Modeling N400 amplitude using vector space models of word representation

Allyson Ettinger[1] (aetting@umd.edu)
Naomi H. Feldman[1,2] (nhf@umd.edu)
Philip Resnik[1,2] (resnik@umd.edu)
Colin Phillips[1] (colin@umd.edu)
[1]Department of Linguistics, [2]Institute for Advanced Computer Studies
1401 Marie Mount Hall, University of Maryland, College Park, MD 20742 USA

Abstract

We use a vector space model (VSM) to simulate semantic relatedness effects in sentence processing, and use this connection to predict N400 amplitude in an ERP study by Federmeier and Kutas (1999). We find that the VSM-based model is able to capture key elements of the authors' manipulations and results, accounting for aspects of the results that are unexplained by cloze probability. This demonstration provides a proof of concept for use of VSMs in modeling the particular context representations and corresponding facilitation processes that seem to influence non-cloze-like behavior in the N400.

Keywords: N400, vector space models, semantic relatedness

Humans process words in relation to a context. During sentence comprehension, incoming words are processed with respect to mental states elicited by preceding words. How words and their preceding contexts are represented, and how these representations interact during processing, is not yet understood—but questions underlying this puzzle are relevant both to cognitive neuroscience and to computer science. In the present paper we bring together lines of research from both domains to enable explicit modeling of a particular class of context representation and corresponding influence on word processing: the representations that seem to underlie facilitation by lexical semantic relatedness.

In cognitive neuroscience, a measure often used to study the effects of context on word processing is the N400 component, a brain response detectable by the event-related potential (ERP) technique. The N400 is elicited by every word of a sentence, occurring approximately 400 milliseconds after the word is encountered. Its amplitude appears to be modulated by the relation of the word to its context: the worse the fit to context, the larger the N400 amplitude. However, the exact nature of the relation reflected by the N400 is complex and likely varies based on particular circumstances. A widespread generalization is that the N400 amplitude tracks "cloze" probability (Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984), a measure based on the proportion of people who choose a word in a given context during an untimed fill-in-the-blank task. To the extent that N400 amplitudes track cloze probability, we may assume that all information used in an untimed fill-in-the-blank task can exert processing influence very shortly after the context is presented (400 milliseconds after arrival of the next word). This plausibly includes a rich and syntactically-composed representation of the context.

In other cases, however, N400 amplitude does not track cloze probability, suggesting that it does not straightforwardly reflect fit between the incoming word and a fully composed representation of the context. For instance, adding negation to a context should dramatically change likely continuations, but negation has little effect on the N400 (Fischler, Childers, Achariyapaopan, & Perry, 1985) unless additional contextual support is provided (Nieuwland & Kuperberg, 2008). Similarly, in the absence of extended processing time, the N400 appears to be less sensitive to structural information about the agent and recipient of an event (Chow, Smith, Lau, & Phillips, 2015). Specifically, in such cases the N400 amplitude can fail to reflect the low cloze probability of completions such as "A robin is not a bird" (negation violation) or "He forgot which waitress the customer had served" (agent/recipient swap). In these cases, the N400 seems to reflect fit to a less structured context representation: more like general lexical fit, or collective facilitation by semantically related words. Evidence has long indicated that the N400 is sensitive to semantic relatedness facilitation, both by sentence contexts (Kutas & Hillyard, 1980, 1984) and by single-word contexts (e.g. Bentin, McCarthy, & Wood, 1985; Kutas & Hillyard, 1989; Holcomb, 1988; Brown & Hagoort, 1993). Semantic relatedness effects on the N400 are broadly accepted and frequently cited, but to our knowledge no explicit models generating quantitative predictions of these effects have yet been proposed.

Meanwhile, in computer science, explicit models have emerged that allow straightforward computation of relations between words—and by extension, of relations between individual words and groups of words. Vector space models (VSMs), now widely used for natural language processing in computer science, use distributional characteristics of words in text (that is, the types of contexts that words tend to occur in) to form representations for individual words in the form of numeric vectors. A prominent early example of this concept in cognitive science is that of latent semantic analysis (LSA), discussed extensively by Landauer and Dumais (1997). Since the development of LSA, much continued progress has been made in building and optimizing such VSMs.

Once a word is represented by a vector, we can think of this representation as being situated in a multi-dimensional space, and we can compute relations between different word representations based on their orientations or locations in that space. Most common is the use of cosine similarity, a measure based on the angle between vectors. VSMs also allow flexible computation of the relation between a word and a group of words—for instance, using a very simple combination function such as an average, one can generate a single vector representation reflecting characteristics of multiple words, without additional information (such as agent/recipient roles) that would be encoded in a representation arrived at through full syntactic composition. Such a representation could be used to simulate the types of representations hypothesized to underlie non-cloze-like behavior in the N400, when amplitude seems to reflect sensitivity to general semantic relatedness and not to other information present in fully composed representations.

There is one study to our knowledge that finds a correspondence between VSM relations and N400 amplitude: Parviz, Johnson, Johnson, and Brock (2011) assess a suite of variables as possible predictors of N400 amplitude in a non-linear regression model, including cosine similarity from an LSA-type VSM, which they find to be significant. Several studies have also shown correspondence between semantic priming and VSM measures (Mandera, Keuleers, & Brysbaert, 2016; Lapesa & Evert, 2013; Jones, Kintsch, & Mewhort, 2006; Padó & Lapata, 2007; Herdağdelen, Erk, & Baroni, 2009; McDonald & Brew, 2004), suggesting that VSMs can predict cognitive semantic relatedness effects more generally.

In the present paper we build on this foundation to implement a VSM-based model of aspects of sentence processing, intended to test whether a model with simple, averaging-based context representations can successfully simulate non-cloze-like N400 results, as the above reasoning would predict. We choose to model the results of the Federmeier and Kutas (1999) N400 study. This study is one in which certain results deviate from predictions of cloze probability, making it a valuable testing ground for a model intended to capture semantic relatedness effects believed to underlie non-cloze-like N400 behavior. The study is also ideal because it explicitly manipulates relations between target words and their contexts, but bases assumptions about these relations on measures such as cloze probability and plausibility ratings. We use this opportunity to test whether relations computed based simply on collective effects of context words will generate better predictions of the observed results. We show that this VSM-based model captures many major characteristics of the study's N400 results, including the key result that deviates from cloze predictions. The model's relation computations also largely align with the assumptions made by the authors about their stimulus manipulations—with one main exception, which is also the key factor accounting for the deviation from cloze. This suggests that the model has successfully captured aspects of semantic relatedness-based processes exerting influence on the N400, and that access to predictions based on such a model can lend useful perspective in interpreting N400 results.

Federmeier and Kutas (1999)

Federmeier and Kutas (1999) investigated N400 effects in contexts that predict a particular completion word, and are then completed by words with varying levels of similarity to that predicted word. The authors found that unpredicted (zero-cloze) words elicit larger N400s, as expected. However, when the unpredicted item is similar to the predicted word in strongly predictive contexts, the N400 amplitude is reduced.

Table 1: Sample stimuli.

Stimulus (**expected/within/between**)
He caught the pass and scored another touchdown. There was nothing he enjoyed more than a good game of **football/baseball/monopoly**.
The day before the wedding, the kitchen was just covered with frosting. Annette's sister was responsible for making the **cake/cookies/toast**.
He complained that after she kissed him, he couldn't get the red color off his face. He finally just asked her to stop wearing that **lipstick/mascara/earring**.

To accomplish this, Federmeier and Kutas constructed two-sentence contexts with three possible ending types: "expected", "within-category", and "between-category". Expected targets are predicted by the context, with high cloze probability. Within-category and between-category targets are both unexpected in the context—cloze probability of approximately zero—but within-category targets share a category with the expected target.[1] If N400 amplitude were to track cloze probability, then we would see reduced N400 amplitude for the expected target condition, and roughly identical, unreduced N400 amplitude for the two unexpected target types, regardless of category relationship to the expected target.

The stimuli were furthermore binned into two conditions based on the extent to which the context constrained toward the expected word: stimuli were classified as either "high-constraint" and "low-constraint", according to a median split on cloze probability of the expected target.

Figure 1 shows the results of Federmeier and Kutas's study. Negative voltages are plotted upward, with higher N400 amplitude (corresponding to a word that is less expected or facilitated) represented by a greater negativity. In both constraint conditions, we see that the expected target has extremely low N400 amplitude, compatible with strong facilitation. Additionally, in both constraint conditions, between-category targets show very high N400 amplitude, compatible with lack of facilitation. There is also a main effect of constraint level, but the key difference emerges for within-category targets: in high-constraint contexts only, within-category targets show reduced N400 amplitude—despite the fact that within-category targets (like between-category targets) have roughly zero cloze probability. Federmeier and Kutas interpret this result as evidence of a mediating influence of the expected target in high-constraint contexts: strong prediction of the expected target in these contexts causes features of that target to be pre-activated, and because of semantic overlap between expected and within-category targets, the latter targets are facilitated as well.

Federmeier and Kutas's interpretation operates on the as-

[1] Federmeier and Kutas explain that "Categories were chosen to be those at the lowest level of inclusion for which the average undergraduate student could be expected to readily differentiate several exemplars." See Table 1 for examples.

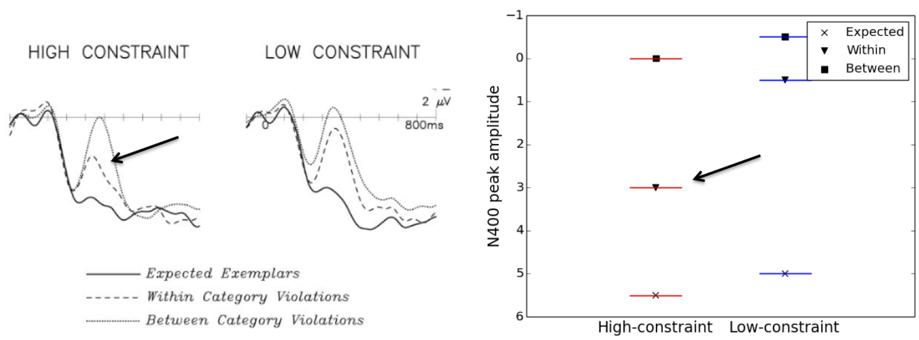

Figure 1: Federmeier and Kutas (1999) N400 results. Left: original results as reported by the authors. Right: Results re-plotted as points representing peak N400 amplitude, for greater ease of comparison to simulation results below. Arrows indicate key facilitation in high-constraint within-category condition.

sumption that high- and low-constraint contexts are equally related to the within-category targets, such that the observed facilitation of within-category targets in high-constraint contexts must be explained by some additional factor. This need motivates their hypothesized mediation by pre-activation of expected target features in high-constraint contexts. An alternative explanation would offer itself if high-constraint contexts were directly more facilitative of within-category targets than are low-constraint contexts. Federmeier and Kutas assume that this is not the case, based on cloze probability measures and plausibility ratings. However, there are other ways that we might conceive of relation to context—in particular, we should consider relations based simply on the collective effect of context words (as opposed to the fully structured and compositional context representations likely to be informing untimed cloze and plausibility decisions). In the next section, with the help of VSMs, we explore whether a notion of fit to context based on collective semantic relatedness can explain the facilitation where cloze and plausibility do not.

Federmeier and Kutas make available a sample of 40 of their experimental stimuli; we run our simulation on that sample.

Model

For testing assumptions and modeling the results of this study, we choose a VSM generated by the word2vec model (Mikolov, Chen, Corrado, & Dean, 2013). Unlike LSA, word2vec uses a neural network to optimize word vectors based on their ability to predict nearby words. In systematic comparisons of VSM performance on various semantic tasks, this model has shown consistently strong and often superior performance (Baroni, Dinu, & Kruszewski, 2014; Levy, Goldberg, & Dagan, 2015). For this reason, we select word2vec as a state-of-the-art VSM of word representations. We train the model on approximately 2 billion words of semantically diverse web data from the ukWaC corpus (Ferraresi, Zanchetta, Baroni, & Bernardini, 2008), training vectors of 100 dimensions using the skip-gram architecture, which maximizes the probability of surrounding words given the current word.

Once we have trained this VSM, each word of the vocabulary is represented as a point within the resulting vector space. For a sentence context, we will refer to vectors for the expected target, within-category target, and between-category target as vectors E, W, and B, respectively.

We model the mental state induced by preceding context words through a simple averaging procedure: vectors for selected context words are averaged to obtain a single context vector C. This representation reflects the collective effect of the included words, without many of the additional structural cues that might inform a cloze decision. In selecting context words, we attempt to isolate the most informative words, which we hypothesize will have the strongest influence upon the context representation. We try two selection methods: *anchored* and *agnostic*.

In the anchored setting, we use relation to the expected target as a proxy for informativeness: using the expected target as an anchor, we select the four context words with highest cosine similarity to that expected target.[2] We employ a minimum cosine similarity of 0.2 (chosen by examination of context word cosine similarities in a small subset of stimuli) to further filter words bearing little relation to the target.

In the agnostic setting, we take the top four words based on negative log frequency (that is, the least frequent words), excluding person names (e.g., *Annette*). This is equivalent to choosing words based on maximum surprisal (information content) as determined by a unigram probability model.

The modeling results suggest *prima facie* that the anchored setting is more successful in isolating the most significant words of the context. If so, this is likely due to the fact that the frequency metric underlying the agnostic setting, while reasonable, is a rather blunt tool for assessing informativeness.[3]

Within these settings, we test two types of average: unweighted, and weighted inversely by linear distance. The latter average aims to instantiate the hypothesis that the effect of a context word would decay over time, with earlier words having less influence than later words.

[2] One target. *polar bear*, is made up of two words; this is represented as the average of the two separate word vectors.

[3] As we caution below, however, at the current stage we should not be overzealous in making fine-grained modeling decisions based on the linear fit of only six datapoints.

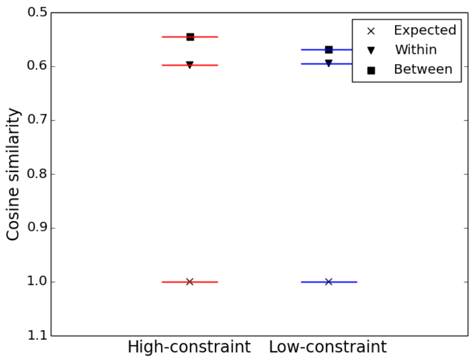

Figure 2: Cosine similarity to expected target

Once we have obtained this context vector C, it can be represented as a point within the space that contains vectors E, W, and B, and its relation to these vectors can be computed. For every stimulus, we take the cosine similarity between C and each of E, W, and B, and we average these cosine similarity values across stimuli within each condition, in order to simulate average N400 amplitude.

We also compute cosine similarity between E and W and between E and B. This allows us to assess the model's representation of the relations between different completion words.

Simulation Results

Figure 2 shows the results of the comparison between target types E, W, and B—this test simply serves as a control, to compare the model's relation computations against those assumed by Federmeier and Kutas, and to check for confounds. In Figure 2 and those that follow, cosine similarity is plotted on the y-axis with the negative direction upward, to facilitate comparison to N400 plots in Figure 1: higher cosine similarity predicts lower N400 amplitude. Note in Figure 2 that the expected word vector E is at cosine similarity of 1, as this is a comparison of a vector to itself. As for the other two comparisons, we see that the model predicts on average a nearly identical level of relation between expected words and within-category words in both constraint conditions. We see a slightly greater distance between the expected word E and the between-category word B in the high- than the low-constraint condition. In both cases the model's relations are roughly consistent with the categorical relations assumed by the experimental manipulation: within-category items are indeed represented as being closer to the expected targets than are the between-category items. The lack of any discernible difference in the expected/within-category target relation between constraint conditions also rules out the possible confound of differing relation strengths between the targets themselves.

Figure 3 shows the full simulations under the anchored and agnostic settings, respectively. (The right hand side of Figure 1 presents Federmeier and Kutas's results in the same plotting format, for ease of comparison.) In these figures we see several things. First, we see a main effect of constraint consistently captured across settings: for each ending type, average cosine similarity to context is higher in the high-constraint condition, corresponding to greater facilitation (lower N400 amplitude). This is consistent with the main effect observed in Federmeier and Kutas's N400 results.[4]

In addition, we see that for the most part, looking independently at the high- and low-constraint conditions, the three ending types pattern as the experimental paradigm predicts: expected targets are most facilitated by the context, while within- and between-category targets are less facilitated. We also see that under all settings, in the high-constraint condition the within-category target falls at an intermediate position between the other two target types. In the low-constraint condition, however, three of the four settings have within- and between-category conditions in reversed or roughly identical positions. The fact that between-category targets in the low-constraint condition fail to fall farthest from the context, often switching with within-category targets, may reflect similar factors to those that lead to within- and between-category conditions having statistically indistinguishable N400 amplitudes in Federmeier and Kutas's results.

Returning to our main effect of constraint: recall Federmeier and Kutas's assumption that facilitation of high-constraint within-category targets cannot be explained by direct relation to context. We see in Figure 3 that the VSM-based representation of context—under both anchored and agnostic word selection settings—does predict greater facilitation of within-category targets in the high-constraint as compared to the low-constraint condition, suggesting that direct relation to context could offer a valid explanation for this deviation from cloze probability. This result both lends support for the explanatory power of our simple non-syntactically-composed context representations, and gives us reason to consider direct facilitation by contextual semantic relatedness as an alternative account for Federmeier and Kutas's results.

Discussion

In this study, we used a vector space model to predict N400 amplitudes observed in Federmeier and Kutas (1999). We find that by representing words in a vector space, averaging vectors of informative context words, and taking cosine similarity measures between the averaged context vector and each of its possible completions, we are able to simulate key aspects of Federmeier and Kutas's N400 results: the basic patterning of expected, within-category, and between-category items within constraint conditions, as well as the main effect of constraint. Our model accounts for the deviation from predictions of cloze probability in the high-constraint condition, and in doing so calls into question the assumption that the key result of this study cannot be explained by a direct facilitation between context words and the within-category targets.

[4]Having access only to 40 items of the original Federmeier and Kutas study, we are not making claims of statistical significance for this pattern of results in the models. This simulation is intended instead as an exploratory proof of concept.

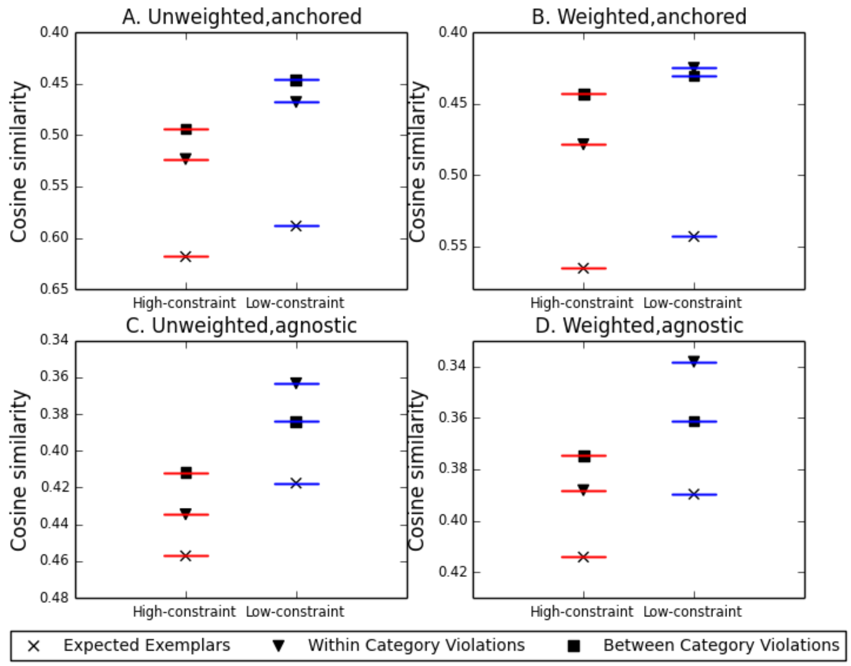

Figure 3: Simulations in four settings. A) Context average unweighted by linear distance and words selected with expected target as anchor. B) Context average weighted by linear distance and words selected with expected target as anchor. C) Context average unweighted by linear distance and words selected by low frequency. D) Context average weighted by linear distance and words selected by low frequency.

At face value, if we assume a linear relation between cosine similarity and N400 amplitude, then Figure 3B is the most faithful simulation of the N400 results. We might take this as evidence in favor of a cognitive model in which activation spreads from informative words (with relation to expected target being a better proxy for informativeness), and in which a word's influence decays over time. However, we caution against drawing strong cognitive conclusions from this single set of simulations. First, we are modeling only six datapoints, without claims of statistical significance. Second, we are for the moment assuming a linear relation between cosine similarity and N400 amplitude, which is very likely an oversimplification. Consider ceiling and floor effects, which are understood to influence N400 amplitude. Floor effects, at least, are likely a factor in Federmeier and Kutas's results, as the study finds no significant effect of constraint on N400 to expected targets, despite the fact that high- and low-constraint contexts are defined precisely by how predictive they are of the expected target. The fact that our cosine similarity measure does reflect an effect of constraint on expected targets suggests that we are capturing important aspects of the context-to-target relation with this measure. However, it also suggests that we will need a nonlinear linking hypothesis to predict the N400 with more precision. This means that we should not be quick to dismiss the other settings in Figure 3, as they could ultimately prove to be the more accurate simulations once we identify the proper linking hypothesis.

We see these simulations as a valuable proof of concept. To understand what the N400 can tell us about contextual representations and their influence on incoming words, we need to be able to tease apart the contributing factors at play when N400 amplitude tracks untimed measures such as cloze, versus the factors at play when it deviates from such measures. VSMs allow for explicit modeling of collective word relations, and as a result they are a promising tool for generating quantitative predictions from a range of hypotheses regarding the semantic relatedness-based processes that may underlie deviations of N400 amplitude from cloze. In the above simulations, we indeed find support for the ability of these models to use averaging-based context representation and simple relation computations to capture aspects of N400 behavior that deviate from the predictions of cloze probability.

It should be noted that our results need not be in direct conflict with Federmeier and Kutas's general framework. Since cosine similarity is computed by dimension-wise comparison of one vector to another, one could think of higher cosine similarity between context and target as representing greater pre-activation of target features as a result of context. As for Federmeier and Kutas's hypothesis of mediating influence by pre-activation of the expected target: in our anchored setting, one might argue that selecting context words based on the expected target instantiates a version of Federmeier and Kutas's mediation account. However, though the expected target does have an increased role in this setting, it is still the context words, and not the expected target, that have the relevant relations to incoming target words in our simulations.

So although this demonstration does not discredit the validity of Federmeier and Kutas's account, it does illustrate a genuine alternative.

It is also important to clarify that our claim is not that our averaging procedure—and the representation that it produces—is an appropriate reflection of the full extent of language processing. We are, however, positing that less structured representations of this kind are likely to underlie the N400 under some circumstances. As discussed above, many aspects of language processing that we know, a priori, will be overlooked by this averaging process, are also aspects of language processing that we have seen the N400 at times to be insensitive to: for instance, this averaging process will not encode agent/recipient information, and it will also fail to capture effects of negation. Such selective insensitivities are in line with N400 studies cited above. It seems not unreasonable, therefore, to suppose that this simple averaging procedure may be approximating a real representational stage tapped into by the N400.

Using the N400 as a probe into online language processing, our results suggest that VSMs are well positioned to capture elements of language interpretation that are driven by lexical semantic relatedness effects. A question that arises now is whether VSMs can also help us to model the more structured compositional processes that seem to underlie the N400 when it does track cloze probability. Structured semantic composition with VSMs is an active area of current research (e.g., Mitchell & Lapata, 2008; Socher, Huval, Manning, & Ng, 2012; Fyshe, Wehbe, Talukdar, Murphy, & Mitchell, 2015), and as progress continues in this area, it will be interesting to investigate whether the influences of more structured context representations can also be captured through VSMs. Other interesting questions will include whether these results extend to other types of VSMs, or to different approaches to semantic similarity, such as manual feature generation.

Acknowledgments This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1322106. This work benefited from helpful comments of three anonymous reviewers, and discussions with Tal Linzen as well as audiences and visitors across the University of Maryland language science community.

References

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 238–247).

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*(4), 343–355.

Brown, C., & Hagoort, P. (1993). The processing nature of the n400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44.

Chow, W.-Y., Smith, C., Lau, E., & Phillips, C. (2015). A 'bag-of-arguments' mechanism for initial verb predictions. *Language, Cognition, and Neuroscience*.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495.

Ferraresi, A., Zanchetta, E., Baroni, M., & Bernardini, S. (2008). Introducing and evaluating ukWaC, a very large web-derived corpus of English. In *Proceedings of the 4th Web as Corpus Workshop (wac-4) Can we beat Google* (pp. 47–54).

Fischler, I., Childers, D. G., Achariyapaopan, T., & Perry, N. W. (1985). Brain potentials during sentence verification: Automatic aspects of comprehension. *Biological Psychology*, *21*(2), 83–105.

Fyshe, A., Wehbe, L., Talukdar, P. P., Murphy, B., & Mitchell, T. M. (2015). A compositional and interpretable semantic space. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.*

Herdağdelen, A., Erk, K., & Baroni, M. (2009). Measuring semantic relatedness with vector space models and random walks. In *Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing* (pp. 50–53).

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language*, *35*(1), 66–85.

Jones, M. N., Kintsch, W., & Mewhort, D. J. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534–552.

Kutas, M., & Hillyard, S. (1989). An electrophysiological probe of incidental semantic association. *Cognitive Neuroscience, Journal of*, *1*(1), 38–49.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*.

Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 217–237).

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*(2), 211.

Lapesa, G., & Evert, S. (2013). Evaluating neighbor rank and distance measures as predictors of semantic priming. In *Proceedings of the ACL Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2013)* (pp. 66–74).

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, *3*, 211–225.

Mandera, P., Keuleers, E., & Brysbaert, M. (2016). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*.

McDonald, S., & Brew, C. (2004). A distributional model of semantic context effects in lexical processing. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics* (p. 17).

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mitchell, J., & Lapata, M. (2008). Vector-based models of semantic composition. In *Acl* (pp. 236–244).

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, *19*(12), 1213–1218.

Padó, S., & Lapata, M. (2007). Dependency-based construction of semantic space models. *Computational Linguistics*, *33*(2), 161–199.

Parviz, M., Johnson, M., Johnson, B., & Brock, J. (2011). Using language models and latent semantic analysis to characterise the N400m neural response. In *Proceedings of the Australasian Language Technology Association Workshop 2011* (pp. 38–46).

Socher, R., Huval, B., Manning, C. D., & Ng, A. Y. (2012). Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 1201–1211).