

John Benjamins Publishing Company



This is a contribution from *The Mental Lexicon* 9:3
© 2014. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible only to members (students and faculty) of the author's/s' institute. It is not permitted to post this PDF on the internet, or to share it on sites such as Mendeley, ResearchGate, Academia.edu.

Please see our rights policy on <https://benjamins.com/#authors/rightspolicy>

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Time and again

The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences

Inbal Arnon and Uriel Cohen Priva

Hebrew University / Brown University

There is growing evidence that multiword information affects processing. In this paper, we look at the effect of word and multiword frequency on the phonetic duration of words in spontaneous speech to (a) extend previous findings and (b) ask whether the relation between word and multiword information changes across the frequency continuum. If highly frequent sequences are stored holistically, then the effect of word frequency should disappear. If alternatively, increased sequence usage causes a change in the prominence of word and multiword information, we should see reduced effects of word frequency, and increased effects of sequence frequency for high frequency sequences. We first extend previous findings by showing that trigram frequency affects single word duration, even when controlling for word predictability. We then show that the effect of trigram frequency increases while the effect of word frequency decreases — but does not disappear — for highly frequent sequences. The findings provide further support for the effect of multiword information on processing and document the growing prominence of multiword information with repeated usage.

Keywords: speech production, phonetic duration, multiword frequency, formulaic sequences, usage-based, holistic storage

There is growing evidence that speakers (children and adults) are sensitive to the distributional properties of multi-word sequences and draw on such information in production, comprehension and learning (Arnon & Clark, 2011; Arnon & Snider, 2010; Arnon & Cohen Priva, 2013; Bannard & Matthews, 2008; Reali & Christiansen, 2007; Siyanova-Chanturia et al., 2011; Tremblay & Baayen, 2010). Adult speakers are faster to recognize higher frequency four-word phrases (Arnon

& Snider, 2010) and show better memory of them (Tremblay & Baayen, 2010) even when part frequencies are controlled for. Young children (two- and three-year-olds) are faster and more accurate at producing higher frequency phrases compared to lower frequency ones (Bannard & Matthews, 2008) while slightly older children show better production of irregular plurals inside frequent frames (e.g., *Brush your – teeth*, Arnon & Clark, 2011).

Finding that speakers are sensitive to the distributional properties of multiword sequences has theoretical and empirical implications. Finding multiword frequency effects across the frequency continuum (not only for very high frequency sequences) highlights the parallels between words and larger sequences (Snider & Arnon, 2012) and makes it harder to argue, as traditional words-and-rules models do, that the forms are generated and processed by two qualitatively different cognitive systems (Pinker & Ullman, 2002; Ullman & Walenski, 2005). The findings expand the range of information speakers are sensitive to and call for production and comprehension models that can capture and predict the effect of multiword information on processing (Arnon & Cohen Priva, 2013).

The findings also raise a novel grain-size challenge: seeing that speakers are sensitive to both word and multiword information, we can ask how these different frequency measures interact in online processing, and whether there are cases where there is competition between word and multiword information. This question has implications for two different bodies of literature: the psycholinguistic literature, and the literature on formulaic language. Grain-size challenges have a long history in psycholinguistics. The debate was made prominent in the 90's when it became clear that speakers are affected by both structural and lexical properties when processing ambiguous sentences (e.g., Trueswell et al., 1994). In an important paper, Mitchell et al. (1995) raised the grain-size challenge: if speakers are sensitive to distributional information at various grain-sizes (structure, structure given verb), how is that information integrated and weighed in online processing? While the empirical findings since then have made this challenge more complicated — we now know of many more distributional measures that affect processing (predictability, bigram frequency, multiword frequency) — the challenge remains unresolved. We do not have a fully operational model of how different types of information are weighed in online processing. Examining the relative influence of word and multiword information on processing adds another layer to the grain-size challenge.

At the same time, the relation between word and multiword information is highly relevant for studies of formulaic language (Schmitt, 2004; Wray, 2002, 2008). This literature highlights the formulaic aspect of language use. The fact that despite its productive potential, actual language use is restricted, containing a considerable amount of multiword expressions: word combinations that co-occur in

predictable and restricted ways (Sinclair, 1991; Wray, 2002). Such expressions are common in the speech and writing of native speakers (e.g., Conklin & Schmitt, 2008; De Cock, 1998) and are seen as marker of native-like fluency (Pawley & Syder, 1983). The term formulaic language is used differently by different researchers and is applied to many different types of sequences: from short collocations and binomials, through longer multiword expressions, to non-compositional idiomatic expressions (see Wray, 2008 for in depth summary of this research).

While an exact definition of formulaic language is hard to come by (see Ellis, 2012; Wray, 2013), a commonly held assumption is that frequently co-occurring sequences are stored and retrieved as a holistic whole. This assumption is made explicit in Sinclair's Idiom Principle (Sinclair, 1991), and is also found in more current formulations. To take an example, one of the most cited definitions of formulaic language is the one provided by Alison Wray in her seminal book on formulaic language: "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, *stored and retrieved whole from memory at the time of use*" (Wray, 2002:9).

The claim that certain multiword sequences are stored and retrieved as wholes is common in the literature, but it is not clear what evidence supports it (see Siyanova-Chanturia & Martinez, 2014; Siyanova-Chanturia, under review). What does it mean to be stored holistically? Such a definition implies that the sequence has a lexical representation that is independent of its parts. To argue for holistic storage, we would have to show not only that speakers are sensitive to the properties of the sequence (semantic, syntactic, distributional), but also that this knowledge is independent of the properties of the parts. While many studies find a processing advantage for formulaic sequences (see Ellis, 2012; Siyanova-Chanturia & Martinez, 2014 for reviews), this does not mean they are stored holistically. In these studies, formulaic sequences are usually defined as highly recurring sequences (they are formulaic because they are highly frequent sequences). Finding that higher frequency sequences are easier to process means that speakers are sensitive to the distributional properties of sequences (and not just words). However, such findings do not mean that the entire sequence is represented as a holistic whole they tell us nothing about the relation between the parts and the whole. Moreover, multiword frequency effects are found across the frequency continuum, not only for high frequency sequences (Arnon & Snider, 2010), which is unexpected if such effects are a marker of the special lexical status of formulaic sequences.

How then can we capture the intuition reflected in the work on formulaic language that there is something special about frequently occurring sequences? The idea that words can fuse into larger units if they co-occur together often is also advocated in usage-based approaches to language, where such chunking processes are thought to play an important role in language learning, use and change

(Bybee, 1995, 2002, 2010; Bybee & Hopper, 2001). The claim here is that repeated usage of sequences makes them more chunked over time, and that this leads to a processing advantage for the sequence, and to a growing prominence of the sequence relative to the individual words. Importantly, this does not mean that the parts are no longer available, but that there is a weakening of the parts relative to the whole — their effect on use and processing is reduced relative to that of the whole (Bybee, 2002).

The usage-based notion of chunkedness makes two empirical predictions that can be tested by looking at the relative effect of word and multiword frequency on processing. The first prediction is that word frequency will still affect the processing of highly frequent sequences, suggesting that they are not stored holistically. The second prediction is that the effect of the two measures will change across the frequency continuum: the effect of word frequency should weaken for highly frequent sequences (i.e. have less of an effect on processing) while the effect of multiword frequency should increase (i.e. have more of an effect on processing). Taken together, such findings would illustrate the growing prominence of the whole relative to the part with increasing frequency. Interestingly, only two existing studies have looked at the changing prominence of words in highly frequent sequences (Kapatsinki & Radicke, 2009; Sosa & MacFarlane, 2002), both have serious limitations, and neither examined the interplay between word and multiword information across the frequency range.

Sosa & MacFarlane (2002) used a word-monitoring task to examine speakers' ability to detect the function word *of* in two-word collocations of varying frequency (e.g., *because of*, *kind of*). They found that speakers were slower and less accurate in the highest frequency bin and took this as evidence that high frequency collocations are stored holistically. However, because the test items were naturally produced utterances taken from the Switchboard corpus, and were not controlled for the phonetic duration of the test word (or for any other acoustic factors), detection could have been slowed down simply because the duration of *of* was shorter in the higher frequency sequences. In a more controlled study, Kapatsinki & Radicke (2009) asked participants to detect the word *up* in verb-particle collocations of varying frequency (e.g., *give up*, *throw up*, *sign up*). They found a u-shaped pattern: detection got faster as the particle became more predictable and then slowed down in the highest frequency bin. These findings show the predicted strengthening of the bigram information. However, since they only looked at one particle, the findings cannot inform us of the relation between word and sequence frequency. Moreover, since they only looked at two-word sequences, the findings could reflect word-to-word transitional probabilities rather than multiword information.

In this paper, we look at the effect of word and multiword frequency on the phonetic duration of words in spontaneous speech to ask whether the relation between word and multiword information changes across the frequency continuum. If highly frequent sequences are stored holistically, then the effect of word frequency should disappear. If alternatively, increased sequence usage causes a change in the prominence of word and multiword information, we should see reduced effects of word frequency, and increased effects of sequence frequency for higher frequency sequences. There are several advantages to looking at duration in spontaneous speech. First, both word and multiword frequency effects have already been found in this domain, but the interplay between them has rarely been examined (see the following subsection). Second, using spontaneous speech gives us a large number of sequences to look at, a large frequency range to examine, and avoids the possible confounds of having people produce language out of context in the lab (e.g., Bresnan et al., 2007).

Previous Findings on the Effect of Multi-Word Frequency on Phonetic Reduction

Many studies have looked at factors affecting word duration. Phonetic duration is affected by properties of the word itself: words are phonetically reduced when they are more frequent (Bell et al., 2009; Gahl, 2008), have a higher neighborhood density (Gahl et al., 2012), and are repeated in a given conversation (Bell et al., 2009). Word duration is also strongly affected by contextual factors: words are reduced when they appear in more predictable semantic and syntactic environments (Jurafsky et al., 2001; Gahl & Garnsey, 2004; Tily et al., 2009), as well as when they are more predictable given the immediately preceding and immediately following word (Bell et al., 2003, 2009; Jurafsky et al., 2001; Gahl et al., 2012).

Fewer studies have looked at the effect of multiword frequency (as opposed to bigram frequency) on phonetic duration. In a seminal study, Bybee & Schiebman (1999) found that *don't* was phonetically reduced in frequently recurring phrases like *I don't know*. Ayelett & Turk (2004) showed that higher trigram syllable frequency leads to shorter duration in a set of items that included (but was not limited to) three monosyllabic word sequences. Until recently, the few studies that have explicitly asked whether duration is affected by multi-word frequency (as opposed to bigram frequency) have yielded mixed results. A repetition study done with children found an effect of multiword frequency on duration (Bannard & Matthews, 2008) while two elicitation studies with adults did not find such effects (Ellis, Simpson-Vlach & Maynard, 2008; Tremblay & Tucker, 2011). This

pattern is surprising given the propensity of multiword frequency effects on other aspects of language use (recognition speed, voice onset, omission rates).

In a recent study we examined the effect of multiword frequency on duration in elicited and spontaneous speech, and found an effect for both speech types (Arnon & Cohen Priva, 2013). Unlike the previous studies, we used a large sample of spontaneous speech (over 4000 trigrams) and controlled carefully for part frequencies. We found that duration is affected by multiword frequency and that syntactic constituency does not modulate the effect: duration was shorter for more frequent trigrams within and across syntactic boundaries. These findings show that duration is affected by multiword frequency, and raise several additional questions. First, because we looked at the duration of the entire sequence, we do not know if all the words in the trigram were affected by sequence frequency. In particular, we cannot rule out the possibility that the effect we saw was driven by the final word and consequently reflects the last word's increased predictability and not the frequency of the whole trigram. Second, we want to look at the relation between multiword frequency and the predictability of the words given the preceding and following words, two measures that are known to affect duration (e.g., Bell et al., 2003) but were not examined in our previous study.

The Current Studies

In the current paper, we look at the effect of multiword frequency on phonetic duration with several goals in mind. Our first goal is to validate and extend our previous findings. First, we want to ensure that the effect was not just carried by the predictability of the final word. To do so, we look at the effect of trigram frequency on the duration of a single word — the middle word in a trigram. Second, we ask if the effect of sequence frequency persists when we take into account various predictability measures, which were not included in previous studies. We look at predictability given the preceding word, the following word, and given both the preceding and following word — a measure that has only been found previously for function words (Bell et al., 2003, 2009). Our second goal is to go beyond previous findings and look at the prominence of word and multiword information across the frequency continuum as a way of examining the characteristics of formulaic sequences, defined here as highly frequent sequences.¹ Specifically, we want to test the predictions that the impact of multiword information

1. We follow other studies in using frequency as a proxy for formulaicity (e.g., Underwood et al., 2004), though we do not think this is the only factor to affect the formulaicity of a sequence. We return to this topic in the General Discussion.

will increase, and the impact of word information will decrease (but not vanish) for highly frequent sequences.

We examine these questions in three studies. In the first study, we ask if the effect of multiword frequency on duration persists when looking at the middle word of a trigram. In the second study, we ask if the effect remains significant when controlling for a host of predictability measures known to affect duration. In the third study, we examine the prominence of word and multiword information for highly frequent sequences compared to less frequent ones.

Materials and Method

For all the reported studies we used the same sample extracted from the Buckeye Speech Corpus (Pitt et al., 2007). This corpus contains conversational speech recorded from 40 different native speakers of American English (size: ~300,000 words). The recordings are orthographically and phonetically transcribed and are time-aligned enabling us to align each word and phoneme with their durations (see Cohen Priva 2012; §2.7 for more details). We aligned each word with its duration and syntactic properties. Each word was matched with (a) its part-of-speech tag taken from the Buckeye corpus, (b) its duration in the corpus, and (c) its number of phonemes and syllables using the Carnegie Mellon University pronunciation dictionary (CMU dictionary, Weide, 1998).² We calculated rate of speech as the number of lexical segments per second (including omitted and reduced segments). To control for phrase-final lengthening, we had to identify which of the sequences appeared at an end of an utterance. Each sequence was matched with its location in a prosodic phrase and coded as either phrase-final or not phrase-final: phrases were considered to be continuous speech, allowing gaps of less than 0.1 seconds.

We extracted the trigrams of interest in the following way. First, we extracted all the words that were tagged as nouns (part of speech NN or NNS), had a CMU dictionary transcription, and were preceded and followed by same-utterance words (27,202 words). We restricted our selection to content words in general, and to nouns in particular, since both word type (content vs. function) and syntactic category are also known to affect duration (Bell et al., 2003, 2009). We chose nouns rather than other content word categories because nouns are the most frequent and most diverse word category (number of types) in the Buckeye corpus (and in English more generally). We then extracted the preceding and

2. Number of syllables was estimated as the number of vowels in the word.

following word to create three-word sequences where the variable of interest is the duration of the second word.

We excluded sequences that contained filled pauses and interjections (part of speech UH in the Buckeye corpus, e.g. *um, uh, oh, yeah*), as well as sequences that appeared at the end of an utterance (to remove phrase-final lengthening effects). Because we are interested in the effect of part and sequence frequencies on word duration, we removed sequences whose bigrams and trigrams appeared only once in the corpus (to make our calculation of conditional probabilities more reliable). We also removed sequences whose bigram and trigram frequencies were extremely high (over 2.5 sds from the mean frequency in the sample) because they consisted of a few items that appeared very frequently in the corpus (*a lot of, kind of*) and we wanted to ensure any effects of frequency we find are not driven by just a few items. This selection resulted in 8294 trigrams. We then also excluded items where whose word duration was 1.5 IQR lower than the 1st quartile, or 1.5 IQR higher than the 3rd quartile (153 observations were excluded due to this criterion). We ended up with 8141 observations coming from 40 different speakers. We used the corpus to calculate the expected duration for each word by summing the expected duration of each sound in the word – calculated using the geometric mean duration of every occurrence of that sound in the corpus. We calculated the log frequencies of every word, two-word, and three-word sequences using the combined Buckeye, Switchboard (Godfrey, Holliman, & McDaniel, 1992) and Fisher (Cieri, Miller, & Walker, 2004) corpora, to compensate for Buckeye's relatively small size. We used word and sequence frequencies from the corpus without smoothing (after removing sequences that appeared only once in the three corpora).

We used mixed-effects linear models to conduct our analyses. In all the reported models we reduced collinearity in two ways: residualization and centering. Word and sequence frequencies are highly collinear. Therefore, sequence log frequencies were residualized based on their respective substrings when those appeared in the model (see full details in Arnon & Cohen Priva, 2013). To reduce collinearity further, we mean-centered all the variables by subtracting the mean of each variable from its values.

All models were evaluated using R (R Core Team 2014). Mixed effects linear regressions were conducted using R's *lme4* and *lmerTest* libraries (Bates et al., 2013; Kuznetsova et al., 2013). For every variable in our studies we report its beta, standard error, *t*-value, and *p*-value as evaluated by the *lmerTest* library. In the tables, we also report for each variable its model-comparison *p*-value and Wald χ^2 -test, estimated using R's *car* library's *Anova()* function (Fox & Weisberg, 2011), which performs Type II model comparisons. Model comparison was performed against unresidualized versions of the models. We report both residualized beta

values and unresidualized model comparisons and their p -values in light of recent criticism of residualization (Wurm & FisiCaro, 2014).

Study 1: The Effect of Word and Multiword Frequency on Word Duration

In the first study, we replicate and extend previous findings on the effect of multiword frequency. In particular, we look at the middle word in the trigram to ensure that the effect of multiword frequency was not driven by the predictability of the last word in the sequence. Before examining the effect of multiword frequency on duration we wanted to make sure our sample shows the basic effect of word frequency on duration that has been found in previous studies (Bell et al., 2003, 2009; Gahl et al., 2012).

Results and Discussion

1a — Word Frequency Affects Word Duration

As a first step, we wanted to ensure our sample (the 8141 extracted items) shows the basic effect of word frequency on word duration, before looking at larger frequency measures. To do this, we looked at the middle words in our trigrams. We used a mixed-effects linear regression to estimate the log duration of the word based on its expected duration. We used several phonological and phonetic controls: speakers' logged speech rate, and the log number of phones in the word (residualized using baseline duration). Because we had a mix of singular and plural nouns, and because morphological factors have also been shown to affect duration (Kuperman et al., 2007; Schuppler et al., 2011) we also added plurality to the model. We had speaker as random effect, and logged word frequency, and plurality (singular vs. plural) as fixed effects. Correlations between every two variables were < 0.3 , and the kappa of the model after random effects were removed was 14 (low).

As expected, higher word frequency led to shorter word duration in spontaneous speech, ($\beta = -0.034$, $SE = 0.002$, $t = -18.675$, $p < 10^{-15}$, see Table 1 for the full model). The significance of word frequency was not dependent on the inclusion of speech rate, expected duration, expected duration given part-of-speech, or number of phones (though their inclusion significantly improved the model, all p -values $< .001$ in model comparisons). Not surprisingly, higher speech rate led to shorter durations ($\beta = -0.576$, $SE = 0.016$, $t = -35.042$, $p < 10^{-15}$) and having a longer expected duration led to longer word duration ($\beta = 0.63$, $SE = 0.009$, $t = 70.507$, $p < 10^{-15}$). Having a higher number of phones in the word led to shorter

Table 1. Regression coefficients and *p* values, effects of word frequency on word duration.

Variable	Estimate	SE	<i>t</i>	<i>p</i>	Model comparison χ^2	Model comparison <i>p</i> -value
Expected duration (cent., log)	0.63	0.009	70.507	<10 ⁻¹⁵	4971.506	<10 ⁻¹⁵
Speech rate (cent., log)	-0.58	0.016	-35.042	<10 ⁻¹⁵	1229.273	<10 ⁻¹⁵
Number of phones (cent. resid.)	-0.11	0.027	-4.082	4.5e-05	16.666	4.46e-05
Word2-Plural	0.024	0.007	3.499	0.00047	12.242	0.000467
Word2-Freq (log, cent.)	-0.034	0.002	-18.675	<10 ⁻¹⁵	348.777	<10 ⁻¹⁵

durations ($\beta = -0.11$, SE = 0.027, $t = -4.082$, $p < 10^{-4}$) probably due to *compression* (see Katz, 2012 for a discussion of compression effects in words). The effect of plurality was also significant — controlling for word length and word frequency, word duration was longer for plural nouns ($\beta = 0.024$, SE = 0.007, $t = 3.499$, $p < .001$). In sum, our sample shows the expected effect of word frequency on duration: higher frequency led to shorter word duration.

1b — Multiword Frequency Affects Word Duration

We now look at the effect of multiword frequency on word duration. Since much of the variability in word duration emerges from purely phonetic and phonological factors and since phonological factors such as number of sounds per word are highly collinear with frequency measures (Piantadosi, Tily & Gibson, 2011), we created a purely phonetic model of word duration by predicting duration using the sum of the geometric mean duration of the sounds that comprise each word, rate of speech, and two length measurements — number of sounds and number of syllables. We then used the model's predictions to create a single control variable — the phonetically based approximation of the word's duration (we label this variable phonetic fit).

We used a mixed-effects linear regression to estimate the log duration of the word based on its expected duration (using our phonetic fit variable that takes phonetic, phonological and speech rate variables into account). We had plurality as an additional variable since it significantly affected duration in the previous model. To look at the effect of the different frequency measures we included the log frequencies of the three words, the residual frequencies of the first and second bigram, and the residual frequency of the trigram. We used speaker as a random

Table 2. Regression coefficients and *p* values multiword frequency and word duration.

Variable	Estimate	SE	<i>t</i>	<i>p</i>	Model comparison χ^2	Model comparison <i>p</i> -value
Expected phonetic duration (cent., log)	0.9	0.012	76.789	<10 ⁻¹⁵	5896.587	<10 ⁻¹⁵
Word2-Plural	0.026	0.006	4.072	4.7e-05	16.584	4.65e-05
Word1-Freq (log, cent.)	0.008	0.001	5.782	7.6e-09	33.432	7.38e-09
Word2-Freq (log, cent.)	-0.036	0.002	-20.642	<10 ⁻¹⁵	426.086	<10 ⁻¹⁵
Word3-Freq (log, cent.)	0.024	0.002	15.994	<10 ⁻¹⁵	255.813	<10 ⁻¹⁵
Bigram1-Freq (log, cent. resid.)	-0.014	0.002	-6.683	2.5e-11	44.664	2.34e-11
Bigram2-Freq (log, cent. resid.)	-0.036	0.002	-17.589	<10 ⁻¹⁵	309.375	<10 ⁻¹⁵
Trigram-Freq (log, cent. resid.)	-0.039	0.004	-9.005	<10 ⁻¹⁵	81.089	<10 ⁻¹⁵

effect (intercept and slope for our variable of interest — trigram frequency). Correlations between every two variables were <0.25, and the kappa of the model after random effects were removed was 6.04 (low).

As predicted, the effect of trigram frequency on the duration of the middle word in the trigram was significant when controlling for part frequencies: higher trigram frequency led to shorter duration trigram frequency ($\beta = -0.039$, SE = 0.004, $t = -9.005$, $p < 10^{-9}$, see Table 2 for full model). The frequency of the word itself also affected duration: higher frequency led to shorter duration ($\beta = -0.036$, SE = 0.002, $t = -20.642$, $p < 10^{-15}$), as did the frequency of the first bigram (word12: $\beta = -0.014$, SE = 0.002, $t = -6.683$, $p < 10^{-10}$) and the last bigram (word23: $\beta = -0.036$, SE = 0.002, $t = -17.589$, $p < 10^{-15}$). The effect of the control measures was as expected. Longer expected duration led to longer actual duration ($\beta = 0.9$, SE = 0.012, $t = 76.789$, $p < 10^{-15}$), and being a plural noun also led to longer word duration (even though word length was controlled for, $\beta = 0.026$, SE = 0.006, $t = 4.072$, $p < 10^{-4}$).

The frequency of the first and the third word had an unexpected, inverse effect on duration: duration was longer when the first and third words were more frequent (first: $\beta = 0.008$, SE = 0.001, $t = 5.782$, $p < 10^{-8}$, last: $\beta = 0.024$, SE = 0.002, $t = 15.994$, $p < 10^{-15}$). Neither effect is dependent on the inclusion of other variables. We believe both effects stem from (a) the relatively skewed distribution of words surrounding the nouns: the most frequent 10 words appearing in both

positions comprise more than 50% of the data, and (b) the fact that many of the words appearing before and after the nouns are function words whose frequency may be less predictive of the duration of the following and preceding word compared to content words. This assumption is supported by the fact that when we exclude sequences whose first word is a function word — the effect of the first word's frequency on the duration of the middle word is as expected: higher frequency words lead to shorter duration.

In this study, we replicate the effect of trigram frequency on duration found in previous studies, and show it is not driven by the predictability of the final word but is present even in the middle word of the sequence. Next we ask how various predictability measures affect duration, and whether trigram frequency is still predictive when those measures are taken into account.

Study 2: The Effect of Multiword Frequency on Duration when Controlling for Predictability Measures

In this study, we want to see if the effect of multiword frequency persists when we control for predictability measures that are known to affect word duration, and which were not controlled for in our previous study. Word duration is affected by the predictability of the word given the preceding and following words (e.g., Bell et al., 2003). The effect of predictability given *both* the preceding and the following word is less clear: it has been found for frequent function words (Bell et al., 2003) but not for content words (Bell et al., 2009). Here, we want to both examine the different predictability measures and ask whether the effect of multiword frequency persists when all three measures (predictability given the previous word, given the following word, and given both the previous and following words) are taken into account.

Results and Discussion

We use the same sample as the previous studies, and like before, we use a mixed-effects linear regression to estimate the log duration of the word based on the expected duration (using our phonetic fit variable that takes phonetic, phonological and speech rate variables into account). We had plurality in the model, and had speaker as a random effect. We included three predictability measures, all of which were calculated using maximum likelihood estimates of the logged probability of the middle word given the context in question: (a) negative log predictability of the word given the previous word, $-\log\text{Pr}(\text{middle word}|\text{first word})$,

(b) negative log predictability of the word given the following word, $-\log\text{Pr}(\text{middle word}|\text{last word})$, and (c) negative log predictability of the word given both the preceding and the following words, $-\log\text{Pr}(\text{middle word}|\text{first and last words})$. All probability measurements were residualized based on the log frequency of word2. The conditional probabilities cannot be estimated alongside the frequencies of their parts (e.g., $\log\text{Pr}(\text{middle word}|\text{last word})$ is equal to $\log\text{Pr}(\text{second bigram}) - \log\text{Pr}(\text{last word})$). Therefore, the log frequencies of word1 and word3 were not included in the model, and were not used for residualization. Correlations between every two variables were ≤ 0.3 , and the kappa of the model after random effects were removed was 7.96 (low).

The effect of trigram frequency on the duration of the middle word in the trigram remained significant when controlling for all three predictability measures: higher frequency sequences led to shorter middle word durations ($\beta = -0.011$, $\text{SE} = 0.002$, $t = -6.116$, $p < 10^{-8}$, see Table 3). The predictability measures themselves also had an effect on duration. Words were shorter when they were more predictable given the following word ($\beta = 0.023$, $\text{SE} = 0.001$, $t = 20.785$, $p < 10^{-15}$), and when they were more predictable given the previous word ($\beta = 0.008$, $\text{SE} = 0.001$, $t = 7.589$, $p < 10^{-13}$). Words were also shorter when they were more predictable given *both* the preceding and following word ($\beta = 0.015$, $\text{SE} = 0.002$, $t = 8.235$, $p < 10^{-15}$), an effect that was previously found only for function words, and which provides converging evidence to the influence of distributional properties larger than bigrams on word duration.

The frequency of the word itself also affected duration: words were shorter when they were more frequent, ($\beta = -0.035$, $\text{SE} = 0.002$, $t = -20.209$, $p < 10^{-15}$). The effect of the control measures was also as expected. Longer expected duration led to longer actual duration ($\beta = 0.91$, $\text{SE} = 0.012$, $t = 76.829$, $p < 10^{-15}$) and being a plural noun also led to longer word duration (even when word length was controlled, $\beta = 0.028$, $\text{SE} = 0.006$, $t = 4.391$, $p < 10^{-4}$).

In sum, the second study found robust effects of multiword frequency on phonetic duration even when the word in question was the middle word of a sequence (and not the final word), and even when controlling for a host of predictability measures. Interestingly, the predictability of the word given its surrounding context had an effect on duration — an effect that was not found for content words previously (Bell et al., 2009). This difference may be due to our more uniform subset of data — we used only nouns whereas previous research used either function words (for which a similar effect was found) or many kinds of content words. Importantly, the frequency of the sequence appears to have an effect on duration beyond the frequency and predictability of the word itself. Having established the effect of multiword frequency on phonetic duration, we can

Table 3. Regression coefficients and *p* values for effect of multiword frequency on word duration, controlling for predictability.

Variable	Estimate	SE	<i>t</i>	<i>p</i>	Model comparison χ^2	Model comparison <i>p</i> -value
Expected phonetic duration (cent., log)	0.91	0.012	76.829	<10 ⁻¹⁵	5902.726	<10 ⁻¹⁵
Word2-Plural	0.028	0.006	4.391	1.1e-05	19.282	1.13e-05
Word2-Freq (log, cent.)	-0.035	0.002	-20.209	<10 ⁻¹⁵	408.411	<10 ⁻¹⁵
Pred (word2 word1) (neg. log, cent. resid.)	0.008	0.001	7.589	3.6e-14	57.588	3.23e-14
Pred (word2 word3) (neg. log, cent. resid.)	0.023	0.001	20.785	<10 ⁻¹⁵	432.029	<10 ⁻¹⁵
Pred (word2 word13) (neg. log, cent. resid.)	0.015	0.002	8.235	<10 ⁻¹⁵	67.817	<10 ⁻¹⁵
Trigram-Freq (log, cent. resid.)	-0.011	0.002	-6.116	1.96e-09	37.405	9.6e-10

now ask whether part and whole information play a similar role across the frequency continuum as a way to ask if they are really stored holistically.

Study 3: The Changing Effects of Word and Multiword Frequency Across the Frequency Continuum

The first two studies replicated and expanded the effect of multiword frequency on duration. In the third study, we test the usage-based predictions about part and whole prominence by looking at the effect of word and trigram frequency across the frequency continuum. If, as suggested in the formulaic language literature, highly frequency sequences become stored holistically, then we would expect not to find an effect of word frequency for highly frequent sequences. If, alternatively, repeated usage leads to a change in the prominence word and multiword information, as suggested in usage-based approaches, we would expect the effect of whole frequency to increase and the effect of part frequency to decrease (but not vanish entirely) for highly frequent sequences. We explore this question by comparing the effect of word, bigram and trigram frequency for items in the highest frequency quartile, to their effect in the entire sample.

Results and Discussion

We used the same sample of 8141 observations used in the previous studies. As in the previous studies, we used a mixed-effects linear regression to estimate the log duration of the word based on its' expected duration (using our phonetic fit variable that takes phonetic, phonological and speech rate variables into account), and had plurality as a fixed effect and speaker as random effect. We used the same methods to reduce collinearity and to assess the significance of the different variables.

To examine the possible change in the effect of word and multiword information on duration for very frequent sequences, we divided our items according to trigram frequency. We defined very frequent as being in the 4th quartile in terms of trigram frequency: there were 2031 such items, with a mean trigram frequency of 39 occurrences per million (in the combined Switchboard, Fisher and Buckeye corpus). This is higher than the 10-per-million threshold often used as a cut off for formulaic sequences or lexical bundles (e.g., Biber et al., 1999). The range of word frequencies was similar for the very frequent sequences compared to the rest of the sample (highly frequent sequences: words ranged from 11–8277 per million, mean 889. 1st-2nd-3rd-quartiles: words ranged from 0.3–8277 per million, mean 674). This is important since we want to compare the effect of word frequency in the two bins, and need to ensure that there is similar variability in word frequencies.

We examined the way sequence-frequency (as a binary factor: high — in the 4th quartile, low — all the rest) interacted with the logged frequency of the word and the trigram. If the effect of word frequency decreases and the effect of trigram frequency increases for highly frequent sequences, then we expect both interactions to (a) be significant, and (b) go in opposite directions. We use a mixed-effects linear regression to estimate the log duration of the vowel based on the expected duration (using our phonetic fit variable that takes phonetic, phonological and speech rate variables into account). We had plurality in the model, and had speaker as a random effect (intercept and slope for our variable of interest). The model had the log frequencies of the three words, the residual frequencies of the first and second bigram, and the residual frequency of the trigram. It also had two interaction terms: between sequence-frequency (high vs. the rest) and logged word frequency, and between sequence-frequency (high vs. the rest) and trigram frequency. Correlations between the non-interaction terms were <0.3 , and the kappa of the unresidualized model was 8.52 (low).

As predicted, both interactions were significant, and went in opposite directions. The effect of word frequency on duration was *smaller* for highly frequent

Table 4. Regression coefficients and p values, comparing word and multiword information for highly frequent sequences.

Variable	Estimate	SE	t	p	Model comparison χ^2	Model comparison p -value
Expected phonetic duration (cent., log)	0.9	0.012	76.715	<10 ⁻¹⁵	5893.851	<10 ⁻¹⁵
Word2-Plural	0.024	0.007	3.627	0.000289	14.019	0.000181
Word1-Freq (log, cent.)	0.008	0.001	5.796	7.06e-09	66.017	<10 ⁻¹⁵
Word2-Freq (log, cent.)	-0.038	0.002	-19.401	<10 ⁻¹⁵	50.387	1.26e-12
Word3-Freq (log, cent.)	0.025	0.002	15.99	<10 ⁻¹⁵	375.674	<10 ⁻¹⁵
Bigram1-Freq (log, cent. resid.)	-0.013	0.002	-5.635	1.81e-08	10.863	0.000981
Bigram2-Freq (log, cent. resid.)	-0.034	0.002	-15.086	<10 ⁻¹⁵	3.625	0.0569
Trigram-Freq (log, cent. resid.)	-0.034	0.005	-6.567	8.7e-09	121.193	<10 ⁻¹⁵
Word2-Freq X 4th-quartile	0.02	0.005	4.017	5.94e-05	18.573	1.64e-05
Trigram-Freq X 4th-quartile	-0.036	0.011	-3.181	0.00148	5.964	0.0146

sequences compared to the rest of the sample ($\beta = 0.02$, $SE = 0.005$, $t = 4.017$, $p < 10^{-4}$, see Table 4). At the same time, the effect of trigram frequency on duration was *larger* for highly frequent sequences ($\beta = -0.036$, $SE = 0.011$, $t = -3.181$, $p < 0.01$). In other words, the effect of multiword information increased while the effect of word information decreased. Importantly, the effect of word frequency remained significant ($\beta = -0.038$, $SE = 0.002$, $t = -19.401$, $p < 10^{-15}$) and this was true also when we looked only at the highly frequent sequences ($p < .03$). Word information still affects the processing of highly frequent sequences.

The effect of trigram frequency was still significant across the entire sample ($\beta = -0.034$, $SE = 0.005$, $t = -6.567$, $p < 10^{-8}$), as was the effect of the first bigram ($\beta = -0.013$, $SE = 0.002$, $t = -5.635$, $p < 10^{-7}$) and second bigram ($\beta = -0.034$, $SE = 0.002$, $t = -15.086$, $p < 10^{-15}$). All control variables behaved as expected.

Looking at the effect of word and multiword frequency on duration in highly frequent sequences, and comparing those effects to the ones in the entire sample revealed several interesting patterns. The effect of multiword information increased with repeated usage: word duration was more affected by trigram frequency for

sequences in the highest frequency quartile. At the same time, the effect of word information on duration decreased: word frequency affected duration less for highly frequent sequences. Both findings are consistent with usage-based predictions and illustrate the growing prominence of whole information with repeated usage. The findings also show that part information still plays a role in the processing of even highly frequent sequences.

General Discussion

In a series of three studies, we set out to examine the effect of word and multiword frequency on the phonetic duration of words in spontaneous speech. In the first study, we show that trigram frequency affects the duration of the second word of the trigram in spontaneous speech — ensuring that previous findings were not driven solely by the final word in the sequence. In the second study, we show that the effect of multiword frequency persists when controlling for the predictability of the word given the words around it, variables not taken into account in our previous work. This study also shows that the duration of content words is affected by the predictability of the word given *both* the preceding and following word, a finding previously reported only for function words (Bell et al., 2003, 2009). In the third study, we document a change in the prominence of word and multiword information with growing sequence frequency: for highly frequent sequences, the effect of word frequency on duration decreases while the effect of trigram frequency on duration increases. This novel finding shows that repeated usage leads to a growing prominence of multiword information without obliterating the effect of word information.

These findings join the growing literature on the effect of multiword information on processing. They show that such effects are robust, found across a range of different types of trigrams, and are not only predictability effects in disguise. Theoretically, finding multiword frequency effects across a wide range of linguistic tasks and items further demonstrates the parallels between words and larger sequences, and makes it harder to argue that the two are processed by qualitatively different systems (Pinker, 1999; Pinker & Ullman, 2002). Instead, the findings provide support for a single-system view of language (Elman, 2009; McClelland et al., 2010) where all linguistic experience is processed by a similar cognitive mechanism. The findings also have theoretical and practical implications for how we think about formulaic sequences.

Implications for Studies of Formulaic Language

We looked at the relative effect of word and multiword information on duration as a way to contrast two interpretations of what happens to the representation of sequences when they recur frequently. One view — found in the formulaic language literature — suggests that increased repetition leads to holistic storage for formulaic sequences (e.g., Wray, 2002). Being stored holistically implies that speakers no longer have access to the parts making up the whole. Taken seriously, this claim leads to the prediction that the effect of part frequency will not be present for such sequences. On the other hand, the view expressed in usage-based approaches (e.g., Bybee, 2002) suggests that repeated usage leads to a growing prominence of the sequence relative to the individual words, while still leaving word information available. Importantly, neither interpretation has been tested empirically in a satisfactory way. The few studies that have looked at the relation between word and multiword information in production (Kapatsinski & Radicke, 2009; Sosa & MacFarlane, 2002) did not systematically examine items varying in both word and multiword frequency.

Our findings support the usage-based interpretation of the effect of repeated usage. We found that the effect of multiword information increased for highly frequent sequences while the effect of word information decreased. However, the effect of word frequency remained significant even for highly frequent sequences. It is important to note that our highly frequent sequences had a mean frequency of 40-per-million: four times higher than the ten-per-million that is often taken to indicate lexical storage. These findings demonstrate the increased prominence of multiword information with increased frequency, while indicating that part information is still available.

The claim that formulaic sequences are stored holistically is quite common in the formulaic language literature, but it is not entirely clear what holistic storage entails or what the supportive evidence is (Sivanova-Chanturia, under review). As discussed in the introduction, finding processing advantages for formulaic sequences does not mean that these sequences are stored as a holistic whole. To substantiate the claim, the empirical predictions have to be made clearer: What does it mean to be stored holistically? How are such sequences empirically distinguishable from ones that are not stored holistically? On a more conceptual level, there may not be a need to argue for holistic storage. The intuition that there is something special about the processing and use of formulaic sequences — and the resulting processing and acquisition benefits — can be captured without assuming that they are stored as one holistic unit, simply by allowing the whole to become more prominent with increased usage.

One limitation of the current findings is that we used frequency as a proxy for formulaicity: our formulaic sequences were simply highly frequent sequences. While this is not uncommon in studies of formulaic language (e.g., Biber, 2009; Conklin & Schmitt, 2008), it is clear that other factors beyond frequency play a role in how ‘chunked’ or cohesive a sequence is. Recent studies suggest that a host of measures (and their combination) impact the perceived formulaicity of a sequence — among them mutual information, discourse function, semantic compositionality, and more (Simpson-Vlach & Ellis, 2010; Martinez & Schmitt, 2012; O’Donnell et al., 2013). Looking at the relation between part and whole frequency may contribute to this debate. We can ask if sequences that are defined as formulaic on other dimensions (MI, discourse-function) also display an increased prominence of multiword information relative to less formulaic sequences. We may also be able to use the increased prominence of multiword information as a diagnostic measure: find those sequences that are more affected by multiword information and examine their semantic, syntactic and distributional properties. Another limitation is that we looked only at the highest frequency quartile — further work needs to examine more subtle changes that may occur across the frequency continuum.

Implications for Grain-Size Challenges in Psycholinguistics

The changing relation between word and multiword information is also relevant for the grain-size debate in the psycholinguistic literature. Despite the increase in knowledge about the different distributional measures that affect processing, there has been relatively little work attempting to integrate the different measures in one predictive model. In particular, what is lacking is a principled model of how different frequency measures are weighted in online processing, and how that relation may be impacted by item characteristics (e.g., highly frequent sequences) and task characteristics (e.g., production vs. comprehension). Bayesian models may be one useful tool for exploring these questions because they allow us to model both the information itself and speakers’ confidence about it. How can we explain the growing effect of trigram frequency for more frequent trigrams at the expense of unigram frequency? One possibility is that speakers draw on multiple sources of information for production, but their confidence in the sources differs. A word in isolation is a good source of information when little is known about its surrounding context, but if larger contexts are reliable enough (= frequent enough), then the importance of the word in isolation diminishes. A related approach has been used in statistical natural language processing since the

mid-1990s — where the relative importance of a word's frequency is discounted when it appears in restricted contexts (Kneser & Ney, 1995).

Conclusion

In this paper, we find robust effects of multiword frequency on single word duration when controlling for the predictability of the word given its context. We go on to show that the effect of word and multiword information changes for highly frequent sequences: while the effect of trigram frequency increases, the effect of word frequency decreases. The findings provide further evidence for the effect of multiword information on processing. From the perspective of formulaic language research, they show that repeated usage of sequences leads to a growing prominence of the whole, but that part information is still available.

Author Note

The studies were partially funded by ISF grant 527-12 given to the first author. Both authors contributed equally to this paper. We would like to thank two anonymous reviewers and the editors for helpful comments.

References

- Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth – children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, 7(2), 107–129. DOI: 10.1080/15475441.2010.505489
- Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, 56(3), 349–371. DOI: 10.1177/0023830913484891
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1), 67–82. DOI: 10.1016/j.jml.2009.09.005
- Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1), 31–56. DOI: 10.1177/00238309040470010201
- Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19(3), 241–248. DOI: 10.1111/j.1467-9280.2008.02075.x
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4*.

- Bell, A., Brenier, J., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60, 92–111. DOI: 10.1016/j.jml.2008.06.003
- 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. *Journal of the Acoustical Society of America*, 113, 1001–1024. DOI: 10.1121/1.1534836
- Biber, D. (1999). *Longman grammar of spoken and written English*. Harlow, England; [New York]: Longman.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3), 275–311. DOI: 10.1075/ijcl.14.3.08bib
- Bresnan, B., Cueni, A., Nikitina, T., & Baayen, H. (2007). Predicting the dative alternation. In G. Boume, I. Kraemer & J. Zwarts (Eds), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10(5), 425–455. DOI: 10.1080/01690969508407111
- Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14(03), 261–290. DOI: 10.1017/S0954394502143018
- Bybee, J. L. (2010). *Language, usage and cognition*. Cambridge; New York: Cambridge University Press. DOI: 10.1017/CBO9780511750526
- Bybee, J. L., & Hopper, P. J. (2001). *Frequency and the emergence of linguistic structure*. Amsterdam; [Great Britain]: John Benjamins. DOI: 10.1075/tsl.45
- Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics*, 37(4), 575–596. DOI: 10.1515/ling.37.4.575
- Cieri, C., Miller, D., & Walker, K. (2004). The fisher corpus: A resource for the next generations of speech-to-text. In *Proceedings of the Language Resources and Evaluation Conference*.
- Cohen Priva, U. (2012). *Sign and signal: Deriving linguistic generalizations from information utility*. Stanford University dissertation.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72–89. DOI: 10.1093/applin/amm022
- De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics*, 3(1), 59–80. DOI: 10.1075/ijcl.3.1.04dec
- Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 41, 375–396.
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17–44. DOI: 10.1017/S0267190512000025
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33(4), 547–582. DOI: 10.1111/j.1551-6709.2009.01023.x
- Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks CA: Sage.

- Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3), 474–496. DOI: 10.1353/lan.0.0035
- Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, 80(4), 748–775. DOI: 10.1353/lan.2004.0185
- Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4), 789–806. DOI: 10.1016/j.jml.2011.11.006
- Godfrey, J., Holliman, E., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of ICASSP-92*, 517–520.
- Jurafsky, D., Bell, A., Gregory, M. L., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. L. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam, the Netherlands: John Benjamins. DOI: 10.1075/tsl.45.13jur
- 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. DOI: 10.1075/tsl.83.14kap
- Katz, J. (2012). Compression effects in English. *Journal of Phonetics*, 40, 390–402.
- Kneser, R., & Ney, H. (1995). Improved backing-off for M-gram language modeling. In *1995 International Conference on Acoustics, Speech, and Signal Processing, 1995. ICASSP-95* (Vol. 1, pp. 181–184).
- Kuperman, V., Pluymaekers, M., Ernestus, M., & Baayen, H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *The Journal of the Acoustical Society of America*, 121(4), 2261–2271. DOI: 10.1121/1.2537393
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)*.
- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299–320. DOI: 10.1093/applin/ams010
- McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, 14(8), 348–356. DOI: 10.1016/j.tics.2010.06.002
- Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, 24(6), 469–488. DOI: 10.1007/BF02143162
- O'Donnell, M. B., Romer, U., & Ellis, N. C. (2013). The development of formulaic sequences in first and second language writing. *International Journal of Corpus Linguistics*, 18, 83–108. DOI: 10.1075/ijcl.18.1.07odo
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication*, 191, 225.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9), 3526–3529. DOI: 10.1073/pnas.1012551108
- Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: Harper Collins.
- Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6(11), 456–463. DOI: 10.1016/S1364-6613(02)01990-3

- Pitt, Mark, Dilley, Laura, Johnson, Keith, Kiesling, Scott, Raymond, William, Hume, Elizabeth, and Fosler-Lussier, Eric. (2007). Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu]. Department of Psychology, Ohio State University.
- R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Real, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57(1), 1–23.
DOI: 10.1016/j.jml.2006.08.014
- Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins Publishing. DOI: 10.1075/llt.9
- Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, 39(1), 96–109. DOI: 10.1016/j.wocn.2010.11.006
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487–512. DOI: 10.1093/applin/amp058
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. (2011). Seeing a phrase ‘time and again’ matters: The role of phrasal frequency in the processing of multi-word sequences. *Journal of Experimental Psychology: Language, Memory, and Cognition*, 37(3), 776–784.
DOI: 10.1037/a0022531
- Siyanova-Chanturia, A., & Martinez, R. (2014). The idiom principle revisited. *Applied Linguistics*, amt054. DOI: 10.1093/applin/amt054.
- Siyanova-Chanturia, A. (under review). On the holistic nature of multiword expressions: A word of caution.
- Snider, N., & Arnon, I. (2012). A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon. In S. Gries & D. Divjak (Eds.), *Frequency effects in language* (pp. 127–164). Berlin: Mouton de Gruyter.
- Sosa, A. V., & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word of. *Brain and Language*, 83(2), 227–236.
DOI: 10.1016/S0093-934X(02)00032-9
- Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, 1(2), 147–165. DOI: 10.1515/LANGCOG.2009.008
- 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: The Continuum International Publishing Group.
- Tremblay, A., & Tucker, B. V. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon*, 6(2), 302–324.
DOI: 10.1075/ml.6.2.04tre
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33(3), 285–318. DOI: 10.1006/jmla.1994.1014
- Ullman, M. T., & Walenski, M. (2005). Moving past the past tense. *Brain and Language*, 93(2), 248–252. DOI: 10.1016/j.bandl.2004.10.004

- Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Eds.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 153–172). Amsterdam: John Benjamins.
- Weide, R. (1998). *The CMU pronunciation dictionary, release 0.6*.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge; New York: Cambridge University Press. DOI: 10.1017/CBO9780511519772
- Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: OUP.
- Wray, A. (2013). Formulaic language. *Language Teaching*, 46(03), 316–334. DOI: 10.1017/S0261444813000013
- Wurm, L. H., & FisiCaro, S. A. (2014). What residualizing predictors in regression analyses does. *Journal of Memory and Language*, 72, 37–48. DOI: 10.1016/j.jml.2013.12.003

Corresponding Address

Inbal Arnon
Department of Psychology
Hebrew University, Mount Scopus
Jerusalem, Israel
Inbal.arnon@gmail.com