You can take a noun out of syntax...: Syntactic similarity effects in lexical priming

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

Laurie B. Feldman (lfeldman@albany.edu)
Haskins Laboratories, New Haven, CT, & Department of Psychology,
State University of New York, Albany, NY, USA

Fermín Moscoso del Prado Martín (fmoscoso@linguistics.ucsb.edu)
Department of Linguistics,
University of California, Santa Barbara, CA, USA

Abstract

Usage-based theories of syntax predict that words and syntactic constructions are probabilistically interconnected. If this is true, then words that occur in similar distributions of syntactic constructions should prime each other. These effects should be fine-grained; even small differences between the syntactic distributions of pairs of words of the same grammatical category should cause variation in priming. Prior research from production suggests that this prediction should hold even in tasks without any syntactic requirement. In this study, we introduce a measure of the similarity between the syntactic contexts in which two nouns occur. We show that this similarity measure significantly predicts visual lexical decision priming magnitudes between pairs of nouns. This finding is consistent with the predictions of usage-based theories where fine-grained similarity of syntactic usages between prime-target pairs affects decision latencies, over and above any effects attributable to semantic similarity.

Keywords: syntax; priming; usage-based linguistics; visual lexical decision; information theory

Background

Lexical priming experiments have a long history in psycholinguistic research. Though the bulk of this research has focused on semantic and orthographic effects, some studies have considered the role of syntax (henceforth *grammatical priming*). Early work looked at the effects of inflectional congruity across word classes. For example, in Serbian, inflected nouns are recognized faster when primed by case-appropriate adjectives (e.g., Gurjanov, Lukatela, Moskovljević, Savić, & Turvey, 1985). More recent work has looked at contextualized reading effects. Nouns and verbs that are biased to occur in congruent syntactic constructions (e.g., direct-object vs. subordinate clause continuations; *I need some coffee/to go to the market*) facilitate processing of later content (Novick, Kim, & Trueswell, 2003). Thus, accessing a noun primes expectations about its syntactic context. Congruity effects have been interpreted as evidence for robust, probabilistic syntactic specifications for lexical items.

The empirical evidence outlined so far is complemented by work in theoretical linguistics. Usage-based linguistic theories argue that *all* facets of grammar, including words and syntactic structures, are potentially interconnected on the basis of one's experience with language (e.g., Diessel, 2015). Let us refer to this position as the *probabilistic network* hypothesis. Results such as those reported by Novick et al. (2003) are easily accounted for under this framework. To use the connectionist metaphor, connections between lexical and syntactic nodes are tuned as a function of their frequency of distinctive co-activation (e.g., Gries & Stefanowitsch, 2004). Stronger connections are processed more efficiently. Further support for this hypothesis comes from work on word production: the probability distributions of words in particular syntactic structures influence picture naming latencies (Lester & Moscoso del Prado Martín, 2016).

Direct, probabilistic relationships between words and syntactic structures are not universally accepted across linguistic models. Many models argue that syntax only enters the lexicon through general categorical specifications (i.e., most generative approaches to syntax). Accordingly, words may have a feature indicating the part-of-speech category to which they belong (noun, verb, adjective, and so on). More recent work in this vein has expanded the syntactic content of the lexicon to include more fine-grained syntactic categories. For example, in current mainstream generativist syntax (the Minimalist Program; Chomksy, 1995), words contain information about the syntactic frames with which they can combine as functional head (sometimes called *subcategorization* or *c-selection*; for a similar approach, see Bresnan, 2001). Crucially, these syntactic specifications represent categorical constraints on the *possible* distributions of words. We will call this the *categorical constraint* hypothesis. Under this account, *probabilistic* relationships are simply not available to the grammar. Any effects of probability are designated "extra-grammatical" (Stabler, 2013) and are instead usually attributed to relationships in other mental systems, such as the Conceptual-Intentional system.

This theoretical distinction leads to different predictions about the nature of grammatical priming. The probabilistic network hypothesis predicts that probabilistic information about the semantic and syntactic similarity of words should produce independent priming effects. The categorical constraint hypothesis predicts that probabilistic effects should only arise for semantic similarity (as the syntactic system does not encode such relationships). We test this contrast using a simple lexical priming paradigm.

Research on grammatical priming has largely relied on syntactic or pseudo-syntactic contexts (e.g., using an adjective as a prime for a noun). However, the predictions of usage-based theory, along with recent evidence from production (e.g., Lester & Moscoso del Prado Martín, 2016), suggest that syntactic information –all of it– should be automatically activated every time a word is accessed. This should be true even when the word is presented in isolation for purposes of the task, as in visual lexical decision (see also Durán and Pillon, 2011). We therefore use a simple overt lexical priming paradigm with visual lexical decision. We restrict our analysis to nouns to guard against intercategorical effects. We predict RTs based on the similarity of semantic and syntactic distributions across a range of words words. The probabilistic network hypothesis would be supported by evidence of priming for similar syntax and semantics, independently. The categorical constraint hypothesis would be supported by priming only in the domain of semantics.

Methods

Data

We used the response latencies contained in the Semantic Priming Project (SPP; Hutchison, et al., 2013). The SPP contains response times and accuracies, along with a host of norming data, that were collected using a visual lexical decision task with overt orthographic priming. On each trial, participants were shown a centered fixation cross for 500 ms, followed by a prime word (all caps) for 150 ms. The prime was followed by a blank screen lasting either 50 or 1050 ms. The target word was displayed (all lowercase) until a either decision was made or 3,000 ms elapsed, at which point the experiment would advance to the next trial.

We used only those trials containing primes and targets that also appear both in the British Lexicon Project (BLP; Keuleers, Lacey, Rastle, & Brysbaert, 2012) and the age of acquisition norming database of Kuperman, Stadthagen-Gonzalez, & Brysbaert (2012). We limit the data in this way to take advantage of the additional lexical controls afforded by these databases. To ensure that all stimuli were understood primarily as nouns, we further limited the trials to include only those in which both prime and target received majority noun tags in the British National Corpus (BNC). In this way, we obtained a dataset consisting 1,305 unique primes and 821 unique targets (a total of 1,670 unique nouns).

Syntactic space

To measure the relationship between the noun-pairs in the syntactic system, we first need to operationalize the syntactic system itself. Decades of research have failed to produce an exhaustive list of the syntactic constructions of English (let alone any other language), and we do not presume to offer such a list here. Instead, we rely on the set of low-level relations as defined within Dependency Grammar formalisms (e.g., Mel'čuk, 1988; Nivre, 2005). Dependency Grammars model only relations (*dependencies*) between pairs of words. These relations are asymmetric: each extends from a *head* (the syntactic and conceptual core word) to a *modifier* (whose syntactic role is contingent on the head). Each dependency is labeled to reflect its syntactic function. For example, *the* and *waffle* in the noun phrase *the waffle* would be bound by the *det* relation, which attaches a determiner (*the*, the modifier) to a noun (*waffle*, the head). Other examples include the *nsubj* relation, which binds a noun (modifier) to a verb (head) as its subject, and the *pobj* relation, which binds a noun (modifier) to a preposition (head) as its object. A further detailed description and discussion of Dependency Grammar formalism is beyond the scope of this study. We adopt the dependency formalism implemented in the *spaCy* parser (Honnibal & Johnson, 2015), one of the fastest and most accurate dependency parsers available.

We define the syntactic space for nouns as the set of dependencies for which at least one noun from our sample of SPP primes and targets has been attested either as head or as modifier. For each noun in our dataset, we extracted all sentences containing that noun from the BNC. We conditioned the search to include only sentences in which the word form was indeed tagged as a noun. Those sentences were parsed using *spaCy*. We then compute the frequency distribution of each noun across the dependencies for which it serves as head or modifier. To increase the reliability of our frequency estimates, we discard vectors for all nouns that occurred in fewer than 100 sentences in the BNC (~1 per million words). The total syntactic space is defined as a vector in which each column reflects one among the set of unique dependencies occurring across all nouns. Finally, we merge the individual frequency distribution of each noun into the total syntactic space, creating a matrix of n rows by m columns, where n = the number of total unique dependency types (46) and m = the number of unique SPP/BLP nouns (1,241). The result is therefore a uniform syntactic space for all nouns, where individual nouns may or may not be attested in each possible dependency. In theoretical terms, we treat these vectors as reflecting the statistical connectivity between each noun and the syntactic structures in which it takes part, as is proposed in the usage-based literature. Psycholinguistic support for this treatment comes from an earlier study showing that these and similar dependency vectors affect processing latencies in noun production over and above the effects of other known factors (Lester & Moscoso del Prado Martín, 2016).

Measuring syntactic similarity

We are interested in the possibility that pre-activation of shared syntactic representations will affect the speed of word recognition. Therefore, we need some measure of the similarity between the syntactic distributions of primes and targets in our behavioral data. Note that similarity in syntactic space outlined above does not reduce solely to shared *types* of dependencies. For example, consider two words, *w1* and *w2*, that occupy the same set of 20 dependency types. Suppose that *w1* and *w2* have roughly equivalent overall frequencies and that those frequencies are distributed equally across the dependency types for both words. In this case, we would call them syntactically similar, and consider the number of overlapping types as an appropriate measure of the strength of their similarity. Now suppose that the two words have similar overall frequencies, but that these frequencies are distributed over complementary sets of the dependencies that they share, such that *w1* has a frequency of 1 wherever *w2* has a frequency >100 and vice versa. In this case, we would call them dissimilar. For this, we need to simultaneously account for shared types, as well their probability distributions. One measure well suited to this task is the Jensen-Shannon Divergence (JSD; Lin, 1991). JSD is a symmetric variant of the Kullback-Leibler Divergence (KLD). The KLD between two probability distributions P and Q is defined in Eq. 1.

$$KLD(P||Q) = \sum_{\mathrm{i}} P(\mathrm{i}) \log \frac{P(\mathrm{i})}{Q(\mathrm{i})} \quad (1)$$

This measure captures the average amount of additional information that one would need in order to recode an event from distribution P as if it belonged to distribution Q. Importantly, $\mathrm{KLD}(P||Q) \neq \mathrm{KLD}(Q||P)$, meaning that one must decide *a priori* in which direction to take the distance. JSD provides a solution to the asymmetry problem by taking the midpoint between the two distributions, then taking the mean distance of the distributions to the midpoint. Formally, JSD is expressed as follows (Eqs. 2 and 3).

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

where

$$M = \frac{1}{2}(P + Q) \quad (3)$$

This measure has the advantage of being both symmetrical [$\mathrm{JSD}(P||Q) = \mathrm{JSD}(Q||P)$] and bounded ($0 \leq \mathrm{JSD} \leq 1$).

JSD measurements depend on *estimates* of the probability distributions of events within a distribution, rather than on their *actual* probability distributions. Maximum-likelihood estimates of information-theoretical measures are known to be biased. To guard against this bias we apply a bias-reducing frequency correction to our syntactic vectors, using the plug-in James-Stein shrinkage estimator (Hausser & Strimmer, 2009).

The methods above provide an operationalization of syntactic similarity between primes and targets. For each prime—target pair in the sample, we compute the JSD between their syntactic vectors. A value of 0 indicates identity; a value of 1 indicates complete independence. According to usage-based theories, (at least the bulk of) syntactic structure is meaningful– that is, directly linked to semantic representations in the same way as words (e.g., Diessel, 2015). This means that any effect we uncover for our measure may actually reflect *semantic* similarity, which is well known to affect priming magnitudes (e.g., Neely, 1991). Moreover, the contrast between the probabilistic network and categorical constraint hypotheses depends on a direct comparison of syntactic and semantic sources of similarity. Fortunately, the SPP contains annotation of the degree of semantic similarity between prime and target, indicated by cosine similarities in Latent Semantic Analysis space (LSA). LSA measures the extent to which words occur in similar texts, with higher cosine values indicating greater similarity (Landauer & Dumais, 1997). We transformed the cosine similarities into distances (i.e., 1-cos).

Figure 1 plots the relationship between the syntactic distances (JSD) between pairs of words as a function of their semantic distances (LSA) values. As one would expect, there is a significant positive (linear)[1] correlation between both measures, meaning that words that are similar in meaning tend to occur in similar syntactic contexts. However, an important feature of Figure 1 is the triangular shape of the variance: words that are very close in meaning vary only slightly in syntactic similarity, while words that are distant in meaning vary more widely. This relationship supports the account of Jackendoff (2013), who argues for the existence of syntactic generalizations (i.e., constructions) that allow structural inheritance among sets of semantically heterogeneous sub-constructions. In other words, nouns that are extremely similar in meaning (e.g., synonyms) will always appear in extremely similar syntactic contexts. However, there is large variability in the syntactic similarities of words with different meaning (or there is large variability in the semantic similarity between pairs of words that appear in very different syntactic contexts). This suggests that syntax and semantics are not as tightly coupled as some would argue (e.g., Goldberg, 1995), and their contributions can indeed be considered separately.

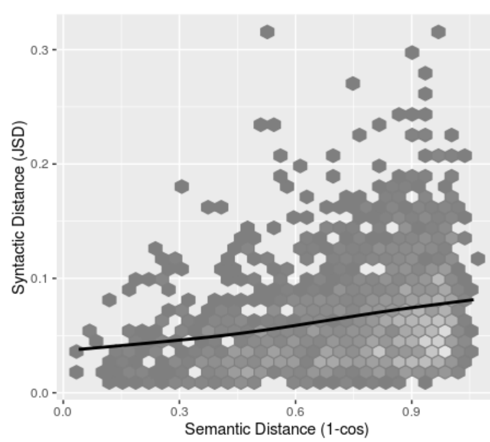

Figure 1: Relationship between syntactic and semantic distance measures

[1] The linear nature of this relation was confirmed using a Generalized Additive Model with penalized spline-based smoothers.

To disentangle the purely syntactic aspects of lexical similarity from what can be attributed to similarity in meaning, we residualized the semantic measure out of the syntactic measure. This was achieved by fitting a linear regression predicting the JSDs as a function of the LSA distances, and using the residuals of this regression as our measure of syntactic difference. This measure captures the information in JSD that is not attributable to semantics (cf., Hendrix, Bolger, & Baayen, 2017; responding to the concerns expressed by Wurm & Fisicaro, 2014).

Further controls

A number of other factors are known to impact recognition latencies in the primed lexical decision paradigm. These fall into three categories: effects related to recognizing individual words, (other) effects based on the relationship between prime and target, and effects related to the nature of the task itself. From the first set, the most important predictor is the surface frequency of the target: i.e., more frequent words are recognized faster. We use the SUBTLEX-UK frequencies, which are based on movie subtitles and known to outperform estimates drawn from other corpora, including the BNC (van Heuven, Mandera, Keuleers, & Brysbaert, 2014). We also include a measure of the density of the orthographic neighborhood of the target known as OLD20 (Yarkoni, Balota, & Yap, 2008). The more similar the spelling of the word to its closest neighbors, the faster it is recognized. Another predictor that has been proposed is age of acquisition: the earlier a word is acquired in the lifespan, the faster it is recognized (e.g., Kuperman et al., 2012). Less important, but nevertheless known to exert an effect, is the orthographic length of the word: longer words take longer to recognize (New, Ferrand, Pallier, & Brysbaert, 2006).

Besides our residualized syntactic measure, we included two additional predictors relating the prime and target: We included semantic distance (i.e., the LSA distances), as semantic similarity is known to facilitate access to targets (i.e., semantic priming). In addition, we considered the Levenshtein distance (LD; Levenshtein, 1966; van der Loo, 2014) between prime and target to account for possible effects of orthographic relatedness. We expect orthographically similar prime-target pairs to result in slower recognition latencies (cf., Adelman, et al., 2014). In addition to these main effects, we tested two-way interactions between the inter-stimulus interval (ISI) on the one hand, and LSA distance, LD, and residualized JSD on the other. This was done to account for the possibility that priming effects might change with the different offsets between prime and target.

Finally, we included the (log) sequential position of each trial in the overall experimental order of presentation. As participants move through the trials, we expected some degree of fatigue to set in (each participant performed over 800 trials).

Results

We fitted a linear mixed-effect regression model predicting response latencies from the SPP primed lexical decision database as a function of the variables outlined above. In addition to the fixed effects, we included random effects for participants and prime-target pairs (i.e., random slopes). We discarded 6.7% of all trials as outliers (all latencies falling below 400 ms or 2 standard deviations above the mean). In addition, we corrected for a strong positive skew in the response times by taking the logarithm of RTS (as suggested by a Box-Cox power analysis; Box & Cox, 1964). Visual inspection of the model residuals with and without the corrections confirmed the adequacy of these steps.

All main effects for the control predictors besides OLD20 surfaced as significant at the α=.05 level, and in the expected direction. The model also revealed a significant (p<.001) effect of the two-way interaction between LD and ISI: at 50 ms ISI, LD had a negative impact on response times (-2.5 ms per unit increase in LD), with no effect at 1050 ms. Importantly, the model revealed a significant interaction (p<.01) between ISI and LSA distance, consistent with what one would expect. Response latencies increased by about 5 ms per .1 increase in cosine distance at a short ISI. At a long ISI, this effect was reduced to ~3 ms per .1 increase. As semantic distance between prime and target increased, so did target recognition latencies, with stronger effects at the shorter ISI.

Over and above the effects of the controls, and crucially over that of semantic similarity, the model revealed a statistically independent significant main effect (p<.001) of the residualized syntactic distance. For every .1 increase in residualized syntactic distance, response latencies were increased by ~4 ± ~3 ms. As predicted by the probabilistic network hypothesis, the less related the prime and target in syntactic space, the longer it takes to recognize the target. There was also a marginal interaction of JSD with ISI (p=.07). The trend resembled that observed for LSA: longer ISIs lead to an attenuated contribution of syntactic similarity. However, given the marginal status of the effect, we do not interpret it further.

Discussion

The present study finds a relatively strong effect of syntactic similarity on lexical priming magnitudes. In fact, the effect was similar in strength –if anything stronger– to that of semantic similarity. To our knowledge, this study is the first to demonstrate that pre-activating a word's syntactic space affects access to that word in a *prima facie* non-syntactic task. This effect provides support for the probabilistic network hypothesis, which predicts that words and syntactic structures are interdependent, and that these connections are forged and tuned by experience. Crucially, these probabilistic relationships are at the core of the grammatical apparatus – they are not simply attributable to the extra-grammatical conceptual system. If that were the case, we should have found no effect of syntactic similarity once semantics was accounted for.

The data we rely on here do not provide us with a non-primed baseline, meaning that we cannot distinguish a facilitation effect of syntactic similarity from an inhibitory effect of syntactic dissimilarity. We therefore leave this question for further research. However, the similarity in shape between the syntactic and semantic effects suggest that syntax –as is argued for semantics (e.g., Lam et al, 2015)– constrains the set of lexical candidates prior to the lexicality judgment. Furthermore, it suggests that syntax, like semantics, is *obligatorily* accessed as soon as lexical forms become active. Crucially, the relationships between words and syntax become active even when (overt) syntactic structure is not built into the stimuli and not really necessary for performing the task. Recent psycholinguistic work on single-word production has echoed this point. For example, Lester and Moscoso del Prado Martín (2016) report chronometric findings suggestive of large-scale interactivation between syntax to lexicon in a bare-noun picture-naming task. Other studies have found that syntactic category information is likewise obligatorily activated in non-syntactic production tasks (e.g., Durán and Pillon, 2011). The present study extends these findings from production to comprehension, from spoken language to written language, and from a simple to a primed paradigm. Hence, the converging evidence suggests that obligatory syntactic access, along with bi-directional activation between syntax and lexicon, is a general, modality-independent property of language processing.

These data also speak to linguistic representation. Branigan and Pickering (*in press*) argue that, in order for priming to take place, some common connection must exist between the prime and target on the one hand, and the representations underlying the measurement of their similarity. This notion is applied to the relationship between words and conceptual content in the semantic priming literature (e.g., Lam, Dijkstra, & Rueschemeyer, 2015). Likewise, our results can be interpreted as reflecting that each noun's representation is explicitly connected to the set of syntactic structures in which it participates and that these representations are shared across words. Moreover, the probabilistic nature of our measure suggests that connection weights –not just the set of shared syntactic types– are represented in the lexico-syntactic network, exactly as predicted by usage-based models of linguistic representation (Diessel, 2015) and as evidenced in sentence-reading paradigms (Novick et al., 2003). Importantly, these findings are *not* consistent with modular-syntactic models (e.g., Chomsky, 1995), which posit only categorical relationships between words and syntax. Adapting the old adage, “you can take the noun out of syntax, but you can't take the syntax out of the noun.”

A possible limitation is that we used Latent Semantic Analysis as a proxy for semantic related when 'cleaning' our syntactic measure of its semantic component. It remains possible –albeit, in our opinion, unlikely– that part, or even all, of the effects of syntactic similarity could be accounted for by a more fine-grained measure of semantic relatedness or similarity than that provided by LSA.

Another possible limitation concerns the morphological structure of the words in our study. While we only included monosyllabic and disyllabic nouns, some of the tokens contained derivational morphology (e.g., *actor*). Morphology is known to interact with priming from other domains (e.g., semantics; Feldman et al., 2015). Therefore, it remains unclear to what extent morphology was contributing to both the shapes of the distributions we computed from the corpus and/or aspects of the priming relationship. In future research, it will be necessary to account for possible derivational relationships between target and prime, and to explore how morphological structure impacts syntactic diversity.

The interaction between our measure and the temporal offset between the prime and the target was only marginally significant. The SPP contains only two such offsets: extremely fast and extremely slow. We suspect that a more robust interaction might arise if one considers offsets intermediate between these extremes. Furthermore, by incrementally increasing the offset between 50 and 1050 ms, we would allow considering the ISI as the numerical magnitude it is (cf., Feldman et al., 2015), rather than as a bi-valued factor.

In sum, our results suggest that, in line with the predictions of usage-based theories of grammar, the representation of words is inextricably tied to the grammatical contexts in which these words are encountered. The results indicate that even the extremely fine-grained differences in syntactic use that can be found between words of a single class (nouns) have detectable effects on their processing and representation. This is true even in tasks –such as visual lexical decision– that do not to involve any explicit involvement of the syntactic system. In other words, in comprehension, the activation of the syntactic properties of a word is automatic. The word comes with its whole syntactic baggage. Furthermore, this syntactic baggage goes well beyond mere grammatical category information, and includes a rich, fine-grained account of the syntactic contexts in which each particular noun is used.

References

Adelman, J. S., Johnson, R. L., McCormick, S. F., McKague, M., Kinoshita, S., Bowers, J. S., Perry, J. R., Lupker, S. J., Forster, K. I., Cortese, M. J., Scaltritti, M., Aschenbrenner, A., J., Coane, J. H., White, L., Yap, M. J., Davis, C., Kim, J., & Davis, C. J. (2014). A behavioral database for masked form priming. *Behavioral Research Methods, 46*, 1052-1067.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistics Society, Series B (Methodological)*, *26*, 211-252.

Branigan, H. & Pickering, M. (*in press*). An experimental approach to linguistic representation. *Behavioral and Brain Sciences*.

Bresnan, J. (2001). *Lexical Functional Syntax*. Oxford: Blackwell Publishers.

Chomsky, N. (1995). *The minimalist program*. Cambridge: MIT Press.

Diessel, H. (2015). Usage-based construction grammar. In E. Dabrowska and D. Divjak (Eds.), *Handbook of Cognitive Linguistics* (pp. 295-321). Boston: De Gruyter.

Duràn, C. P. & Pillon, A. (2011). The role of grammatical category information in spoken word retrieval. *Frontiers in Psychology*, *2*, 1-20.

Feldman, L. B., Milin, P., Cho, K. W., Moscoso del Prado Martín, F., & O'Connor, P. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience*, *11*, 1-19.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar approach to argument structure*. Oxford: Oxford University Press.

Gries, S. Th. & Stefanowitsch, A. (2004). Extending Collostructional Analysis: A corpus-based examination of 'alternations.' *International Journal of Corpus Linguistics*, *9*, 97-129.

GurJanov, M., Lukatela, G., Moskoljević, J., Savić, M. & Turvey, M. T.. (1985). Grammatical priming of inflected nouns by inflected adjectives. *Cognition, 19*, 55-71.

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.

Hendrix, P., Bolger, P. and Baayen, R. H. (2017). Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 128-149.

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.

Hutchison, K.A., Balota, D.A., Neely, J.H., Cortese, M.J., Cohen-Shikora, E. R., Tse, Chi-Shing, Yap, M. J., Bengson, J. J., Niemeyer, D., & Buchanan, E. (2013). The Semantic Priming Project. *Behavior Research Methods*, *45*, 1099-1114.

Jackendoff, R. (2013). Constructions in the parallel architecture. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 70-92), Oxford: Oxford University Press.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*, 287-304.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods*, *44*, 978-990.

Lam, K. J. Y., Dijkstra, T., & Rueschemeyer, S-A. (2015). Feature activation during word recognition: action, visual, and associative-semantic priming effects. *Frontiers in Psychology*, 6, 1-8.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.

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.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Doklady Akademii Nauk SSSR*, *163*, 845-848.

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

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

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theory. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsadale, NJ: Erlbaum.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin and Review*, *13*, 45-52.

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

Novick, J. M., Kim, A., Trueswell, J. C. (2003).Studying the grammatical aspects of word recognition: Lexical priming, parsing, and syntactic-ambiguity resolution. *Journal of Psycholinguistic Research*, *32*, 57-75.

Stabler, E. P. (2013). Two models of minimalist, incremental syntactic analysis. *Topics in Cognitive Science*, *5*, 611-633.

Stefanowitsch, A. & Gries, S. Th. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, *8*, 209-243.

van der Loo, M. P. J. (2014). The *stringdist* package for approximate string matching. *The R Journal*, *6*, 111-122.

Van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*, 1176-1190.

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression models does (and what it does not do). *Journal of Memory and Language*, *72*, 37-48.

Yarkoni,T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's *N*: A new measure of orthographic similarity. *Psychonomic Bulletin and Review*, *15*, 971-979.