Title
Phonetic duration of nouns depends on de-lexicalized syntactic distributions:Evidence from naturally occurring conversation

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Authors
Lester, Nicholas A
Baum, Daniel
Biron, Tirza

Publication Date
2018

Phonetic duration of nouns depends on de-lexicalized syntactic distributions: Evidence from naturally occurring conversation

Nicholas A. Lester (nlester@umail.ucsb.edu)
Department of Linguistics, SH 3432
University of California, Santa Barbara
Santa Barbara, CA 93106 USA

Daniel Baum (daniel.baum@weizmann.ac.il) & Tirza Biron (tirza.biron@weizmann.ac.il)
Department of Physics of Complex Systems, 234 Herzl Street
Weizmann Institute of Science
Rehovot 7610001 Israel

Abstract

We explore whether de-lexicalized syntactic information impacts the phonetic duration of nouns. The motivating expectation is that nouns that carry more syntactic information will be more difficult to produce *in situ*, leading to longer durations. We approach this question from two perspectives: pure diversity of a noun's distribution across its available syntactic relations, and distance of this distribution from the average distribution of nouns in the language at large. The former measure is designed to capture the interconnectivity between the lexical and syntactic tiers of linguistic representation. The latter measure targets how well an individual noun fits the behavior expected for the noun class. We find that durations are sensitive to both measures in complementary fashion: nouns with more diverse syntactic distributions are produced with longer durations, and nouns that have distinctive (non-prototypical) distributions have shorter durations.

Keywords: phonetic duration, entropy, syntax-lexis interface, nouns, naturally occurring speech

Introduction

The last two decades have seen a sharp increase in the attention paid to how language use – in particular, the probability of use – shapes linguistic representation and processing. From phonology, to morphology, to words, and syntactic structures, probability has repeatedly surfaced as a significant predictor of performance in a number of experimental and observational paradigms. A common index used to gauge the effects of probability on performance is the time that it takes to articulate a given linguistic unit. One major finding in this line of research is that words that appear in less predictable contexts take longer to produce. This finding has been observed for syntagmatic (collocational), syntactic, phonological, and prosodic contexts (e.g., Arnon & Cohen Priva, 2013; Aylett & Turk, 2004; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Gahl & Garnsey, 2004; Jurafsky, Bell, Gregory, & Raymond, 2001; Kuperman & Bresnan, 2012; Mahovald, Fedorenko, Piantodosi, & Gibson, 2013; Michelas & D'Imperio, 2012; Piantodosi, Tily, & Gibson, 2011; Seyfarth, 2014). These findings have been explained in terms of the informational load carried by these words. Words that are less predictable given their context contribute more information to the signal than words that are contextually redundant (e.g., Aylett & Turk, 2004; Jaeger, 2010). This logic can also be expressed in terms of processing difficulty: informationally heavy (i.e., surprising) units are more difficult to access, and speakers respond to this difficulty by increasing phonetic duration. On the comprehension side, speakers may anticipate that listeners would benefit from slower production of high-information content. Cohen Priva (2017) provides evidence that this response applies throughout the language production process, rather than as a late-stage modulation of pre-selected material. That is, articulation is sensitive to the information carried by multiple levels of representation at multiple stages during speech production.

Syntax and phonetic duration

The relationship between syntax and phonetic duration has been studied most extensively in terms of how the predictability of a word given its immediate context affects efficiency of processing. Studies in this vein argue that speakers attempt to maximize communicative efficiency. One way they accomplish this is to control the flow of information within the signal. When the prior context generates predictions that are not satisfied by the following word, the information carried by that word increases. Faced with this situation, the speaker risks overloading the bandwidth of the communicative channel (e.g., Shannon, 1948), which could lead to a breakdown in message integrity (e.g., disfluency; Jaeger, 2005). One way that the speaker can mitigate these local spikes in information is to pronounce the word with a longer duration. In this way, the information is spread out in time, helping the language processor to handle the heavier burden. On the other hand, when the prior context renders the production of a word redundant, the speaker may contract pronunciation of the word to preserve articulatory effort and promote informative communication (Aylett & Turk, 2004; Jaeger, 2010).

The predictability of words given prior syntactic context has been operationalized in different ways. Many studies have relied on collocation as a proxy for syntactic variability. These studies find that words that are highly predictable given the preceding two or three words tend to have shorter durations than words that are less predictable (Arnon & Cohen Priva, 2013; Aylett & Turk, 2004; Jurafsky et al., 2001). While not all of these studies interpret these as syntactic effects, Harald Baayen and colleagues have recently suggested that surface predictability of this kind may alone suffice for the learning of "syntactic" generalizations (e.g., Baayen, Milin, Filipović-Đurđević, Hendrix, & Marelli,, 2011).

Other studies have looked at the likelihood of a given syntactic continuation based on the prior context, and how this relationship effects the durations of words that instantiate those completions. Gahl & Garnsey (2004) look at verbs that could be completed by either a direct object (e.g., *The woman saw the package*) or a subordinate clause (e.g., *The woman saw the package was from her mother*). When verbs were biased towards the direct-object completion but produced with a subordinate clause, the verbs and following nouns had longer durations than when they were produced as part of the direct-object completion. Similar effects have been replicated for other syntactic alternations (e.g., Jaeger, 2010; Kuperman & Bresnan, 2012; Tily, Gahl, Arnon, Snider, Kothari, & Bresnan, 2009).

So far, we have discussed work that uses estimates of prior probabilities of linguistic units to predict variable phonetic duration in contexts that are favored or avoided within those prior distributions. Other work has demonstrated that prior probabilities can have consistent effects independent of local context. A good example is the baseline probability of words. All else being equal, more frequent words are produced with shorter durations than less frequent words (e.g., Gahl, 2008). Other effects relate to the structure of words. For example, Piantodosi et al. (2011) show that contextually diverse words tend to be shorter (based on *n*-gram models of the aggregate distributions of prior context, as sampled from large corpora; see also Mahowald, et al., 2013, and Seyfarth, 2014). the segmental structure of words can affect duration: words that are built from more frequent segmental *n*-grams or whose segmental structure is more similar to other words in the language are produced with shorter durations (a lower-level replication of the syntagmatic effects discussed above; Arnon & Cohen Priva, 2013; Aylett & Turk, 2004; Fox, Reilly, & Blumstein, 2016).

All of the existing studies which claim to relate the diversity of aggregate syntactic distributions to word durations rely on *n*-gram models at the level of words (i.e., lexical context; e.g., Piantodosi et al., 2011; Seyfarth, 2014). However, lexical contexts are known to carry semantic information (Bullinaria & Levy, 2012). Any measure based on such distributions is therefore ambiguous: should the observed effects be attributed to the syntactic or semantic component of the underlying informational signal? Moreover, such effects could easily be accounted for by the a-syntactic assumptions of the discriminative learning approach (Baayen, et al., 2011). We therefore need some way to measure syntax in the absence of the purely collocational signal, which either introduces a confound, or potentially eliminates the syntactic interpretation in the face of a purely surface-driven explanation from learning.

In the present study, we attempt to address this gap in the literature. We introduce a novel information-theoretic measure of syntactic diversity that can distinguish syntactic from lexical information. If aggregate lexical contexts correlate with phonetic duration when words are embedded in syntactic contexts, then de-lexicalized syntactic information may also play a role in determining contextualized phonetic duration. If so, we expect two things. First, words with diverse distributions, which carry more syntactic information, should have longer average durations (Mahowald, et al., 2013; Piantodosi et al., 2011). From the production side, such words should be more difficult to access and integrate (for similar effects at the level of phonotactics, see Shaw & Kawahara, 2017). From the comprehension side, speakers may intuitively slow production of such words to accommodate comprehension given the general informational load borne by these words (e.g., Cohen Priva, 2017).

We also introduce a measure of the typicality of syntactic distributions. Other work finds that phonological typicality facilitates processing in sentential contexts (Farmer, Mortensen, & Monaghan, 2006). Such effects are commonly reported for lexical processing in isolation, as well (e.g., Meyer & Schriefers, 1991). However, words that activate multiple semantic or syntactic neighbors are known to show interference effects (Cubelli, Lotto, Paolieri, Girelli, & Job, 2005; Meyer, 1996). Therefore, we expect words with more distinctive distributions to have shorter average durations. This effect should arise because words with more prototypical distributions share their distributed syntactic representation with more other words, and so should resonate with many other words, hindering the lexical selection process. We test these predictions for nouns in a corpus of naturally occurring conversation.

Data and Methods

The data for this analysis come from two sources. First, the syntactic distributions of nouns are estimated using the Open American National Corpus (OANC; based on the second release of the full American National Corpus; Reppen, Ide, & Suderman, 2005). The OANC contains approximately 15 million words of speech and writing produced by native speakers of American English. We parse the corpus for syntactic dependencies using the *spaCy* dependency parser (Honnibal & Johnson, 2015). Timing estimates for the phonetic durations of words were collected from a word-level force-aligned acoustic analysis of the Santa Barbara Corpus of Spoken American English (SBCAE; Du Bois, Chafe, Meyer, Thompson, Englebretson, & Martey, 2000-2005). Forced alignment was carried out using the Montreal Forced Aligner (McAuliffe, Socolof, Mihuc, Wagner, and Sonderegger, 2017),

Measures

To measure syntactic diversity and atypicality, we use the dependency grammar formalism as implemented in *spaCy.* Dependency grammar treats syntactic relationships in terms of typed dependency relations between pairs of words (see, e.g., Nivre, 2005). For example, the sentence *A fish swam* would consist of two dependencies: DET(*a*, *fish*) and NSUBJ(*fish*, *swam*), where the first word in each tuple is dependent on the second for its syntactic function within the given relation. We define a syntactic vector space based on the set of dependency relations that appear in the *spaCy* parses, with the caveat that at least one noun has been observed in that dependency relation. However, using dependencies alone conflates two factors: the syntactic relations and the associated non-target lexical items (the other halves of the linked pairs of words). Other work has relied solely on the latter for characterizing syntactic information; however, lexical co-distributions are known to reflect semantics (Bullinaria & Levy, 2012). To avoid this confound, we compute frequency distributions based on tuples of dependency relations and non-target words for each noun. For example, we would measure the frequency of (DET, *a*) and (NSUBJ, *swam*) for *fish* in the previous example. We compute these frequency distributions using stemmed (*lemmatized*) forms of the target nouns. Conditional entropy is formalized as follows (Eqs. 2-5):

$$H(D|W)=H(D,W)-H(W) \quad (2)$$

$$H(D,W)=H(D)+H(W)-I(D;W) \quad (3)$$

$$H(X)=\sum_{i=1} p(X_i)\log p(X_i) \quad (4)$$

$$I(D;W)=\sum_{i=1}\sum_{k=1} p(D_i,W_k)\log\frac{p(D_i,W_k)}{p(D_i)p(W_k)} \quad (5)$$

Eqs. 2-5 state that the information carried by the set of dependencies *D* given the non-target words *W* is equal to the information carried by the joint distribution of dependencies and words minus the information carried by the words alone. If we apply Eq. 2 directly to the observed probabilities found in a corpus, we arrive at the *maximum-likelihood* estimate of the entropy. However, maximum-likelihood estimates are known to be biased: they underestimate the true entropy (Miller, 1955). We correct the component entropies of Eq. 2 prior to subtraction using the estimator proposed by Chao, Wang, & Jost (2013). This estimator, based on the species accumulation curve, has been proposed to be the optimal (i.e., least biased) means for correcting linguistic entropies (Moscoso del Prado Martín, 2016).

To capture the prototypicality of nouns, we apply a different method. We rely solely on the distributional space defined by *D* (i.e., without taking the joint probabilities of dependency relation and non-target word). To our knowledge, there is no straightforward information-theoretic method for computing the distances between conditional spaces.

We first construct a syntactic prototype for nouns based on their distribution across dependencies *D*. We define the prototype as the sum of the syntactic distributions across all noun types. Similar approaches have been adopted elsewhere in the lexical processing literature (Baayen et al., 2011; Milin et al., 2009). We then measure the distance of each noun's distribution from the prototypical distribution. For this purpose, we employ a symmetrical version of the Kullback-Leibler Divergence (KLD; i.e. *relative entropy*), known as the Jensen-Shannon Divergence (JSD; Lin, 1991), which is formalized in Eq. (6-8).

$$JSD(T\|P)=\frac{1}{2}KLD(T\|M)+\frac{1}{2}KLD(P\|M) \quad (6)$$

$$KLD(X\|M)=\sum_i p(X_i)\log\frac{p(X_i)}{p(M_i)} \quad (7)$$

$$M=\frac{1}{2}(T+P) \quad (8)$$

In Eqs. 6 and 8, *T* refers to the distribution of the target noun across the set of dependencies *D*, and *P* refers to the prototypical distribution of nouns across *D*. In Eq. 7, X refers to either *P* or *T*. To account for bias in the frequency estimates, we apply the James-Stein plug-in smooth to the estimated frequencies (Hausser & Strimmer, 2009), which is optimal for cases in which the number of outcomes are known a priori (here, we know the number of outcomes based on the dependencies available to the *spaCy* parser).

Results

To avoid issues of undersampling, we first removed all words with token frequency of less than 100 in the OANC (~7 pMw). We then fit a linear mixed-effect model predicting phonetic duration (seconds) as dependent variable. These durations showed a strong positive skew. Such skew violates the assumptions of the linear model. A Box-Cox power analysis suggested that a logarithmic transformation of the durations would best approximate normality. In addition, several points (*n* = 168) were identified as outliers, defined as 1.5 times the interquartile range. Computing the model with these points resulted in strongly non-normal residuals, indicating problems with model fit. Removing these points produced normally distributed residuals, suggesting that this trim was necessary.

Besides our critical measure of syntactic diversity and aytpicality, we include a number of control variables that can affect the phonetic duration of words[1]:

[1]Lexical measures of the phonological structures of words come from the Irvine Phonotactic Online Dictionary (IPhOD; Valden, Halpin, & Hickok, 2009). For all variables, we used the unstressed variants (correlation between stressed and unstressed versions of all variables was greater than .95 for the nouns in this sample). Further we only used the raw measures (unweighted by corpus frequency).

Lexical variables

- log frequency
  (based on the SUBTLEX-US subtitle corpus)
- length in phonemes
- phonological density (PLD20)
- average biphone probability
- average positional segment probability
  (adjusted for length)
- inflectional entropy[2]

Repeated Measures

- length in seconds of the intonation unit (IU) in which the noun was observed
- relative position of the word within the intonation unit (a proportion: 0 = first word, 1 = last word)
- speech rate within the intonation unit
  (number of words divided by length in seconds)

The total set of variables is highly collinear (κ>40), which can negatively impact the reliability and interpretability of parameter estimates in linear models. We therefore decorrelate the predictors using Independent Component Analysis (ICA). ICA allows us to produce a set of fully syntactically decorrelated components. The original variables load either positively or negatively and with different magnitudes for each component. These loadings allow us to interpret the components relative to the original variables. Prior to applying the ICAs, we used z-scores to transform each variable to put them on comparable scales. We applied two ICAs. The first was based on the purely lexical variables (frequency, phonological length, phonological density, biphone probability, positional segment probability, and our critical measures) using the unique sample of words (i.e., without repetition based on the SBCSAE sample). The second was based on the repeated measures per word type (IU length, IU position, and speech rate). For each ICA, we produced as many components as ther were variables, yielding eleven total variables. This procedure reduced collinearity of the entire dataset almost entirely (κ<4). We entered these eleven variables as surrogates for the raw measures in the final model.

Table 1 reports the results of the model for the ICA components. The loadings are approximated by the labels: +/- reflect the direction of correlation between the raw measures and the component scores. Parameter estimates (β**)** and errors (SE) are given in the transformed scale.

Table 1: Parameter estimates and significance

Predictor	**β**	**SE**
positional segment prob (+) phonological density (+)	-0.06***	0.01
phonological density (+) length in phonemes (-) biphone probability (-) positional segment prob (-)	-0.10***	0.01
biphone probability (+)	-0.04***	0.00
length in phonemes (+)	0.07***	0.00
inflectional entropy (+)	0.01*	0.01
speech rate (+)	-0.05***	0.00
IU length (+)	0.04***	0.00
relative IU position (+)	0.08***	0.00
atypicality (+)	-0.02***	0.01
freq and diversity (+)	-0.02***	0.01
freq (-) vs. diversity (+)	0.02***	0.01

For reasons of space, we only comment in detail on those components that correspond to the critical predictors[3]. All other predictors were significant, and in the expected direction: longer words took longer, but words with higher (log) frequency, higher average bigram probability, greater average positional probability, and more dense phonological neighborhoods were produced faster. One unexpected point was that variables related to segmental probability and length were contrasted with phonological density: when density is low (the word is dissimilar from other words), biphone probability and positional probability reverse sign to slow production. Inflectional entropy also exhibited the expected effect. The more information carried by the inflectional paradigm, the slower the word was produced. Likewise, faster speech rates, shorter IUs, and earlier positions within IUs all correlated with faster production.

Syntactic atypicality loaded positively within its own component (other variable loadings were near zero). This component correlated negatively with word durations: the more atypical the prior syntactic distribution, the faster the word was produced.

Syntactic diversity surfaced in two components. In the first of these, both diversity and (log) frequency loaded positively and at comparable magnitudes. This component therefore captures the intuition that more frequent words

[2]Inflectional entropy measures the information carried by the distirbution of nouns between plural and singular instantiations of the nouns. We compute the frequency distribution of each noun as it occurs in either its singular or plural form, and take the Shannon entropy of that distribution (negative sum of the weighted log probabilities of singular and plural). We include this measures to ensure that the syntactic measures we take over lemmas (which ignore number distinctions) are not reducible to morphology.

[3]Given the relative complexity of the relationships expressed by the component loadings, and the lack of space, we forgo reporting specific *F* statistics for the signifcance tests.

will have a greater chance of occurring in a greater number of syntactic dependencies. In other words, this is the information expressed by syntactic diversity that is redundantly carried by lexical frequency. The component correlated negatively with articulation times: words that are simultaneously frequent and diverse are produced faster. But this not the whole story. The final component that was related to the critical variables contrasted frequency (positive loading) and diversity (negative loading). The correlation to production time was positive, which means that (a) frequent words are produced faster when they are not syntactically diverse, and (b) syntactically diverse words are produced slower when they are not particularly frequent. Thus, while frequency – with or without syntactic diversity – speeds production (as expected), syntactic diversity without frequency slows production.

Discussion

The results from diversity and prototypicality present a coherent picture of the effect of prior syntactic distributions on variable production durations of nouns. Increased diversity correlated with longer durations. This fits with the notion that words that bear a higher informational load are more difficult to retrieve and prepare for production. A similar effect was observed for inflectional entropy. These findings provide strong support that morphosyntactic diversity – fully divorced from lexical context – is predictive of word durations. We also found that nouns that diverged more in their syntactic distributions from the average noun were produced with shorter durations. Crucially, these effects held irrespective of a number of other predictors known to impact speed of articulation. Taken together, these findings suggest that information introduced by choices made at both the morphological and syntactic levels has consequences for downstream processes in the production pipeline (i.e., articulation).

We interpret the diversity effect as a straightforward replication of prior work that has found longer durations for greater informational content (e.g., Piantodosi et al., 2011; Mahowald, et al., 2013). But why should these nouns be difficult to process? Perhaps they spread activation to a greater degree within the syntactic network. Crucially, this spreading activation would need to be automatic, even when words are selected to be integrated into particular syntactic contexts. Thus, when one wants to embed such nouns into a syntactic structure, one must tamp down a greater number of semi-active but incompatible syntactic frames to arrive at the desired syntactic linkage. By this account, speakers compensate for increased production cost by stretching their articulation of the problem word.

Turning to prototypicality, nouns that behave similarly to the greatest proportion of other nouns in the syntactic space are more likely to be pronounced with longer durations. This effect is expected if syntactic representations are shared across nouns. Prototypical nouns would produce the most potential for interference from other nouns by activating the most densely populated sub-space within the syntax-lexis network. In this case, the interference stems from co-activation at the level of words rather than syntactic dependencies. This interpretation requires that we demonstrate that syntactically similar words indeed behave similarly. Prior work from lexical priming supports this interpretation. Lester, Feldman, & Moscoso del Prado Martín (2017) find that nouns with similar syntactic distributions prime each other independently of orthographic and semantic similarity in visual lexical decision. We leave it to future research to investigate such priming relationships in production.

This research also suggests that there may be more to lexical representation than a history of surface variability as modeled in prior studies of aggregate syntactic context, as well as the *naïve discriminative learning* model of Baayen et al. (2011). We do not deny the suitability of these models; however, our data suggest that other, more abstract levels carry information completely divorced from the surface string.

The most important next step is to pit the measures of syntagmatic and syntactic contextual predictability against our de-lexicalized measures directly in a single model. We are especially interested in testing for interactions between the contextual and de-lexicalized measures. We predict that de-lexicalized diversity and prototypicality exert their strongest effects when *in situ* predictability is at its lowest.

Another important step is to examine the potential for different aspects of the syntactic system to affect duration differently (after Lester & Moscoso del Prado Martín, 2016). More fine-grained analyses might compare headship and directionality of syntactic dependency (is the word bound to another word to the left or right).

References

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

Baayen, R. H., Milin, P., Filipović-Đurđević, D., Hendrix, P. & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*, 438-482.

Bell, A., Brenier, J. M., 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

Bullinaria, J. A., & Levy, J. P. (2012). Extracting Semantic Representations from Word Co-occurrence Statistics: Stop-lists, Stemming and SVD. *Behavior Research Methods*, *44*, 890-907.

Chao, A., Wang, Y. T., & Jost, L. (2013). Entropy and the species accumulation curve: a novel estimator of entropy via discovery rates of new species. *Methods in Ecology and Evolution*, *4*, 1091–1110.

Cohen Priva, U. (2017). Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition*, *160*, 27-34.

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture-word interference paradigm. *Journal of Memory and Language*, *53*, 42–59.

Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., Englebretson, R., & Martey, N. (2000-2005). *Santa Barbara Corpus of Spoken American English, Parts 1-4*. Philadelphia: Linguistic Data Consortium.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, *103*, 12203-12208.

Gahl, S. (2008). “Time” and “thyme” are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*, 474-496.

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language, 80*, 748-775.

Gahl, S. & Garnsey, S. M. (2006). Knowledge of grammar includes knowledge of syntactic probabilities. *Language*, *82,* 405-410.

Hausser, J. & Strimmer, K. (2009). Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *Journal of Machine Learning Research*, *10*, 1469-1484.

Honnibal, M. & Johnson, M. (2015). An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1373-1378). Lisbon, Association for Computational Linguistics.

Jaeger, T. F. (2005). “Optional that indicates production difficulty: Evidence from disfluencies”. *Proceedings of DiSS '05: The Disfluency in Spontaneous Speech Workshop*. France: Aix-en-Provence, 103–109.

Jaeger, T. F. (2010). “Redundancy and reduction: Speakers manage syntactic information density”. *Cognitive Psychology, 61*, 23–62.

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: John Benjamins.

Kuperman, V. & Bresnan, J. (2012). The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory & Language*, *66*, 588-611.

Lester, N. A., Feldman, L. B., & Moscoso del Prado Martín (2017). You can take a noun out of syntax...: Syntactic similarity effects in lexical priming. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 2537-2542).

Lester, N. A. & Moscoso del Prado Martín, F. (2016). Syntactic flexibility in the noun: Evidence from picture naming. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 2585-2590). Austin, TX: Cognitive Science Society.

Lin, J. (1991). Divergence measures based on the Shannon Entropy. *IEEE Transactions on Information Theory*, *37*, 145-151.

Mahowald, K, Fedorenko, E, Piantadosi, ST, Gibson, E (2013). Info/information theory: speakers choose shorter words in predictive contexts. *Cognition*, *126*, 313-318.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017). Montreal Forced Aligner: trainable text-speech alignment using Kaldi. In *Proceedings of the 18th Conference of the International Speech Communication Association.*

Mel'čuk, I. (1988). *Dependency syntax: Theory and practice*. Albany: The SUNY Press.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*, *35*, 477-496.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *17*, 1146-1160.

Milin P, Filipović Đurđević, D., & Moscoso del Prado Martín, F. (2009). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language*, *60*, 50-64.

Miller, G. A. (1955). Note on the bias of information estimates. In H. Quastler (Ed.), *Information theory in psychology* (pp. 95–100). Glencoe, IL: Free Press.

Moscoso del Prado Martín, F. (2016). Vocabulary, grammar, sex, and aging. *Cognitive Science*, 1-26. DOI: 10.1111/cogs.12367.

Nivre, J. 2005. Dependency grammar and dependency parsing. *Technical Report MSI 05133*, Växjö University: School of Mathematics and Systems Engineering.

Piantadosi, ST, Tily, H, Gibson, E (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*, 3526-3529.

Reppen, R., Ide, N. & Suderman, K. (2005). *American National Corpus (ANC), Second Release*. Philadelphia: Linguistic Data Consortium.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictivity on lexical representation. *Cognition*, *133*, 140-155.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *28*, 379-423.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A. & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language & Cognition*, *1*, 147-165.

Valden, K. I., Halpin, H. R., & Hickok, G. S. (2009). *Irvine Phonotactic Online Dictionary*. Version 2.0. Available from http://www.iphod.com.