Journal of Experimental Psychology: Learning, Memory, and Cognition

Production of Familiar Phrases: Frequency Effects in Native Speakers and Second Language Learners

Anna Siyanova-Chanturia and Niels Janssen

Online First Publication, April 12, 2018. http://dx.doi.org/10.1037/xlm0000562

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Learning, Memory, and Cognition

0278-7393/18/$12.00

http://dx.doi.org/10.1037/xlm0000562

RESEARCH REPORT

Production of Familiar Phrases: Frequency Effects in Native Speakers and Second Language Learners

Anna Siyanova-Chanturia
Victoria University of Wellington

Niels Janssen
Universidad de La Laguna

Current evidence suggests that native speakers and, to a lesser degree, second language learners are sensitive to the frequency with which phrases occur in language. Much of this evidence, however, comes from language comprehension. While a number of production studies have looked at phrase frequency effects in a first language, little evidence exists with respect to the production of phrases in a second language. The present study addressed this gap by examining the production of English binomial expressions by first and late second language speakers. In a phrase elicitation task, participants produced binomial expressions (*bride and groom*) and their reversed forms (*groom and bride*), which are identical in form and meaning but differ in frequency. Mixed-effects modeling revealed that native speakers' articulatory durations were modulated by phrase frequency, but not the type of stimulus (binomial vs. reversed). Nonnative speakers' articulatory durations were not affected either by phrase frequency or stimulus type. Our findings provide further evidence for the effect of multiword information on language production in native speakers, and raise important questions about the effects of phrase frequency on language production in second language learners.

Keywords: multiword expressions, speech production, phrase frequency, first language speakers, second language learners

Frequency effects have long been recognized as a key factor affecting the speed of processing (Balota & Chumbley, 1984). Until recently, the bulk of relevant literature focused almost entirely on lexical (single word) frequency. Recent years, however, have seen a surge of interest in frequency effects at the phrase level in language processing in first language (L1) and, to a lesser extent, second language (L2) speakers. Comprehension studies employing behavioral paradigms showed that L1 speakers are sensitive to the frequency with which phrases, commonly known as multiword expressions (MWEs), occur in a language (Arnon & Snider, 2010; Kapatsinski & Radicke, 2009; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Tremblay & Baayen, 2010; Tremblay, Derwing, Libben, & Westbury, 2011). Electrophysiological evidence further suggests that MWEs are processed qualitatively and quantitatively differently from novel language, in that they are associated with faster processing, easier semantic integration, and template matching mechanisms (Molinaro & Carreiras, 2010; Siyanova-Chanturia, Conklin, Caffarra, Kaan, & van Heuven, 2017; Vespignani, Canal, Molinaro, Fonda, & Cacciari, 2010).

Crucial evidence for L1 speakers' sensitivity to distributional properties of MWEs comes from production studies. These studies too suggest a processing advantage for frequent phrases over less frequent ones. In the earliest such studies, Bybee and Scheibman (1999) and Bell et al. (2003) demonstrated that words were more likely to be phonetically reduced when they appeared in predictable contexts, such as frequent phrases (*I don't know, middle of the*). Bybee (2000) found that in recurrent word pairs, the boundary between the two words was akin to that between word-internal segments. Interestingly, a production advantage for frequent phrases versus infrequent combinations has also been observed in young children. Bannard and Matthews (2008) found that 2- and 3-year-old children's phrase production was modulated by the phrase frequency; higher frequency phrases were articulated more quickly than lower frequency ones.

Anna Siyanova-Chanturia, School of Linguistics and Applied Language Studies, Victoria University of Wellington; Niels Janssen, Psychology Department, Institute of Biomedical Technologies, and Institute for Neurosciences, Universidad de La Laguna.

The research leading to the findings reported in the present study was supported by a Research Establishment grant (Victoria University of Wellington) to Anna Siyanova-Chanturia, and by the Spanish Ministry of Economy and Competitiveness (Grants RYC2011-08433 and PSI2013-46334) to Niels Janssen.

Correspondence concerning this article should be addressed to Anna Siyanova-Chanturia, School of Linguistics and Applied Language Studies, Victoria University of Wellington, P.O. Box 600, Wellington 6140, New Zealand. E-mail: anna.siyanova@vuw.ac.nz

More recently, the studies looking at the production of n-grams varying in size, syntactic structure and the level of abstractness have reaffirmed the earlier findings reported in the literature, as well as provided a richer account of the frequency effects involved in the production of units above the word level (Arnon & Cohen Priva, 2013, 2014; Janssen & Barber, 2012; Tremblay & Tucker, 2011). Using an elicitation task, Janssen and Barber (2012) had Spanish participants produce noun-adjective, noun-noun, and determiner-noun-adjective phrases (*el coche rojo*: "the red car"). Naming latencies were shorter for higher frequency phrases than for lower frequency ones, implying that the language processor is sensitive to the distribution of linguistic information beyond individual words. In a similar vein, Arnon and Cohen Priva (2013) investigated the effect of phrase frequency on phonetic duration of automatically extracted n-grams (*do not have to worry*). Akin to Janssen and Barber (2012), these authors demonstrated that phonetic durations were significantly reduced in higher frequency phrases relative to lower frequency phrases, in spontaneous as well as elicited speech.

The findings attesting to phrase frequency effects in L1 (child and adult) comprehension and production point to the similarities in the processing of (single) words and compositional word combinations, implying that all linguistic information—at the word or phrase level—is governed by analogous cognitive mechanisms (Arnon & Snider, 2010; Arnon, McCauley, & Christiansen, 2017; Christiansen & Chater, 1999; Elman, 2009; Snider & Arnon, 2012). These findings further emphasize parallels between words and phrases, implying that MWEs are fundamental building blocks of language on a par with single words. An important question, however, that has, to date, received little attention, is whether or not proficient L2 speakers are sensitive to phrase frequency distributions during language processing. With over half of the world's population being bilingual (Grosjean, 1994), it is critical from a theoretical and practical standpoint that we address the question of phrasal processing in L1 as well as L2 speakers.

Overall, it has been proposed that L2 learners experience difficulties acquiring and using a rich repertoire of MWEs (Wray, 2002). Given that much of second language learning happens in a classroom rather than a naturalistic setting, most L2 learners will have had little experience encountering the thousands of MWEs that exist in English (or, other languages). As a result, many MWEs that are treated as prefabricated chunks by L1 speakers are treated as novel propositional speech by L2 learners. A learner may well know the meaning of the individual components—for example, *research* and *development*—but is likely to have had little experience encountering the phrase *research and development* for it to be entrenched in their mental lexicon. For such a learner, the more frequent and canonical *research and development* may well be as good as the less frequent *development and research.* In addition, in L2 learning, the focus has traditionally been on the amassing of single words and the acquisition of grammatical rules to be able to produce a seemingly infinite number of novel utterances. Learners, especially in foreign language teaching contexts, are rarely made aware of combinatorial mechanisms in language. As a result, it has been proposed that L1 learners and adult speakers, and L2 learners differ in their dependence on multiword information; while L1 speech is highly chunked in nature, L2 discourse is often described as less idiomatic, formulaic, or chunked. According to Wray (2002), L2 learners rely on linguistic creativity and make "overliberal assumptions about the collocational equivalence of semantically similar items" (pp. 201–202). Employing a computational model on the data from L1 learners, L1 adult speakers, and L2 adult speakers, McCauley and Christiansen (2017) found that L2 speech was characterized by a lesser use of MWEs relative to the other participant groups. Consequently, multiword information may play a different role in L1 and L2 learning (Arnon & Christiansen, 2017; McCauley & Christiansen, 2017; Wray, 2002). Because of the differences—in terms of conditions and outcomes—between L1 and L2 learning, there are reasons to expect dissimilarities not only in the use but also in the online processing of MWEs in native versus nonnative speakers.

Although the psycholinguistic evidence is limited and somewhat mixed, a handful of comprehension studies have demonstrated a reliable processing advantage for higher versus lower frequency phrases in L2 speakers (Hernández, Costa, & Arnon, 2016; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Sonbul, 2015). Interestingly, while some show that late second language learners are sensitive to distributional properties of MWEs akin to native speakers (Hernández et al., 2016), others point to an important role of language proficiency (Siyanova-Chanturia, Conklin, & Schmitt, 2011; Siyanova-Chanturia, Conklin, & van Heuven, 2011). Using a phrasal decision task and four-word combinations (*I have to say*), Hernández et al. (2016) examined phrase frequency effects in L1 speakers and late L2 learners. In line with earlier research, L1 speakers were found to be sensitive to multiword frequency. Crucially, L2 learners demonstrated multiword frequency effects on a par with natives, irrespective of the learning setting experienced (classroom or naturalistic) or English language proficiency (advanced or intermediate), suggesting parallels between phrasal processing in L1 and L2 speakers. On the contrary, in an eye-tracking reading experiment by Siyanova-Chanturia, Conklin, and van Heuven (2011), L1 speakers and more—but not less—proficient L2 speakers read English binomial expressions (*bride and groom*) faster than their reversed forms (*groom and bride*). At the same time, L1 and L2 speakers, irrespective of proficiency, exhibited general phrase frequency effects, suggesting that phrasal configuration (binomial vs. reversed) and phrase frequency are distinct cognitive phenomena (more on this in the General Discussion section).

The Present Study

Both native and proficient nonnative speakers have been found to comprehend MWEs faster and more easily than novel phrases. On the contrary, the findings of the production studies pertain largely to native speakers. Whether or not proficient nonnative speakers show phrase frequency effects at the production level remains an unanswered question. It is well established in language acquisition literature that receptive vocabulary knowledge (listening and reading) precedes productive knowledge (speaking and writing) both in first (L1) and second (L2) language acquisition (Clark, 1993; Nation, 2013). In addition, receptive learning and use is considered to be easier than productive learning and use (Ellis & Beaton, 1993; Nation, 2013). While native speakers would have had years of exposure to their L1 to be able to show a processing advantage for MWEs at the receptive as well as productive level, the processing advantage observed for MWEs in L2 comprehension studies may not necessarily hold in language production, because L2 learners would have had far less experience producing language than comprehending it. This especially applies to foreign, rather than second, language contexts where opportunities for output are limited.

Thus, an important empirical question and one that we sought to address in the present investigation is whether or not L1 speakers and, critically, L2 speakers are sensitive to phrase frequency distributions at the production level. The experiment relied on a phrase-elicitation task in which articulation durations were measured. In addition, articulation durations were measured for visually presented phrases that were produced after a delay of 1,700 ms. The delay between visual presentation and articulation allowed us to separate stages of comprehension and production, and ensure a production locus of potential phrase frequency effects. The finding of L2 speaker sensitivity to multiword frequency, or lack thereof, will have important implications for our forging a better understanding of the nature of the L2 mental lexicon, and the role that exposure and frequency play in (late) second language acquisition, processing, and use.

Method

Participants

Twenty-four native speakers (16 female, mean age: 21.1 years, $SD = 5.2$) and 24 late second-language learners (15 female, mean age: 26.5 years, $SD = 5.6$) of English took part in the experiment. One nonnative participant was excluded from the analysis as she reported having been brought up bilingual (Arabic and English) from birth. Thus, the data from 24 native speakers and 23 second language learners were included in the analysis reported below. All participants were full-time students studying a variety of subjects at Victoria University of Wellington. The nonnative speakers came from a wide range of L1 backgrounds and all had successfully passed either the International English Language Testing System (IELTS) test or Test of English as a Foreign Language (TOEFL) in order to study in an English-speaking country. The time the nonnative participants had spent in an English-speaking country ranged from 4 to 120 months ($M = 34.7$, $SD = 33.6$), while their reported first contact with English (in their home countries) ranged from 3 to 14 years ($M = 8.9$, $SD = 3.2$). All nonnative speaker participants reported to have studied English as a foreign language (EFL) at school; thus, the participants can be said to come from a variety of EFL contexts. Participants received a $10.00 gift voucher for their participation. All participants were informed of their rights and gave written informed consent for participation in the study, according to the Declaration of Helsinki. The research was carried out fulfilling ethical requirements in accordance with the standard procedures of Victoria University of Wellington.

Materials

The experimental and filler items were borrowed from Siyanova-Chanturia, Conklin, and van Heuven (2011). These authors used the British National Corpus (BNC) to extract 42 binomial expressions and their reversed forms (*bride and groom* vs. *groom and bride*), as well as their frequency of occurrence. These were considered to be well-matched phrases as, by definition, binomials and their reversed forms are matched in frequency of the individual words (content Word 1, the conjunction "and," and content Word 2), length, and part of speech. However, binomial expressions and their reversed forms differ considerably in their phrase frequency. On average, Siyanova-Chanturia, Conklin, and van Heuven (2011) estimated 247.3 occurrences in the BNC (per 100 million words) for binomials and 27.4 occurrences for their reversed forms. The binomials and reversed forms were also matched for association strength. To this aim, the Edinburgh Associative Thesaurus database was used. The mean strength of the forward association was 0.29, whereas the backward association was 0.25. The difference between the two was not found to be statistically significant: $t(37) = 0.73$, $p = .47$.

In addition to the experimental items, two types of fillers were used. The first group of fillers contained 42 meaningful and grammatically correct low frequency phrases ($M = 3.2$ occurrences in the BNC) that were matched with the binomials and their reversed forms in word length and part of speech (*fluid and fumes*). The second group of fillers was made of 63 meaningful and grammatically correct phrases of varying frequency (*tennis and badminton*), not matched with the target items in any of the above properties ($M = 20.7$ occurrences in the BNC). The two types of fillers served to prevent participants from noticing the presence of the reversed forms, which might have stood out as salient due to their inverted, and thus marked, order. The syntactic structure of both filler types was identical to that of binomials and reversed forms ('A and B').

Procedure

Binomials and their reversed forms were presented across two presentation lists. Thus, no participant saw both versions of the same phrase (each participant saw either a binomial or its reversed form). In each list, 42 experimental items (21 binomials and 21 reversed forms) were intermixed with 21 (half of the) fillers of the first type and all 63 fillers of the second. The total number of items seen by each participant was 126. Both native and nonnative speaker participants were randomly assigned to one of the two experimental lists. The experiment was divided into two blocks (63 items in each block) with a short break in between, and took about 25 min to complete from start to finish.

The experiment was run using Eprime and high-quality recording equipment. Participants sat in a soundproof laboratory in front of a computer and completed a phrase-elicitation task modeled on Arnon and Cohen-Priva (2013). A phrase-elicitation task was used because we wanted to focus on articulatory durations, and thus sought to eliminate the influence of the processes normally associated with comprehension. It is noteworthy that production studies often employ reaction time paradigms, for example, measuring the time of the vocal onset upon seeing a stimulus (Levelt, Roelofs, & Meyer, 1999). However, because reading a text out loud involves a perception stage and a production stage (Bock, 1996), our aim was to use a paradigm that would allow us to focus solely on the latter. The experiment started with a practice session consisting of five trials, followed by experimental trials. Each trial started with a 500 ms fixation point in the middle of the screen that participants had to fixate. Following the fixation point, a three-word phrase appeared across one line in the middle of the screen and stayed there for 1,700 ms. Participants were instructed to read the phrase silently while it was on the screen, and then produce (articulate) it out loud as soon as it disappeared from the screen. Thus, the participants produced the phrase once they could no longer see it on the screen. They were instructed to produce the phrase in the their most natural way.[1] The participants had 4,000 ms to articulate the phrase (this was the maximum time allowed). Following this, a new trial started.

Analysis and Results

Articulatory durations were automatically extracted from the individual wave files using a Matlab script (adapted from Bansal, Griffin, & Spieler, 2001). According to the authors, the script achieves an accuracy of around 85–99%, and so it has been used in various studies (Griffin & Oppenheimer, 2006; Jacobs & Dell, 2014). In order to further check its accuracy, visual checks were performed on every tenth wave file to ensure that the script was accurately detecting the onset and offset of speech in the audio wave files. These checks largely confirmed the accuracy indications cited above. Articulatory durations were calculated by subtracting the onset from the offset time for each trial and were subsequently log transformed.

Trials on which the participant produced an incorrect response (the produced phrase differed from the one presented on the screen) were removed from the analyses (38 trials out of a total of 1,974 trials, or 1.9% of the data), as well as those trials on which the articulatory durations were longer than 2,500 ms or shorter than 200 ms (four trials or 0.2%), and trials on which the onset time was shorter than 200 ms (one trial or 0.1%). The remaining set of 1,931 trials was analyzed using mixed-effects methodology (Bates, 2005; Pinheiro & Bates, 2000). The software used to analyze the data was R (v3.4.0), using the packages lme4 (v1.1–13) and afex (v0.17–8).

For statistical modeling of the articulatory durations, we considered the following fixed-effect predictors: a control variable called trial, frequency of the phrase (freq_phrase), frequency of the first word (freq_w1), frequency of the second word (freq_w2), frequency of the first bigram (freq_w1_and), frequency of the second bigram (freq_and_w2), phrase length in phonemes (len_phrase), nativeness (native), stimulus type (stim_type), and the interaction between the variables nativeness and the frequency of the phrase.[2]

The variable trial codes the ordinal position of a trial in the experiment, and was included to capture variance in the articulatory durations due to practice or fatigue (Baayen, 2008). The frequency variables were log-transformed and centered in order to remove skewness and to approximate a normal distribution (Figure 1). Inspection of the correlation matrix revealed a strong correlation between the phrase frequency variable and the two bigram frequencies (Table 1). To avoid problems with parameter estimation, and because our main hypothesis concerned distinguishing a phrase frequency from a simple word frequency effect, we decided to only model the phrase frequency variable in conjunction with the two unigram frequencies (freq_w1 and freq_w2). The question of whether phrase frequency and bigram frequency have independent contributions to articulatory durations was addressed in a different model and will be discussed separately. All models used the default dummy coding scheme in R (i.e., treatment coding).

The general modeling strategy followed recommendations outlined in Bates, Kliegl, Vasishth, and Baayen (2015). Specifically, the appropriate random effect structure was determined by incrementally removing terms starting from a full model that converged to a final model in which removal of a given term was no longer justified. Justification for removal was determined on the basis of model comparisons using the chi-square tests implemented in the anova function of R. The same strategy was applied to the fixed-effect structure. An overview of the random and fixed-effects terms and their corresponding model parameters when they were removed from the model is presented in Table 2. The final random effect structure included by-subject and by-item random intercepts. The fixed-effect structure included trial, phrase length, nativeness, phrase frequency, and the interaction between nativeness and phrase frequency. A model comparison between the initial full converging model and the final model was not significant suggesting that this simpler final model of the data was justified (no accumulation). The final model did not display a high degree of collinearity (condition index = 1.12). The *p* values corresponding to the parameter estimates in the final model were obtained using the Kenward-Rogers approximation for degrees of freedom implemented in the afex package.

As can be seen in Table 3 and Figure 2, the analyses revealed a positive effect of trial (Figure 2A), where later positions in the experiment were associated with longer durations, suggesting a fatigue effect. In addition, there were effects of phrase length, where longer phrases were predictably associated with longer durations (Figure 2B), and of nativeness, with shorter articulatory durations for native than for nonnative speakers (Figure 2C). Further, there was an effect of frequency of the phrase, with shorter durations associated with higher frequency. Importantly, nativeness significantly interacted with frequency of the phrase (Figure 2D). Further exploration of this interaction revealed that in a model with a comparable model structure to the model above, the effect of frequency of the phrase was significant for native speak-

[1] Of note is that response times in our task are not the same as response times in other production tasks. Specifically, in our task, the participants were asked to silently rehearse the three-word phrase that appeared on the screen and, as soon as the phrase disappeared, the participants were required to pronounce it out loud. In other words, response time in our task is a delayed naming response time. As noted above, the rationale behind this task procedure is that it attempts to minimize the impact of reading and comprehension that would have contaminated articulation times had the task been to simply read aloud the phrases as soon as they appeared on the screen.

[2] Note that phrase frequency and stimulus type are distinct. It is not the case that reversed forms are always lower frequency than binomial phrases. For example, *east and west* occurs 380 times in the British National Corpus, while its reversed form is attested 63 times; *sweet and sour* occurs 36 times, while its reversed form is unattested. Thus, some binomials are less frequent than some reversed forms (note, however, that a binomial is always more frequent than its own reversed form). See Siyanova-Chanturia, Conklin, and van Heuven (2011) for further discussion of this point.

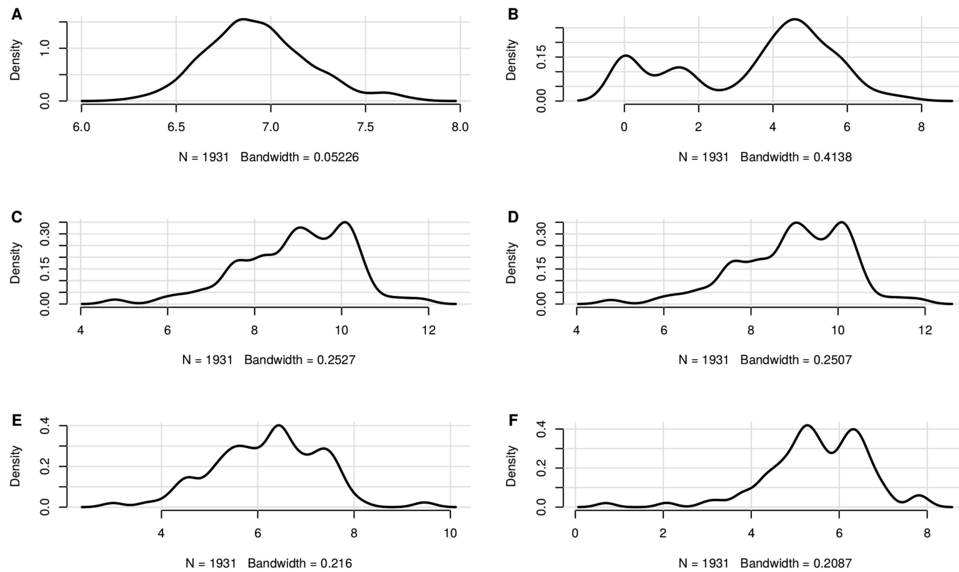

Figure 1. (A) Overview of the density distribution of the log transformed articulatory durations, (B) phrase frequency, (C) frequency of w1, (D) frequency of w2, (E) bigram frequency of "w1_and," and (F) bigram frequency of "and_w2" as used in the experiment (after artifact removal). Note that given the nature of the stimuli (binomials and their reversed forms), the unigram frequency distributions for "w1" and "w2" are highly similar.

ers, $t = -3.14$, $p < .003$, but not for nonnative speakers, $t = -1.50$, $p = .14$.[3]

Additional Analysis of Phrase Frequency and Bigram Frequency

The analyses above revealed that articulatory duration was determined by phrase frequency and not by unigram frequency. However, this finding does not provide information about the effect of bigram frequency on articulatory durations and whether these durations are more accurately modeled with bigram frequency than with phrase frequency. One problem in addressing this issue is that the bigram frequencies strongly correlate with the phrase frequency (see Table 1), which introduces problems related to collinearity. For this reason, we attempted to address this issue in a simplified model using all the available data, the target items as well as both types of filler items ($N = 5{,}711$). The statistical model included the fixed-effect predictors for trial, phrase length, nativeness, the two bigram frequencies (frequencies for "w1_and" [e.g., "bride and"], and "and_w2" [e.g., "and groom"]) and the phrase frequency. Using the same modeling strategy as detailed above to determine the final model, we found that whereas bigram frequency for and_w2 did not predict articulatory durations, $t = -0.70$, $p = .48$, both bigram frequency w1_and, $t = -2.13$, $p < .05$ and phrase frequency, $t = -2.0$, $p < .05$ remained significant predictors of articulatory durations.[4] These results are in line with our hypothesis that the frequency of units larger than single words contribute to articulatory durations.

General Discussion

In the present investigation, using experimentally elicited speech, we set out to investigate the effect of multiword frequency on articulatory durations in a L1 and L2. Specifically, we asked the following question: Are L1 speakers and, crucially, late L2 learners sensitive to phrase frequency distributions during language production? The following findings emerged.

Native speakers but not nonnative speakers were found to show phrase frequency effects during online language production. While native speaker participants clearly articulated more frequent target phrases faster than less frequent ones, second language users showed no articulatory advantage for frequent phrases over less frequent ones. The results specific to L1 speakers are in line with earlier comprehension (Arnon & Snider, 2010; Kapatsinski & Radicke, 2009; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Tremblay & Baayen, 2010) and production literature (Arnon & Cohen Priva, 2013, 2014; Janssen & Barber, 2012; Tremblay & Tucker, 2011). As such, these results add to the existing body of research supporting the view according to which speakers are sensitive to the frequency with which linguistic units, single words as well as longer phrases, occur in language. Interestingly, L2 results do not appear to support the findings of Hernández et al. (2016) and other comprehension studies that have reported phrase frequency effects for nonnative participants. It appears that while L2 speakers have been shown to be attuned to distributional properties of MWEs in language comprehension, they may not

[3] Further analyses revealed that the absence of the phrase frequency effect for nonnative speakers was not modulated by the variables length of stay in an English speaking country or the age of exposure to English (all *p*s >0.05, see the script for further details https://github.com/iamnielsjanssen/ArticulatoryDurations).

[4] In this additional analysis, all of the data (including the fillers) were used. It was not possible to do this for the main analyses reported above, since these analyses relied on the stim_type effect.

Table 1
Correlation Matrix of the Frequency Variables Included in the Experiment, Comprising the Frequency of the Phrase, the Unigram Frequencies as Well as the Bigram Frequencies. Note the High Correlation Between Frequency of the Phrase and the Bigram Frequencies

Variables	freq_phrase	freq_w1	freq_w2	freq_w1_and	freq_and_w2
freq_phrase	1				
freq_w1	.22	1			
freq_w2	.17	.59	1		
freq_w1_and	.47	.82	.50	1	
freq_and_w2	.58	.48	.69	.56	1

Note. freq_phrase = frequency of the phrase; freq_w1 and freq_w2 = unigram frequencies; freq_w1_and, freq_and_w2 = bigram frequencies.

necessarily show a comparable processing advantage in language production. We offer two possible explanations for why this might be the case. These explanations, which are complementary rather than competing, draw on an important distinction between receptive and productive knowledge in language learning. First, as Corson (1995) argues, the description of productive and receptive vocabulary should be based on the idea of language use, rather than solely on degrees of knowledge. Extending Corson's proposition, Nation (2013) suggests that some receptive vocabulary may be well known but rarely used and, therefore, never productive. As he points out, although some people may be able to curse and swear, they never do.

Second, it is generally accepted that receptive learning and use is easier than productive learning and use (Ellis & Beaton, 1993; Nation, 2013). Various accounts have been proposed in support of this proposition, two of which are of relevance to the present study. The first one is the "amount of knowledge" explanation (Crow, 1986; Nation, 2013). Productive learning and use is more difficult because it requires "extra learning of new spoken and written output patterns" compared with receptive learning and use (Nation, 2013: 51). Indeed, for receptive purposes, a second language learner may only need to know a few distinctive features of the form; while for productive purposes, the learner's knowledge (of a word or a phrase) has to be more complex, complete, and precise (in the case of the spoken output, knowing how to pronounce the target item as well as the ease with which it is pronounced are key). As Nation (2013) further elaborates, the form of an item is more likely to cause difficulty than the meaning. In our study, this "difficulty" might have manifested itself as lack of automaticity or proceduralization (Segalowitz & Hulstijn, 2005), and hence the absence of phrase frequency effects. In addition, the amount of knowledge explanation is connected with contextual knowledge, such as learning which word(s) a given word does or does not go with (collocational knowledge). The development of such knowledge, as Nation (2013) notes, requires great amounts of exposure to the language, which may be more essential for productive learning and use than for receptive learning and use. For example, having never come across the binomial *alive and well* before, a second language learner who knows the individual words of this phrase will not have any difficulty understanding the meaning of the phrase. On the contrary, not being familiar with the binomial as a phrasal configuration, a learner is unlikely to be able to produce it from scratch. If we continue with this logic, it appears that more exposure to and experience with the language might be required to exhibit frequency effects in production than in receptive use.

Table 2
Overview of the Random and Fixed Effect Terms That Were Removed From the Model. Listed Are the Values That These Terms Had at the Moment When They Were Removed

Random parts	Beta				
$Slope_{item}$ native	.0000				
$Slope_{subject}$ w2 freq	.0000				
$Slope_{subject}$ w1 freq	.0000				
$Slope_{subject}$ phrs freq	.0000				
$Slope_{subject}$ phrs len	.0038				
$Slope_{subject}$ stim type	.0238				
Fixed parts	**Beta**	***SE***	***df***	***t***	***p***
Freq w2	−.0008	.0087	1,921	−.11	.92
Stim type	−.0023	.0027	1,922	−.91	.37
Native: freq w1	−.0078	.0079	1,923	−.98	.32
Native: freq w2	.0066	.0064	1,924	1.03	.30
Freq w1	−.0092	.0069	1,925	−1.33	.19

Note. *SE* = standard error; *df* = degrees of freedom; w1 = first word; w2 = second word; freq = frequency; phrs freq = phrase frequency; phrs len = phrase length; stim type = stimulus type.

Table 3
Results of Mixed-Effects Modeling Detailing Aspects of the Random Effects and Fixed Effect Structure

Random parts	Beta				
SD $intercept_{item}$	.0682				
SD $intercept_{subject}$	.1132				
SD residual	.18				
N_{item}	84				
$N_{subject}$	47				
$N_{observations}$	1,931				
Fixed parts	**Beta**	***SE***	***df***	***t***	***p***
Intercept	6.9310	.0186	64	373.44	<.0001
Trial	.0012	.0002	83	5.25	<.0001
Phrase length	.0418	.0029	83	14.39	<.0001
Nativeness	.2121	.0340	47	6.24	<.0001
Phrase frequency	−.0109	.0041	83	−2.67	<.009
Native: phrase frequency	.0088	.0038	1,807	2.31	<.03

Note. *SD* = standard deviation; *SE* = standard error; *df* = degrees of freedom.

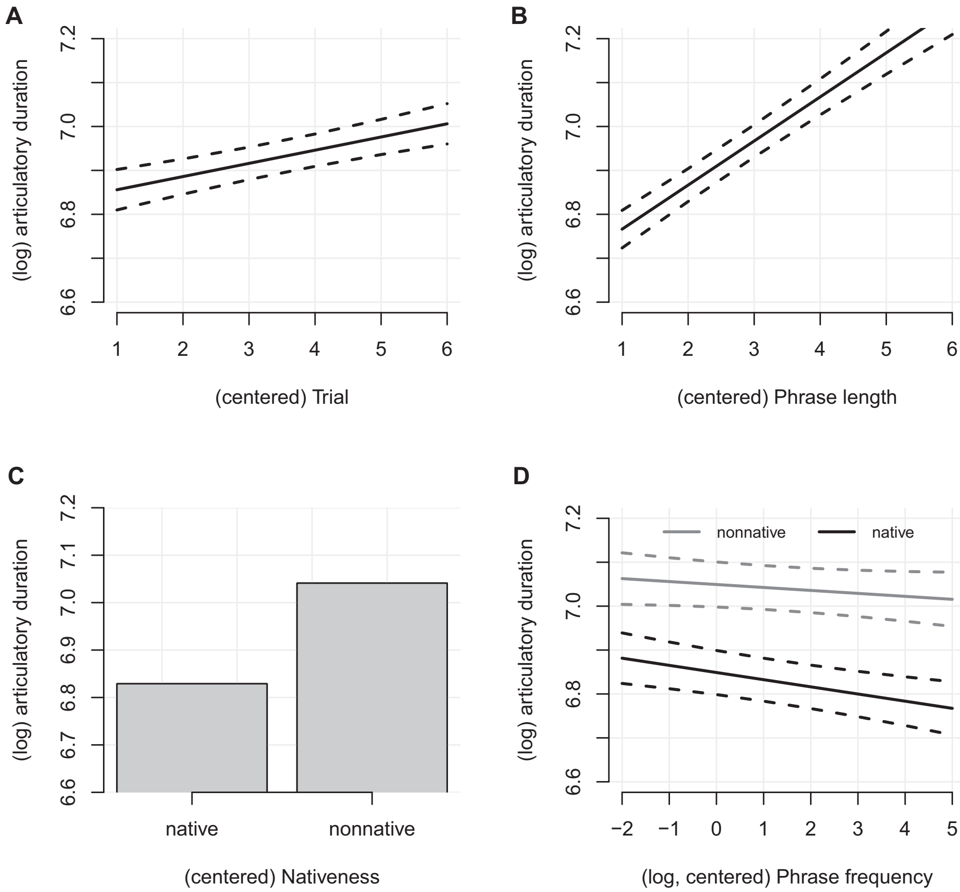

Figure 2. (A) Graphical presentation of the effect of trial, (B) phrase length, (C) nativeness, and (D) the interaction between nativeness and phrase frequency. In all panels the *y* axis displays the log transformed articulatory durations, and the *x* axis the relevant variable information.

The other explanation is the "practice" explanation. As Nation (2013) argues, in typical language-learning conditions (be they classroom or naturalistic settings), receptive use of language (listening and reading) is more common and hence gets more practice than productive use (speaking and writing). This might be particularly the case for those of our participants who had only recently arrived in New Zealand. In fact, 12 of the 23 nonnative speaker participants reported having spent under two years in an English-speaking country. It may well be that our L2 speakers simply had too few opportunities to use English binomial expressions productively, despite undoubtedly knowing them receptively, either as three-word set phrases or as individual words. It is noteworthy that both the amount of knowledge and the practice explanations are further substantiated by the differences in productive versus receptive vocabulary size reported in the literature. It is well established that both first and second language speakers have larger receptive vocabularies than productive ones (Nation, 2013; Webb, 2008), with some estimates suggesting that productive knowledge is less than half receptive knowledge (Brysbaert, Stevens, Mandera, & Keuleers, 2016). As a result, receptive tests are typically easier than productive tests, with both L1 and, especially, L2 speakers scoring more on the former than the latter (Ellis & Beaton, 1993; Stoddard, 1929). In sum, all things being equal, it appears that more time, effort, and repeated exposure is required to learn vocabulary for productive use than for receptive use, especially when it comes to units above the word level. Given this consideration, we argue that it might be easier to observe phrase frequency effects at the level of comprehension than production. Future research employing both modalities on the same group of native and nonnative participants and using the same materials should be able to test this hypothesis further.

Our findings may also be suggestive of the possibility that there *is* an effect of phrase frequency, but that it is simply more difficult to detect than the effect for L1. The study employed 24 L1 and 23 L2 speakers. It is conceivable that, due to generally greater variability among nonnative speakers and individual differences associated with L2 learning (e.g., L1 background, quality and quantity of exposure to L2, learning context, motivation and anxiety using L2; see Ellis, 2015; Ortega, 2009) compared with native speakers who are more homogenous in their characteristics and experience with the target language, more L2 participants might be needed for the effect to become detectable quantitatively. Such an interpretation of the data, however, would also predict that there should be a phrase frequency effect for L2 speakers who have spent more time in an L2-speaking country, or who have had greater exposure to an L2. Yet, as noted in Footnote 3, we did not find a modulation of the phrase frequency effect by these variables in our analyses. Thus, although we acknowledge that there must be future studies with a larger participant pool to address this issue directly, our current results are not consistent with an explanation that assumes that the lack of phrase frequency effect for the L2 speakers is simply due to an increased variability in the L2 data.

A further finding in our study is that, contrary to our predictions, the analyses showed that while the frequency of occurrence affected articulatory durations of the target phrases (for native speakers), the type of the phrasal configuration—binomial versus reversed—was not found to be a significant factor. More so, the two phrasal configurations were found to be articulated in a very similar way. And because no interaction with the learners' L1 background (native vs. nonnative) was observed, we can conclude that this was the case for both participant groups. On the surface, this finding goes against Siyanova-Chanturia, Conklin, and van Heuven (2011), who found that L1 speakers and higher proficiency L2 users were sensitive to whether a phrase was presented as a binomial or a reversed form. What might account for such differences between comprehension (Siyanova-Chanturia, Conklin, & van Heuven, 2011) and production (the present study)? We believe the explanation is task-related. The participants in our study were required to first read the target phrase presented in isolation and then articulate it, once it disappeared from the screen. On the contrary, the participants in Siyanova-Chanturia, Conklin, and van Heuven (2011) read the phrases embedded in sentence contexts as soon as they appeared on the screen. One possibility is that the phrase elicitation task used in the present study, with its delayed articulation component, was able to eliminate the effect of stimulus type, but not the effect of phrase frequency. That is, during the articulation delay period, participants were able to prepare the phrases such that they no longer exhibited differences in the durations of binomials and reversed forms. However, the delay period was not sufficient to eliminate the phrase frequency effect.

Delayed naming tasks are frequently used in the language production literature to examine the locus of effects in the cognitive system, where a long delay between stimulus presentation and articulation is thought to eliminate any contribution from memory retrieval (Janssen, Schirm, Mahon, & Caramazza, 2008). Applying this logic to the present data suggests that the effect of phrase type arises during the retrieval of words from memory. The presence of this effect in the direct reading studies (Siyanova-Chanturia, Conklin, & van Heuven, 2011) and the absence of this effect in the delayed articulatory durations may indicate that this effect arises during memory retrieval. In addition, the presence of the effect of phrase frequency on the delayed articulatory durations observed here suggests that this effect arises relatively late in the production of the phrases, presumably during articulatory planning. Note that this explanation does not preclude other loci for the phrase frequency effect (in comprehension), a conclusion that is consistent with the idea that frequency effects arise at multiple levels in the language production system (Kittredge, Dell, Verkuilen, & Schwartz, 2008; Knobel, Finkbeiner, & Caramazza, 2008). In short, the results observed here point toward a multilevel representation of the phrase frequency effect where at least part of the effect arises at an articulatory level. This explanation, however, is ad hoc and further targeted investigation is necessary to confirm this possibility.

In summary, the results reported in the present study advance our understanding of multiword statistics and its effect on language production in a number of ways. First, they confirm that phrase frequency affects articulatory durations in native speakers, reflecting language users' sensitivity to the distribution of linguistic information at various grain sizes. Second, our findings with nonnative speakers suggest that late L2 learners might not be attuned to phrase frequency distributions at the level of production. This finding lends support to the propositions of Wray (2002), Arnon and Christiansen (2017), and McCauley and Christiansen (2017) that MWEs play a different role in L1 and L2 learning and, consequently, processing, with nonnative speakers being less reliant on multiword information. We further take this result to support the idea that productive learning and use is more laborious than receptive learning and use, and, as a result, it might be harder to observe phrase frequency effects in elicited production than in receptive L2 language usage. Third, the results revealed different effects of phrase frequency and phrasal configuration on articulatory durations, possibly suggesting different loci for these two effects: A memory retrieval locus for the effect of phrasal configuration, and an articulatory locus of the effect of phrase frequency. In line with earlier research (Bod, 2006; Bybee, 1998, 2006; Bybee & Beckner, 2009; Bybee & McClelland, 2005; Christiansen & Chater, 1999; Ellis, 2002; Rumelhart & McClelland, 1986), our data further highlight the important role of experience in a theory of language acquisition, processing, and use.

References

Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science, 9,* 621–636. http://dx.doi.org/10.1111/tops.12271

Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech, 56,* 349–371. http://dx.doi.org/10.1177/0023830913484891

Arnon, I., & Cohen Priva, U. (2014). The changing effect of word and multiword frequency on phonetic duration for highly frequent phrases. *The Mental Lexicon, 9,* 377–400.

Arnon, I., McCauley, S., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-Acquisition effects for multiword phrases. *Journal of Memory and Language, 92,* 265–280. http://dx.doi.org/10.1016/j.jml.2016.07.004

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62,* 67–82. http://dx.doi.org/10.1016/j.jml.2009.09.005

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511801686

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340–357. http://dx.doi.org/10.1037/0096-1523.10.3.340

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. http://dx.doi.org/10.1111/j.1467-9280.2008.02075.x

Bansal, P., Griffin, Z. M., & Spieler, D. H. (2001). Epd: Matlab wave file parsing software. Retrieved from http://homepage.psy.utexas.edu/homepage/faculty/Griffin/resources.html

Bates, D. M. (2005). Fitting linear mixed models in R. *R News, 5,* 27–30.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint, arXiv,* 1506.04967.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America, 113,* 1001–1024. http://dx.doi.org/10.1121/1.1534836

Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3,* 395–421. http://dx.doi.org/10.3758/BF03214545

Bod, R. (2006). Exemplar-based syntax: How to get productivity from exemplars. *The Linguistic Review, 23,* 291–320. http://dx.doi.org/10.1515/TLR.2006.012

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology, 7,* 1116. http://dx.doi.org/10.3389/fpsyg.2016.01116

Bybee, J. (1998). The emergent lexicon. In M. C. Gruber, D. Higgins, K. S. Olson, and T. Wysocki (Eds.), *CLS 34: The panels* (pp. 421–435). Chicago, IL: University of Chicago, Chicago Linguistic Society.

Bybee, J. (2000). The phonology of the lexicon. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 65–85). Stanford, CA: CSLI.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82,* 711–733. http://dx.doi.org/10.1353/lan.2006.0186

Bybee, J., & Beckner, C. (2009). Usage-based theory. In B. Heine & H. Narrog (Eds.) *The Oxford handbook of linguistic analysis* (pp. 1–30). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199544004.013.0032

Bybee, J., & McClelland, J. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review, 22,* 381–410. http://dx.doi.org/10.1515/tlir.2005.22.2-4.381

Bybee, J., & Scheibman, J. (1999). The effect of usage on degree of constituency: The reduction of *don't* in American English. *Linguistics, 37,* 575–596. http://dx.doi.org/10.1515/ling.37.4.575

Christiansen, M., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23,* 157–205. http://dx.doi.org/10.1207/s15516709cog2302_2

Clark, E. (1993). *The lexicon in acquisition*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511554377

Corson, D. J. (1995). *Using English words*. Dordrecht, the Netherlands: Kluwer Academic. http://dx.doi.org/10.1007/978-94-011-0425-8

Crow, J. T. (1986). Receptive vocabulary acquisition for reading comprehension. *Modern Language Journal, 70,* 242–250. http://dx.doi.org/10.1111/j.1540-4781.1986.tb05271.x

Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 46,* 533–558. http://dx.doi.org/10.1080/14640749308401062

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24,* 143–188. http://dx.doi.org/10.1017/S0272263102002024

Ellis, R. (2015). *Understanding second language acquisition*. New York, NY: Oxford University Press.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science, 33,* 547–582. http://dx.doi.org/10.1111/j.1551-6709.2009.01023.x

Griffin, Z. M., & Oppenheimer, D. M. (2006). Speakers gaze at objects while preparing intentionally inaccurate labels for them. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 943–948. http://dx.doi.org/10.1037/0278-7393.32.4.943

Grosjean, F. (1994). Individual bilingualism. *The encyclopedia of language and linguistics*. Oxford, UK: Pergamon Press.

Hernández, M., Costa, A., & Arnon, I. (2016). More than words: Multiword frequency effects in non-native speakers. *Language, Cognition and Neuroscience, 31,* 785–800. http://dx.doi.org/10.1080/23273798.2016.1152389

Jacobs, C. L., & Dell, G. S. (2014). "hotdog", not "hot" "dog": The phonological planning of compound words. *Language, Cognition and Neuroscience, 29,* 512–523. http://dx.doi.org/10.1080/23273798.2014.892144

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PLoS ONE, 7,* e33202. http://dx.doi.org/10.1371/journal.pone.0033202

Janssen, N., Schirm, W., Mahon, B. Z., & Caramazza, A. (2008). Semantic interference in a delayed naming task: Evidence for the response exclusion hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 249–256. http://dx.doi.org/10.1037/0278-7393.34.1.249

Kapatsinski, V., & Radicke, J. (2009). Frequency and the emergence of prefabs: Evidence from monitoring. In R. Corrigan, E. Moravcsik, H. Ouali, & K. Wheatley (Eds.), *Formulaic language* (pp. 499–522). Amsterdam, the Netherlands: John Benjamins.

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology, 25,* 463–492. http://dx.doi.org/10.1080/02643290701674851

Knobel, M., Finkbeiner, M., & Caramazza, A. (2008). The many places of frequency: Evidence for a novel locus of the lexical frequency effect in word production. *Cognitive Neuropsychology, 25,* 256–286. http://dx.doi.org/10.1080/02643290701502425

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38. http://dx.doi.org/10.1017/S0140525X99001776

McCauley, S. M., & Christiansen, M. H. (2017). Computational investigations of multiword chunks in language learning. *Topics in Cognitive Science, 9,* 637–652. http://dx.doi.org/10.1111/tops.12258

Molinaro, N., & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology, 83,* 176–190. http://dx.doi.org/10.1016/j.biopsycho.2009.12.006

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

Ortega, L. (2009). *Understanding second language acquisition*. London, UK: Hodder.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models and S and S-PLUS*. New York NY: Springer-Verlag. http://dx.doi.org/10.1007/978-1-4419-0318-1

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Mechanisms of language acquisition* (Vol. 2, pp. 216–271). Hillsdale, NJ: Erlbaum.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). New York, NY: Oxford University Press.

Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., & van Heuven, W. J. B. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language, 175,* 111–122. http://dx.doi.org/10.1016/j.bandl.2017.10.004

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and nonnative speakers. *Second Language Research, 27,* 251–272. http://dx.doi.org/10.1177/0267658310382068

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 776–784. http://dx.doi.org/10.1037/a0022531

Snider, N., & Arnon, I. (2012). A unified lexicon and grammar? Compositional and noncompositional phrases in the lexicon. In S. Gries & D. Divjak (Eds.), *Frequency effects in language* (pp. 127–164). Berlin, Germany: Mouton de Gruyter. http://dx.doi.org/10.1515/9783110274073.127

Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition, 18,* 419–437. http://dx.doi.org/10.1017/S1366728914000674

Stoddard, G. D. (1929). An experiment in verbal learning. *Journal of Educational Psychology, 20,* 452–457. http://dx.doi.org/10.1037/h0073293

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London, UK: Continuum International.

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. http://dx.doi.org/10.1111/j.1467-9922.2010.00622.x

Tremblay, A., & Tucker, B. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon, 6,* 302–324. http://dx.doi.org/10.1075/ml.6.2.04tre

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience, 22,* 1682–1700. http://dx.doi.org/10.1162/jocn.2009.21293

Webb, S. (2008). Receptive and productive vocabulary size. *Studies in Second Language Acquisition, 30,* 79–95. http://dx.doi.org/10.1017/S0272263108080042

Wray, A. (2002). *Formulaic language and the lexicon*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511519772

Received October 31, 2016
Revision received January 8, 2018
Accepted January 9, 2018 ■