Title
N400 amplitudes reflect change in a probabilistic representation of meaning:Evidence from a connectionist model

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors
Rabovsky, Milena
Hansen, Steven
Hansen, James L.

Publication Date
2016

Peer reviewed

N400 amplitudes reflect change in a probabilistic representation of meaning: Evidence from a connectionist model

Milena Rabovsky (rabovsky@stanford.edu)
Steven S. Hansen (sshansen@stanford.edu)
James L. McClelland (mcclelland@stanford.edu)
Department of Psychology, Stanford University

Abstract

The N400 component of the event-related brain potential is widely used in research on language and semantic memory, but the cognitive functions underlying N400 amplitudes are still unclear and actively debated. Recent simulations with a neural network model of word meaning suggest that N400 amplitudes might reflect implicit semantic prediction error. Here, we extend these simulations to sentence comprehension, using a neural network model of sentence processing to simulate a number of N400 effects obtained in empirical research. In the model, sequentially incoming words update a representation capturing probabilities of elements of sentence meaning, not only reflecting the constituents presented so far, but also the model's best guess at all features of the sentence meaning based on the statistical regularities in the model's environment internalized in its connection weights. Simulating influences of semantic congruity, cloze probability, a word's position in the sentence, reversal anomalies, semantic and associative priming, categorically related incongruities, lexical frequency, repetition, and interactions between repetition and semantic congruity, we found that the update of the predictive representation of sentence meaning consistently patterned with N400 amplitudes. These results are in line with the idea that N400 amplitudes reflect *semantic surprise*, defined as the change in the probability distribution over semantic features in an integrated representation of meaning occasioned by the arrival of each successive constituent of a sentence.

Keywords: neural network model; sentence comprehension; language; event-related potentials; N400; semantic surprise

Introduction

Language and meaning processing have been investigated with event-related brain potentials (ERPs), providing continuous time-resolved measures of electrical brain activity, and with models that characterize the processing of language and meaning, either in terms of the principles that govern these processes or the processes that implement them. Integration of these approaches could constrain selection among alternative models of the computations and the processes that implement them, while also providing for a possible integrated explanation of the diverse set of empirical phenomena that have been discovered through ERP research. In this spirit, the current work builds on other work described below, aiming to contribute to a better understanding of the computational principles and functional processes underlying the N400 ERP component, an electrophysiological indicator of meaning processing (see Kutas & Federmeier, 2011, for review).

The N400 is a negative deflection at centro-parietal electrode sites peaking around 400 ms after the presentation of a meaningful stimulus. N400 amplitudes have been shown to be modulated by a wide variety of variables. For example, the N400 is modulated by contextual fit, with larger amplitudes to incongruent as compared to congruent sentence continuations, such as when the sentence "He spread the warm bread with…" is completed by the word "socks" instead of "butter". The N400 also shows larger amplitudes to congruent continuations with lower as compared to higher cloze probability; decreasing amplitudes over the course of a sentence; smaller amplitudes for targets after semantically or associatively related as compared to unrelated primes; smaller amplitudes for repeated words as compared to a first presentation; and smaller amplitudes for words of high as compared to low lexical frequency.

Despite the large body of data on N400 amplitude modulations and widespread agreement that the N400 is related to semantic processing, the computational principles and processing mechanisms underlying N400 amplitude generation are as yet unclear. Various verbally-formulated theories are currently under active debate proposing, e.g., that N400 amplitudes reflect lexical and/or semantic access, semantic integration/ unification, semantic binding, or semantic inhibition (reviewed by Kutas & Federmeier, 2011). The development of an explicit computational account that addresses this diverse range of phenomena would thus appear to be a worthwhile goal.

There are at least two ways in which one could seek to better understand the N400 component by means of computational modeling. First, one might try to implement a neurobiologically realistic model of the brain processes underlying the N400 component, an approach that makes it possible to also model the morphology of the ERP waveform (Laszlo & Plaut, 2012; Laszlo & Armstrong, 2014). Another approach is to directly relate variations in N400 amplitudes to measures obtained from functional-level models of cognitive processes. While this approach may entail losing some possibly interesting information with respect to neural realization, it allows the modeling process to focus on the goal of better understanding the cognitive functions underlying N400 amplitudes. Many neural network models are of this functional type, in that the model is viewed as conforming to a computational principle characterized at the functional level. The principle is often articulated in terms of the goal to maximize consistency with information in the environment which can then be formalized in terms of the more specific goal of minimizing *prediction error* (Hinton, 1987; Rumelhart et al., 1995). Focusing on this functional level, Rabovsky & McRae (2014) used an attractor network model of word meaning to investigate which measure in the model covaries with N400 amplitudes over a series of typical N400 word processing paradigms. They consistently observed a close correspondence between N400 amplitudes and network error at the semantic layer of representation, and took this correspondence to suggest that N400 amplitudes reflect implicit semantic prediction error. Here we extend this approach to the processing of words in sentences.

What kind of model would be most appropriate to simulate N400 amplitudes in sentences? Simple recurrent network models (SRNs; Elman, 1990) are typically trained to predict the next word in sentences based on the preceding context so that network error in these models reflects an implicit prediction error which could correlate with N400 amplitudes. Indeed, such a correlation was recently reported by Frank et al. (2015) who used four information measures derived from three probabilistic language models as predictors for six ERP deflections (including the N400). However, the prediction error in SRNs trained to predict the next input based on the preceding context is not specific to semantics but rather reflects *word surprisal* (the negative log of the probability of a word in a specific context) and is affected by both syntactic and semantic expectation violations (Levy, 2008). As the N400 is a functionally specific indicator of meaning processing (Kutas & Federmeier, 2011) while syntactic violations typically modulate different ERP components, we therefore decided against using such an SRN, and instead simulated N400 amplitudes using a model that is specifically trained to understand and predict sentence meaning, the Sentence Gestalt (SG) model (McClelland et al., 1989).

The SG model minimizes the mismatch between its estimates of the probability of semantic features of events given the words presented so far in a sentence and the observed probabilities of these features in the meanings of sentences, such that, once it has learned, its estimates after each new word encountered as a sentence unfolds should come close to matching the true probabilities. Thus the model can be characterized as an implicit probabilistic model of sentence comprehension: The model's outputs can be seen as representing conditional probability distributions over possible semantic features of the events described by the sentence up to and including the latest word. Furthermore, the magnitude of the update of the hidden unit state produced by the presentation of the latest word can be characterized as reflecting the change in this probability distribution produced by the word. We use a measure of this update we call *semantic surprise,* based on a measure that has been called Bayesian surprise (Itti & Baldi, 2005). Formally, the semantic surprise (*SemS*) produced by the *nth* word in a sentence is defined as the difference between the probability distribution over semantic feature representations consistent with the sentence through the *nth* word and the distribution consistent with the sentence through the preceding word, as measured by the Kullbach-Leibler divergence:

$$SemS_n = \sum_r p(r|n) \log_2(p(r|n)/p(r|n-1))$$

Here r indexes alternative possible patterns of semantic features of the event being described by the sentence and $p(r|n)$ and $p(r|n\text{-}1)$ denotes the probability of that pattern given the sentence up through word n and n-1 respectively.

The change in activation at the hidden (SG) layer of the model reflects this semantic surprise, and, as shown in a series of simulations of empirical N400 effects described below, models the N400, suggesting that the N400 is itself a measure of semantic surprise.

The Sentence Gestalt model

The SG model does not assume that sentences are represented in a specific propositional format. Instead, it is based on the idea that the task of sentence processing consists in processing sequences of incoming words to build representations enabling correct responses to various probes, and the model is allowed to find the best way to build these representations in order to meet the imposed demands through adjustments of connections between simple processing units organized in layers. A detailed description of the model is provided elsewhere (McClelland et al., 1989); we briefly sketch it here. For the current simulations, the model was re-implemented in the PDPTool software, V3 (http://web.stanford.edu/group/pdplab/pdphandbookV3/).

Architecture. The model can be conceptualized as consisting of two parts (see Fig. 1). The first part sequentially processes each incoming word (presented at the input layer) to update activation in the SG layer which represents the model's best guess interpretation of the meaning of the sentence as a whole, using the previous activation of the SG layer together with the activation induced by the new incoming word to produce the updated SG layer activation. The second part of the model is used primarily for performance assessment and training, decoding the content from the SG layer by probing it concerning the event described by the sentence.

Environment and training. It is important to note that the statistical regularities underlying the model's best guess interpretation of the meaning of the sentence at a given point in its presentation are determined by the training set so that the effects on semantic surprise depend on the training set as well. There are two different approaches to training which are complementary in that they each have strengths and weaknesses. First, models can be trained on large-scale training corpora approximating real life language environments of human participants. While this approach allows for simulation of empirical experiments with the exact same stimuli on a single-trial basis, the factors responsible for the effects produced by the model may remain somewhat opaque. A second possible approach is to train models on a synthetic training set which implements variation among certain dimensions considered to be relevant for the target empirical phenomenon (the N400, in our case) and/or the theory advanced to explain the empirical phenomenon (semantic surprise, in our case). While this approach is limited in its capacity to fully explain specific empirical data points, it is more transparent concerning the general factors and principles responsible for the effects produced by the model. Because the main aim of the current study is to advance a theory concerning the functional basis of N400 amplitudes by highlighting the common core shared by the different dimensions that have been shown to modulate them, this transparency concerning the responsible factors is of primary importance to our goals. Thus, we trained our model on a small synthetic training set, aiming to create statistical regularities in the training set that allowed us to run simulation experiments containing manipulations corresponding to manipulations in empirical N400 experiments. We observe, based on these simulation experiments, that variables or dimensions that influence N400 amplitudes in the world influence semantic surprise in our model in the same way, suggesting that N400 amplitudes reflect semantic surprise.

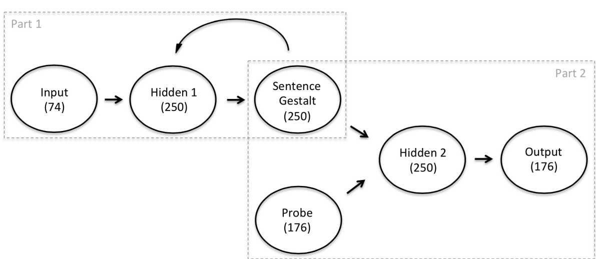

Fig. 1: The Sentence Gestalt (SG) model.

The model environment consists of sentences (presented word by word at the input layer) such as 'At breakfast, the man eats eggs' each paired with a corresponding event (a set of role-filler pairs, e.g., agent – the man), probabilistically generated online during training according to pre-specified constraints. After each presented word (represented by a word-specific unit at the input layer), the model is probed concerning the event described by the sentence. Responding to a probe consists in completing a role-filler pair when probed with either a thematic role (i.e., agent, action, patient, location, or situation; each represented by an individual unit at the probe and output layer) or a filler of a thematic role (e.g., the man, to eat, the eggs, etc.). For the filler concepts, we used feature-based semantic representations that were handcrafted so that members of the same semantic category shared some semantic features. For example, somewhat similar to the representations used by Rogers and McClelland (2008), all living things shared a semantic feature ('can grow'), all animals shared an additional semantic feature ('can move'), all birds shared one more semantic feature ('can fly') and then the canary had two individuating features ('can sing' and an item-unique individuating feature) so that the robin and the canary shared three of their five semantic features while the salmon and the canary shared two features, the rose and the canary shared only one feature, and the jeans and the canary did not share any features. While the labels for the features are irrelevant for model behavior, the aim in constructing the representations was to create graded similarities between concepts roughly corresponding to real world similarities. A comparison between a similarity matrix of the concepts based on the hand-crafted semantic representations and representations based on semantic word vectors derived from co-occurrences in large text corpora (Pennington, Socher, & Manning, 2014) suggested a reasonable correspondence (r = .73). Such feature-based semantic representations were also employed in the original version of the model; this allows us to capture the influence of semantic similarity over and above the influence of co-occurrence in language as implemented via the presented sentences (enabling simulation of categorically related semantic incongruities; Sim. 1).

After each presented word, the model is probed for each thematic role and each filler of each role-filler pair involved in the described event, and the model's activation at the output layer is compared with the correct output. Error is then back-propagated through the entire network and connections are adjusted to minimize the difference between model-generated and correct output (we used cross-entropy error, a learning rate of 0.00005 and momentum of 0.9). Because the model is probed concerning the described event after every single presented word, it anticipates the meaning of each sentence as early as possible, so that the activation at the SG layer (and accordingly at the output layer in response to the presented probes) becomes tuned to the regularities in the corpus. For example, the model learns that a sentence beginning "The woman writes…" more often describes the woman writing an email than an sms, and encodes this regularity in the connection weights, resulting in probabilistic pre-activations of units in the SG layer before email or sms appear in the sentence. Indeed the model's connection weights capture the base-rate probabilities of the semantic features of each of the roles in the sentence, so that when probed with a role prior to the presentation of the first word of a sentence the pattern over the filler units corresponds approximately to the overall probability across the entire environment that the feature will be present in the filler of the probed-for role.

Since the minimum of the cross-entropy error is reached when the network's estimates of feature probabilities match the actual probabilities of those features, the change in the network's estimates occasioned by each successive word should match the change in these actual probabilities (Rumelhart et al., 1995). Treating the semantic feature probabilities as conditionally independent given the words seen so far, this change in estimates of feature probabilities can be shown to correspond to the $SemS_n$ measure defined in the introduction.

Thus far we have described how changes in the activation of semantic features in the model's output layer should correspond to the semantic surprise. However, we do not assume that these semantic feature activations are actually computed during sentence processing. Instead, we propose that the pattern of activation over the SG layer (together with the connection weights in Part 2 of the model) implicitly represent this probability distribution in such a way that the update at the SG layer mirrors the update in the actual probability distribution over features. We use the following cross-entropy measure to characterize this update:

$$\sum_i a_i(n)\log\left(\frac{a_i(n)}{a_i(n-1)}\right) + \left(1 - a_i(n)\right)\log\left(\frac{1 - a_i(n)}{1 - a_i(n-1)}\right)$$

Here i ranges over all of the SG layer units, $a_i(n)$ represents the activation of unit i based on the current word and $a_i(n-1)$ represents the activation of unit i based on the previous word. Similar results are obtained using the sum over SG units of the absolute value of the difference between $a_i(n)$ and $a_i(n-1)$.

Simulations

Sim. 1: Categorically related semantic incongruities. Federmeier and Kutas (1999) presented sentence pairs such as "They wanted to make the hotel look more like a tropical resort. So along the driveway they planted rows of…" and observed gradually increasing N400 amplitudes from congruent sentence continuations ("palms") to unexpected continuations which were members of the same semantic category as the expected continuation and thus shared semantic features ("pines") to incongruent continuations.

To simulate these data, we trained the model such that one member of each semantic category (i.e., trees, drinks, etc.) was never presented in the same sentence context (i.e., as a patient of the same action) as the other category members so that it was completely unexpected in that context. For the simulation experiment, we presented the model with 10 such unexpected sentence continuations which were categorically related to the congruent continuations, as well as 10 congruent continuations, presented with a probability of .8 during training when the specific combination of agent and action had been presented, and 10 incongruent sentence continuations which were never presented as patients of the specific action during training and did not share any semantic features with the congruent continuations.

As shown in Fig. 2A, semantic surprise induced by the critical words gradually increased from congruent continuations to unexpected continuations categorically related to the congruent continuations, $t_{(9)} = 5.23$, $p < .0001$, and from those to unexpected continuations unrelated to the congruent continuations, $t_{(9)} = 6.30$, $p < .0001$, in line with N400 amplitudes.

The model also successfully captures other typical N400 effects in sentences such as cloze probability effects, with larger N400 for low cloze as compared to high cloze probability sentence continuations (**Sim. 2**; Kutas & Hillyard, 1984), sentence position effects with decreased N400 amplitudes over the course of a sentence (**Sim. 3**; Van Petten & Kutas, 1990), and influences of so-called semantic illusions or reversal anomalies, i.e. only a very slight increase of N400 amplitudes in sentences such as 'For breakfast, the eggs <u>eat</u>…' as compared to 'For breakfast, the boys <u>eat</u>…' while the increase in 'For breakfast, the boys <u>plant</u>…' is much larger (**Sim. 4**; Kuperberg et al., 2003; also simulated by Brouwer, 2014, PhD thesis, see below). Even though the SG model is designed as a model of sentence processing, word pairs and isolated words should be processed by the same system so that we also used the model to simulate N400 effects outside of sentence context. We describe the simulation of semantic priming in detail.

Sim. 5: Semantic priming. Bentin et al. (1985) observed smaller N400 amplitudes to target words presented after semantically related primes (i.e., primes from the same semantic category as the targets) as compared to unrelated primes. To simulate these data, we presented the model with 10 word pairs where the referenced concepts were members of the same semantic category and thus shared semantic features at the output layer (e.g., monopoly – chess) and 10 word pairs where the primes and targets from the related pairs were re-assigned such that there was no semantic similarity between prime and target. As shown in Fig. 2B, semantic surprise was smaller for targets after semantically related as compared to unrelated primes, $t_{(9)} = 5.14$, $p < .0001$, in line with N400 amplitudes (Bentin et al., 1985). The SG model additionally captures several other N400 effects in word processing such as associative priming, with smaller amplitudes to targets after associatively related (e.g., play – chess) as compared to unrelated primes (**Sim. 6**; Kutas & Hillyard, 1989), repetition priming, i.e. smaller N400 amplitudes to immediately repeated words as compared to target words presented after unrelated primes (**Sim. 7**; Nagy & Rugg, 1989), and smaller amplitudes to words of high as compared to low lexical frequency (**Sim. 8**; Van Petten & Kutas, 1990), captured through the encoding of base rate probabilities of features in the model's connection weights.

Finally, probability distributions can change and anticipatory preparedness to likely upcoming features depends on constant adaptation of represented probabilities based on new experiences. In neural network models, this adaptation is driven by the difference between expected and observed outcomes. Thus, if N400 amplitudes reflect this difference then larger N400 amplitudes should entail enhanced adaptation. Simulation 9 focuses on this relation.

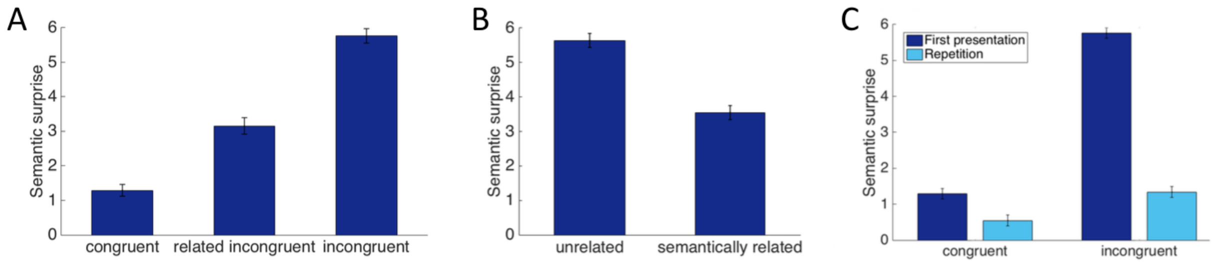

Fig. 2: Influences of (A) categorically related semantic incongruities, (B) semantic priming, and (C) repetition X congruity.

Sim. 9: Semantic incongruity and repetition. The influence of semantic incongruity on N400 amplitudes is reduced by repetition due to a stronger repetition-induced reduction of amplitudes for incongruent continuations (Besson et al., 1992). Repetition effects have been simulated as consequences of connection adjustments induced by the prior presentations of the item within the experiment. We thus presented the 10 congruent and 10 incongruent sentences from Sim. 1 twice while learning was operative (learning rate = 0.0005) so that the first presentation served not only as an experimental condition but also produced connection adjustments.

As shown in Fig. 2C, the difference in semantic surprise between the critical congruent vs. incongruent words was smaller during repetition. A rmANOVA confirmed a significant interaction between repetition and congruity, $F(1,9) = 87.18$, $p < .0001$, reflecting a stronger influence of repetition for incongruent, $F(1,9) = 82.35$, $p < .0001$, $\eta_p^2 = .90$, as compared to congruent, $F(1,9) = 16.47$, $p = .003$, $\eta_p^2 = .65$, sentence continuations, in line with N400 data.

These results nicely illustrate the intrinsic relationship between semantic surprise and adaptation. However, there is a subtle issue with this simulation, namely that the output layer which drives learning represents the events described by the sentences such that the simulation assumes these events to be observed while processing the sentences. This is not true for the empirical experiment. Thus, in a prospective version of this simulation the error signal that drives learning should be derived from the difference between SG activation before and after the critical words.

Discussion

The goal of the present study was to investigate the functional basis of the N400 ERP component by relating N400 amplitudes to a computational model of sentence comprehension, the Sentence Gestalt (SG) model. Across a series of simulations of N400 effects, we consistently observed a correspondence between N400 amplitudes and semantic surprise as represented by the magnitude of the update of hidden unit activations that implicitly represent probability distributions over semantic features.

N400 amplitudes have been previously linked to changes in lexical activation (Brouwer, 2014, PhD thesis; see also Rabovsky & McRae, 2014, for discussion, and Crocker et al., 2010, for a model of the P600 component). Brouwer (2014) focused on reversal anomalies and took the very small increase of N400 amplitudes in sentences such as 'For breakfast, the eggs eat…' (Kuperberg et al., 2003) to indicate that the N400 component does not reflect semantic integration but the retrieval of lexical information. He further suggests that semantic integration is linked to the P600 component (which is increased in reversal anomalies). Our account differs from Brouwer's in that it specifies a single integrated representation of meaning which is updated whenever a word is presented, with the extent of this update reflected in N400 amplitudes. When the word is presented in a sentence, the update is seen as an update of an integrated representation of the meaning of the sentence. The integration process is relatively heuristic and may not accord with syntactic constraints in constructions such as reversal anomalies. Indeed, analysis of our model's output layer suggests that it experiences a 'semantic illusion' in that it continues to assign the eggs to the patient instead of the agent role even after the word 'eat', in line with the suggestion that language processing can be shallow (Ferreira et al., 2002). Our model does not address the P600. It is possible that the P600 effect in reversal anomalies reflects a re-assignment of the eggs to an agent role. Alternatively, the comprehender may continue to see the eggs as being eaten, with the P600 reflecting detection of a syntax error. As a third possibility, the P600 may reflect conflict monitoring triggered by competing interpretations, one arising from a heuristic process and the other arising from a controlled process (van Herten et al., 2006). Further research seems required to better understand the P600.

As noted above, the SG representation together with the model's weights latently predict the semantic features of each role filler in the sentence based on prior constituents, and the update of the SG due to the next constituent adjusts these latent predictions. Latent prediction in that sense means that the SG model (and presumably the brain) becomes tuned through experience to be prepared to respond to likely upcoming semantic features with little additional effort. This kind of latent prediction seems to range from the pre-activation of specific semantic features in sentence context (e.g., for the categorically related incongruities, where less semantic update is necessary when an unexpected sentence continuation shares semantic features with an expected continuation; Sim. 1) to the latent structure in connection strengths and default activation that leads to less semantic update when processing a high frequent as compared to a low frequent word in isolation

(Sim. 8). This characterization should make clear that prediction in that sense does not refer to explicit intentional prediction of specific items but rather to a general configuration of the system optimized for upcoming semantic information. This entails that semantic activation changes induced by new incoming input primarily reflect the discrepancy between probabilistically anticipated and encountered features, i.e. semantic surprise, in accordance with predictive coding (Friston, 2005). In line with this view, N400 amplitudes in many paradigms appear to have one thing in common, namely that they seem to be a function of the fit between the semantic features that are implicitly expected based on previously experienced regularities and those activated by the current stimulus.

In sum, the present study aimed to contribute to a better understanding of the functional basis of the N400 ERP component by relating N400 amplitudes to an implemented model of sentence comprehension. Across a series of simulations of N400 effects, we consistently observed a correspondence between N400 amplitudes and the update of conditional probability distributions over semantic features. Besides demonstrating that the SG model naturally captures electrophysiological indicators of internal cognitive dynamics during language comprehension, these results are in line with the idea that N400 amplitudes reflect semantic surprise as the extent of change induced by an incoming stimulus in a probabilistic representation of meaning.

Acknowledgments

Funded by the European Union's Horizon 2020 programme, Marie Sklodowska-Curie grant 658999 to Milena Rabovsky.

References

Bentin, S., McCarthy, G., & Wood, C.C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroen Clin Neuro, 60,* 343-355.

Besson, M., Kutas, M., & Van Petten, C. (1992). An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences. *JoCN, 4,* 132-149.

Brouwer, H. (2014). *The electrophysiology of language comprehension: A neurocomputational model.* PhD thesis.

Crocker, M.W., Knoeferle, P., & Mayberry, M.R. (2010). Situated sentence processing: The coordinated interplay account and a neurobehavioral model. *Brain Lang, 112,* 189-201.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science, 14,* 170-211.

Federmeier, K. & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *J Mem Lang, 41,* 469-495.

Ferreira, F., Bailey, K.G.B., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11,* 11-15.

Frank. S.L., Otten, L.J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain Lang, 140,* 1-11.

Friston, K. (2005). A theory of cortical responses. *Philos T R Soc B, 360,* 815-836.

Hinton, G. E. (1989). Connectionist learning procedures. *Artif Intell*, *40*(1), 185-234.

Itti, L. & Baldi, P. F. (2005). Bayesian surprise attracts human attention. In *NIPS* (Vol. 19), 547-554, Cambridge, MA: MIT Press.

Kuperberg, G.R., Sitnikova, T., Caplan, D., & Holcomb, P.J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cogn Brain Res, 17,* 117-129.

Kutas, M. & Federmeier, K. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol, 62,* 621-647.

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

Kutas, M. & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161-163.

Kutas, M. & Hillyard, S.A. (1989). An electrophysiological probe of incidental semantic association. *JoCN, 1,* 38-49.

Laszlo, S. & Armstrong, B. (2014). PSPs and ERPs: Applying the dynamics of post-synaptic potentials to individual units in simulation of ERP reading data. *Brain Lang, 132,* 22-27.

Laszlo, S. & Plaut, D. (2012). A neurally plausible parallel distributed processing model of event-related potential word reading data. *Brain Lang, 120,* 271-281.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

McClelland, J.L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Lang Cognitive Proc, 4,* 287-335.

Nagy, M.E. & Rugg, M.D. (1989). Modulation of event-related potentials by word repetition: The effects of inter-item lag. *Psychophysiology, 26,* 431-436.

Pennington, J., Socher, R., & Manning, C.D. (2014). GloVe: Global vectors for Word Representation. *EMNLP.*

Rabovsky, M. & McRae, K. (2014). Simulating N400 amplitudes as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition, 132,* 68-89.

Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1995). Backpropagation: The basic theory. *Backpropagation: Theory, Architectures and Applications*, 1-34.

Van Herten, M., Chwilla, D.J., & Kolk, H.H.J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *JoCN, 18,* 1181- 1197.

Rogers, T.T. & McClelland, J.L. (2008). Précis of semantic cognition: A parallel distributed processing approach. *Behav Brain Sci, 31,* 689-749.

Van Petten, C. & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Mem Cognition, 18,* 380-393.