschat & J. Williams

(Eds.), *Statistical learning and language acquisition*. Berlin, Germany: Mouton de Gruyter.

Ellis, N. C., & Simpson-Vlach, R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory, 5*, 61–78.

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*, 375–396.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *Quarterly Journal of Experimental Psychology A, 49A*, 234–250.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Eskildsen, S. W. (2009). Constructing another language—Usage-based linguistics in second language acquisition. *Applied Linguistics, 30*, 335–357.

Eskildsen, S. W. (in press). L2 negation constructions at work. *Language Learning*.

Eskildsen, S. W., & Cadierno, T. (2007). Are recurring multi-word expressions really syntactic freezes? Second language acquisition from the perspective of usage-based linguistics. In M. Nenonen & S. Niemi (Eds.), *Collocations and idioms 1: Papers from the First Nordic Conference on Syntactic Freezes*. Joensuu, Finland: Joensuu University Press.

Evert, S. (2005). *The statistics of word cooccurrences: Word pairs and collocations*. Stuttgart, Germany: University of Stuttgart.

French, L., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics, 29*, 463–487.

Gerken, L. (2007). Acquiring linguistic structure. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 173–190). Oxford, UK: Blackwell.

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

Granger, S. (2001). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications*. Oxford, UK: Oxford University Press.

Gries, S. T. (2008). Phraseology and linguistic theory: A brief survey. In S. Granger & F. Meunier (Eds.), *Phraseology: an interdisciplinary perspective*. Amsterdam, the Netherlands: John Benjamins.

Gries, S. T. (2009). *Quantitative corpus linguistics with R: A practical introduction*. London, UK: Routledge/Taylor Francis.

Gries, S. T. (2012). Data in construction grammar. In G. Trousdale & T. Hoffmann (Eds.), *Oxford handbook of construction grammar*. Oxford, UK: Oxford University Press.

Gries, S. T. (in press). Frequencies, probabilities, association measures in usage-/exemplar-based linguistics: Some necessary clarifications. *Studies in Language*.

Gries, S. T., & Wulff, S. (2005). Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics, 3*, 182–200.

Gries, S. T., & Wulff, S. (2009). Psycholinguistic and corpus linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics, 7*, 164–187.

Gurevich, O., Johnson, M. A., & Goldberg, A. E. (2010). Incidental verbatim memory for language. *Language and Cognition, 2*, 45–78. doi: 10.1515/langcog.2010.003

Hall, T. (2010). L2 Learner-made formulaic expressions and constructions. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 10*, 1–18.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics, 4*, 237–260.

Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. *Cognitive Science, 31*, 765–814. doi: 10.1080/03640210701530755

Hebb, D. O. (1949). *The organization of behaviour*. New York, NY: Wiley.

Hilpert, M. (2008). New evidence against the modularity of grammar: Constructions, collocations, and speech perception. *Cognitive Linguistics, 19*, 491–511.

Hoey, M. (2004). The textual priming of lexis. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and language learners*. Amsterdam, the Netherlands: John Benjamins.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London, UK: Routledge.

Hopper, P. J. (1987). Emergent grammar. *Berkeley Linguistics Society, 13*, 139–157.

Hummel, K. M. (2009). Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics, 30*, 225–249.

Jiang, N. A. N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal, 91*, 433–445. doi: 10.1111/j.1540-4781.2007.00589.x

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Kapatsinski, V., & Radicke, J. (2008). Frequency and the emergence of prefabs: Evidence from monitoring. In R. Corrigan, E. A. Moravcsik, H. Ouali, & K. M. Wheatley (Eds.), *Formulaic language: Vol. 2. Acquisition, loss, psychological reality, functional explanation* (pp. 499–522). Amsterdam, the Netherlands: John Benjamins.

Kiss, G. R. (1973). Grammatical word classes: A learning process and its simulation. *Psychology of Learning and Motivation, 7*, 1–41.

Klein, W., & Perdue, C. (1992). *Utterance structure: Developing grammars again*. Amsterdam, the Netherlands: John Benjamins.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11,* 261–271.

Krashen, S., & Scarcella, R. (1978). On routines and patterns in language acquisition and performance. *Language Learning, 28*, 283–300.

Lieven, E. V. M., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage based approach. *Journal of Child Language, 30*, 333–370.

MacDonald, M. C., & Seidenberg, M. S. (2006). Constraint satisfaction accounts of lexical and sentence comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 581–611). London, UK: Elsevier.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological STM and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*.

Maslen, R., Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2004). A dense corpus study of past tense and plural overgeneralizations in English. *Journal of Speech, Language, & Hearing Research, 47*, 1319–1333.

McDonald, S. A., & Shillcock, R. C. (2003). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research, 43*, 1735–1751.

McDonald, S. A., & Shillcock, R. C. (2004). Eye-movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science, 14*, 648–652.

McDonough, K., & De Vleeschauwer, J. (in press). Prompt type frequency, auditory pattern discrimination, and EFL learners' production of wh-questions. *Studies in Second Language Acquisition, 34*.

McDonough, K., & Kim, Y. (2009). Syntactic priming, type frequency, and EFL learners' production of wh-questions. *Modern Language Journal, 93*, 386–398.

McDonough, K., & Mackey, A. (2008). Syntactic priming and ESL question development. *Studies in Second Language Acquisition, 30*, 31–47.

McDonough, K., & Trofimovich, P. (2008). *Using priming methods in second language research*. London, UK: Routledge.

Mellow, J. D. (2008). The emergence of complex syntax: A longitudinal case study of the ESL development of dependency resolution. *Lingua, 118*, 499–521.

Mintz, T. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*, 91–117.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning, 62*, 302–331. doi: 10.1111/j.1467–9922.2010.00626.x

Myles, F. (2004). From data to theory: The over-representation of linguistic knowledge in SLA. *Transactions of the Philological Society, 102*, 139–168.

Myles, F., Mitchell, R., & Hooper, J. (1999). Interrogative chunks in French L2: A basis for creative construction. *Studies in Second Language Acquisition, 21*, 49–80.

Naigles, L. R., Hoff, E., & Vear, D. (2009). Flexibility in early verb use: Evidence from a multiple-n diary study. *Monographs of the Society for Research in Child Development, 74*.

Ninio, A. (2011). *Syntactic development, its input and output*. Oxford, UK: Oxford University Press.

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second-language oral production by adult learners. *Applied Psycholinguistics, 27*, 377–402.

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition, 29*, 557–581.

O'Donnell, M. B., Römer, U., & Ellis, N. C. (2012). The development of formulaic language in first and second language writing: Investigating effects of frequency, association, and native norm. *International Journal of Corpus Linguistics, Special issue on Phraseology*.

Oakes, M. (1998). *Statistics for corpus linguistics*. Edinburgh, UK: Edinburgh University Press.

Oaksford, M., & Chater, N. (1999). *Rational models of cognition*. Oxford, UK: Oxford University Press.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–225). London, UK: Longman.

Perdue, C. (Ed.). (1993). *Adult language acquisition: Crosslinguistic perspectives*. Cambridge, UK: Cambridge University Press.

Peters, A. M. (1983). *The units of language acquisition*. New York, NY: Cambridge University Press.

Pickering, M. J. (2006). The dance of dialogue. *Psychologist, 19*, 734–737.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin, 134*, 427–459.

Pickering, M. J., & Garrod, S. C. (2006). Alignment as the basis for successful communication. *Research on Language and Computation, 4*, 203–228.

Reali, F., & Christiansen, M. H. (2009). Sequential learning and the interaction between biological and linguistic adaptation in language evolution. *Interaction Studies, 10*, 5–30.

Rebuschat, J., & Williams, J. N. (Eds.). (forthcoming). *Statistical learning and language acquisition*. Berlin, Germany: Mouton de Gruyter.

Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes, 13*, 129–192.

Robinson, P., & Ellis, N. C. (Eds.). (2008). *A handbook of cognitive linguistics and second language acquisition*. London, UK: Routledge.

Saffran, J. R., & Thiessen, E. D. (2007). Domain-general learning capacities. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 173–190). Oxford, UK: Blackwell.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13*, 206–226.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology, 45A*, 21–50.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16*, 155–172.

Shanks, D. R. (1995). *The psychology of associative learning*. New York, NY: Cambridge University Press.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list (AFL). *Applied Linguistics, 31*, 487–512.

Solé, R. V., Murtra, B., Valverde, S., & Steels, L. (2005). Language networks: Their structure, function and evolution. *Trends in Cognitive Sciences, 12*.

Sugaya, N., & Shirai, Y. (2009). Can L2 learners productively use Japanese tense-aspect markers? A usage-based approach. In R. Corrigan, E. Moravcsik, H. Ouali, & K. Wheatley (Eds.), *Formulaic language: Vol. 2. Acquisition, loss, psychological reality, functional applications*. Amsterdam, the Netherlands: John Benjamins.

Taguchi, N. (2007). Chunk learning and the development of spoken discourse in a Japanese as a foreign language classroom. *Language Teaching Research, 11*, 433–457.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. New York, NY: Cambridge University Press.

Tomasello, M. (2000). The item based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4*, 156–163.

Tomasello, M. (2003). *Constructing a language*. Boston, MA: Harvard University Press.

Tomasello, M., & Stahl, D. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language, 31*, 101–121.

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*, 569–613. doi: 10.1111/j.1467–9922.2010.00622.x

Trousdale, G., & Hoffmann, T. (Eds.). (2012). *Oxford handbook of construction grammar*. Oxford, UK: Oxford University Press.

Wen, Z. (2011). *Working memory and second language learning*. Bristol, UK: Multilingual Matters.

Wiechmann, D. (2008). On the computation of collostruction strength. *Corpus Linguistics and Linguistic Theory, 4*, 253–290.

Williams, J. N. (2011). Working memory and SLA. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 427–441). Oxford, UK: Routledge/Taylor Francis.

Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning, 53*, 67–121.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, 463–489.

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: MIT Press.