Title
Constructional paradigms affect visual lexical decision latencies in English

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

Authors
Lester, Nicholas A
del Prado, Fermin Moscoso

Publication Date
2015

Peer reviewed

Constructional paradigms affect visual lexical decision latencies in English

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

Fermín Moscoso del Prado Martín (fmoscoso@linguistics.ucsb.edu)
Department of Linguistics, South Hall 3432
Santa Barbara, CA 93106 USA

Abstract

Previous research on morphological processing suggests that the probability distribution of a word across its inflected variants influences the recognition of that word. Recently, similar effects have been reported for relations between prepositions and definite-noun-phrase heads in English trigrams (e.g., *in the bucket*). In the present study, we test whether both effects could be accounted for in terms of string proximity and/or semantic similarity alone, or whether the findings for English trigrams should be attributed to syntactic paradigm effects. We 'fake' a case system for English using syntactic positions and prepositions as proxies for the relational meanings expressed inflectionally in other languages. Based on these syntactic factors, we define a syntactic inflectional entropy to parallel the morphological entropy measures used in prior studies. We found that this new measure correlates negatively with visual lexical decision RTs. However, unlike prior studies, we did not find a semantic priming effect between nouns with similar distributions in our paradigmatic vectors. This finding suggests that abstract constructional distributions facilitate lexical access while obscuring semantic relations between similarly distributed words.

Keywords: inflectional entropy; case systems; morphology; lexicon.

Introduction

Many recent studies of morphological processing have investigated how *inflectional paradigms* influence lexical recognition. An inflectional paradigm is the structured set of grammatically licensed word-form variants available to a given word in a given grammatical domain. Earlier work on inflectional paradigms and lexical processing focused on the frequencies of inflected and uninflected variants of words, either together (stem-frequency) or individually (surface-frequency). For example, Baayen, Dijkstra, and Schreuder (1997) showed that visual lexical decision response times for Dutch singulars (unmarked) and plurals (marked with –*en*) were sensitive to different frequencies within the number paradigm. Expanding on these findings, Kostić, Marković and Baucal (2003) showed that, beyond the mere frequencies, the distribution of forms within their inflectional paradigms was of crucial importance. This point was further elaborated by Moscoso del Prado Martín, Kostić, and Baayen (2004) for Serbian. These paradigmatic effects have also been shown to co-determine the order in which children acquire words (Stoll et al., 2012).

Looking beyond the inflectional paradigms of individual lexemes in isolation, Milin, Filipović-Đurđević, and Moscoso del Prado (2009) introduced the additional notion of an *inflectional class*. An inflectional class represents the 'prototypical' or average distribution of word forms within an inflectional paradigm taken across all lexemes that participate in that paradigm. Milin and colleagues showed that the more divergent the inflectional paradigm of a word is from the prototype of the inflectional class, the more difficult it is for people to recognize that word in visual lexical decision.

One question that has yet to be resolved concerns the extent to which these effects are semantic in nature. Kostić et al. (2003) argued that paradigmatic effects could not be fully separated from those of the meanings of the particular affixes, suggesting that semantic and formal properties are ultimately entangled during lexical processing. Moscoso del Prado et al. (2004) provided further evidence of the importance of semantic relations for the representation of morphological paradigms. In the most extreme case, Moscoso del Prado Martín (2007) showed that the effects of a word's inflectional paradigm could not be distinguished at all from those attributable to its positioning in semantic space, as estimated by the variability of contexts in which it was used.

Relatively little attention has been paid to the presence of the above effects in languages with much less pervasive inflectional morphology. However, Baayen, Milin, Filipović Đurđević, Hendrix, & Marelli (2011) report that the paradigmatic relations between English prepositions and their head nouns – defined relative to trigrams of the form *preposition* + *article* + *noun* – affect word recognition latencies in ways similar to those reported for morphologically defined paradigms in inflectionally rich languages. Specifically, they found that the divergence between a noun's distribution across prepositions (its syntactic paradigm) and the overall distribution of prepositions in the sample (the class) correlated positively with visual lexical decision reaction times. Moreover, when trained on the same data, a connectionist model with no intermediary representations for paradigms whatsoever yielded similar results. Baayen and colleagues interpret their findings as evidence that morphological effects arise from the direct association of form and meaning.

A possible caveat needs to be mentioned here. The patterns of usage of the prepositional trigrams do indeed capture *some* of the structural mechanisms by which English encodes information for which other languages would use inflectional morphology. However, in English, sheer co-occurrence at short distances is not the sole way in which relations are encoded. More sophisticated structural factors also encode these relations. Furthermore, it is not clear whether, by examining only the short-distance relations between nouns and prepositions, one is not just tapping into purely semantic factors, which are known to be strongly reflected at such short distances (cf., Bullinaria & Levy, 2007, 2012). If this were the case, the possibility of circularity could arise, in that one would in fact be arguing from semantic effects to semantic effects.

Another question concerns what information must be stored in order to generate the observed paradigmatic effects. Traditionally, frequency effects for linguistic units of a certain complexity (lexical roots, inflected stems, *n*-grams, etc.) that are not reducible to similar effects at lower levels of complexity have been taken as evidence for those units being stored independently in the lexicon. For example, Arnon & Snider (2010) found a facilitatory effect of frequency for compositional 4-word *n*-grams that was not correlated with frequencies of the constituent uni-, bi- or tri-grams. They concluded that we must store frequency information linked to multi-word units, regardless of compositionality (but see Baayen, Hendrix, & Ramscar, 2013, for a replication with an alternative explanation of the findings). This finding meshes well with current usage-based linguistic theories (Goldberg, 2006). However, it also introduces some potential problems. It does not seem realistic to argue that the all instances of all multi-word expressions (at least up to 4 words) are explicitly represented in memory. Recognizing this, Baayen et al. (2011) were able to replicate the paradigmatic effects previously reported for Serbian (cf., Milin et al., 2009) by making reference solely to the orthographic content of the strings (that is, without defining class membership *a priori* and without representations of paradigms). Therefore, frequency effects alone cannot be taken as evidence for explicit exemplar storage.

One feature common to all of the previous studies which have considered the effects of morphological paradigms on lexical processing is their reliance on non-discontinuous strings, either in the form of individual words or fully specified sequences (i.e., *n*-grams). However, if English exhibits anything close to a proper case system, it does so syntactically, through prepositional phrases and privileged positions within the clause. Prepositional phrases may dominate the 'case-relevant' head of a noun phrase (NP) across a theoretically infinite amount of interceding material of any degree of hierarchical complexity (much more than a single determiner, as tested in Baayen et al., 2011). Non-prepositional, syntactic 'cases' (e.g., nominative case) are defined in terms of specialized mother-daughter and sibling relations within the sentence (S) and verb phrase (VP). Therefore, any satisfactory account of paradigmatic processing for English must allow for possibly largely discontinuous *and* purely positional specification of functional roles.

Present Study

In the present study, we effectively 'fake' a case system for English on analogy with the semantically informed paradigms of languages with robust morphological case systems. More specifically, we substitute prepositions and some syntactic positions for the inflectional cases of morphologically rich languages. We expand on the effect reported for phrasally specified, but sequentially contiguous, trigrams of Baayen et al. (2011) by allowing discontinuous dependencies at much greater distances. To do so, we rely explicitly on the grammatical relation between the preposition or syntactic position and the head noun. This provides a closer approximation of the actual distribution of nouns across functional categories and side-steps the possible confound of short-scale semantic effects.

Data

We constructed a semi-arbitrary case system for English loosely based (and expanding) on that of the Balto-Finnic languages (e.g., Finnish). Twenty-two such cases were defined on the basis of prepositional meaning, while two (nominative and accusative) were defined by specific positions within the phrase-structure tree. In addition, the genitive case was linked to both the possessive clitic ***'s*** (as in *John**'s** car*) and to the prepositional phrase with *of*. Some cases have been linked to multiple prepositions in an attempt to capture the variation in meaning ('polysemy') found for those cases cross-linguistically (see Kostić et al., 2003, for evidence that such affixal polysemy is relevant for lexical recognition). The entire fake case system, including the associated English prepositions (or syntactic positions) is presented in Table 1.

We further divided each case into singular and plural to distinguish local, truly morphological effects from syntactic 'fake-inflectional' effects. Including morphological number is important because if, for example, the nouns in our sample were biased in their singular or plural realization (e.g., many singular but few plural instances appearing in one case but many plurals and few singulars in another for a single word), any effect of entropy across the cases might be reducible to the effects of a stem's inflectional paradigm for number. Dividing the cells in this way left us with a total of 48 paradigm cells.

All singular and plural common nouns occurring in each case were extracted from the parsed version of the Open American National Corpus (Reppen, Ide, & Suderman., 2005). This corpus contains approximately 14 million words of written and transcribed-spoken American English which were automatically parsed using a phrase-structure grammar. For the prepositionally defined cases, we took

Table 1: A 'fake' case system for English

Case	Preposition/Position
nominative	NP immediately governed by S node
accusative	NP immediately governed by VP node
adessive	*near, at, beside*
inessive	*inside, in, within*
intrative	*between*
subessive	*under, beneath, below*
superessive	*above, on*
ablative	*from*
allative	*to*
sublative	*onto*
illative	*into*
perlative	*through, along*
prosecutive	*across, over*
instrumental	*with, via, by*
benefactive	*for*
distributive	*per*
genitive	*'s, of*
privative	*without*
semblative	*like, as*
adversative	*against, despite*
essive	*concerning, regarding, about*
co-temporal	*during, throughout*
pre-temporal	*before*
post-temporal	*after*

any terminal noun immediately dominated by an NP which was itself dominated – possibly through an (unbroken) chain of NPs – by a PP whose head was one of the target prepositions. For the syntactically defined cases, we followed a similar procedure. We collected all nouns dominated by an NP (or an unbroken chain thereof) which were dominated by an S (nominative) or a VP (accusative). In this manner, we obtained 1,389,619 tokens.

We then converted all words to lowercase and lemmatized them (i.e., we collapsed singulars and plurals together, so that *cat* and *cats* both counted as an instance of *cat*). This process resulted in the identification of 4,613 unique noun types (i.e., *lemmas*). This list of lemmas was crossed with the 48 case/number-inflected categories, forming a vector whose cells contained the frequency of occurrence of each lemma in each case. We refer to these as *paradigmatic vectors*, which describe the frequency distribution of the grammatical usages of each lemma. These frequencies estimate the probability of usage of each noun in each case/number.

Following Moscoso del Prado, et al. (2004), we computed the entropy (Shannon, 1948) of the probability distribution of the cells in the paradigm. Moscoso del Prado and colleagues refer to this measure as the *inflectional entropy* of the paradigm. Inflectional entropy is the minus average log-probability (i.e., pointwise informativity) of the cells in the paradigm. In our case, given our 48 paradigm cells, this 'fake' inflectional entropy (denoted by H_f) is defined as,

$$H_f(w) = -\sum_{i=1}^{48} p(w_i) \log p(w_i) \ .$$

Moscoso del Prado and collaborators showed that this measure correlates negatively with visual lexical decision reaction times. This has been further confirmed in other studies (e.g., Baayen, Feldman & Schreuder, 2007; Baayen & Moscoso del Prado, 2007).

Notice that the entropy shown above is defined in terms of probabilities. However, our frequency-based measures are *estimates* of the probabilities themselves, not their actual values. These maximum-likelihood probability estimates are biased; they are certain to underestimate the true entropy values (cf., Cover & Thomas, 1991). Crucially, this underestimate will be directly related to sample size (i.e., the lemma frequency). Therefore, using the uncorrected maximum-likelihood estimates will result in a larger underestimation for low frequency words than for high frequency words. This introduces a potential confound between frequency and inflectional entropy effects (a confound that was not addressed in any of the previously published studies using inflectional entropy).

In order to correct the bias of the entropy estimator, we used the James-Stein Shrinkage Estimator (Hauser and Strimmer, 2009).[1] This estimate is particularly well suited to correct entropy underestimates in extremely under-sampled situations, in which the number of possible alternatives (48 in our case) is known *a priori*. As we suspected, whether or not this smoothing was used was of crucial importance for the presence or absence of correlations between frequency and inflectional entropy: the correlations were rather strong with the uncorrected estimates, but disappeared after we applied the smoother.

In order to assess to what extent our fake inflectional measure added independent predictive value over and above morphological inflection, we computed the inflectional entropy in actual morphological terms (i.e., contrasting singular and plural frequencies, as done for instance by Baayen et al., 2007). We denote this 'true' inflectional entropy by H_m (for 'morphological' entropy). H_m was computed by the same method as H_f but with only two cells in the sum (instead of 48).

By pitting this traditional measure against our 'fake' inflectional entropy, we can ensure that a significant effect of the latter is not reducible to number morphology. Further, to normalize the scales of H_f and H_m, we measured both of them on a normalized [0,1] scale by taking the logs to the base of the number of alternatives (i.e., either two or 48). Based on previous studies, we expect that H_m will correlate negatively with visual lexical decision reaction times. In addition, if the purely grammatical factors are also of importance to the noun's representation, we predict that H_f should also correlate negatively with the reaction times, over and above any effects of H_m.

[1] This method outperforms Good-Turing based methods (e.g., Chao & Shen, 2003) in extreme undersampling situations.

Materials

We obtained reaction times from the visual lexical decision data provided as part of the British Lexicon Project[2] (BLP; Keuleers, Lacey, Rastle, & Brysbaert, 2012). The BLP contained response time data for 2,051 nouns for which we had reliable vectors. Vectors were considered reliable if they had a lemma frequency of at least 50 occurrences in our corpus. The resulting dataset consisted of 80,056 trials across 78 subjects. We added token and lemma frequencies for each target from the CELEX database (Baayen, Piepenrock & Gulikers, 1995). The reason to include both frequency counts (which can induce some collinearity in the dataset) was that both frequencies have been found to separately influence reaction times in visual lexical decision. In addition, both have been treated as evidence for a layer of representation at the level of inflectional paradigms (Baayen et al., 1997; Taft, 1979). In order to minimize the impact of the collinearity between both frequency counts, we residualized lemma frequency from surface frequency using a linear model, and used this residual instead of the lemma frequency measure in the later models.

Procedure

We fit a structured additive regression (STAR; Umlauf, Adler, Kneib, Lang & Zeileis, 2012)[3] model with the log-transformed reaction times as dependent variable and the log-transformed token frequency and residualized lemma frequency from CELEX, along with our normalized measures of inflectional entropy H_m and H_f as fixed effects, and lemma and subject as random effects. We considered the possibility of including non-linear smoothers (i.e., penalized splines) for each of the fixed-effect terms. We used the Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin & van der Linde, 2002) to select the most parsimonious combination of smoothers.

Model selection suggested keeping a non-linear effect of surface frequency and simple linear effects for both H_f and H_m. The residualized lemma frequency variable was not found to provide any significant contribution over and above the other predictors, and was thus excluded from the final model. The final model was validated using model criticism.

Results and Discussion

As one would expect, we found facilitatory effects of surface frequency. This frequency effect was non-linear (even in the log scale) in the sense that the strength of the effect was attenuated for the highest frequency words (β=.0002, 95% C.I. [.0000, .0007]). The estimated effect is plotted in the leftmost panel of Figure 1.

Crucially both our main variables of interest, H_m, and H_f, had significant facilitatory effects on the reaction times (H_m: β=-.0209, 95% C.I. [-.0349, -.0079]; H_f: β=-.0766, 95% C.I. [-.1262, -.0261]). The middle and rightmost panels in Fig. 1 plot the estimated effects of both entropy measures. Both effects are clearly significant over and above each other. However, notice that the effect of H_f is much stronger than that of H_m. This is evidenced by the β values for the former being over three times stronger than for the latter, and by the much larger differences in estimated RTs (about 20 ms. for Hf versus some 6 ms. for H_m). It is important to note here that these differences are not due to different scales: both entropy measures range from zero to one.

On the one hand, we have replicated previous results for a negative correlation between inflectional entropy and visual lexical decision response times (Baayen. Feldman & Schreuder. 2007; Moscoso del Prado, et al., 2004). On the other hand, we have extended the results of Baayen et al. (2011) for English prepositional phrases. The more evenly distributed a noun is across our English 'cases.' *regardless of contiguity of the noun and preposition*, the faster it is recognized. Importantly, this new effect is not only much stronger than that of morphological inflectional entropy; it is also clearly separate from it. The truly morphological inflectional entropy neither subsumes, nor is it subsumed by, the new effect of the fake paradigm. This indicates that the nature of the grammatical paradigm effects is to some degree distinct from the morphological effects. Note that number was also included in our new syntactic paradigmatic measure, such that the new measure could have absorbed the whole predictive power of the traditional inflectional entropy. As we will see below, it is not clear that both of these entropy measures are tapping into the same representational levels.

The effects of inflectional entropy within inflectional paradigms are typically interpreted as involving semantic relations. This raises an interesting question: are the facilitatory effects of H_f yet another consequence of the semantic content of the target nouns? The fake inflectional entropy was computed on the basis of the paradigmatic vectors described above. It is possible that these distributions themselves reflect aspects of lexical meaning, as has recently been shown for the equivalent vectors in inflectionally rich languages (Wdzenczny & Moscoso del Prado, 2014). To investigate this possibility, we analyzed whether the degree of semantic priming between pairs of nouns correlates with how similar their paradigmatic vectors are. In addition, we also used word co-occurrence similarity metrics to assess whether our vectors could simply be considered to capture another type of semantically driven word co-occurrence.

[2] Although we are comparing data from a corpus of American English with behavioral data from British speakers of English, we cannot see how the effects we observe could be attributable to dialectal variation.

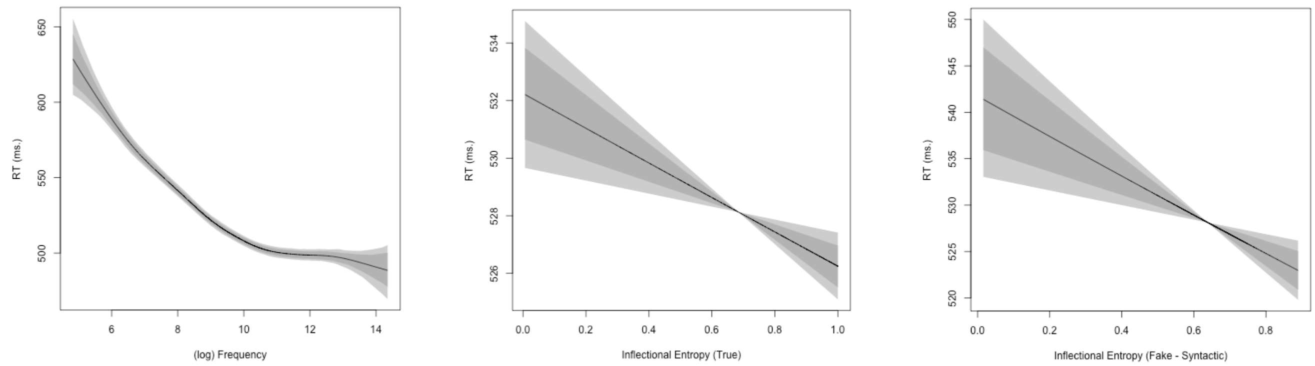

Figure 1: Effects of log surface frequency (left panel), H_m (center panel) and H_f (right panel) on RTs in a visual lexical decision task. Note the different scales of the y-axes, which reflect the different effect sizes of each predictor on RTs.

We collected overt visual lexical decision priming responses from the Semantic Priming Project (SPP; Hutchison, et al., 2013), and a measure of semantic relatedness for pairs of words using Latent Semantic Analysis (LSA; Landauer & Dumais, 1997). We followed Wdzenczny and Moscoso del Prado (2014) in using the Jensen-Shannon Divergence (JSD; Lin, 1991) to assess the degree of similarity between paradigmatic vectors. The JSD is a symmetrical extension of the Kullbach-Leibler Divergence (KLD), useful for summarizing the symmetric point-wise similarity between probability distributions.

We fit a linear mixed effect model with log-transformed priming RTs as dependent variable and token frequency, LSA score, JSD, and stimulus onset asynchrony (200 ms or 1200 ms) as fixed effects. We found no effect of JSD (F<1). However, we did find the usual facilitatory semantic priming effect of LSA similarity ($F[1, 1986.4] = 69.27$, $p < .0001$). These analyses do not support the hypothesis that distances in the paradigmatic space reflect distances in semantic space. Furthermore, we do not even find *any* correlation between distances in paradigmatic (i.e., JSDs) and semantic (i.e., LSA) measures. This null correlation, illustrated in Figure 2, further suggests that our paradigmatic measure is not tapping into the semantics of the nouns.

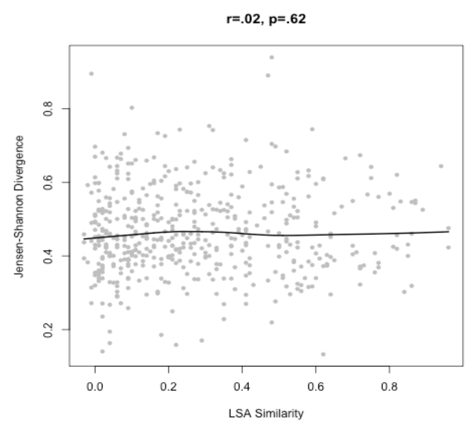

Figure 2: Scatterplot of LSA similarity scores and Jensen-Shannon Divergence values for prime/ target pairs in a primed lexical decision task. The regression line was calculated using a non-parametric smoother.

Conclusions

Using a 'fake' syntactic case system for English, we showed that languages that tend to encode grammatical information on nouns non-morphologically (i.e., syntactically) may also develop paradigmatic relations between words, in this case defined over abstract, possibly discontinuous syntactic distributions. Previous studies have shown that distributional similarities between the inflectional paradigms of nouns reflect semantic similarities between those nouns (Wdzenczny & Moscoso del Prado, 2014; replicated here for English number). However, the grammatical space that we have defined does not seem to correlate with the semantic space occupied by English nouns. Crucially, we do not interpret this finding as evidence for an explicit and autonomous level of representation for lexico-syntactic relations. One reason not to accept such a conclusion is simply that the majority of our cases were not *strictly* syntactic, but also included morphological and lexical information (the former in terms of number contrasts, the latter in terms of the specific prepositional heads used to define cases within the PP frame). Instead, we take these findings as support for the tracking of constructionally (in the sense of Goldberg, 2006) mediated dependencies at quite abstract levels of representation. The distributions of words within and among directly linked networks of constructional templates (including partially lexically specified PP constructions) could generate the orthogonal entropy effect observed here. But if we have found a constructional paradigm effect, which therefore entails the involvement of semantics, why do we not find a lexical priming effect? The answer may be that abstract constructions like PP provide for a wide variety of possible syntactic embeddings – constructional interleavings – both internally and externally. Each layer of the constructional assemblage will contribute some meaning to the whole. The 'semantic noise' produced by such diversity may interfere with the formation of stable semantic associations between lexemes. Work in progress has revealed that increasing the size of the co-occurrence window (operationalized as the distance of the head noun from the preposition) decreases – but does not overturn – the potency of the constructional paradigmaticity effect. These findings suggest that reducing the semantic noise within words' abstract-constructional distributions sharpens their 'paradigmatic resolution,' allowing for stronger semantic associations to develop.

References

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

Baayen, R. H., Dijkstra, T. & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language, 37*, 94-117.

Baayen, R. H., Feldman, L. B., Schreuder, R., (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, *55*, 290–313.

Baayen, R. H., Hendrix, P. & Ramscar, M. (2013). Sidestepping the combinatorial explosion: Towards a processing model based on discriminative learning. *Language and Speech, 56* , 329-347)

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.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database. Philadelphia: Linguistic Data Consortium.

Bullinaria, J. A. & Levy, J. P. (2007). Extracting Semantic Representations from Word Co-occurrence Statistics: A Computational Study. *Behavior Research Methods*, *39*, 510-526.

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. & Shen, T-J. (2003). Non-parametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environmental and Ecological Sciences*, *10*, 429-443.

Cover, T. M., & Thomas, J. A. (1991). *Elements of Information Theory*. New York: Wiley.

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

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.

Hutchison, K.A., Balota, D.A., Neely, J.H., Cortese, M.J., Cohen-Shikora, E.R., Tse, C.-S., Yap, M.J., Bengson, J.J., Niemeyer, D., Buchanan, E. (2013). The Semantic Priming Project. *Behavior Research Methods*, *45*(4), 1099-1114.

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.

Kostić, A., Marković, T., & Baucal, A. (2003). Inflectional morphology and word meaning: Orthogonal or co-implicative domains? In R. H. Baayen & R. Schreuder (Eds.), Morphological structure in language processing (pp. 1-44). Berlin, Germany: Mouton de Gruyter.

Landauer, T. & Dumais, S. (1997), A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Lund, K. & Burgess, C. (1996), Producing high-dimensional semantic spaces from lexical co-occurrence. *Behaviour Research Methods, Instruments, and Computers, 28*(2), 203-208.

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

Moscoso del Prado, F. (2007). Co-occurrence and the effect of inflectional paradigms. *Lingue e Linguaggio*, *2*, 1-16.

Moscoso del Prado, F., Kostić , A., & Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition, 94*, 1-18.

Reppen, R., Ide, N., & Suderman, K. (2005). *Open American National Corpus*. Retrieved from http://americannationalcorpus.org.

Schütze, H. (1992). Dimensions of meaning. *Proceedings of Supercomputing '92*, 787–796.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, *64*(4), 583-639.

Stoll, S., Bickel, B., Lieven, E., Paudyal, N. P., Banjade, G., Bhatta, T. N., Gaenszle, M., Pettigrew, J., Rai, I. P., Rai, M. and Rai, N. K. 2012. Nouns and verbs in Chintang: Children's usage and surrounding adult speech. *Journal of Child Language 39*, 284–321.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, *7*, 263-272.

Umlauf, N., Adler, D., Kneib, T., Lang, S. & Zeileis, A. (2012). Structured Additive Regression Models: An R interface to BayesX. Working Papers in Economics and Statistics No. 2012-12, University of Innsbruck

Wdzenczny ,D., & Moscoso del Prado, F. (2014). Inflectional Affordances: The semantic implications of Alutor and Estonian inflectional paradigms. Presented at the 16th International Morphology Meeting, Buapest, Hungary.