# Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events

Gina R. Kuperberg

Routledge
Taylor & Francis Group

COMMENTARY

# Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events

Gina R. Kuperberg[a,b,c]

[a]Department of Psychology and Center for Cognitive Studies, Tufts University, Medford, MA, USA; [b]Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; [c]Athinoula A. Martinos Center for Biomedical Imaging, Boston, MA, USA

**ABSTRACT**
Since the early 2000s, several event-related potential studies have challenged the assumption that we always use syntactic contextual information to influence semantic processing of incoming words, as reflected by the N400 component. One approach for explaining these findings is to posit distinct semantic and syntactic processing mechanisms, each with distinct time courses. While this approach can explain specific datasets, it cannot account for the wider body of findings. I propose an alternative explanation: a dynamic generative framework in which our goal is to infer the underlying event that best explains the set of inputs encountered at any given time. Within this framework, combinations of semantic and syntactic cues with varying reliabilities are used as evidence to weight probabilistic hypotheses about this event. I further argue that the computational principles of this framework can be extended to understand how we infer situation models during discourse comprehension, and intended messages during spoken communication.

The question of when and how we extract higher order meaning from semantic and syntactic contextual information has been central to psycholinguistic theory since early descriptions of effects of semantic context on syntactic parsing (Frazier, 1987; Ferreira & Clifton, 1986; Rayner, Carlson, & Frazier, 1983; Stowe, 1989; Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell, Tanenhaus, & Kello, 1993). One approach is to posit the existence of distinct semantic and syntactic mechanisms of processing, sometimes with distinct time courses, which interact at particular stages of processing. Another is to posit that syntactic and semantic information can interact very quickly but that, at any particular time, their relative influence on processing depends on the strength of evidence offered by the bottom-up input. In this article, I discuss what the N400 event-related potential (ERP) can tell us about these interactions.

In Section 1, I characterise the modulation of the N400 as reflecting the retrieval or access of semantic features associated with incoming words that have not already been pre-activated by the context. I introduce a set of ERP studies that challenge the assumption that we necessarily or always use all syntactic information within a given context to influence semantic processing of incoming words, as reflected by the N400 (Kuperberg, 2007), including the study reported by Chow, Smith, Lau,

Phillips, and House (this issue). In Section 2, I summarise the account that Chow et al. offer to explain their findings, which proposes distinct syntactic and semantic mechanisms, each with distinct time courses. I argue that, while this approach can explain subsets of data, its challenge is to explain the wider set of ERP and behavioural data available to us. I point out that it was this challenge that led to the development of constraint satisfaction frameworks of language processing in the 1990s. While *lexical* constraint-based models fall short of explaining the findings of these ERP studies, including Chow et al.'s data, I argue that the principles of constraint satisfaction approaches should not be abandoned.

In Section 3, I summarise a dynamic generative framework of event comprehension, which retains these constraint-based assumptions, and offers an alternative explanation for this set of ERP findings. I illustrate the principles of this framework by discussing how they can potentially account for the pattern of data reported by Chow et al. In Section 4, I consider how this dynamic generative architecture might be extended beyond the comprehension of specific events, to understand how we infer general event structures in relatively impoverished contexts, wider situation models during text and discourse comprehension, and intended messages and speech acts during spoken communication. In Section 5,

---

**CONTACT**  Gina R. Kuperberg  ✉ kuperber@nmr.mgh.harvard.edu

I discuss several open questions that this framework raises, and I suggest ways in which its predictions can be empirically tested. I conclude by suggesting that the computational principles of this framework have the potential to explain a large body of ERP and behavioural data on event comprehension, and potentially unify this literature with broader theories of language adaptation and learning.

## 1. The N400: what can it tell us?

(1) He spread the warm bread with …

The N400 ERP component was first described by Kutas and Hillyard (1980), who reported that, following contexts like (1), a negative-going waveform, peaking at approximately 400 ms, was larger (more negative) to semantically incongruous words like "socks" than to congruous words like "butter". Since this initial report, the N400 has been viewed as a valuable index of how we derive meaning from language. Its amplitude can pattern with real-world plausibility (Hagoort, Hald, Bastiaansen, & Petersson, 2004; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003) and with discourse-level incoherence (e.g. Van Berkum, Hagoort, & Brown, 1999); it can also be highly sensitive to an incoming word's lexical probability, as operationalised by its cloze probability (DeLong, Urbach, & Kutas, 2005; Kutas & Hillyard, 1984; Wlotko & Federmeier, 2012).

There are, however, important exceptions to both of these findings, which suggest that the N400 is neither a direct index of constructing higher-level meaning (contra Hagoort, Baggio, & Willems, 2009), nor a reflection of access to specific stored lexical entries (contra Lau, Phillips, & Poeppel, 2008). Rather, it is best characterised as reflecting the ease of accessing semantic features and properties associated with incoming words, stored within semantic memory (Federmeier & Kutas, 1999; Kutas & Federmeier, 2011; see also Van Berkum, 2009). On this account, prior to encountering an incoming target word, comprehenders use multiple sources of contextual information to predict or pre-activate upcoming semantic features in a graded fashion, thereby changing the state of semantic memory (here, and throughout this article, I use the term *prediction* simply to refer to some effect of context on the state of activation at a particular level of representation, ahead of new bottom-up information becoming available at this level of representation). The degree to which the N400 evoked by an incoming word is attenuated depends on the degree to which its preceding context has pre-activated that word's semantic features prior to these features becoming available from the bottom-up input. In

information theoretic terms (Shannon, 1948), it can be conceptualised as reflecting surprisal at the level of semantic features (see Kuperberg, 2015; Rabovsky & McRae, 2014 for discussion) – the amount of new semantic *information* that is yielded by an incoming stimulus that has not already been predicted by the context.

According to this semantic access account, the reason the N400 evoked by a target word is often sensitive to that target's lexical probability is because lexically constraining contexts are also semantically constraining, and because lexically expected words are also semantically expected, even though lexical and semantic probability can be dissociated (e.g. a lexically unexpected word can still be semantically expected if it is semantically related to a lexically expected word, see Federmeier & Kutas, 1999; Kutas & Hillyard, 1984). And the reason the N400 is often sensitive to real-world plausibility is that comprehenders draw upon their fine-grained stored knowledge about the likelihood of real-world events and states to predict upcoming semantic features that are congruous with this knowledge.

Both these explanations, however, rest on an important assumption – that because language comprehension is highly interactive (Elman, Hare, & McRae, 2004; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), comprehenders are able to quickly use all sources of contextual information available to them to build a rich higher-level representation of meaning, and that they can use this high-level representation to predict upcoming semantic features prior to new semantic information becoming available from the bottom-up input. For example, as we read the context in (1), we build a partial representation of the event, <spreading something on bread>, and, ahead of encountering the next word, we use this event representation to predict semantic features associated with <butter>, leading to more semantic facilitation, and a smaller N400, when "butter" (rather than "socks") is then encountered.

(2a) Every morning at breakfast the boys would eat …
(2b) Every morning at breakfast the eggs would eat …

In the early 2000s, a set of ERP studies called this assumption into question. These studies reported cases of small N400s produced by lexically unexpected targets that were neither semantically related to lexically predictable words, nor plausible in relation to their preceding contexts. For example, in 2003, my colleagues and I reported that the amplitude of the N400 was just as small to words like "eat" in sentences like (2b) as in sentences like (2a), although this contrast was associated with modulation of a later positive-going posteriorly distributed ERP component – the P600

(Kuperberg et al., 2003, see also Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005; Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006, and see Kuperberg, 2007 for a review).

These findings led to some soul searching amongst ERP researchers. One account, for example, was that, rather than revealing anything about how an event representation of the context influences semantic processing, the N400 simply indexes the semantic relatedness between the "bag of words" that constitutes this context and the incoming target word (see Kuperberg, Paczynski, & Ditman, 2011, and Van Berkum, 2009, for discussion). This idea was not as far-fetched as at first it might seem because most of the original studies examining how the N400 was modulated in sentences and discourse contexts did not match simple semantic relationships between content words across experimental conditions. It was therefore possible that, following context (1), any differential pre-activation of the semantic features associated with <butter>, relative to <socks>, had nothing to do with the prior build-up of a specific event representation, <spreading something on bread>, but rather reflected the effects of a more general schema associated with the words, <warm bread spread>.[1]

The story, however, is not so simple. There are now many studies contrasting sentence or discourse scenarios that either contain the identical sets of content words (e.g. Nieuwland & Kuperberg, 2008; Otten, Nieuwland, & Van Berkum, 2007; Urbach, DeLong, & Kutas, 2015; Xiang & Kuperberg, 2015), or that contain content words whose semantic relatedness is fully matched across conditions using computational measures such as Latent Semantic Analysis (Kintsch, 2001; Landauer & Dumais, 1997; see e.g. Kuperberg et al., 2011; Paczynski & Kuperberg, 2011, 2012). The results of these studies reveal variable patterns of N400 modulation: sometimes there are no effects of either lexical predictability or contextual coherence, consistent with the idea that the N400 is indeed driven primarily by the schema describing the bag of words in the context; at other times, however despite schema-based semantic relationships being matched across conditions, the N400 patterns with both these factors.

This type of variability in N400 modulation suggests that not all sources of information within a context are necessarily or always available to pre-activate upcoming semantic features during online language comprehension. It therefore follows that, to better understand when and how different types of contextual information interact to influence semantic prediction, we need to systematically vary their influences and examine how the N400 is modulated in response to incoming words.

(3a) The exterminator inquired which neighbour the landlord had evicted …
(3b) The neighbour inquired which exterminator the landlord had evicted …
(3c) The restaurant owner forgot which customer the waitress had served …
(3d) The restaurant owner forgot which waitress the customer had served …

This is the approach taken by Chow et al. (this issue), who use the N400 to examine when and how semantic and syntactic information interact to predict the semantic properties of upcoming verbs. The authors report a smaller (less negative) N400 on verbs like "evicted" in sentences like (3a) relative to (3b). However, they show no difference in N400 amplitude on verbs like "served" in sentences like (3c) versus (3d), telling us that, in (3d), comprehenders had not used all the syntactic information available to them in the context to predict upcoming semantic information. What exactly this tells us more generally about when and how semantic and syntactic information interact during online language comprehension, however, is debated, and it is this question that I turn to next.

## 2. Building event meaning: distinct mechanisms or multiple constraints?

Chow et al. explain their findings by appealing to three distinct mechanisms of processing. They propose that a rudimentary syntactic parsing mechanism initially delineates a "bag of arguments", and that a bag-of-arguments mechanism then draws upon simple semantic relationships across these arguments to pre-activate upcoming semantic information. They argue that the reason why the N400 is smaller to verbs like "evicted" in sentences like (3a), relative to (3b), is because the bag-of-arguments mechanism generates stronger predictions for <evicted> in (3a) than (3b), on the basis of its closer semantic relationship with <neighbour-landlord> than with <exterminator-landlord>.

Chow et al. distinguish both these fast mechanisms from a third mechanism that builds the meaning of whole events (representations of "who does what to whom") by syntactically assigning lexical items to structural thematic roles, but that acts too slowly to influence semantic prediction and the N400. They argue that the reason why the amplitude of the N400 evoked by "served" is the same in (3c) and (3d) is because, ahead of encountering "served", comprehenders had not yet built a higher-level event representation that differentially predicted upcoming semantic information. And,

because the arguments were identical in (3c) and (3d), the bag-of-arguments mechanism also failed to differentiate between the two conditions.

This move to explain particular patterns of data in terms of distinct mechanisms of processing, each with distinct time courses, has a long history in psycholinguistics: in the 1980s, to account for effects of semantic context on the syntactic garden path effect, a fast syntactic processor, which used a "minimal attachment" strategy to identify the major syntactic phrases of a sentence (Frazier & Fodor, 1978; see also Bever, 1970), was distinguished from a slower semantic processor that evaluated the semantic plausibility of a set of arguments around the head of a syntactic phrase (Rayner et al., 1983; see also Frazier, 1987 and Ferreira & Clifton, 1986).

More recently, within the ERP literature several researchers have proposed distinct mechanisms to explain why the N400 does not always or necessarily pattern with either lexical predictability or high-level coherence. Many different mechanisms or "streams" of processing have been invoked, including a semantic memory-based mechanism that is sensitive to relationships between content words (Kuperberg, 2007), a semantic–thematic mechanism that is sensitive to the thematic structure of verbs and coarse semantic features of arguments such as animacy (Kuperberg, 2007), a mechanism that computes structure on the basis of animacy, case markings and/or the linear order of constituents (Bornkessel & Schlesewsky, 2006), and mechanisms that derive overall meaning on the basis of thematic fit, regardless of syntactic structure, either using a plausibility heuristic (van de Meerendonk, Kolk, Chwilla, & Vissers, 2009) or through independent semantic combination (Kim & Osterhout, 2005; see also Hoeks et al., 2004). These different accounts vary in their assumptions. Some, similar to Chow et al., assume that the N400 indexes the retrieval of features from semantic memory (e.g. Kuperberg, 2007), while others assume that it indexes the construction of overall meaning (e.g. Hoeks et al., 2004; Kim & Osterhout, 2005). Some accounts, similar to Chow et al., attribute distinct time courses to each of these processing mechanisms, resulting in discrete stages of processing (e.g. Bornkessel & Schlesewsky, 2006; see also Friederici & Weissenborn, 2007), while others have argued that these mechanisms interact continuously, with additional analysis ensuing only if their outputs conflict (e.g. Kim & Osterhout, 2005; Kuperberg, 2007; van de Meerendonk et al., 2009).

These various stream-based frameworks have been quite successful in explaining the specific patterns of ERP findings within the particular datasets that inspired them. The major challenge that they all face, however,

is in accounting for the wider body of ERP and behavioural data. The first issue is that, as more data become available, showing that online sentence processing is sensitive to more and more different types of contextual information, one is forced to posit more and more distinct mechanisms to account for these data. At its extreme, therefore, this approach simply offers a re-description of various data patterns, rather than a deeper underlying explanation.

The second issue, which is particularly problematic for accounts that attribute inherently distinct time courses to distinct mechanisms of processing, is that they can lead to conflicting conclusions. This is because the same mechanism that needs to be slow to explain one dataset needs to be fast to explain another. To illustrate this point, it is worth noting that Chow et al.'s semantic bag-of-arguments mechanism is quite reminiscent of Rayner, Carlson and Frazier's original proposal of a mechanism that evaluates arguments on the basis of their semantic plausibility (Rayner et al., 1983). However, to explain Chow et al.'s ERP findings, this mechanism is posited to act quickly (prior to the syntactic assignment of thematic roles), whereas to explain the early findings of a delayed effect of semantic context on the syntactic garden path effect (Ferreira & Clifton, 1986; Rayner et al., 1983), it needed to act at a later stage of processing.

It is precisely these types of issues that, in the 1990s, inspired the development of constraint satisfaction frameworks of sentence comprehension to account for effects of semantic and discourse contexts on syntactic parsing (MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994; see also Marslen-Wilson, 1987; Marslen-Wilson, Brown, & Tyler, 1988). The central tenet of these constraint-based approaches is that multiple types of contextual information can, in principle, interact very quickly, but the degree to which each type of information influences processing in any given situation will depend on the *strength of evidence* offered by the bottom-up input. Thus, within this framework, if, at a given stage of processing a particular sentence, a comprehender appears to have used syntactic but not semantic contextual information, this does not imply an inherently fast syntactic mechanism and an inherently slow semantic mechanism; it simply means that, in this particular sentence, the semantic evidence offered by contextual input is weaker than the syntactic evidence (for discussion, see McRae, Spivey-Knowlton, & Tanenhaus, 1998; Trueswell et al., 1994; Trueswell & Tanenhaus, 1994).

The most developed of these constraint-based accounts are models that posit the use of rich *lexical* representations that encode graded semantic and syntactic information about how words combine with other words

(e.g. MacDonald et al., 1994; Trueswell & Tanenhaus, 1994; Trueswell et al., 1993, 1994). Depending on the strength of bottom-up evidence, comprehenders activate these lexical representations in a graded fashion, combine them together to construct propositions, and map these propositions on to event representations, which, in turn, modulate lexical expectations for upcoming words. These models successfully explain the effects of semantic context on syntactic parsing (e.g. Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Hare, Tanenhaus, & McRae, 2007; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Stowe, 1989; Trueswell et al., 1993, 1994; Wilson & Garnsey, 2009). On the other hand, by assuming that comprehenders necessarily need to achieve access to the syntactic properties of each word in a context in order to build a higher-level event representation that can be used to predict lower-level information, lexical constraint-based models run into problems in explaining the types of ERP findings described above (see Kuperberg, 2007 for discussion).

In the next section, I describe a dynamic, generative[2] framework of event comprehension that can potentially explain these ERP data. I first give a broad overview of this framework (see Brown & Kuperberg, 2015; Kuperberg, 2015; Kuperberg & Jaeger, 2015 for more general discussion)[3]. I then illustrate its principles by showing how it can account for Chow et al.'s findings.

## 3. A dynamic generative framework of event comprehension

A generative framework of language comprehension takes as its starting point that the comprehender seeks to infer, with as much certainty as possible, the underlying *cause* of an observed set of sensory inputs that unfold incrementally over time. To infer this underlying cause, the comprehender draws upon an internal *generative model* – a set of hierarchically organised internal representations which, at any given time, she believes can best *explain* the statistical properties of the bottom-up input that she has encountered, given her current communicative goals as well as her current beliefs about the statistical structure of her broader communicative environment (for a general discussion of hierarchical generative models, see Clark, 2013; Friston et al., 2015; Hinton, 2007; for discussion in relation to language processing, see Brown & Kuperberg, 2015; Kuperberg & Jaeger, 2015). At the top of this generative hierarchy lies a set of hypotheses that the comprehender currently holds with varying degrees of belief about the underlying cause of the sensory input she has encountered.

The comprehender can then test these hypotheses by *actively generating* probabilistic predictions that are propagated down to lower levels of the generative model, thereby changing the distribution of prior beliefs at these lower levels of representation before new bottom-up input becomes available. The degree to which beliefs are propagated down from higher to lower levels of the generative model will depend on the *certainty* with which they are held. Then, at the next moment in time when new bottom-up evidence becomes available, the comprehender learns whether her probabilistic predictions are supported. Any bottom-up evidence that is consistent with these predictions is "explained away", while any evidence that is not explained away – prediction error – is propagated back up the generative model, and used to update the comprehender's high-level beliefs about the underlying cause of the sensory input through Bayesian inference.[4] In this way, by iteratively cycling between probabilistic inference and prediction, with the aim of reducing prediction error across the entire generative model, the comprehender will converge, with increasing certainty and specificity, upon the particular higher-level representation that causes and best explains the full set of her observations.

This dynamic generative framework potentially provides an alternative explanation for Chow et al.'s findings. Within this framework, the underlying "cause" that the comprehender seeks to infer is the *event or event structure* that best explains the sequence of words that she has encountered at any given time, and the structure of her generative model is defined by her goal of correctly judging the plausibility of each sentence, as well as her beliefs about the statistical regularities of the wider experimental environment.

To explain why the N400 is smaller to verbs like "evicted" in sentences like (3a) than (3b), this framework posits that, just prior to encountering the verb, the comprehender entertains multiple hypotheses about the event or event structure being conveyed. These beliefs were inferred on the basis of all the semantic *and* syntactic information encountered thus far in the context, which, together provided the comprehender with reliable evidence to support her prior beliefs (stored real-world knowledge) about the most likely semantic roles that particular participants or entities play around a given action or state (see McRae, Ferretti, & Amyote, 1997; McRae et al., 1998; McRae & Matsuki, 2009). To test these high-level hypotheses, the comprehender has generated predictions at lower levels of her generative model, including a level of representation that encodes the semantic properties associated with upcoming verbs. Because the comprehender's belief for the event <landlords evict neighbours> in (3a) is stronger than her belief for the event <landlords evict exterminators> in (3b), her predictions of the semantic properties

of <evicted> are stronger in (3a) than (3b). This, in turn, leads to a smaller N400 to "evicted" in (3a) than (3b). In effect, the comprehender's generative model "explains away" more of the new incoming semantic evidence in (3a) then (3b).

To explain the absence of N400 modulation to verbs like "served" in sentences like (3c) versus (3d), this generative framework posits that, in both (3c) and (3d), before "served" is encountered, the comprehender has inferred (predicted) the entire specific event <waitress served customer> with fairly high probability. In (3c), this inference was based on all the semantic and syntactic evidence available in the context, just as described above. In (3d), however, it was based on a *subset* of contextual cues – the combination of "waitress" and "customer", in that linear order, and following a clause that established a restaurant schema (see below for further discussion). This is because this subset of cues offered the comprehender more *reliable* evidence to support her hypothesis that the event which *caused* the set of inputs that she had just encountered (the event the producer intended to communicate) was, <waitress served customers>, than the full set contextual cues provided for any alternative.

This explanation follows directly from the principles of a resource bound probabilistic framework, which assumes that, given the fast pace at which language unfolds and some limitation on cognitive and metabolic resources, comprehenders will draw upon their previous experience of how reliable particular sets of contextual cues are for predicting particular events or event structures. In the presence of multiple competing cues, those cue combinations that are most *reliable* are weighted during belief updating such that they have relatively more influence on what event is inferred (for general discussion of cue competition within a Bayesian framework, see Jacobs & Kruschke, 2011, and Kruschke, 2008; for discussion in relation to aspects of language comprehension, see Kleinschmidt, Fine, & Jaeger, 2012; see also Kuperberg & Jaeger, 2015, section 3.5). Formally, reliability can be quantified as the variance over previous inferences generated on the basis of a particular set of cues (see Jacobs, 2002). Thus, the assumption here is that, on the basis of all her language and real-world experience, the comprehender has accumulated and stored knowledge that the combination of "waitress" and "customer", in that linear order, and within the broader context of a restaurant schema, reliably predicts the frequent, canonical event <waitress served customer>. In contrast, the full combination of semantic and syntactic cues in the phrase, "which waitress the customer had", within a restaurant schema context, has not been encountered as frequently and is a less reliable

predictor of any alternative event. Therefore, in (3d), just ahead of encountering the verb, the comprehender will use the most reliable combination of cues to weight her inference towards the event, <waitress served customer>. This type of trade-off between different types of probabilistic predictive cues on the basis of their reliabilities has been well described in theories of animal learning (e.g. Courville, Daw, & Touretzky, 2004; Gershman & Niv, 2012; Kruschke, 2008), in other cognitive domains such as vision (e.g. Knill & Saunders, 2003) and visual–haptic interaction (e.g. Ernst & Banks, 2002), and it has been linked to theories of attentional gain and biased competition (Dayan, Kakade, & Montague, 2000; Feldman & Friston, 2010). As discussed further below, it is also reminiscent of the principles of competition models of language acquisition and processing (e.g. Bates & MacWhinney, 1987; McRae et al., 1998; Spivey & Tanenhaus, 1998).

Having now inferred the event <waitress served customer> in both (3c) and (3d) with high probability, even before encountering the upcoming verb, the comprehender generates strong predictions for the semantic properties of <serve>. Thus, when these semantic properties become available as "served" is encountered, they are "explained away", and so semantic processing of "served" is minimal in both sentences.

This generative account differs from the explanation offered by Chow et al. in two related ways. First, it does not posit distinct semantic and syntactic mechanisms with different time courses. Specifically, rather than attributing a failure to syntactically assign thematic roles before encountering the verb in (3d) to an inherently slow syntactic mechanism, it attributes this to the combination of a subset of highly reliable cues (the content of the first clause, the semantic properties and relationship between the arguments, and their linear order), which, at least temporarily, outweigh the less reliable evidence provided by the full set of semantic and syntactic cues in influencing the comprehender's hypotheses about the underlying event. Second, this generative account does not assume that, during online sentence processing, the only way a comprehender can derive or predict event-level meaning is to use all the syntactic information available to them to assign thematic roles. Like lexical constraint-based models, it posits that, in most cases, event-level meaning will be based on all semantic and syntactic information in the context. If, however, certain combinations of contextual cues within a context provide more reliable evidence for a particular event than the full set of contextual cues provide for any other event, comprehenders will, at least temporarily, infer this event with high probability. In effect, this means that, in certain situations, semantic

information will appear to temporarily override syntactic information to infer higher-level meaning, just as in certain stream-based models (e.g. Hoeks et al., 2004; Kim & Osterhout, 2005; Kuperberg, 2007; van de Meerendonk et al., 2009).

Importantly, in Chow et al.'s study (as well as in other ERP studies reporting an attenuated N400 to lexically unexpected or incoherent target words), this inference is temporary: in sentence (3d), as more information – the syntactic–thematic properties of "served" – becomes available from the bottom-up input, the full cumulated set of semantic and syntactic information within the context provides enough reliable evidence for the comprehender to further update her beliefs and infer that, in fact, the event being conveyed is more likely to be <customer served waitress> than <waitress served customer>.

This large shift from a previously inferred high certainty belief about one event to a new high certainty belief about another event is likely to also entail some updating of beliefs about the likely statistical dependencies across the semantic and syntactic representations that define the comprehender's generative model. More generally, this type of updating of beliefs about probabilistic statistical contingencies *across* levels of representation is thought to play a crucial role in allowing comprehenders to *adapt* to changes in the broader statistical structures of their ever-changing communicative environments. Such adaptation may involve modifying the comprehender's existing generative model, or it may involve switching to (or inferring) a new, previously stored generative model (for a detailed discussion of language adaptation within this type of generative framework, see Kleinschmidt & Jaeger, 2015). In Chow et al's study, I suggest that, in sentences like (3d), the comprehender's updating of beliefs about the likely statistical dependencies between semantic and syntactic representations enabled her to adapt to the wider statistical structure of the experimental context itself, such that she was able to generate more accurate predictions about the future implausible sentences that she was likely to encounter over the course of the experiment. I further suggest that this adaptation process was reflected by the larger late positivity/P600 evoked by incoming verbs like "served" in (3d) relative to (3c) (see Kuperberg, 2013 and Kuperberg, 2015 for further discussion).[5]

## 4. Extensions: inference beyond specific events

Thus far, I have discussed this generative architecture in relation to the comprehension of quite specific events

(e.g. <waitress serves customer> or <landlord evicted neighbour>), which allow comprehenders to predict, with fairly high probability, finer-grained semantic properties of incoming words. While not all contexts provide enough evidence for comprehenders to infer such specific hypotheses, many do provide enough evidence for comprehenders to reliably infer more general *event structures* – representations that specify the coarser-grained semantic properties that are necessary for participants to play particular semantic roles around certain types of actions or states (Dowty, 1989; Fillmore, 1967; Gruber, 1965; Jackendoff, 1987). Inferences at the level of event structures can be based on several different types of contextual information, including the selection restrictions of verbs (Chomsky, 1965; Katz & Fodor, 1963) and certain prominence cues that appear consistently across different languages, such as animacy in combination with either the linear order of arguments or case markings (Bates & MacWhinney, 1987; Bornkessel & Schlesewsky, 2006; Bornkessel-Schlesewsky & Schlesewsky, 2009; MacWhinney, Bates, & Kliegl, 1984). These event structure inferences can, in turn, lead comprehenders to generate strong predictions for coarse-grained semantic properties of upcoming information. For example, high certainty hypotheses about particular event structures, based on the selection restrictions of verbs (e.g. Paczynski & Kuperberg, 2011, 2012), the combination of the animacy and the linear order of arguments (e.g. Paczynski & Kuperberg, 2011; Weckerly & Kutas, 1999), and the combination of animacy and particular case markings (e.g. Frisch & Schlesewsky, 2001), can generate predictions about the animacy of upcoming arguments, leading to N400 modulation depending whether or not these animacy predictions are fulfilled.

This dynamic generative architecture can also yield insights into the comprehension of sequences of events and states in wider text and discourse contexts. Here, the goal of the comprehender is to infer the entire *situation model* being communicated (e.g. Zwaan & Radvansky, 1998) – an inference that proceeds at a level of representation that is still higher on the generative hierarchy. At times, comprehenders may only be able to infer a broad situation model that encodes likely co-occurrences between events and states but not their specific relationships – a general schema that approximates to a "bag of events". In such cases, the amplitude of the N400 evoked by a given target word will pattern with its semantic fit with this general contextual schema, and the N400 may not necessarily pattern with discourse coherence, as discussed in Section 1 (see Paczynski & Kuperberg, 2012; and Xiang & Kuperberg, 2015 for examples and further discussion).

At other times, however, the comprehender may be able to infer a richer situation model that encodes relationships between events along specific spatial, temporal, referential and causal dimensions (Zwaan & Radvansky, 1998). This effectively "narrows down" predictions to those events that are related to the inferred situation model along these specific dimensions (see also Zacks, Speer, Swallow, Braver, & Reynolds, 2007), which, in turn, leads to stronger predictions of upcoming semantic features and to an attenuation of the N400 on incoming words whose features match these predictions (e.g. Ferretti, Kutas, & McRae, 2007; Metusalem et al., 2012; Paczynski & Kuperberg, 2012). In these cases, the N400 will pattern with overall discourse plausibility or coherence.

To go beyond inferring a broad schema, and infer this type of richer situation model, comprehenders would need to draw upon their prior stored knowledge about likely and possible event sequences. This knowledge has variously been referred to as scripts (Bower, Black, & Turner, 1979; Schank & Abelson, 1977), frames (Fillmore, 2006), or structured event complexes (Wood & Grafman, 2003). Certain types of information within discourse contexts may act as particularly reliable cues, functioning to weight comprehenders' inferences towards particular types of event sequences. For example, connectives such as "because" and "therefore" may cue comprehenders to infer situation models that are characterised by particular causal and temporal relationships. Concessive connectives such as "even so", "however" and "in fact", may explicitly tell comprehenders to infer situation models that are different from the ones that they would have inferred on the basis of real-world knowledge alone (see Xiang & Kuperberg, 2015 for discussion). Certain information structural cues may play a similar role (e.g. Grondelaers, Speelman, Drieghe, Brysbaert, & Geeraerts, 2009). In addition, certain semantic classes of verbs might cue comprehenders to infer discourse structures that are characterised by particular types of referential relationships (see Hartshorne, O'Donnell, & Tenenbaum, 2015 for discussion).

Finally, and more generally, this generative approach can be extended to understand the role of pragmatic information in influencing comprehenders' inferences about the specific messages (Bock, 1987; Bock & Levelt, 1994; Dell & Brown, 1991) or speech acts (Levinson, 2013) that underlie the communicative intentions of the producer. It is well established that various pragmatic factors can influence modulation of the N400. These include the informativeness of a context (Nieuwland & Kuperberg, 2008), communicative cues like commas (Nieuwland, Ditman, & Kuperberg, 2010, Experiment 2), speech dysfluencies (Corley, MacGregor, & Donaldson, 2007), the characteristics of a speaker's voice (Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008), and non-verbal information in the surrounding context (e.g. Knoeferle, Urbach, & Kutas, 2011). Once again, these pragmatic factors may act as reliable cues that weight comprehenders' message-level inferences, leading to semantic predictions that influence how the N400 is modulated to incoming words.

## 5. Implications, open questions and conclusions

The N400 evoked by an incoming word can be characterised as reflecting the retrieval of or access to its semantic features that have not already been pre-activated or predicted by the context. In this article, I have argued against the idea that, during sentence comprehension, contextual semantic predictions result from distinct mechanisms of processing, each with distinct time courses. Rather, I have argued that they stem from a generative process that constantly seeks to infer the underlying higher-level representation – the event, event structure, situation model or overall message – that best explains or causes the full set of inputs that have just been encountered. These hypotheses, in turn, generate probabilistic predictions at lower levels of representation, including semantic features, thereby modulating the N400 evoked by incoming words.

This probabilistic generative framework retains the core insights of constraint satisfaction approaches in that comprehenders are able to use multiple types of information within a context, in conjunction with stored linguistic and real-world knowledge, in a graded fashion to derive higher-level meaning. Where it differs from lexical constraint-based approaches is that it makes no assumption that all these different types of contextual information are necessarily yoked to specific lexical entries, and that comprehenders necessarily need to achieve access to the full lexical structure of each word (its semantic, syntactic and word-form representations) in order to build higher-level meaning. Rather, it assumes that we are able to infer higher-level meaning, with different degrees of belief, based on partial bottom-up evidence, and that these higher-level hypotheses or predictions are the source of graded predictions at lower levels of representation. Moreover, I have also argued that, during belief updating, we use the reliability of the evidence available to us – multiple combinations of contextual cues within and across different levels of representation – to weight our inferences. In this sense, this framework has commonalities with competition models, which also assume that probabilistic cues can be differentially weighted and can compete

with one another depending on their reliabilities during sentence comprehension (e.g. Bates & MacWhinney, 1987; McRae et al., 1998; Spivey & Tanenhaus, 1998). Indeed, the computational principles of this framework can potentially provide a formal account of these observations (see Kleinschmidt et al., 2012 for discussion).[6]

In practice, this means that there will be occasions when comprehenders will not necessarily use all information available to them in a context at any given time to predict upcoming semantic features, resulting in temporary "semantic illusions" (e.g. Hoeks et al., 2004; Kim & Osterhout, 2005; Kuperberg, 2007; van de Meerendonk et al., 2009). There may even be times when comprehenders come up with final interpretations that are not syntactically licensed (so called "good enough" processing: Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, 2003; Ferreira, Bailey, & Ferraro, 2002; see also Sanford & Sturt, 2002; Sanford, Leuthold, Bohan, & Sanford, 2011). However, there is no need to attribute these phenomena to heuristic strategies (Ferreira, 2003), or to distinct mechanisms of processing with inherently different time courses (Bornkessel & Schlesewsky, 2006; Chow et al., this issue; Friederici & Weissenborn, 2007), or even to distinct interacting streams of processing (Hoeks et al., 2004; Kim & Osterhout, 2005; Kuperberg, 2007; van de Meerendonk et al., 2009). Rather, these phenomena follow from the computational principles of this probabilistic generative framework, according to which, if a subset of contextual cues provides more reliable evidence for a particular event or event structure than the entire set of contextual cues provides for an alternative event or event structure, belief updating will be weighted towards inferring this event or event structure with higher probability. This, in turn, will lead to more pre-activation of semantic features that are consistent with this event or event structure and so a smaller N400 to incoming words whose properties match these predictions, even if these words are lexically unexpected or implausible.

Of course, there is much work to be done in order to comprehensively test this theory. For example, it will be important to develop measures that proxy the reliability of a given context for predicting a particular event (and sets of semantic features) to determine whether such measures can predict patterns of N400 modulation better than measures of lexical probability. Formally, the reliability of a context can be quantified as the variance over previous inferences based on this context, and it can be approximated as the predictive constraint of a context. Previous work has focused on the constraint of contexts for predicting specific words (lexical constraint), which is operationalised as the entropy or relative entropy over all words that follow a particular context in a corpus (e.g. McDonald & Shillcock, 2001; Resnik, 1993) or as the probability of the word that is most frequently produced during a cloze procedure (e.g. Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Schwanenflugel & Lacount, 1988; Staub, Grant, Astheimer, & Cohen, 2015). As discussed in Section 1, however, the constraint of a given context for a specific word may not necessarily pattern with its constraint for a particular event or event structure. One way of estimating a context's event constraint may able to adapt the cloze procedure so that, rather than asking participants to produce the most likely word that follows a given context, they are asked to produce the most likely event to which the context refers ("who does what to whom"), and then calculating either the entropy over all responses, or the probability of the modal continuation over all events produced.

In developing these types of measures, it will be important to bear in mind that during real-time language comprehension, any inferences and predictions that influence the N400 must be generated very quickly, before new bottom-up evidence becomes available. There is existing evidence that the rate at which linguistic input unfolds can influence patterns of N400 modulation, (e.g. Camblin, Ledoux, Boudewyn, Gordon, & Swaab, 2007; Wlotko & Federmeier, 2015). It is also likely that these temporal dynamics contribute to discrepancies between patterns of N400 modulation during fast word-by-word processing and measures of lexical predictability as operationalised using untimed cloze tasks (see Xiang & Kuperberg, 2015 for discussion). Therefore, to determine whether the primary factor that drives semantic pre-activation and N400 modulation is indeed the reliability of a context for inferring/predicting events, it will be important to develop measures that approximate fast event inference, for example, by carrying out an event cloze procedure under time pressure (for an example of how timed cloze procedures have been used to estimate lexical contextual constraint, see Staub et al., 2015).

This generative probabilistic framework also raises questions about what particular combinations of linguistic cues, at the same and at different levels of representation, contribute to reliability of a given context for predicting a given event. As already discussed, there is evidence that certain syntactic features like argument ordering or case markings (depending on the language) can act in conjunction with certain coarse-grained semantic features such as animacy to provide particularly reliable evidence for inferring event structures (Bates & MacWhinney, 1987; Bornkessel & Schlesewsky, 2006; Bornkessel-Schlesewsky & Schlesewsky, 2009; MacWhinney et al., 1984). It is possible that some of the same

types of syntactic cues may act in conjunction with particular combinations of content words (e.g. Bicknell, Elman, Hare, McRae, & Kutas, 2010; Matsuki et al., 2011) to provide reliable evidence for inferring more specific events. For example, as noted above, it may be the combination of "waitress" and "customer", and the linear order in which these arguments appeared in the sentence that, in sentence (3d) of Chow et al.'s study, provided comprehenders with particularly reliable evidence that led them to infer the canonical event <waitress served customer>, prior to encountering the verb.

It will also be important to systematically investigate how specific sentence-level cues combine with information in the wider discourse (or non-verbal) context to influence inferences about events at any given time. For example, as discussed above, comprehenders may have been particularly likely to infer the canonical <waitress serves customer> event in sentence (3d) of Chow et al.'s study, given that the initial clause, "The restaurant owner forgot", had already established a restaurant schema. It is possible that comprehenders' belief for this particular event would have been weaker if the first clause had been more neutral (e.g. "The man forgot"), and that it would have been weaker still if the preceding discourse context set up a situation in which a group of customers were discussing how pleasant they found the waitresses, leading comprehenders to infer quite a different event (e.g. <The customer liked the waitress>). There is existing evidence that wider discourse contexts can influence effects of sentence-level incongruities or anomalies on the N400 (e.g. Nieuwland & Van Berkum, 2006; Sanford et al., 2011), and it will be important for follow-up studies to determine whether this is also true of the pattern of ERP responses evoked by critical verbs in Chow et al.'s stimuli.

Finally, because the generative model we employ at any given time is defined by our particular comprehension goals, as well as the statistical structure of our current environment, there remain important questions about when and how we adapt our existing generative models, or switch to new pre-stored models, in response to changes in either these goals and/or changes in our communicative environments. I have suggested that semantic/syntactic adaptation may be reflected by the posteriorly distributed late positivity that follows the N400 – the P600. It will be important to systematically test this hypothesis by determining exactly how the P600 varies in response to changes in the statistical contingencies of the environment (for discussion of existing evidence linking the P600 to adaptation, see Kuperberg, 2013).

There are clearly many open questions, and many of these are only now being addressed in other areas of cognitive science (e.g. Gershman & Niv, 2012; Qian, Jaeger, & Aslin, 2012). In this article, however, I have argued that this type of generative framework may provide a way of potentially tackling them under a broad unifying umbrella, and that its computational principles have the potential to explain a large body of ERP and behavioural data on event comprehension. I have also suggested that, by providing an online index of how probabilistic inference and prediction impact the fluctuating landscape of semantic memory, the N400 can give us unique insights into its temporal dynamics.

## Notes

1. This influence of a schema-based "bag of words" on processing has sometimes been attributed to simple lexico-semantic "priming" (e.g. Brouwer, Fitz, & Hoeks, 2012). The problem with the term *priming*, however, is that it has been used in many different ways, and researchers do not always spell out their assumptions about what it means (see Kuperberg & Jaeger, 2015, p. 9 for recent discussion). Lexico-semantic priming in sentence and discourse contexts was originally taken to imply an automatic spread of activation across semantic associates at a lexical level of representation that was impervious to top-down influences of long-term semantic memory (e.g. Forster, 1981; Norris, 1986; see also Fodor, 1983). This, however, is not easily reconciled with the sensitivity of the N400 to semantic facilitation rather than (necessarily) lexical facilitation (e.g., Federmeier & Kutas, 1999; Kutas & Hillyard, 1984). And, indeed, N400 effects during sentence-level processing cannot be fully explained by this type of pure lexico-semantic association (e.g. Camblin, Gordon, & Swaab, 2007; Coulson, Federmeier, Van Petten, & Kutas, 2005; Van Petten, 1993). The term *priming* has also been used in a mechanistic sense to distinguish a passive spread of activation across long-term semantic memory, from activation that stems from a representation of context that is held within working memory (e.g. Lau, Holcomb, & Kuperberg, 2013). Finally, in memory-based models of text and discourse comprehension (e.g. Kintsch, 1988; Myers & O'Brien, 1998; Sanford, 1990; Sanford & Garrod, 1998), the term priming has sometimes been used refer to any type of semantic facilitation of an incoming word by its context, without specific assumptions about either architecture or mechanism. This, however, leaves open the question under consideration here – what types of information are comprehenders able to use within a given context to facilitate semantic processing of incoming words?

2. The use of the term, *generative* here should not be confused with its use to describe a class of linguistic theories, for example, Generative Syntax (although there are some broad conceptual similarities). In this article, I use the term, *generative* to refer to a class of computational frameworks in which an underlying latent cause is viewed as probabilistically *generating* observations, and in which the goal of the agent is to "discover" this underlying cause on the basis of these observations (for introductory overviews of how probabilistic generative frameworks can be used

to study cognition, see Jacobs & Kruschke, 2011 and Perfors, Tenenbaum, Griffiths, & Xu, 2011).

3. Several formal computational models of language processing have been developed that are generative in nature (for references, see Kuperberg & Jaeger, 2015, sections 1.2 and 2.2). Some of these are dynamic in the sense that beliefs are updated incrementally on the basis of new inputs that unfold over time (e.g. Levy, 2008; Kleinschmidt & Jaeger, 2015). In addition, Kleinschmidt and Jaeger (2015) discuss the idea that comprehenders store multiple generative models that correspond to their beliefs about the distributions of statistical inputs that characterize particular talkers or groups of talkers. The framework that I sketch out below shares many of the same principles of these existing computational models. However, it differs from these existing formalisations in two main ways. First, rather than new bottom-up evidence becoming available only at a single level of representation, it assumes that new evidence becomes available at multiple levels of representation. Second, rather than waiting for new bottom-up evidence to become available before beliefs can be updated at a given level of representation, it assumes that comprehenders can use higher-level beliefs to *actively generate* lower-level predictions, thereby changing prior beliefs at lower levels of representation before new bottom-up evidence is encountered (predictive pre-activation; for discussion, see Kuperberg & Jaeger, 2015, section 3.5).

4. The prior beliefs or *hypotheses* at the highest level of the comprehender's generative model essentially also constitute probabilistic predictions in the sense that they can be based on incomplete bottom-up evidence. Conversely, the *predictions* at lower levels of representation constitute prior beliefs or hypotheses at these lower representational levels. The assumption is that probabilistic inference proceeds not only at the top of the generative hierarchy, but at all levels of representation below it. Within this framework, prediction error is itself a consequence of belief updating or inference. It is formalized as *Bayesian surprise* – the degree to which the comprehender's beliefs *shift* from her *prior* probabilistic predictions at a given representational level to what actually happens after observing new evidence to form a new *posterior* distribution of beliefs, on the basis of Bayes' rule (Doya, Ishii, Pouget, & Rao, 2004). This new posterior distribution then becomes the prior distribution for the next cycle of inference. On some accounts, like predictive coding, prediction error is the *only* signal that is passed up the generative hierarchy (Clark, 2013; Friston, 2005).

5. As Chow et al. point out, a larger late positivity/P600 was also evoked by critical verbs in (3b) than (3a), although this late positivity appeared to begin later than the positivity in (3d). One possibility is that the early divergence of the late positivity in (3d) reflected the large shift in belief from the event that was originally inferred, to the event that was inferred when additional evidence became available, while the later divergence seen in both (3d) and (3b) reflected the adaptation process itself (belief updating about the likely semantic-syntactic statistical contingencies that define the structure of the generative model). An important caveat of this interpretation, however, is that component overlap with the earlier N400 makes it difficult to determine exactly when these late positivities began (see Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007 for discussion).

6. It is also possible that certain cues may point comprehenders *away* from inferring the overall meaning of events, and towards recognizing information that is particularly salient or emotional (for evidence that these types of cues can influence N400 modulation, see Delaney-Busch & Kuperberg, 2013; and Lotze, Tune, Schlesewsky, & Bornkessel-Schlesewsky, 2011).

## References

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York, NY: John Wiley & Sons.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, *63*(4), 489–505. doi:10.1016/j.jml.2010.08.004

Bock, J. K. (1987). Exploring levels of processing in sentence production. In G. Kempen (Ed.), *Natural language generation* (pp. 351–363). Dordrecht: Martinus Nijhoff.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). London: Academic Press.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*(4), 787–821.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass*, *3*(1), 19–58. doi:10.1111/j.1749-818X.2008.00099.x

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*(2), 177–220. doi:10.1016/0010-0285(79)90009-4

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143. doi:10.1016/J.Brainres.2012.01.055

Brown, M., & Kuperberg, G. R. (2015). A hierarchical generative framework of language processing: Linking language perception, interpretation, and production abnormalities in schizophrenia. *Frontiers in Human Neuroscience*, *9*. doi:10.3389/fnhum.2015.00643

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*, *56*(1), 103–128.

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., & Swaab, T. Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVP. *Brain Research*, *1146*, 172–184. doi:10.1016/j.brainres.2006.07.033

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chow, W., Smith, C., Lau, E., Phillips, C., & House, C. (2015). A "bag-of-arguments" mechanism for initial verb predictions. *Language, Cognition, & Neuroscience*. Advance online publication. doi:10.1080/23273798.2015.1066832

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. doi:10.1017/S0140525X12000477

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*(3), 658–668. doi:10.1016/j.cognition.2006.10.010

Coulson, S., Federmeier, K. D., Van Petten, C., & Kutas, M. (2005). Right hemisphere sensitivity to word- and sentence-level context: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 129–147. doi:10.1037/0278-7393.31.1.129

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2004). *Similarity and discrimination in classical conditioning: A latent variable account*. Paper presented at the Advances in Neural Information Processing Systems 17 (NIPS 2004), Vancouver, BC.

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3*(Suppl.), 1218–1223. doi:10.1038/81504

Delaney-Busch, N., & Kuperberg, G. (2013). Friendly drug-dealers and terrifying puppies: Affective primacy can attenuate the N400 effect in emotional discourse contexts. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(3), 473–490. doi:10.3758/s13415-013-0159-5

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener". In D. J. Napoli & J. A. Kegl (Eds.), *Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman* (pp. 105–129). Hillsdale, NJ: Lawrence Erlbaum Associates.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. doi:10.1038/nn1504

Dowty, D. R. (1989). On the semantic content of the notion of thematic role. In G. Cherchia, B. Partee & R. Turner (Eds.), *Properties, types and meaning: Volume II: Semantic issues*. (pp. 69–129). Norwell, MA: Kluwer.

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. N. (2004). *Bayesian brain: Probabilistic approaches to neural coding*. Cambridge, MA: MIT Press.

Elman, J. L., Hare, M., & McRae, K. (2004). *Cues, constraints, and competition in sentence processing beyond nature-nurture: Essays in honor of Elizabeth Bates* (pp. 111–138). Mahwah, NJ: Lawrence Erlbaum Associates.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. doi:10.1038/415429a

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. doi:10.1006/Jmla.1999.2660

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. doi:10.1016/j.brainres.2006.06.101

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215. doi:10.3389/fnhum.2010.00215

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164–203. doi:10.1016/S0010-0285(03)00005-7

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. doi:10.1111/1467-8721.00158

Ferreira, F., & Clifton, Jr. C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368. doi:10.1016/0749-596X(86)90006-9

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 182–196.

Fillmore, C. J. (1967). *The case for case*. Paper presented at the Texas Symposium on Language Universals.

Fillmore, C. J. (2006). Frame semantics. *Cognitive Linguistics: Basic Readings*, *34*, 373–400. doi:10.1515/9783110199901.373

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *33*(4), 465–495. doi:10.1080/14640748108400804

Frazier, L. (1987). Sentence processing: A tutorial review. Attention and performance XII. In M. Coltheart (Ed.), *The psychology of reading* (pp. 559–586). London: Lawrence Erlbaum Associates.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325. doi:10.1016/0010-0277(78)90002-1

Friederici, A. D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax-semantic interface. *Brain Research*, *1146*, 50–58. doi:10.1016/j.brainres.2006.08.038

Frisch, S., & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *Neuroreport*, *12*(15), 3391–3394.

Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360* (1456), 815–836. doi:10.1098/Rstb.2005.1622

Friston, K. J., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, 1–28. doi:10.1080/17588928.2015.1020053

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*(1), 58–93. doi:10.1006/Jmla.1997.2512

Gershman, S. J., & Niv, Y. (2012). Exploring a latent cause theory of classical conditioning. *Learning and Behavior*, *40*(3), 255–268. doi:10.3758/s13420-012-0080-8

Grondelaers, S., Speelman, D., Drieghe, D., Brysbaert, M., & Geeraerts, D. (2009). Introducing a new entity into discourse: Comprehension and production evidence for the status of Dutch er "there" as a higher-level expectancy monitor. *Acta Psychologica*, *130*(2), 153–160. doi:10.1016/j.actpsy.2008.11.003

Gruber, J. S. (1965). *Studies in lexical relations* (Doctoral dissertation). Massachusetts Institute of Technology.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 819–836). Cambridge, MA: MIT Press.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438–441.

Hare, M., Tanenhaus, M. K., & McRae, K. (2007). Understanding and producing the reduced relative construction: Evidence from ratings, editing and corpora. *Journal of Memory and Language*, *56*(3), 410–435. doi:10.1016/j.jml.2006.08.007

Hartshorne, J. K., O'Donnell, T. J., & Tenenbaum, J. B. (2015). The causes and consequences explicit in verbs. *Language, Cognition and Neuroscience*, *30*(6), 716–734. doi:10.1080/23273798.2015.1008524

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*, *11*(10), 428–434. doi:10.1016/j.tics.2007.09.004

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*, 59–73.

Jackendoff, R. (1987). The status of thematic relations in linguistic theory. *Linguistic Inquiry*, *18*(3), 369–411.

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, *6*(8), 345–350.

Jacobs, R. A., & Kruschke, J. K. (2011). Bayesian learning theory applied to human cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(1), 8–21. doi:10.1002/wcs.80

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language*, *39*, 170–210.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225. doi:10.1016/J.Jml.2004.10.002

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182. doi:10.1037/0033-295X.95.2.163

Kintsch, W. (2001). Predication. *Cognitive Science*, *25*(173–202). doi:10.1207/s15516709cog2502_1

Kleinschmidt, D. F., Fine, A. B., & Jaeger, T. F. (2012). A belief-updating model of adaptation and cue combination in syntactic comprehension. In N. Miyake, D. Peebles & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 605–610). Sapporo: Cognitive Science Society.

Kleinschmidt, D. F., & Jaeger, F. T. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. doi:10.1037/a0038695

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539–2558. doi:10.1016/s0042-6989(03)00458-9

Knoeferle, P., Urbach, T. P., & Kutas, M. (2011). Comprehending how visual context influences incremental sentence processing: Insights from ERPs and picture-sentence verification. *Psychophysiology*, *48*(4), 495–506. doi:10.1111/j.1469-8986.2010.01080.x

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*(1), 1–36. doi:10.1016/S0093-934X(02)00548-5

Kruschke, J. K. (2008). Bayesian approaches to associative learning: From passive to active learning. *Learning and Behavior*, *36*(3), 210–226. doi:10.3758/lb.36.3.210

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49. doi:10.1016/j.brainres.2006.12.063

Kuperberg, G. R. (2013). The proactive comprehender: What event-related potentials tell us about the dynamics of reading comprehension. In B. Miller, L. Cutting & P. McCardle (Eds.), *Unraveling reading comprehension: Behavioral, neurobiological, and genetic components* (pp. 176–192). Baltimore, MD: Paul Brookes.

Kuperberg, G. R. (2015). What event-related potentials might tell us about the neural architecture of language comprehension. Manuscript submitted for publication.

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes*, *21*(5), 489–530. doi:10.1080/01690960500094279

Kuperberg, G. R., & Jaeger, T. F. (2015). What do we mean by prediction in language comprehension? *Language, Cognition, and Neuroscience*. doi:10.1080/23273798.2015.1102299

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, *100*(3), 223–237. doi:10.1016/j.bandl.2005.12.006

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, *23*(5), 1230–1246. doi:10.1162/jocn.2010.21452

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*(1), 117–129. doi:10.1016/S0926-6410(03)00086-7

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. doi:10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, *11*(2), 99–116.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. doi:10.1038/307161a0

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), 211–240. doi:10.1037/0033-295x.104.2.211

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484–502. doi:10.1162/jocn_a_00328

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. doi:10.1038/nrn2532

Levinson, S. C. (2013). Action formation and ascription. In T. Stivers & J. Sidnell (Eds.), *The handbook of conversation analysis* (pp. 103–130). Malden, MA: Wiley-Blackwell.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. doi:10.1016/j.cognition.2007.05.006

Lotze, N., Tune, S., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2011). Meaningful physical changes mediate lexical-semantic integration: Top-down and form-based bottom-up information sources interact in the N400. *Neuropsychologia*, *49*(13), 3573–3582. doi:10.1016/j.neuropsychologia.2011.09.009

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703. doi:10.1037/0033-295X.101.4.676

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Memory and Language*, *23*, 127–150.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102. doi:10.1016/0010-0277(87)90005-9

Marslen-Wilson, W. D., Brown, C., & Tyler, L. K. (1988). Lexical representations in spoken language comprehension. *Language and Cognitive Processes*, *3*, 1–16. doi:10.1080/01690968808402079

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 913–934. doi:10.1037/a0022964

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. doi:10.1037//0033-295x.88.5.375

McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, *44*(3), 295–322. doi:10.1177/00238309010440030101

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, *12*(2–3), 137–176. doi:10.1080/016909697386835

McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass*, *3*(6), 1417–1429. doi:10.1111/j.1749-818X.2009.00174.x

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*(3), 283–312. doi:Doi 10.1006/Jmla.1997.2543

van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, *3*(5), 1211–1224. doi:10.1111/j.1749-818X.2009.00163.x

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, *66*(4), 545–567. doi:10.1016/j.jml.2012.01.001

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26*(2&3), 131–157. doi:10.1080/01638539809545042

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. doi:10.1111/j.1467-9280.2008.02226.x

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111. doi:10.1162/Jocn.2006.18.7.1098

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*(3), 324–346. doi:10.1016/j.jml.2010.06.005

Norris, D. (1986). Word recognition: Context effects without priming. *Cognition*, *22*(2), 93–136. doi:10.1016/s0010-0277(86)90001-6

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, *8*(1), 89–97.

Paczynski, M., & Kuperberg, G. R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb argument processing. *Language and Cognitive Processes*, *26*(9), 1402–1456. doi:10.1080/01690965.2011.580143

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, *67*(4), 426–448. doi:10.1016/j.jml.2012.07.003

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321. doi:10.1016/j.cognition.2010.11.015

Qian, T., Jaeger, T. F., & Aslin, R. N. (2012). Learning to represent a multi-context environment: More than detecting changes. *Frontiers in Psychology*, *3*, 228. doi:10.3389/fpsyg.2012.00228

Rabovsky, M., & McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition*, *132*(1), 68–89. doi:10.1016/j.cognition.2014.03.010

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358–374.

Resnik, P. S. (1993). *Selection and information: A class-based approach to lexical relationships* (Doctor of Philosophy). University of Pennsylvania.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89(1), 60–94. doi:10.1037/0033-295x.89.1.60

Sanford, A. J. (1990). On the nature of text-driven inference. In D. A. Balota, F. d'Arcais & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 515–538). Hillsdale, NJ: Erlbaum.

Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, 26(2–3), 159–190. doi:10.1080/01638539809545043

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, 23, 514–523.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, 6, 382–386.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schwanenflugel, P. J., & Lacount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning Memory and Cognition*, 14(2), 344–354. doi:10.1037//0278-7393.14.2.344

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1521–1543. doi:10.1037/0278-7393.24.6.1521

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology*, 47(2), 276–309. doi:10.1037/h0078826

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, 82, 1–17. doi:10.1016/j.jml.2015.02.004

Stowe, L. A. (1989). Thematic structures and sentence comprehension. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structures in language processing* (pp. 318–357). Dordrecht: Kluwer Academic.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Towards a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155–179). Hillsdale, NJ: Erlbaum.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285–318. doi:10.1006/jmla.1994.1014

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19(3), 528–553. doi:10.1037/0278-7393.19.3.528

Urbach, T. P., DeLong, K. A., & Kutas, M. (2015). Quantifiers are incrementally interpreted in context, more than less. *Journal of Memory and Language*, 83, 79–96. doi:10.1016/j.jml.2015.03.010

Van Berkum, J. J. A. (2009). The neuropragmatics of "simple" utterance comprehension: An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Basingstoke: Palgrave Macmillan.

Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6), 657–671.

Van Berkum, J. J. A., Van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4), 580–591.

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. Special Issue: Event-related brain potentials in the study of language. *Language and Cognitive Processes*, 8, 485–531.

Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*, 36(5), 559–570.

Wilson, M. P., & Garnsey, S. M. (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language*, 60(3), 368–392. doi:10.1016/j.jml.2008.09.005

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, 62(1), 356–366. doi:10.1016/j.neuroimage.2012.04.054

Wlotko, E. W., & Federmeier, K. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, 68, 20–32. doi:10.1016/j.cortex.2015.03.014

Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: Processing and representational perspectives. *Nature Reviews Neuroscience*, 4(2), 139–147. doi:10.1038/Nrn1033

Xiang, M., & Kuperberg, G. R. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, 30(6), 648–672. doi:10.1080/23273798.2014.995679

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, 133(2), 273–293. doi:10.1037/0033-2909.133.2.273

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185. doi:10.1037/0033-2909.123.2.162