International Journal of Psychophysiology 83 (2012) 176–190

Contents lists available at SciVerse ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Review

Prediction during language comprehension: Benefits, costs, and ERP components

Cyma Van Petten [a,*], Barbara J. Luka [b]

[a] Department of Psychology, Binghamton University, State University of New York, Binghamton, NY, United States
[b] Psychology Program, Bard College, Annandale-on-Hudson, NY, United States

ARTICLE INFO

Article history:
Received 28 August 2011
Received in revised form 10 September 2011
Accepted 14 September 2011
Available online 19 October 2011

Keywords:
Prediction
Sentence
Event-related potential
N400
P600
P300

ABSTRACT

Because context has a robust influence on the processing of subsequent words, the idea that readers and listeners predict upcoming words has attracted research attention, but prediction has fallen in and out of favor as a likely factor in normal comprehension. We note that the common sense of this word includes both benefits for confirmed predictions and costs for disconfirmed predictions. The N400 component of the event-related potential (ERP) reliably indexes the benefits of semantic context. Evidence that the N400 is sensitive to the other half of prediction – a cost for failure – is largely absent from the literature. This raises the possibility that "prediction" is not a good description of what comprehenders do. However, it need not be the case that the benefits and costs of prediction are evident in a single ERP component. Research outside of language processing indicates that late positive components of the ERP are very sensitive to disconfirmed predictions. We review late positive components elicited by words that are potentially more or less predictable from preceding sentence context. This survey suggests that late positive responses to unexpected words are fairly common, but that these consist of two distinct components with different scalp topographies, one associated with semantically incongruent words and one associated with congruent words. We conclude with a discussion of the possible cognitive correlates of these distinct late positivities and their relationships with more thoroughly characterized ERP components, namely the P300, P600 response to syntactic errors, and the "old/new effect" in studies of recognition memory.

1. Prediction in (behavioral) psycholinguistic theory: Fifty years of variable opinion

The standard usage of *predict* corresponds well to its origins in Latin – *pre* (before or in front of) plus *dicere* (to speak) – to declare what will happen in the future. Outside the realm of verbal acts, a great deal of behavior has a predictive quality, from the rapid timing of movements (predicting the trajectory of a moving object in order to catch or avoid it) to slower acts preceded by more deliberate decision-making (installing solar panels in anticipation of higher fossil-fuel costs). These extended senses of "prediction" retain a core feature of the basic version, namely that predictions can be correct or incorrect and accordingly produce benefits or costs. In some laboratory paradigms, both the benefits and costs of prediction are clearly evident in performance. For instance, as compared to no prior information, a cue signaling the most likely location of an upcoming target leads to faster responses when valid, but slower responses when invalid (Posner et al., 1980).

For language comprehension, the idea that readers and listeners spontaneously make predictions about upcoming words (as opposed to simply processing them when they arrive) has fallen in and out of favor among psycholinguists. In some of the first studies to show that sentence context improved the identification of words in noise or after brief exposure durations, the authors assumed that people generate hypotheses about upcoming words (Miller and Isard, 1963; Tulving and Gold, 1963). For instance, Tulving and Gold (1963 pg 327) concluded that:

> The findings of the present experiments are relevant to the hypothesis theory of perception… According to this theory, perception depends upon two classes of variable — a) stimulus factors, and b) expectancies or hypotheses of the organism. Two basic theorems relate strength of hypothesis to perceptual information. The first says that the greater the strength of the hypothesis, the less the amount of appropriate information necessary to confirm it. The second states that the greater the strength of the hypothesis, the more the amount of inappropriate or contradictory information necessary to infirm it. Two theorems seem to be necessary to specify the relations involved since the theory is focused on hypotheses and their confirmation or infirmation.

By the 1980s, doubts about prediction (or *priming*[1] or *pre-activation*) as a mechanism engaged during sentence processing arose from

* Corresponding author at: Dept. of Psychology, Binghamton University, State University of New York, 4400 Vestal Parkway East, Binghamton, NY, 13902, United States.
E-mail address: cvanpett@binghamton.edu (C. Van Petten).

0167-8760/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpsycho.2011.09.015

[1] We restrict the use of the word *priming* to its mechanistic sense of pre-activation of a mental representation or process, before a physical stimulus that triggers the corresponding process is presented. *Context effect* is used to describe behavioral or neural responses that vary depending on the nature of the prior stimuli, because this phrase does not imply a particular mechanism.

intuitions about the low likelihood of success. Forster (1981) provided an articulate summary of the "low payoff" intuition.

> Suppose it was possible to design an experiment which showed a sentence context effect across a class of words that could not be considered a natural class. How could such an effect occur? Obviously, there would have to be an inferential process which systematically enumerated the possible completions. For example, suppose the context was "The bird flew down onto the…" and our experiment showed that words such as BRANCH, PERCH, LEDGE, LAWN, VERANDAH, etc., were all facilitated, whereas words such as MIRROR, MOON, PENCIL,[2] etc., were not. Since there is no obvious semantic property common to the facilitated words (other than being plausible completions of the context sentence), it follows that each plausible completion would have to be independently discovered prior to the target word actually occurring. Such an enumeration procedure would surely require considerable time to reach completion, but in order for it to serve any purpose at all, it must be completed on line before the target word is presented. Even more puzzling is the fact that this enumeration must in some sense be a more efficient procedure than relying solely on the stimulus properties of the target word. (pg 468)

Jackendoff (2002, pg 59) repeated essentially the same intuition some twenty years later: "One might well predict that what comes after 'The big star's beside a little…' is likely to be a noun (though it might be BLUE or VERY OLD), but that still leaves open some tens of thousands of choices." Implicit in these critiques of prediction is the assumption that infrequent payoffs are not worth pursuing because 1) the process of prediction consumes some resource that could be applied more fruitfully, and/or 2) a failed prediction is worse than none at all — like an invalid attentional cue, it hinders the processing of an unexpected item.

Behavioral psycholinguists in the 1970s and 1980s examined the impact of sentence contexts on speeded processing of their final words (naming or lexical decision times) with an eye to both the benefits for predictable words, and the costs for unpredictable words. *Predictability*, both then and now, is quantified by behavior in an unspeeded task, namely a paper-and-pencil procedure in which a normative group (separate from the experimental subjects) is given some sentence frames and asked to fill in the most likely final word for each. For a given sentence frame, each word in the set of completions offered by the group is assigned a *cloze probability* — the percentage of subjects who used that particular word, from zero to 100%. (The fact that research subjects are universally able to perform cloze probability tasks with minimal instruction indicates, of course, that predicting words is not a difficult chore; at issue is what readers and listeners do spontaneously, in real time). Assessing both the benefits accruing to high-cloze words and the costs accruing to low-cloze words (or the extreme case of anomalous words) required a neutral baseline. Across studies, the most-favored sort of baseline consisted of *neutral sentence frames* such as "The next word is…", "At the end of this sentence, please say…", "They were thinking about the…", etc. As compared to this sort of baseline, the dominant pattern of results was facilitation (faster RTs) for high-cloze sentence completions without corresponding inhibition (slower RTs) for anomalous or low-cloze completions (Forster, 1981; Schwanenflugel and LaCount, 1988; Schwanenflugel and Shoben, 1985; Stanovich and West, 1981, 1983; see Traxler and Foss, 2000 for similar more recent results). Occasional observations of slow responses to words that might disconfirm a prediction were restricted to situations that appeared more distant from normal reading: visually-degraded stimuli that yielded high error rates (Stanovich and West, 1983), or contrasts to baselines that consisted of rows of x's (Fischler and Bloom, 1985) or random word strings with no syntactic structure (Forster, 1981).

Given the apparent absence of processing costs for unpredictable words, the idea that sentence comprehension included a predictive mechanism was largely abandoned in the 1980s. The benefits of supportive semantic context were instead attributed to *automatic spreading activation* from the sentence context to lexical entries (West and Stanovich, 1982) or greater *ease of integration* when the semantic properties of the context match those of the target word (Schwanenflugel and Shoben, 1985; Traxler and Foss, 2000). Automatic spreading activation was popular as an account of context effects in word pairs, and appeals to the idea that related words are "nearby" in an organized mental lexicon. It can be discarded as an account of sentence context effects because sentences are, by definition, an infinite set that cannot be pre-stored (see Van Petten, 1993; Coulson et al., 2005 for longer critiques). The dominant descriptions of "integration" in the 1980s were offered by reading researchers, who stipulated that this process was slow to begin, and occurred only after the meanings of all the words had been completely retrieved (e.g., Kintsch, 1988; Seidenberg et al., 1982; see Tabossi, 1991; Van Petten and Kutas, 1991b for reviews and critiques). "Integration" thus meant *delayed* integration, perhaps even waiting for clause or sentence boundaries to catch up with word recognition. In contrast, researchers working with spoken language could more readily track the amount of physical information available to a listener simply because auditory input accrues over time. Some of these latter investigators favored more inclusive descriptions of "integration" that incorporated an early combination of semantic constraints from the prior context with incomplete perceptual information about the current word (e.g., Moss and Marslen-Wilson, 1993; Tyler and Wessels, 1983; see Van Petten et al., 1999 for a longer summary that includes conflicting views). However, even these theories with more extensive interactions between top–down contextual information and bottom–up perceptual input did not include a role for prediction per se, as candidate words were held to be generated from perceptual information only — context was allowed to speed or improve selection from the set of candidates that were consistent with the auditory input, but not to suggest words *de novo*.

In the late 1990s, views among reading researchers underwent a dramatic shift, such that some of the same researchers who had espoused delayed integration of word meanings with prior context now proposed "a language processing system in which semantic interpretation, as well as syntactic processing, is conducted incrementally, with early integration of contextual information" (Sedivy et al., 1999, pg 109; but see also Altmann and Steedman, 1988 for early adopters). In part, this shift of opinion was due to the use of a new behavioral measure – eye tracking – that allowed the visualization of data across time (Altmann and Kamide, 1999; Tanenhaus et al., 1995). These incremental eye-tracking measures are, of course, more like ERPs in offering a continuous stream of data, rather than a single discrete response when subjects press a button. Even in behavioral studies without eye-tracking, it is now more standard to argue for immediate application of all sources of information during the processing of sentences and discourse (Matsuki et al., 2011), although this continues to be a topic of debate (Bornkessel and Schlesewsky, 2006).

2. The N400

2.1. Rapid, graded, and incremental benefits of semantic context

In contrast to the dominant stream of thought among behavioral psycholinguistics in the 1980s and early 1990s, ERP sentence experiments in the same era showed early influences of semantic context on

[2] Throughout this review, we enclose incomplete sentence contexts (sentence frames) in quotes, and use small capital letters to flag critical words. Quotes from other investigators have been edited to create a standard format. In actual experiments, critical items are presented in the same font as other sentence words.

the processing of individual words. In Kutas and Hillyard's (1980a) initial comparisons between semantically congruent and incongruent sentence-final words and in many subsequent studies, the larger negative wave (N400) elicited by the incongruent endings began at roughly 200 ms after visual word presentation. Studies using spoken sentences showed that the onset (and usually the peak) latency of the N400 context effect was well before the acoustic offset of the eliciting words (McCallum et al., 1984; Holcomb and Neville, 1991). This was not especially surprising given that, even when presented in isolation, most English words can be identified well before their offsets (Grosjean, 1980). A later study examined the timecourse of spoken word identification before embedding the critical words in sentence contexts, via the *gating* method. In this method, listeners are presented with only the first 50 ms of word, or the first 100 ms, etc. (in increments of 50 ms), and forced to guess/decide what the word might be. With brief amounts of acoustic input, the number of words generated may be as large as the number of participants, but at some point, the large majority of participants correctly specify the actual word. The signal duration when the acoustic information is sufficient to uniquely identify one word, and eliminate alternatives with similar onsets is the *isolation point* (e.g., the duration that allows CAPTAIN to be distinguished from CAPTIVE and CAPSULE and CAPTION). When the words served as congruent or incongruent sentence completions, the average ERP sentence congruity effect began some 200 ms *before* the isolation point, if the initial phonemes of the spoken words were inconsistent with the semantically preferred completion (Van Petten et al., 1999; see also van den Brink et al., 2006). When the initial phonemes were consistent with the preferred completion, but the auditory signal continued to form some other word (i.e., CAPTIVE at the end of a sentence for which CAPTAIN was congruent), the ERPs elicited by the congruent and incongruent words diverged at a time very close to the isolation point — the moment when a listener could determine they were not hearing the congruent word. These studies show that there is no delay in utilizing semantic context.

Other early ERP studies showed that the benefits of semantic context for reducing N400 amplitude are graded, rather than showing a sharp cutoff between congruent and incongruent sentence completions. Using the cloze probability measure described above, Kutas and colleagues showed that congruent sentence completions elicited N400s whose amplitudes scaled inversely with their cloze probability. A word offered by 90% of the normative group thus elicited a smaller N400 than one offered by 70% of the normative group, which in turn elicited a smaller N400 than one offered by only 30% of the normative group (Kutas and Hillyard, 1984; Kutas et al., 1984). Although the cloze probability effect on N400 amplitude has been replicated numerous times (e.g., Besson et al., 1997; Diaz and Swaab, 2007; Moreno et al., 2002; Van Petten et al., 1999), its exact interpretation is subject to an ambiguity. Some authors seem to assume that cloze probabilities can be mapped onto the expectations of individual subjects on a trial-by-trial basis (DeLong et al., 2005). The idea is that if a sentence frame elicited four completions from the normative group with frequencies of 50, 35, 10 and 5% for Words A, B, C and D, a subject in the ERP experiment would generate a probabilistic expectation that the sentence was 50% likely to end with Word A, 35% likely to end with Word B, etc. This assumption strikes us as slightly strange because it is akin to mapping the results of an election onto the mind of an individual voter (i.e., Jane Smith from New York was 62.2% in favor of Obama and 36.7% in favor of McCain, given that those were the statewide results). Because ERPs are formed by averaging multiple trials, the data are equally compatible with the idea that expectations are probabilistic only across sentences and not for individual trials. For example, if confronted with 100 different sentences with four possible completions, a given participant might expect completion A on 50 of those trials, completion B on 35 of those trials, etc.

Early ERP studies also showed that application of sentence context is incremental and does not await "wrap-up" processes at the end of the sentence. This was first demonstrated via the similarity between N400 effects for incongruent words in sentence-medial and sentence-final positions (both as compared to congruent words; Kutas and Hillyard, 1983). More critically, *congruent* sentence words also elicit sizeable N400s when they occur early in a sentence — a point at which only minimal context is available to exert either a beneficial or a detrimental influence (e.g., "The…", "Because they …"). Examination of the ERPs elicited by the intermediate words of sentences presented one at a time in serial order showed large N400s for the first open-class words, which became progressively smaller as the sentence context (specifically semantic) accrued and constrained subsequent words (Van Petten and Kutas, 1990, 1991a; Van Petten, 1993; see Dambacher et al., 2006 for recent confirmation). This sentence-position effect on N400 amplitude was observed only in isolated sentences for which readers had no prior inkling of the sentence topic, and not for sentences in connected text (discourse) which do not introduce completely new topics (Van Petten, 1995).

Overall, the first twenty or so years of sentence-processing research using ERPs did not suggest that incongruent or unrelated words elicit unusual brain activity, but rather that a large negativity between 200 and 500 ms or so (the N400) is the default response to words, and that its amplitude is reduced to the degree that context aids in their interpretation (see Hagoort and van Berkum, 2007; Kutas et al., 2006; Kutas and Federmeier, 2011 for more extensive reviews). The brief review above is consistent with either of two views: that prior context stimulates the retrieval of knowledge from semantic memory and this information is rapidly integrated with (even partial) perceptual input about the current word, or that words are actively predicted such that reduced N400 amplitudes reflect the benefits of confirmed predictions.

2.2. Prediction versus rapid integration: mixed evidence

One robust finding argues that readers rapidly compare the meanings of words they read or hear to the prior context, but do not predict specific words in advance. In some of the first N400 studies published, Kutas introduced the *related anomaly* paradigm in which she contrasted high-cloze congruent completions, anomalous completions, and anomalous completions that were semantically related to the congruent words (e.g., "The pizza was too hot to …" EAT/CRY/DRINK; Kutas et al., 1984; Kutas and Hillyard, 1984). The related anomalies elicited an N400 that was larger than the congruent endings, but substantially smaller than the unrelated anomalies. A parallel to the related anomaly effect is observed for sentence completions that are congruent, but not highly favored in a cloze probability procedure. For instance, if 80% of a normative group offer BRAIN as the completion for the sentence frame "He was afraid that doing drugs would damage his…", 10% offer MIND and 10% offer REPUTATION, MIND will elicit a smaller N400 than REPUTATION when presented to a separate group of ERP subjects. Fig. 1 illustrates this effect. Both sets of results argue for a featural semantic representation of words, and suggest that a sentence context facilitates the processing of words containing at least some features that can be matched to the specifications of the preceding sentence fragment.

Federmeier refined the related anomaly design by constructing contexts that more narrowly constrained the semantic features that would form a good fit (Federmeier, 2007; Federmeier and Kutas, 1999a, 1999b; Federmeier et al., 2002). For instance, although both EARRING and NECKLACE are types of jewelry and thus share many semantic features, they also differ in multiple properties such that EARRING is a better completion for the context "I guess his girlfriend really encouraged him to get it pierced. But his father sure blew up when he came home wearing that…" In contrast, NECKLACE is a better completion for the context "She keeps twirling it around and around under her collar. Stephanie seems really happy that Dan gave her that …" The ERPs thus showed larger N400s for the wrong

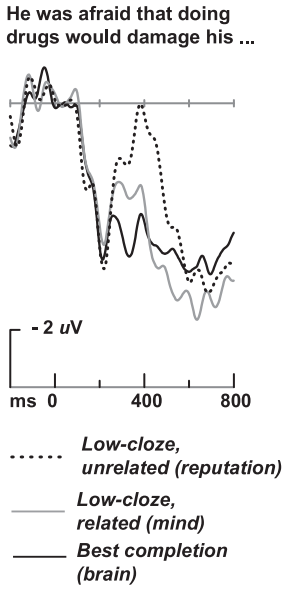

Fig. 1. Grand average ERPs at a midline parietal scalp site (Pz) from 24 subjects. Completions of high-constraint sentence frames were presented visually. Note that, despite having the same cloze probability (~4%), words semantically related to the best completion elicit a smaller N400 than words unrelated to the best completion.
Data from Thornhill and Van Petten (submitted for publication).

variety of jewelry (or the wrong team sport, wrong hand tool, etc.) than for the congruent word. More critically however, these incongruent words that had high featural overlap with the congruent word elicited smaller N400s than words with low featural overlap (such as MASCARA or LIPSTICK for the examples above).

Overall, the related-anomaly experiments indicate that context leads to expectations about the meaning of upcoming words, and suggest that the meanings of words actually presented are rapidly compared with those expectations. The results argue against the idea that readers predict specific words – or at least that the N400 reflects such predictions – given that the related anomalies would never appear on a list of hypothesized completions. The results are thus more consistent with the idea that N400 sentence context effects reflect rapid integration than confirmation/disconfirmation of a prediction. For the remainder of the review, we will reserve the word "prediction" to mean that a reader or listener is expecting a specific word (lexical item) to occur in the future. We will use "expectation" as a broader umbrella term to indicate that a reader/listener anticipates some semantic content, and may or may not have narrowed that expectation to a particular word. The flexibility of natural languages is such that the same concept can usually be expressed in numerous ways, by different words. There is little debate that comprehenders frequently form hypotheses about upcoming content; at issue is whether these expectations take the form of specific lexical predictions. Note that we will also refer to "more predictable" and "less predictable" words, by which we mean *potentially* predictable (i.e., that if comprehenders actually make predictions, the "more predictable" items should be the targets that benefit from this activity).

Although the related anomaly experiments suggest that sentence and discourse contexts act to specify the meanings of plausible continuations, and not their physical forms, other results argue strongly that context can be used to predict particular words. Some recent experiments have created sentences in which semantic plausibility was linked with a nonsemantic lexical feature. For instance, DeLong et al. (2005) exploited the A/AN alternation in English. Although these variants of the indefinite article have identical (minimal) meaning, the correct choice varies with the phonology of the subsequent word. DeLong et al. constructed sentence frames such that the cloze-probability of possible completions (all nouns) ranged from 10% to 90%, and half of those nouns began with a vowel sound (calling for AN) and half began with a consonant sound (calling for A). Sentences were visually presented one word at a time, so that ERPs elicited by the articles could be examined contingent on whether they matched a potentially predictable sentence completion, or a less predictable (but congruent) completion. The articles elicited small N400s, whose amplitude was strongly (inversely) correlated with the cloze probability of the following word. For example, if KITE was the most favored completion of a sentence about flying, then the word A elicited a smaller N400 than the word AN (as in AN AIRPLANE, an acceptable but less preferred ending). Because the two articles have the same meaning, this result cannot be attributed to the ease or difficulty of integrating that meaning with the prior context. The results instead suggest that the subjects were actively predicting the full noun phrase of article-plus-noun, and that the benefits of a confirmed prediction were evident as reduced N400 activity. Studies in Dutch and Spanish have used analogous designs, with results that also suggest that comprehenders sometimes make specific predictions about upcoming words (van Berkum et al., 2005; Wicha et al., 2004). We postpone discussion of these other studies to Section 3.3 because their results show modulation of positive components of the ERP rather than the N400.

The results reviewed in this section present strong but conflicting evidence for the integration versus prediction debate as regards the N400. The data seem to be telling us that this is not an either/or choice, but that readers and listeners can pursue both strategies. However, at the outset of this review, we suggested that one hallmark of prediction, as conventionally understood, is that any particular prediction can yield either benefits or costs depending on whether it is correct. All of the studies reviewed above were interpreted by their authors in terms of the *benefits* offered by supportive semantic context, and this is the dominant interpretation within the larger N400 literature. In the next section, we ask whether there is any evidence that the N400 reflects the costs of a disconfirmed prediction.

2.3. Is the N400 augmented by failed predictions?

Teasing apart the possible benefits offered by a good match between a word and its context from the possible costs created by a poor match is a surprisingly difficult challenge within the existing N400 literature. Below we describe some promising approaches that appear less promising on closer examination, but also other data that suggest that the N400 does not reflect the costs of a disconfirmed prediction. Each of the empirical approaches considered below relies on the same general logic: if we can experimentally manipulate the strength of a subject's expectation (or hypothetical prediction) for a given word, and then present a different word, we can then compare violations of strongly-held expectations to violations of weakly-held expectations. By this logic, disconfirmation of a strong expectation should elicit a larger N400 than disconfirmation of a weak expectation *if* 1) readers and listeners make predictions, 2) our manipulation of prediction strength was successful, and 3) the N400 is sensitive to failed predictions.

2.3.1. Discrepancy between sentence constraint and cloze probability of the word actually presented

Above, we defined the cloze probability metric of expectancy for a particular word in a particular sentence: the percentage of subjects who offer that word as a completion when given the sentence frame and a blank line in a paper-and-pencil procedure. From the same data, the *constraint* of a sentence frame is also calculated, as identical to the cloze of the most-favored completion (or *best completion*, BC). Cloze and constraint are thus related but partially independent measures. Consider two sentence frames: Frame A elicited two different completions from the normative group, with frequencies of 90 and 10%; Frame B elicited fifteen different completions, with frequencies of 30, 10, 10, 10, 7, 3 and thirty other words offered by only 1% of the subjects. Frame A is thus of higher constraint than Frame B (90% versus 30%).

However, during an ERP experiment both frames could be completed with equally favored words chosen by 10% of the normative group. Alternatively, both sentence frames could be completed by incongruent words, with a de facto cloze probability of zero.

Kutas first reasoned that the discrepancy between the constraint of experimental sentences and the cloze probability of the words actually presented could serve to quantify the extent to which an expectation was violated (Kutas and Hillyard, 1984; Kutas et al., 1984). The idea is that high constraint sentences encourage strong predictions and low constraint sentences encourage weak predictions. The discrepancy between contextual constraint and cloze probability has intuitive appeal for quantifying the strength with which a prediction is disconfirmed, so that this idea has been applied in many other papers, including one from our lab (Federmeier, 2007; Federmeier et al., 2002, 2007; Hoeks et al., 2004; Otten and van Berkum, 2008; Van Petten et al., 1999; Vissers et al., 2006; Wlotko and Federmeier, 2007). With one exception (Hoeks et al., 2004), the results have been uniform: N400 amplitude is closely tied to the match between the context and the word presented – cloze probability – but independent of the discrepancy between constraint and cloze. Given the starting assumption about constraint and strength of prediction, the conclusion is that the N400 is insensitive to disconfirmed predictions. However, we argue below that the validity of that assumption looks questionable on closer examination.

The most transparent mapping between the cloze probability procedure performed by the normative group and the on-line cognitions of the ERP subjects is to imagine that they are doing the same thing: generating the first sentence completion that comes to mind. Let's reconsider the two sentence frames A (90% constraint) and B (30% constraint), both completed by words with 10% cloze probability during an ERP experiment. For Frame A, 90% of the subjects will have been thinking of a non-presented word (the *same* non-presented word) when they encounter the actual ending. What of Frame B? If each subject generates a unitary prediction, the outcome is no different: 90% of the participants will have been thinking of a word different than the one actually delivered, although the exact identity of that internal hypothesis will differ across subjects. Under this scenario, the contextual constraint of an experimental sentence is irrelevant for the experience of a given participant, so that constraint/cloze discrepancy is not a useful measure of violated expectancies (although cloze probability remains the best metric of *confirmed* expectancies).

As noted above, the equation of "violation" or "failed prediction" or "disconfirmed expectancy" with the constraint/cloze discrepancy is intuitive enough to have been adopted by multiple researchers. Can that intuition be salvaged? To our thinking, the only possibility requires a more complex mapping between the cloze probability procedure and the on-line cognitions of readers and listeners. In a very small set of studies, reaction times have been collected during cloze probability procedures and participants were faster to offer completions for high-constraint than low-constraint sentence frames (Cohen and Faulkner, 1983; Nebes et al., 1986). During real-time reading and listening, perhaps people only make predictions when a possible sentence continuation is highly constrained and can be quickly generated, and otherwise adopt a *laisse-faire* "wait and see" strategy. Under this scenario, high-constraint sentence frames are more likely than low-constraint frames to elicit predictions about upcoming words. In other words, constraint may influence not the *strength* of a prediction, but the *likelihood* of a prediction. If we adopt this thinking, the magnitude of constraint/cloze discrepancy is relevant to the issue of whether failed predictions influence N400 amplitude. Extant data would then suggest that the disconfirmation of a prediction has no impact on the N400.

2.3.2. Absent semantic context versus violated context

If we adopt the idea that contextual constraint influences the likelihood of making any prediction about upcoming words, other comparisons between weak and strong contexts suggest themselves. Above, we reviewed the fact that the first content words of isolated sentences ("The… [word]…") elicit large N400s relative to later congruent words. Initial content words may thus serve as a neutral baseline — words that are processed with very little potential benefit from supportive context or potential cost from misleading context. Is this sentential-initial N400 different from the N400 elicited when a strongly-constraining context is violated by the presentation of a semantically anomalous word? Surprisingly, this comparison has not been made in the published literature, to our knowledge.

2.3.3. Semantic plus perceptual context versus semantic context alone

The published literature does allow a different sort of comparison between violations of stronger versus weaker contexts. Several studies with auditory materials have compared incongruent sentence completions that are perceptually similar or dissimilar to the potentially predictable congruent completions. For instance, a sentence frame like "It was a pleasant surprise to find that the car repair bill was only seventeen…" can be completed by DOLLARS (congruent), DOLPHINS (sharing initial phonemes with the congruent ending), or CAPTAINS (dissimilar from the first phoneme). In this comparison, the N400 elicited by the perceptually similar incongruity is delayed in onset relative to the perceptually dissimilar word, simply because the auditory information signaling the incongruity is also delayed. But what of the amplitude of the late effect? One might think that the perceptually-similar incongruities (DOLPHINS) violate a stronger expectancy than the dissimilar incongruities. For the perceptually-similar endings, two sources of information converge to support an incorrect hypothesis about the word's identity: the semantic constraint of the sentence and the compatibility of the first few phonemes with the semantically-generated hypothesis. In contrast, for the perceptually-dissimilar incongruities (CAPTAINS), semantically-based expectations are not bolstered by additional misleading perceptual input. The data show that the N400 elicited by the violations of both semantic and perceptual context is never larger than that elicited by violations of semantic context alone (Connolly and Phillips, 1994; Connolly et al., 1995; Van Petten et al., 1999; see Fig. 2 for illustration).

2.3.4. Interim conclusion: little evidence for costs of failed predictions on the N400

As we suggested at the outset of this section, it is surprisingly difficult to define experimental contrasts that can separate the benefits of confirmed predictions from the costs of disconfirmed predictions about upcoming words in sentences or discourse. However, to the extent that published studies have successfully manipulated the strength of readers' or listeners' expectations, current data suggest only that N400 amplitudes are reduced in the presence of supportive semantic context and provide little hint that amplitudes are increased when an hypothesis/expectation/prediction is disconfirmed. From our starting premise that predictions should generate both benefits and costs (on different occasions), the apparent absence of costs is problematic. We might be driven to conclude that readers and listeners do not make predictions, but recall that there is strong evidence that they do, at least sometimes (DeLong et al., 2005). An alternative conclusion is that readers and listeners do make predictions, attended by both costs and benefits, but that those are indexed by different ERP components (see Bendixen et al., in press (this issue) for a review of multiple ERP components sensitive to confirmed and disconfirmed expectancies during nonlinguistic auditory processing). In a few recent papers, the authors have suggested that disconfirmed predictions about upcoming words lead to enhanced late positive components rather than N400s (DeLong et al., 2011; Federmeier et al., 2007; Otten and van Berkum, 2008). In Sections 3.2–3.4, we review late positive ERPs elicited by problematic words in sentence contexts, observe that there seem to be two distinct positive components differentiated by their parietal versus frontal scalp topographies, and speculate that these are associated with distinct sorts of

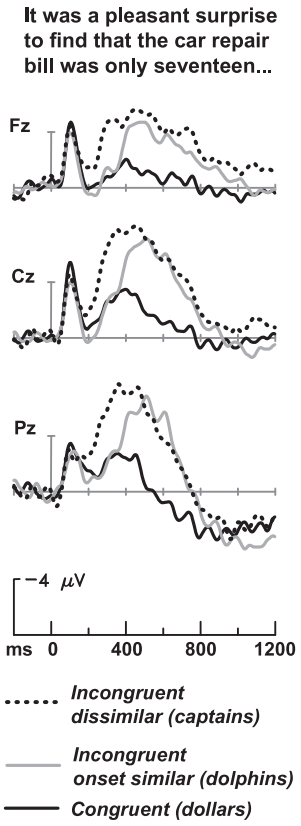

Fig. 2. Grand average auditory ERPs at midline frontal, central and parietal scalp sites from 24 subjects.
Data from Van Petten et al., 1999.

cognitive "costs". Before examining the sentence experiments, we provide a brief background on expectancy and a late positive potential – the P300 – in nonlinguistic paradigms.

3. Late positive ERP components

3.1. Prediction, surprise, and the P300

The P300 was the first component of the event-related potential (ERP) to attract substantial attention from researchers interested in cognition. An intensive effort from the late 1960s through the late 1970s established this component as a reliable response to unpredictable stimuli in all modalities, via the use of nonlinguistic stimuli (or single words). The P300 is commonly divided into two subcomponents with different scalp distributions: a frontally-maximal P3a elicited when perceptually novel (typically unique) stimuli are interspersed among more common stimuli, and a parietally-maximal P3b whose amplitude is more strongly driven by the relevance of the stimuli to the participant's assigned task (Courchesne et al., 1975; Debener et al., 2005; Dien et al., 2004; Donchin, 1981; Friedman et al., 2001; Goldstein et al., 2002; Johnson, 1988; Polich and Comerchero, 2003; Spencer et al., 1999, 2001; Squires et al., 1975).

Multiple findings indicate a close link between the P3b and the disconfirmation of an expectation. The simplest of these is the well-known sensitivity of P3b amplitude to the probability of stimulus categories: stimuli from a rare category elicit a larger positive wave than those from a more frequent category. The probability effect is clearly cognitive (rather than sensory) in nature because the relevant definition of "category" is determined by how subjects are asked to classify the items rather than by the physical identity of the stimuli (Johnson and Donchin, 1980; Kutas et al., 1977; Breton et al., 1988; see Folstein and Van Petten, 2011 for a recent summary). Stronger evidence came from a careful examination of sequential stimulus structure when the global probabilities of two stimulus types were equated (high and low pitched tones, or blue and orange flashes). Squires et al. (1976, 1977) found that much of the variance in P3b amplitude could be explained by the number of dissimilar stimuli that preceded the eliciting stimulus, so that stimulus type "B" in an A–A–A–A–B sequence elicited a larger P3b than in an A–A–A–B sequence which, in turn, was larger than the response to A–A–B. Critically however, the accuracy of the model was improved by incorporating a subjective expectation for regular patterns, such that a large P3b was also elicited by stimuli that violated a repetitive alternation (A–B–A–B–A–B–*B*). Even more compelling evidence that violation of an expectation leads to the emission of a P300 came from paradigms in which late positive waves were observed in the absence of stimuli — namely, situations in which a stimulus was omitted from a regularly-timed sequence and a P300 was recorded at roughly 300 ms after the moment at which an item might have occurred (Sutton et al., 1967; Ruchkin and Sutton, 1973; Ruchkin et al., 1975).[3] The link between the P300 and disconfirmation of an expectation was further strengthened by feedback paradigms in which participants generated a response in a cued-recall task, rated their confidence in that response, and then received accuracy feedback. Independent of its valence, surprising feedback – finding that an answer thought to be correct was wrong or finding that an answer thought to be incorrect was actually right – elicited a larger P300 than feedback which merely confirmed participants' assessment of their own performance (Horst et al., 1980; see Butterfield and Mangels, 2003 for a more recent version of this paradigm).

The most concise summary of the cognitive process reflected by the P3b is that offered by Donchin and Coles (1988): updating of working memory. The brief review above emphasizes a frequent trigger for such updating — disconfirmation of an expectancy about an upcoming event.

3.2. Syntactic and semantic P600s: reprocessing, repair, and retrieval

In contrast to the modulation of N400 amplitude by the semantic fit between a word and its context, sentence words that are erroneous or dispreferred on syntactic grounds reliably elicit a large late positive wave — the *P600* (e.g., Friederici et al., 1996; Hagoort et al., 1993; Osterhout and Holcomb, 1992). This response often lacks a clear peak but is evident in an interval some 500–900 ms after the onset of problematic word, and is typically largest at centroparietal scalp sites. Larger P600s for errors than for correct words have been observed for a variety of agreement errors, violations of local phrase structure, and for errors involving higher-level syntactic structure (Hagoort et al., 1993; McKinnon and Osterhout, 1996; Osterhout and Holcomb, 1992; Osterhout and Mobley, 1995). In contrast, errors of agreement in word pairs do *not* elicit P600s (Barber and Carreiras, 2005; Münte et al., 1993; Münte and Heinze, 1994). Within sentences, P600 effects do not require outright syntactic violations. Correct but syntactically complex sentences elicit larger P600s than simpler sentences, low-frequency verb argument structures elicit larger P600s than preferred argument structures, and garden path sentences elicit substantial P600s as compared to straightforward sentences (Kaan et al., 2000; Osterhout et al., 1994; Phillips et al., 2005; van Berkum et al., 1999). Together, these last two observations support the widely-accepted view that the P600 component does not reflect the detection of a syntactic anomaly, but rather a process that

[3] The "300" in P300 comes from the peak latency of this component in early studies using very simple stimuli. Subsequent research showed that P300 latency varies with the time need to evaluate a stimulus with respect to the classification rule provided in the experimental instructions, at least for simple classification rules. With more complex stimuli and classification rules, such as judging whether names are male or female, P300 latency can be considerably later than 300 ms (Kutas et al., 1977). See Folstein and Van Petten (2011) for recent summary and a limitation of the relationship between P300 latency and classification time.

is initiated *after* encountering difficulty, such as a subsequent attempt to revise one's initial parse of a sentence (Friederici et al., 1996; Hahne and Friederici, 1999; O'Rourke and Van Petten, 2011).[4]

The pleasing dichotomy of "semantic N400 and syntactic P600" has recently been challenged by several studies in which semantic incongruities have elicited late positivities instead of N400s. Across laboratories, ERP effects that closely resemble the syntactic P600 have appeared in response to sentences with plausible elements arranged in thematically implausible ways, such as "The hearty meal was DEVOURING the kids", "For breakfast, the eggs would only EAT toast and jam", and "The cat that from the mouse FLED ran through the room" (Kim and Osterhout, 2005; Kuperberg et al., 2003; van Herten et al., 2005, respectively). These *semantic P600s* have been attributed to processes similar to those thought to underlie clearly syntactic P600s, namely re-analysis, "re-attending", or prolonged analysis of problematic sentences (Kolk et al., 2003; Kuperberg, 2007; van Herten et al., 2006). This interpretation is consistent with a recent study using semantic incongruities that are easily overlooked due to their relationship with the global semantic schema of a sentence (as in "After an air crash, where should the survivors be buried?"). A P600 was elicited only when participants overtly detected (and presumably revisited) the semantic problem (Sanford et al., 2011).

The relationship between the P600 elicited by syntactic violations and the more general P300 elicited by stimuli that are unpredictable and task relevant has received substantial discussion (Coulson et al., 1998; Friederici et al., 2001; Osterhout and Hagoort, 1999). Overall, it seems likely that the syntactic P600 receives a substantial contribution from the P300, given that it is sensitive to some of the same factors that influence P300 amplitude in nonlinguistic paradigms, namely the probability of a violation (larger when erroneous sentences are a small proportion of the stimulus set; Coulson et al., 1998; Gunter et al., 1997; Hahne and Friederici, 1999) and the relevance of the violation to the participants' assigned task (larger during sentence acceptability/grammaticality tasks than in reading for comprehension, Gunter et al., 1997; Osterhout and Mobley, 1995). Similar manipulations of task and probability have not yet been conducted for the paradigms that yield semantic P600s. However, attributing either variety of P600 to surprise or unpredictability does not, in itself, explain why some sorts of problematic sentences should elicit N400-dominant patterns of results and others P600-dominant patterns. It is likely to be fruitful to focus on the cognitive processes that *follow* the registration of an unexpected event. Below, we suggest that one of the cognitive underpinnings of attempted re-analysis of a problematic sentence is memory retrieval.

An ERP effect that closely resembles both the syntactic and semantic P600 is observed in studies of episodic memory. When participants are asked to judge whether words or pictures had occurred in a prior study list, items recognized as old elicit more positive ERPs than new items correctly rejected, new items falsely judged as old, and old items that are unrecognized. This *old/new effect* thus reflects successful retrieval, and for words, is observed in a latency range of roughly 400–800 ms after stimulus onset (Rubin et al., 1999; Van Petten and Senkfor, 1996; Senkfor and Van Petten, 1998). The old/new effect is also robust when participants are not explicitly asked to make memory judgments but some items may be spontaneously recognized. Namely, if participants are assigned a non-mnemonic task, such as making lexical or perceptual decisions about stimuli, recently-encountered stimuli elicit ERPs like those of the studied items in recognition tests (Kazmerski and Friedman, 1997; Paller et al., 1995; Van Petten and Senkfor, 1996).

The similarity between the old/new effect observed in memory paradigms and the sentence P600 is very compatible with the idea that an attempt to revise one's parse of a sentence triggers retrieval or reactivation of the preceding words in the sentence. During fluent reading or listening, individual words are thought to persist for only a brief time in working memory before being combined into larger, interpreted "chunks" (see Neath and Surprenant, 2003 for review). For troubled sentences that appear potentially salvageable, checking one's initial interpretation will necessitate retrieval of the individual un-chunked words to review their sequential order, determine if one missed a word, if the problem might be a typographical error, etc. Like the sentence P600, the episodic old/new effect is largest at centroparietal scalp sites, although its distribution is broader than the typical P600 (see e.g., Kuo and Van Petten, 2006, 2008). The positivity linked to memory retrieval is thus unlikely to be the sole contributor to sentence P600s, but we suggest that one variety of cognitive "cost" that can occur when sentence-processing hits a snag is the burden of retrieving individual words from recent memory.

3.3. Failed semantic predictions: frontal positivities?

The "semantic P600" described may be triggered by integration difficulty and/or by the disconfirmation of active predictions; studies to date have not been aimed at dissociating these. A small number of experiments have attempted to distinguish integration difficulty from disconfirmed predictions about specific words. In Section 2.2, we described one such study, in which A or AN – words that should never be difficult to integrate – could elicit larger or smaller N400s depending on whether they signaled that the upcoming noun would be a poor or good fit with the sentence context. Other studies have also exploited linguistic rules about agreement between words to examine semantic prediction. In both Spanish and Dutch, nouns have grammatical gender. The gender of a noun is largely unpredictable from its meaning, but in grammatically correct sentences, the genders of articles and adjectives must agree with their nouns. Thus, if readers actively predict that "Little Red Riding Hood carried the food for her grandmother in…" A BASKET, specifically, rather than some sort of container generically, the Spanish reader will also predict the feminine article UNA — to agree with the feminine noun CANASTA (i.e., "Caperucita Roja cargaba la comida para su abuela en una canasta"). Wicha et al. (2004) examined ERPs elicited by articles whose gender agreed or disagreed with the most plausible sentence continuation in written Spanish sentences. van Berkum et al. (2005) used a very similar design with spoken Dutch materials, except that the critical words were gender-marked adjectives (e.g., GROOT versus GROTE) that preceded their nouns by several words. In both cases, articles and adjectives whose gender was inappropriate for the most plausible sentence continuation elicited different ERPs than words whose gender matched. These results are striking because the "inappropriate" articles/adjectives were not semantically problematic themselves; the results thus indicate that participants were anticipating specific nouns of specific genders.

In the Spanish and Dutch studies, the responses to the inappropriate articles or adjectives were not enhanced N400s but larger positive potentials relative to the appropriate control words (van Berkum et al., 2005; Wicha et al., 2004). One might wonder if these results are variants of the syntactic P600 elicited by frank violations of agreement, but the positivities showed distinctive frontal scalp distributions which differ from the typical parietal maximum of the P600. Several authors have now suggested that positive potentials reflect the disconfirmation of semantically-based predictions in sentence or discourse contexts (DeLong et al., 2011; Federmeier et al., 2007; Otten and van Berkum, 2008). If this idea is correct, such results should be fairly common in the very large literature ERP sentence-processing literature, if prediction is, in fact, a common strategy during language comprehension. We thus set out to examine the existing literature.

[4] An ERP effect that precedes the P600 in time – dubbed the "left anterior negativity" or LAN – is thought to reflect initial detection of a syntactic error (see O'Rourke and Van Petten, 2011 for dissociation of LAN and P600 effects).

3.4. Semantic expectancies during sentence processing: Parietal and frontal positivities accompanying N400s

Given their knowledge of the literature concerning the P300 and expectancy, Kutas and Hillyard expected to observe a larger positive wave for incongruent than congruent sentence completions in their groundbreaking (1980a) experiment. Because the larger P300 for rare as compared to predictable stimuli is often preceded by a enhanced negative component (the N2, see Folstein and Van Petten, 2008 for recent review), Kutas and Hillyard were careful to note that their new semantic negativity – the N400 – was not part of a negative–positive complex and that no enhanced positivity followed the N400 elicited by incongruent words. Fig. 2 illustrates this sort of *monophasic* result in which incongruent sentence endings elicit larger N400s than congruent, and the waveforms converge when the N400 response is complete. However, in the profusion of studies over the last 30 years, there are also a substantial number with *biphasic* ERP responses to what appear to be fairly conventional manipulations of semantic predictability: a larger N400 followed by a larger late positive wave. As we noted a few years ago (Van Petten and Luka, 2006), both monophasic and biphasic semantic effects have been observed for many years but the source of this variability has attracted only sporadic attention until recently. It has also been unclear whether there is a single late positive component that responds to semantic manipulations or multiple positive waves with different spatial distributions across the scalp, indicative of different neural generators. Here, we adopt a theoretically neutral term – *Post-N400 Positivity* (PNP) – to refer to any enhanced positivity that is evident immediately after a larger N400 in a comparison between two conditions. Fig. 3 shows one biphasic pattern in which the N400 is followed by a PNP with a parietal scalp distribution. Fig. 4 shows a different biphasic pattern in which an N400 is followed by a PNP with a frontal scalp distribution. Before speculating on the functional nature of these PNP responses and their relationship to late positive ERPs in other paradigms, we first sought to determine how commonly they occur.

We searched the literature for ERP experiments in which congruent sentence completions were compared to incongruent (semantically anomalous) completions, or to other completions that were also congruent, but classified as less predictable on the basis of cloze probability tests. We restricted our search to studies that met the following criteria: 1) healthy adult participants, 2) sentence completions that were syntactically appropriate (as explicitly stated by the authors, or suggested by sample sentences), 3) no task that might require an overt decision during the ERP epoch,[5] 4) display of ERP waveforms from both frontal and parietal (or occipital) scalp sites in an epoch showing at least 900 ms of activity after stimulus onset, and 6) mastoid or earlobe references for scalp sites, so that scalp distributions could be reasonably compared across studies. A sixth criterion was also imposed to restrict the range of potentially extraneous variability among studies, namely that the critical words be sentence-final rather than sentence-intermediate. We similarly excluded comparisons that involve manipulations other than garden-variety semantic predictability, such as sentence repetition, metaphoricity, code switches between languages, etc.

In all of the studies we located, the semantic manipulation led to the predicted result of a larger N400 for the incongruent or lower-cloze completions as compared to the more predictable congruent completions. Our current interest is whether these larger N400s were accompanied by larger late positive waves. Because latency ranges after the N400 have not always been analyzed in studies manipulating semantic factors, and amplitude variance measures (standard deviations or standard errors) are infrequent in the ERP literature, a statistical meta-analysis is not possible, but we believe the survey is informative nonetheless.

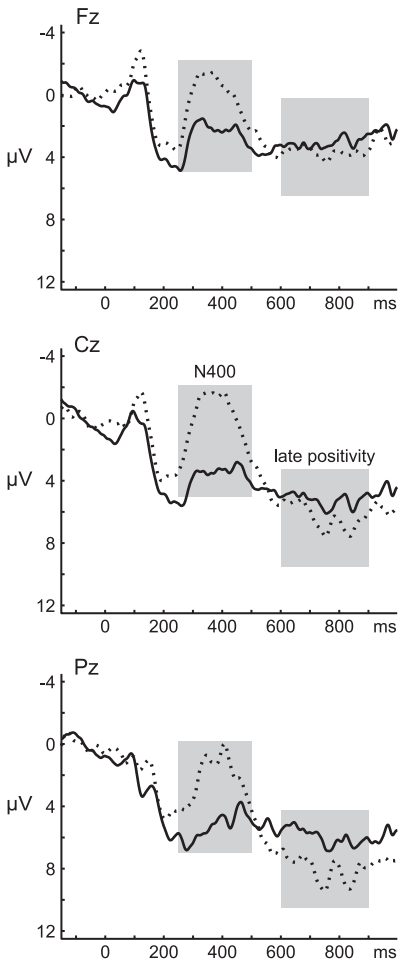

Fig. 3. Grand average ERPs at midline frontal, central and parietal scalp sites from 18 healthy young adults elicited by sentence-final words. Congruent completions were high-cloze (94%) and presented visually, shown in the solid line. Semantically incongruent completions shown in the dotted line. Note that the enhanced late positivity for the incongruent words is larger parietally than frontally.
Data from Pijnacker et al., 2010; we thank Dr. Judith Pijnacker for providing an adaptation of the figure appearing in the published work.

3.4.1. Congruent versus incongruent sentence completions

Table 1 contains one portion of this systematic[6] survey with 45 studies comparing congruent to incongruent sentence completions. Many studies yielded multiple contrasts, such as sentence-final words presented in the right or left visual fields, or incongruent words that were semantically related versus unrelated to the best completion for that sentence, etc. These are listed separately (see "Notes" column), so that Table 1 contains 64 comparisons between congruent and incongruent sentence completions. In some studies, latency windows subsequent to the typical N400 were explicitly analyzed by the authors (typically 600–900 ms after stimulus onset), and the table classifies a PNP effect as present or absent according to those statistical analyses. When no statistical analysis was included in the published article, we judged a PNP effect to be present when waveform figures displayed a positive difference between conditions that appears to be at least half as large as the preceding N400 effect. Similarly, the scalp distribution of a PNP effect was classified as larger at

[5] Although unpredictability is an important eliciting condition for large P300s, a requirement to overtly classify stimuli also increases P300 amplitude (see Johnson, 1988, for an integrative review of the factors influencing P3 amplitude).

[6] The survey presented in Tables 1 and 2 is intended to be an exhaustive list of studies meeting the criteria in the text. Given the large literature, it is possible that we have inadvertently missed some studies.

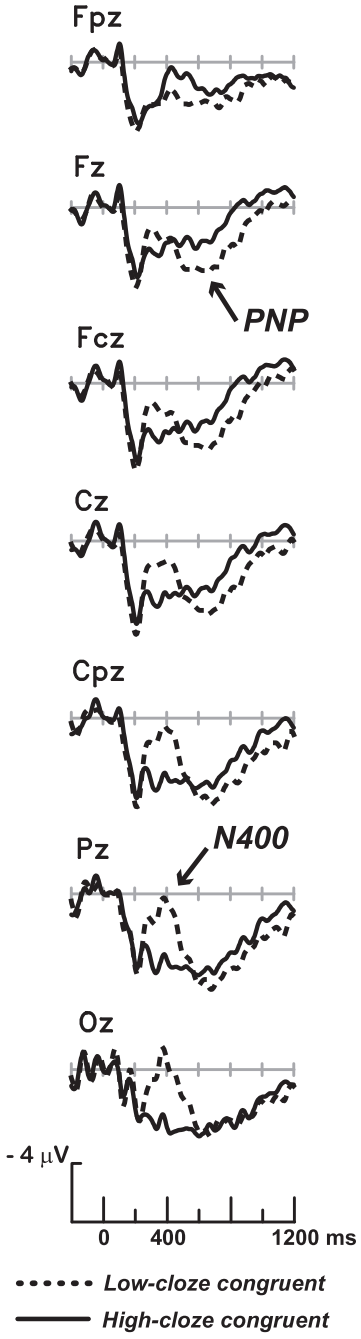

Fig. 4. Grand average visual ERPs at midline prefrontal, frontal, frontocentral, central, centroparietal, parietal and occipital scalp sites from 24 healthy young adults, elicited by sentence-final words. Semantically congruent sentence completions elicited both a larger N400 and a larger positive potential (post-N400 positivity or PNP) when they were less predictable from the sentence context (low-cloze completions in the dotted line as compared to the more predictable high-cloze completions in the solid line). Note that the enhanced late positivity (PNP) for the low-cloze words is largest over frontal (but not prefrontal) scalp.
Data from Thornhill and Van Petten (submitted for publication).

posterior (mostly parietal) scalp sites than at frontal sites or vice versa following the published statistical analyses, or based on visual inspection if no analysis was presented. In two cases, the topography was classified as "broad" because frontal and parietal amplitudes appeared roughly equal.

Even a cursory glance at Table 1 shows substantial variability across studies, and that the presence or absence of a post-N400 positivity does not track any easy-to-classify factor such as the sensory modality or language in which the sentences were presented. The top half of Fig. 5 displays a graphical summary of the table, which also suggests little difference between studies that included statistical analysis of a post-N400 latency range and studies that were subjected to our visual inspection only. With or without formal statistical analysis, roughly a third of the published comparisons show larger late positivities for incongruent than congruent words and two-thirds do not. The source of this variation remains elusive. Given the (debatable) idea that the discrepancy between sentence constraint and the cloze probability of the word actually delivered is a good metric of "violation" (Section 2.3.1), we considered this factor. Sentence constraint per se is rarely reported, but when congruent cloze is greater than 50%, it will necessarily be identical to the contextual constraint of the sentence frames, and the cloze/constraint discrepancy for an incongruent completion will also be identical to the congruent cloze. A substantial number of studies did not report any cloze probability values. For those that do, the large majority used high-constraint sentences with high-cloze congruent endings, leaving little variation to examine. Overall, the reported methods do not, unfortunately, allow conclusions about the impact of cloze/constraint discrepancy in contrasts between congruent and incongruent words.

Table 1 and Fig. 5 do, however, lead to one clear conclusion. When incongruent words elicit a post-N400 positivity, the scalp topography of that effect is predominantly parietal (17 of the 21 comparisons classified as showing a PNP effect). The results shown in Fig. 3 are thus representative of this category of post-N400 positivities. The parietally-maximum topography is much the same as the "semantic P600" elicited when sentence elements that could form plausible statements are arranged to create an incongruent sentence (e.g., "For breakfast the eggs would only EAT…", Section 3.2). This observation suggests that late parietal positivities may not be unique to those carefully-constructed materials, but are also evident – more variably, and more weakly – in the broader range of semantically incongruent materials that have been used by various laboratories over the years. Assuming a continuity between the parietal post-N400 positivity and the semantic P600 implies that the parietal PNP also reflects attempted re-analysis or checking of bad sentences. This process may be only variably invoked by incongruent sentence completions depending on a host of difficult-to-quantify factors: the exact construction of incongruent sentences by different experimenters (i.e., whether there is any hint that a sentence could be re-interpreted in a way that makes sense) and/or the verbal abilities or motivation of individual subjects.

3.4.2. High- versus low-cloze congruent sentence completions

Table 2 is the second portion of our literature survey: studies that have compared semantically congruent sentence completions that were preferred by a normative group of participants (high-cloze endings) to congruent endings that were less preferred by a normative group. Fig. 5 shows that late positive potentials following the N400 are more prevalent for this contrast than for comparisons between congruent and incongruent words. Because there are also substantially fewer high/low cloze comparisons available in the literature, it is not yet obvious whether this difference is reliable. A more definitive conclusion is that, when a PNP is observed in comparisons of high- and low-cloze congruent completions, the topography of the effect is predominantly frontal (17 of 18 cases). The results shown in Fig. 4 are thus representative of this second category of post-N400 positivities, which appear distinct from the more posterior effects observed for incongruent sentence completions.

3.5. Functional correlates of frontal PNP effects

The different topographies of PNP effects for incongruent (parietal) and congruent (frontal) sentence completions suggest that they arise from different brain regions and should be attributed to different functional processes. Above, we suggested that a parsimonious account of

Table 1
Contrasts between congruent and incongruent sentence completions.

Citation	Modality	Language	Cloze probability for congruent	PNP topography	Notes
Kutas and Hillyard (1980a)	V	English	?	**x**	Exp. 1
Kutas and Hillyard (1980a)	V	English	?	**x**	Exp. 2
Kutas and Hillyard (1980b)	V	English	?	x	
Kutas et al. (1984)	V	English	63	Broad	Exp. 1
Kutas et al. (1984)	V	English	92	x	Exp. 2, related to BC
Kutas et al. (1984)	V	English	92	x	Exp. 2, unrelated to BC
McCallum et al. (1984)	A	English	?	**Parietal**	
Besson and Macar (1987)	V	French	75*	**Broad**	
Kutas (1987)	V	English	92	x	
Kutas et al. (1988)	V	English	?	x	
Ardal et al. (1990)	V	English	?	x	
Besson et al. (1992)	V	English	80	**x**	
Besson et al. (1992)	V	English	45	**x**	
Gunter et al. (1992)	V	Dutch	?	**Parietal**	
Nigam et al. (1992)	V	English	?	Parietal	
Andrews et al. (1993)	V	English	>60	**Parietal**	Related to BC
Andrews et al. (1993)	V	English	>60	**Parietal**	Unrelated to BC
Friederici et al. (1993)	A	German	?	**x**	
Mitchell et al. (1993)	V	English	82	**Frontal**	
Woodward et al. (1993)	V	English	?	**Parietal**	
Connolly and Phillips (1994)	A	English	?	x	
Nobre and McCarthy (1994)	V	English	?	Parietal	
Connolly et al. (1995)	V	English	?	x	
Ganis et al. (1996)	V	English	85	**x**	Sentences only
Ganis et al. (1996)	V	English	85	**x**	Mixed with pictures
Juottonen et al. (1996)	A	Finnish	?	**Parietal**	
Swaab et al. (1997)	A	Dutch	60	x	
Revonsuo et al. (1998)	A	Finnish	80	**Parietal**	
Friederici et al. (1999)	V	German	?	**x**	
van Berkum et al. (1999)	V	German	?	x	
Van Petten et al. (1999)	A	English	85	**x**	Exp. 1
Van Petten et al. (1999)	A	English	38	**x**	Exp. 1
Van Petten et al. (1999)	A	English	58	**x**	Exp. 2
Hagoort and Brown (2000)	A	Dutch	60	x	
Johnson and Hamm (2000)	V	English	?	x	
van den Brink et al. (2001)	A	Dutch	84	**Parietal**	
Nixon et al. (2002)	V	English	?	x	Related to BC
Nixon et al. (2002)	V	English	?	x	Unrelated to BC
Robichon et al. (2002)	V	French	49	**x**	Fast rate
Robichon et al. (2002)	V	French	49	**x**	Slow rate
Ruchsow et al. (2003)	V	German	?	**x**	
Brown-Schmidt and Canseco-Gonzalez (2004)	V	Mandarin	?	**x**	
Deldin et al. (2006)	V	English	?	Parietal	Exp. 1
Deldin et al. (2006)	V	English	?	Parietal	Exp. 2
Deldin et al. (2006)	V	English	?	Parietal	Exp. 3
Newman and Connolly (2004)	A	English	?	**Parietal**	
Coulson et al. (2005)	V	English	71	**x**	Related to medial word, RVF
Coulson et al. (2005)	V	English	71	**x**	Related to medial word, LVF
Coulson et al. (2005)	V	English	71	**x**	Unrelated to medial word, RVF
Coulson et al. (2005)	V	English	71	**x**	Unrelated to medial word, LVF
Moreno and Kutas (2005)	V	Eng., Spanish	>70	**x**	Nondominant language
Moreno and Kutas (2005)	V	Eng., Spanish	>70	**Frontal**	Dominant language
van den Brink et al. (2006)	A	Dutch	84	x	
Daltrozzo et al. (2007)	A	French	48	**x**	Women
Daltrozzo et al. (2007)	A	French	48	**Parietal**	Men
Diaz and Swaab (2007)	A	English	71	Posterior**	
Laszlo and Federmeier (2008)	V	English	42	**x**	
Ruchsow et al. (2008)	V	German	?	**x**	
León et al. (2010)	V	Spanish	?	x	
Pijnacker et al. (2010)	V	Dutch	94	**Parietal**	
Arbel et al. (2011)	V	English	94	**x**	Related to BC
Arbel et al. (2011)	V	English	94	**x**	Unrelated to BC
Grose-Fifer et al. (2011)	V	English	34	**x**	Young adults
Grose-Fifer et al. (2011)	V	English	29	**x**	Adolescents

Note. In the Modality column, "A" for auditory and "V" for visual. In Cloze probability column, "?" means that cloze probability for congruent completions not reported; * cloze probability estimated from description of stimuli. "PNP" means "post-N400 positivity". Bold entries in the "PNP topography" column indicate that a latency window subsequent to the N400 peak (most typically 600–900 ms after word onset) was statistically analyzed. An "x" in that column means that the waveform for incongruent words was not more positive than that for congruent words (i.e., no PNP was observed). Non-bolded entries are based on visual inspection of published figures, as described in the text; **parietal site not shown, PNP larger occipitally than frontally. Some studies generated more than one comparison, as described in the "Notes" column. "BC" means "best completions" of the sentence frames — the words with the highest cloze probabilities for those frames. "RVF": sentence completion presented in right visual field. "LVF": sentence completion presented in left visual field. Please also see text Section 3.4.

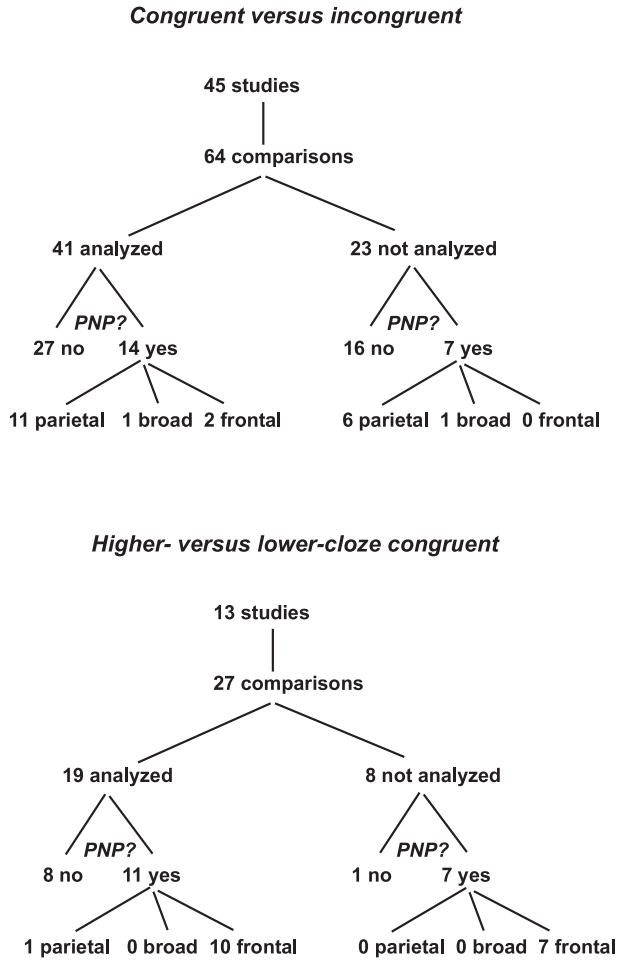

Fig. 5. Top: Graphical summary/analysis of Table 1. Our literature search found 45 studies that compared ERPs elicited by semantically congruent sentence endings to semantically incongruent endings (see text Section 3.4.1 *Congruent* versus *incongruent sentence completions* for inclusion criteria). Many studies included multiple experiments or multiple conditions so that there were 64 contrasts between congruent and incongruent words within the 45 studies. All yielded larger N400s for the incongruent than congruent completions. For 41 of the comparisons, the published work also reports statistical analyses of an epoch after the N400 (most typically 600–900 ms), whereas some articles did not report analyses of semantic congruity effects in a late time window — these are labeled "analyzed" versus "not analyzed", respectively. The next branch point in the figure ("PNP?") summarizes whether the incongruent words elicited larger late positive ERPs than the congruent words. For the results subjected to statistical analyses by their authors (listed in Table 1), "yes" means that incongruent words elicited a significantly larger positive potential than congruent words, in the latency range following the N400. For the "unanalyzed" results, "yes" and "no" refer to the current authors' visual inspection of published figures. Finally, the lowest branch of the figure summarizes the spatial distribution of the observed post-N400 positivities — whether they were largest at parietal scalp sites, largest at frontal scalp sites, or appeared to be of equivalent amplitude at parietal and frontal scalp sites ("broad"). Bottom: Graphical summary of Table 2; see also text Section 3.4.2 *High-* versus *low-cloze congruent sentence completions*. These studies compared ERPs elicited by semantically congruent sentence completions, split by whether the words were offered by a larger ("higher cloze") or smaller ("lower cloze") proportion of a normative group who were given the sentence frames lacking a final word, and asked to fill in the best final word. Other conventions as in the top portion of this figure.

parietal PNP effects is to assume a continuity with positive potentials that occur with little preceding N400 activity, namely syntactic and semantic P600s that have been attributed to re-analysis or checking. Much less is known about the frontal PNP. In particular, it is not yet clear if the frontal positivity following an N400 should be considered the same component as the monophasic frontal positivity that has been more clearly linked to disconfirmed lexical predictions (reviewed in Section 3.3). In one parallel to those results, we have recently found that the frontal PNP effect can be attributed to the presentation of unexpected lexical items rather than unexpected concepts (Thornhill and Van Petten, submitted for publication). Participants read sentences completed by their most preferred ending, a congruent but low-cloze word that was nearly synonymous with that ending, or a low-cloze word that was unrelated to the preferred ending ("On his vacation, he got some much needed REST/RELAXATION/SUN"). As in other studies, the N400 was sensitive to the match between the sentence context and the concept expressed by the final word, so that a smaller N400 was elicited by the related than unrelated low-cloze words (Fig. 1). In contrast, the two varieties of unpredictable final words elicited indistinguishable post-N400 positivities (like that illustrated in Fig. 4), which were maximal at frontocentral scalp sites. These results are consistent with the idea that the frontal PNP reflects disconfirmed predictions, and that the relevant sort of prediction is lexical, rather than for a general concept that can be expressed by alternate words.

Our recent results help to establish the eliciting conditions for a frontal positivity during sentence interpretation, but there is some distance to go in understanding the cognitive function indexed by this ERP component. In an early comment, Kutas (1993) suggested that it might index inhibition of predicted words that were not presented:

> Continuing to speculate in this vein… Frith and his colleagues… have posited an inhibitory interaction between the frontal and temporal areas during word processing. They proposed that successful intrinsic word generation is based on inhibitory modulation of the network of activated …stored word representations in the superior temporal areas by the left prefrontal cortex. A similar inhibitory mechanism might be needed to explain how it is that we can readily interpret an unexpected but congruent ending; on this view, the left prefrontal cortex supports contextual integration of the low cloze ending by inhibiting the activated representation of the ending primed by the sentence context. The engagement of this inhibitory process in the left prefrontal cortex is indexed by the frontal positivity to low cloze endings… A similar mechanism would not be expected for truly anomalous endings, as these cannot be interpreted. (pg 553)

More recent studies are supportive of at least one strand of this argument — that frontal late positive potentials are elicited during the interpretation of plausible sentence completions that are not predictable in advance of their occurrence (Delong et al., 2011; Federmeier et al., 2007; Thornhill and Van Petten, submitted for publication). The second strand of Kutas' (1993) speculation – that the frontocentral PNP reflects inhibition of a predicted-but-not-presented word – is a stronger claim that remains to be tested. In other cognitive models that posit active inhibition of some internal representation, the claim has rested on behavioral measures collected subsequent to the hypothesized inhibition — subsequent measures that suggest reduced accessibility of the inhibited representations. For instance, for the *retrieval induced forgetting* paradigm, it is argued that retrieval practice for some members of a semantic category causes inhibition for members of the same category that are viewed without retrieval practice. This idea was stimulated by the fact that subsequent recall of the unpracticed words from the practiced category is worse than for control items from other categories (Anderson et al., 1994). Determining whether predicted-but-unpresented sentence completions are actively inhibited will require a similar sort of assessment of the fate of those words in subsequent tasks.

4. Summary

Both ERP and behavioral data strongly support the view that readers and listeners interpret input continuously and incrementally, and that interpretation leads to general expectations about the semantic content that will occur later. After that subsequent input is

Table 2
Contrasts between higher- and lower-cloze congruent sentence completions.

Citation	Modality	Language	Higher cloze prob.	Lower cloze prob.	PNP topography	Notes
Kutas et al. (1984)	V	English	63	23	Frontal	Exp. 1
Kutas et al. (1984)	V	English	92	3	Frontal	Exp. 2, related to BC
Kutas et al. (1984)	V	English	92	3	Frontal	Exp. 2, unrelated to BC
Kutas (1993)	V	English	>75	Low	**Frontal**	
Besson et al. (1997)	V	French	>75	<33	**Parietal**	
Van Petten et al. (1999)	A	English	85	38	**x**	Exp. 1
Moreno et al. (2002)	V	Spanish	Mod.	Low	**Frontal**	
Moreno et al. (2002)	V	Spanish	High	Low	**Frontal**	
Federmeier and Kutas (2005)	V	English	85	28	Frontal	Young adults
Federmeier and Kutas (2005)	V	English	85	28	Frontal	Older adults
Federmeier et al. (2005)	V	English	>70	<40	Frontal	RVF
Federmeier et al. (2005)	V	English	>70	<40	Frontal	LVF
Coulson and Van Petten (2007)	V	English	81	2	**Frontal**	RVF
Coulson and Van Petten (2007)	V	English	81	2	**x**	LVF
Diaz and Swaab (2007)	A	English	71	5	x	
Federmeier et al. (2007)	V	English	85	1	**Frontal**	
Federmeier et al. (2007)	V	English	27	1	**x**	
Wlotko and Federmeier (2007)	V	English	85	1	**x**	RVF
Wlotko and Federmeier (2007)	V	English	85	1	**x**	LVF
Wlotko and Federmeier (2007)	V	English	27	1	**x**	RVF
Wlotko and Federmeier (2007)	V	English	27	1	**x**	LVF
Borovsky et al. (2010)	V	English	High	Low	**x**	
DeLong et al. (2011)	V	English	>50	<50	**Frontal**	
Thornhill and Van Petten (submitted for publication)	V	English	78	4	**Frontal**	Related to BC
Thornhill and Van Petten (submitted for publication)	V	English	78	3	**Frontal**	Unrelated to BC
Thornhill and Van Petten (submitted for publication)	V	English	30	4	**Frontal**	Related to BC
Thornhill and Van Petten (submitted for publication)	V	English	30	4	**Frontal**	Unrelated to BC

Note. In the Modality column, "A" for auditory and "V" for visual. In the cloze probability columns, "high", "moderate" (mod.) and "low" reflect the original authors' descriptions. "PNP" means "post-N400 positivity". Bold entries in the "PNP topography" column indicate that a latency window subsequent to the N400 peak (most typically 600–900 ms after word onset) was statistically analyzed. An "x" in that column means that the waveform for incongruent words was not more positive than that for congruent words (i.e., no PNP was observed). Non-bolded entries are based on visual inspection of published figures, as described in the text Section 3.4. Some studies generated more than one comparison, as described in the "Notes" column. "BC" means "best completions" of the sentence frames — the words with the highest cloze probabilities for those frames. "RVF": sentence completion presented in right visual field. "LVF": sentence completion presented in left visual field.

itself interpreted, it may prove easier or more difficult to integrate with what has come before. However, we reserve the term *prediction* for a more specific expectation that a particular word will occur at a particular point in the stream of input. Some psycholinguistics have expressed doubts as to whether predictions of this nature are very frequent during language comprehension, given the intuition that such predictions would be wrong more often than right, and that there would be cognitive costs in recovering from an incorrect guess. We've attempted to survey the ERP language processing literature in search of the benefits and costs that should accompany correct and incorrect predictions which, in turn, may shed light on when, how, and how often comprehenders actually make specific predictions.

An inventory of ERP components sensitive to sentence-processing includes the N400, P600, a frontal positivity, and both parietal and frontal positivities sometimes observed following the N400. The N400 appears to be the most general of these responses, as it is elicited by every word in a sentence, albeit with widely varying amplitude. Results to date strongly suggest that N400 amplitude reflects the benefits of semantically compatible context, but that a misleading or mismatching context is much like no context, so that there is little evidence that the N400 reflects a processing cost incurred from a bad prediction. Most N400 results are consistent with the assumption that readers and listeners rapidly integrate word meanings with prior context without necessarily having formed specific predictions about upcoming words, but a handful of results suggest that broad semantic expectations may sometimes coalesce to a more specific lexical prediction.

Late positive components appear to be less frequently elicited during sentence processing, and more readily attributable to cognitive costs incurred when interpretation does not proceed smoothly. The posterior P600 elicited by syntactic errors and some varieties of semantic errors has been widely suggested as reflective of reprocessing cost — reviewing a prior context to determine what went wrong and if the problem might be repaired. This sort of reprocessing need not imply that participants predicted any particular word in advance, only that they detect a problem during attempted integration. We've suggested that late positive components with more anterior scalp distributions are stronger candidates for the signature of a failed prediction from prior semantic context, but only a modest-size literature has yet accumulated on the prevalence and cognitive correlates of these late frontal positivities. When that literature is further advanced, we may be able to estimate the prevalence of prediction during comprehension.

References

Altmann, G.T.M., Kamide, Y., 1999. Incremental interpretation at verbs: restricting the domain of subsequent reference. Cognition 73, 247–264.

Altmann, G.T.M., Steedman, M.J., 1988. Interaction with context during human sentence processing. Cognition 30, 191–238.

Anderson, M.C., Bjork, R.A., Bjork, E.L., 1994. Remembering can cause forgetting — retrieval dynamics in long-term memory. Journal of Experimental Psychology. Learning, Memory, and Cognition 20, 1063–1087.

Andrews, S., Shelley, A.M., Ward, P.B., Fox, A., Catts, S.V., McConaghy, N., 1993. Event-related potential indexes of semantic processing in schizophrenia. Biological Psychiatry 34, 443–458.

Arbel, Y., Spencer, K.M., Donchin, M., 2011. The N400 and P300 are not all that independent. Psychophysiology 48, 861–875.

Ardal, S., Donald, M.W., Meuter, R., Muldren, S., Luce, M., 1990. Brain responses to semantic incongruity in bilinguals. Brain and Language 39, 187–205.

Barber, H., Carreiras, M., 2005. Grammatical gender and number agreement in Spanish: an ERP comparison. Journal of Cognitive Neuroscience 17, 137–153.

Bendixen, A., SanMiguel, I., & Schröger, E. (in press). Early electrophysiological indicators for predictive processing in audition: A review. International Journal of Psychophysiology.

Besson, M., Macar, F., 1987. An event-related potential analysis of incongruity in music and other non-linguistic contexts. Psychophysiology 24, 14–25.

Besson, M., Kutas, M., Van Petten, C., 1992. An event-related potential analysis of semantic congruity and repetition effects in sentences. Journal of Cognitive Neuroscience 4, 132–149.
Besson, M., Faita, F., Czternasty, C., Kutas, M., 1997. What's in a pause: event-related potential analysis of temporal disruptions in written and spoken sentences. Biological Psychiatry 46, 3–23.
Bornkessel, I., Schlesewsky, M., 2006. The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. Psychological Review 113, 787–821.
Borovsky, A., Kutas, M., Elman, J., 2010. Learning to use words: event-related potentials index single-shot contextual word learning. Cognition 116, 289–296.
Breton, F., Ritter, W., Simson, R., Vaughan Jr., H.G., 1988. The N2 component elicited by stimulus matches and multiple targets. Biological Psychiatry 27, 23–44.
Brown-Schmidt, S., Canseco-Gonzalez, E., 2004. Who do you love, your mother or your horse? An event-related brain potential analysis of tone processing in Mandarin Chinese. Journal of Psycholinguistic Research 33, 103–135.
Butterfield, B., Mangels, J.A., 2003. Neural correlates of error detection and correction in a semantic retrieval task. Brain Research. Cognitive Brain Research 17, 793–817.
Cohen, G., Faulkner, D., 1983. Word recognition: age differences in contextual facilitation effects. The British Journal of Psychiatry 74, 239–251.
Connolly, J.F., Phillips, N.A., 1994. Event-related potential components reflect phonological and semantic processing of the terminal words of spoken sentences. Journal of Cognitive Neuroscience 6, 256–266.
Connolly, J.F., Phillips, N.A., Forbes, K.A.K., 1995. The effects of phonological and semantic features of sentence-ending words on visual event-related potentials. Electroencephalography and Clinical Neurophysiology 94, 276–287.
Coulson, S., Van Petten, C., 2007. A special role for the right hemisphere in metaphor comprehension? ERP evidence from hemifield presentation. Brain Research 1146, 128–145.
Coulson, S., King, J.W., Kutas, M., 1998. Expect the unexpected: event-related brain response to morphosyntactic violations. Language and Cognitive Processes 13, 21–58.
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, 129–147.
Courchesne, E., Hillyard, S.A., Galambos, R., 1975. Stimulus novelty, task relevance and the visual evoked potential in man. Electroencephalography and Clinical Neurophysiology 39, 131–143.
Daltrozzo, J., Wioland, N., Kotchoubey, B., 2007. Sex differences in two event-related potentials components related to semantic priming. Archives of Sexual Behavior 36, 555–568.
Dambacher, M., Kliegl, R., Hoffman, M., Jacobs, A.M., 2006. Frequency and predictability effects on event-related potentials during reading. Brain Research 1084, 89–103.
Debener, S., Makeig, S., Delorme, A., Engel, A.K., 2005. What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. Brain Research. Cognitive Brain Research 22, 309–321.
Deldin, P., Keller, J., Casas, B.R., Best, J., Gergen, J., Miller, G.A., 2006. Normal N400 in mood disorders. Biological Psychology 71, 74–79.
DeLong, K.A., Urbach, T.P., Kutas, M., 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nature Neuroscience 8, 1117–1145.
DeLong, K.A., Urbach, T.P., Groppe, D.M., Kutas, M., 2011. Overlapping dual ERP responses to low cloze probability sentence continuations. Psychophysiology 48, 1203–1207.
Diaz, M.T., Swaab, T.Y., 2007. Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. Brain Research 1146, 85–100.
Dien, J., Spencer, K.M., Donchin, E., 2004. Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. Psychophysiology 41, 665–678.
Donchin, E., 1981. Surprise!.. Surprise. Psychophysiology 18, 493–513.
Donchin, E., Coles, M.G.H., 1988. Is the P300 component a manifestation of context updating? The Behavioral and Brain Sciences 11, 357–374.
Federmeier, K.D., 2007. Thinking ahead: the role and roots of prediction in language comprehension. Psychophysiology 44, 491–505.
Federmeier, K.D., Kutas, M., 1999a. Right words and left words: electrophysiological evidence for hemispheric differences in meaning processing. Brain Research. Cognitive Brain Research 8, 373–392.
Federmeier, K.D., Kutas, M., 1999b. A rose by any other name: long-term memory structure and sentence processing. Journal of Memory and Language 41, 469–495.
Federmeier, K.D., Kutas, M., 2005. Aging in context: age-related changes in context use during language comprehension. Psychophysiology 42, 133–141.
Federmeier, K.D., McLennan, D.B., De Ochoa, E., Kutas, M., 2002. The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an ERP study. Psychophysiology 39, 133–146.
Federmeier, K.D., Mai, H., Kutas, M., 2005. Both sides get the point: hemispheric sensitivities to sentential constraint. Memory & Cognition 33, 871–876.
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.
Fischler, I., Bloom, P.A., 1985. Effects of constraint and validity of sentence contexts on lexical decisions. Memory & Cognition 13, 128–139.
Folstein, J.R., Van Petten, C., 2008. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. Psychophysiology 45, 152–170.
Folstein, J.R., Van Petten, C., 2011. After the P3: late executive processes in stimulus categorization. Psychophysiology 48, 825–841.
Forster, K.I., 1981. Priming and the effects of sentence and lexical contexts on naming time: evidence for autonomous lexical processing. The Quarterly Journal of Experimental Psychology 33A, 465–495.
Friederici, A.D., Pfeifer, E., Hahne, A., 1993. Event-related brain potentials during natural speech processing: effects of semantic, morphological, and syntactic violations. Brain Research. Cognitive Brain Research 1, 183–192.
Friederici, A.D., Hahne, A., Mecklinger, A., 1996. Temporal structure of syntactic parsing: early and late event-related brain potential effects. Journal of Experimental Psychology. Learning, Memory, and Cognition 22, 1219–1248.
Friederici, A.D., Steinhauer, K., Frisch, S., 1999. Lexical integration: sequential effects of syntactic and semantic information. Memory & Cognition 27, 438–453.
Friederici, A.D., Mecklinger, A., Spencer, K.M., Steinhauer, K., Donchin, E., 2001. Syntactic parsing preferences and their on-line revisions: a spatio-temporal analysis of event-related brain potentials. Brain Research. Cognitive Brain Research 11, 305–323.
Friedman, D., Cycowicz, Y.M., Gaeta, H., 2001. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. Neuroscience and Biobehavioral Reviews 25, 355–373.
Ganis, G., Kutas, M., Sereno, M.I., 1996. The search for common sense: an electrophysiological study of the comprehension of words and pictures in reading. Journal of Cognitive Neuroscience 8, 89–106.
Goldstein, A., Spencer, K.M., Donchin, E., 2002. The influence of stimulus deviance and novelty on the P300 and novelty P3. Psychophysiology 39, 781–790.
Grose-Fifer, J., Hoover, S., Zottoli, T., Rodrigues, A., 2011. Expecting the unexpected: an N400 study of risky sentence processing in adolescents. Psychophysiology 48, 1184–1191.
Grosjean, F., 1980. Spoken word recognition processes and the gating paradigm. Perception & Psychophysics 28, 267–283.
Gunter, T.C., Jackson, J.L., Mulder, G., 1992. An electrophysiological study of semantic processing in young and middle-aged academics. Psychophysiology 29, 38–54.
Gunter, T.C., Stowe, L.A., Mulder, G., 1997. When syntax meets semantics. Psychophysiology 34, 660–676.
Hagoort, P., Brown, C.M., 2000. ERP effects of listening to speech: semantic ERP effects. Neuropsychologia 11, 1518–1530.
Hagoort, P., van Berkum, J., 2007. Beyond the sentence given. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 362, 801–821.
Hagoort, P., Brown, C., Groothusen, J.T., 1993. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. Language and Cognitive Processes 8, 439–483.
Hahne, A., Friederici, A.D., 1999. Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. Journal of Cognitive Neuroscience 11, 194–205.
Hoeks, J.C.J., Stowe, L.A., Doedens, G., 2004. Seeing words in context: the interaction of lexical and sentence level information during reading. Brain Research. Cognitive Brain Research 19, 59–73.
Holcomb, P.J., Neville, H.J., 1991. Natural speech processing: an analysis using event-related brain potentials. Psychobiology 19, 286–300.
Horst, R.L., Johnson Jr., R., Donchin, E., 1980. Event-related brain potentials and subjective probability in a learning task. Memory & Cognition 8, 476–488.
Jackendoff, R., 2002. Foundations of Language. Oxford University Press, New York.
Johnson Jr., R., 1988. The amplitude of the P300 component of the event-related potential: review and synthesis. In: Ackles, P.K., Jennings, J.R., Coles, M.G.H. (Eds.), Advances in Psychophysiology, 3. JAI Press, Greenwich, CT, pp. 69–138.
Johnson Jr., R., Donchin, E., 1980. P300 and stimulus categorization: two plus one is not so different from one plus one. Psychophysiology 17, 167–178.
Johnson, B.W., Hamm, J.P., 2000. High-density mapping in an N400 paradigm: evidence for bilateral temporal lobe generators. Clinical Neurophysiology 111, 532–545.
Juottonen, K., Revonsuo, A., Lang, H., 1996. Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effects. Brain Research. Cognitive Brain Research 4, 99–107.
Kaan, E., Harris, T., Gibson, E., Holcomb, P.J., 2000. The P600 as an index of syntactic integration difficulty. Language and Cognitive Processes 15, 159–201.
Kazmerski, V.A., Friedman, D., 1997. Old/new differences in direct and indirect memory tests using pictures and words in within- and cross-form conditions: event-related potential and behavioral measures. Brain Research. Cognitive Brain Research 5, 255–272.
Kim, A., Osterhout, L., 2005. The independence of combinatory semantic processing: evidence from event-related potentials. Journal of Memory and Language 52, 205–225.
Kintsch, W., 1988. The role of knowledge in discourse comprehension: a construction-integration model. Psychological Review 95, 163–182.
Kolk, H.H.J., Chwilla, D.J., van Herten, M., Oor, P.J.W., 2003. Structure and limited capacity in verbal working memory: a study with event-related potentials. Brain and Language 85, 1–36.
Kuo, T.Y., Van Petten, C., 2006. Prefrontal engagement during source memory retrieval depends on the prior encoding task. Journal of Cognitive Neuroscience 18, 1133–1146.
Kuo, T.Y., Van Petten, C., 2008. Perceptual difficulty in source memory encoding and retrieval: prefrontal versus parietal electrical brain activity. Neuropsychologia 46, 2243–2257.
Kuperberg, G., 2007. Neural mechanisms of language comprehension: challenges to syntax. Brain Research 1146, 23–49.
Kuperberg, G.R., Sitnikova, T., Caplan, D., Holcomb, P., 2003. Electrophysiological distinctions in processing conceptual relationships within simple sentences. Brain Research. Cognitive Brain Research 17, 117–129.
Kutas, M., 1987. Event-related brain potentials (ERPs) elicited during rapid serial visual presentation of congruous and incongruous sentences. In: Johnson Jr., R., Rohrbaugh, J.W., Parasuraman, R. (Eds.), Current Trends in Event-related Potential Research. Electroencephalogr. Clin. Neurophysiol. Supp., 40. Elsevier, Amsterdam, pp. 406–411.
Kutas, M., 1993. In the company of other words: electrophysiological evidence for single-word and sentence context effects. Language and Cognitive Processes 8, 533–572.
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.

Kutas, M., Hillyard, S.A., 1980a. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205.
Kutas, M., Hillyard, S.A., 1980b. Event-related brain potentials to semantically inappropriate and surprisingly large words. Biological Psychiatry 11, 99–116.
Kutas, M., Hillyard, S.A., 1983. Event-related brain potentials to grammatical errors and semantic anomalies. Memory & Cognition 11, 539–550.
Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. Nature 307, 161–163.
Kutas, M., McCarthy, G., Donchin, E., 1977. Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. Science 197, 792–795.
Kutas, M., Lindamood, T.E., Hillyard, S.A., 1984. Word expectancy and event-related brain potentials during sentence processing. In: Kornblum, S., Requin, J. (Eds.), Preparatory States and Processes. Erlbaum, Hillsdale, NJ, pp. 217–237.
Kutas, M., Hillyard, S.A., Gazzaniga, M.S., 1988. Processing of semantic anomaly by right and left hemispheres of commissurotomy patients. Brain 111, 553–576.
Kutas, M., Van Petten, C., Kluender, R., 2006. Psycholinguistics electrified II: 1995–2005, In: Traxler, M., Gernsbacher, M.A. (Eds.), Handbook of Psycholinguistics, second edition. Elsevier, New York, pp. 659–724.
Laszlo, S., Federmeier, K.D., 2008. Minding the PS, queues, and PXQs: uniformity of semantic processing across multiple stimulus types. Psychophysiology 45, 458–466.
León, I., Díaz, J.M., de Vega, M., Hernández, J.M., 2010. Discourse-based emotional consistency modulates early and middle components of event-related potentials. Emotion 10, 863–873.
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, 913–934.
McCallum, W.C., Farmer, S.F., Pocock, P.V., 1984. The effects of physical and semantic incongruities on auditory event-related potentials. Electroencephalography and Clinical Neurophysiology 59, 477–488.
McKinnon, R., Osterhout, L., 1996. Constraints on movement phenomena in sentence processing: evidence from event-related brain potentials. Language and Cognitive Processes 11, 495–523.
Miller, G.A., Isard, S., 1963. Some perceptual consequences of linguistic rules. Journal of Verbal Learning and Verbal Behavior 2, 217–228.
Mitchell, P.F., Andrews, S., Ward, P.B., 1993. An event-related potential study of semantic congruity and repetition in a sentence-reading task — effects of context change. Psychophysiology 30, 496–509.
Moreno, E., Kutas, M., 2005. Processing semantic anomalies in two languages: an electrophysiological exploration in both languages of Spanish–English bilinguals. Brain Research. Cognitive Brain Research 22, 205–220.
Moreno, E.M., Federmeier, K.D., Kutas, M., 2002. Switching languages, switching palabras (words): an electrophysiological study of code switching. Brain and Language 80, 188–207.
Moss, H.E., Marslen-Wilson, W., 1993. Access to word meanings during spoken language comprehension: effects of sentential semantic context. Journal of Experimental Psychology. Learning, Memory, and Cognition 19, 1254–1276.
Münte, T.E., Heinze, H.J., 1994. ERP negativities to syntactic violations in written text. In: Heinze, H.J., Münte, T.F., Mangun, G.R. (Eds.), Cognitive Electrophysiology. Birkhäuser, Boston, pp. 211–238.
Münte, T.E., Heinze, H.J., Mangun, G.R., 1993. Dissociation of brain activity related to syntactic and semantic aspects of language. Journal of Cognitive Neuroscience 5, 335–344.
Neath, I., Surprenant, A., 2003. Working memory, In: Neath, I., Surprenant, A. (Eds.), Human Memory, Second ed. Wadsworth/Thomson, Belmont CA, pp. 68–95.
Nebes, R.D., Roller, F., Holland, A., 1986. Use of semantic context by patients with Alzheimer's disease. Psychology and Aging 1, 261–269.
Newman, R.L., Connolly, J.F., 2004. Determining the role of phonology in silent reading using event-related brain potentials. Cognitive Brain Research 21, 94–105.
Nigam, A., Hoffman, J.E., Simons, R.F., 1992. N400 to semantically anomalous pictures and words. Journal of Cognitive Neuroscience 4, 15–22.
Nixon, S.J., Tivis, R., Ceballos, N., Varner, J.L., Rohrbaugh, J., 2002. Neurophysiological efficiency in male and female alcoholics. Progress in Neuro-Psychopharmacology and Biological Psychiatry 26, 919–927.
Nobre, A.C., McCarthy, G., 1994. Language-related ERPs — scalp distributions and modulation by word type and semantic priming. Journal of Cognitive Neuroscience 6, 233–255.
O'Rourke, P.L., Van Petten, C., 2011. Morphological agreement at a distance: dissociation between early and late components of the event-related brain potential. Brain Research 1392, 62–79.
Osterhout, L., Hagoort, P., 1999. A superficial resemblance does not necessarily mean you are part of the family: counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. Language and Cognitive Processes 14, 1–14.
Osterhout, L., Holcomb, P.J., 1992. Event-related potentials elicited by syntactic anomaly. Journal of Memory and Language 31, 785–806.
Osterhout, L., Mobley, L.A., 1995. Event-related potentials elicited by failure to agree. Journal of Memory and Language 34, 739–773.
Osterhout, L., Holcomb, P.J., Swinney, D.A., 1994. Brain potentials elicited by garden path sentences: evidence of the application of verb information during parsing. Journal of Experimental Psychology. Learning, Memory, and Cognition 20, 786–803.
Otten, M., van Berkum, J.J.A., 2008. Discourse-based word anticipation during language processing: prediction or priming? Discourse Processes 45, 464–496.
Paller, K.A., Kutas, M., McIsaac, H.K., 1995. Monitoring conscious recollection via the electrical activity of the brain. Psychological Science 6, 107–111.
Phillips, C., Kazanina, N., Abada, S.H., 2005. ERP effects of the processing of syntactic long-distance dependencies. Cognitive Brain Research 22, 407–428.
Pijnacker, J., Geurts, B., van Lambalgen, M., Buitelaar, J., Hagoort, P., 2010. Exceptions and anomalies: an ERP study on context sensitivity in autism. Neuropsychologia 48, 2940–2951.
Polich, J., Comerchero, M.D., 2003. P3a from visual stimuli: typicality, task, and topography. Brain Topography 15, 141–152.
Posner, M.I., Snyder, C.R.R., Davidson, B.J., 1980. Attention and the detection of signals. Journal of Experimental Psychology. General 109, 160–174.
Revonsuo, A., Portin, R., Juottonen, K., Rinne, J.O., 1998. Semantic processing of spoken words in Alzheimer's disease: an electrophysiological study. Journal of Cognitive Neuroscience 10, 408–420.
Robichon, F., Besson, M., Habib, M., 2002. An electrophysiological study of dyslexic and control adults in a sentence reading task. Biological Psychiatry 59, 29–53.
Rubin, S.R., Van Petten, C., Glisky, E.L., Newberg, W.M., 1999. Memory conjunction errors in younger and older adults: event-related potential and neuropsychological evidence. Cognitive Neuropsychiatry 16, 459–488.
Ruchkin, D.S., Sutton, S., 1973. Visual evoked and emitted potentials and stimulus significance. Bulletin of the Psychonomic Society 2, 144–146.
Ruchkin, D.S., Sutton, S., Tueting, P., 1975. Emitted and evoked P300 potentials and variation in stimulus probability. Psychophysiology 12, 591–595.
Ruchsow, M., Trippel, N., Groen, G., Spitzer, M., Kiefer, M., 2003. Semantic and syntactic processes during sentence comprehension in patients with schizophrenia: evidence from event-related potentials. Schizophrenia Research 64, 147–156.
Ruchsow, M., Groen, G., Spitzer, M., Hermle, L., Buchheim, A., Keifer, M., 2008. Electrophysiological evidence for a syntactic but not a semantic deficit in patients with major depression. Journal of Psychophysiology 22, 121–129.
Sanford, A.J., Leuthold, H., Bohan, J., Sanford, A.J.S., 2011. Anomalies at the borderline of awareness. Journal of Cognitive Neuroscience 23, 514–523.
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, 344–354.
Schwanenflugel, P.J., Shoben, E.J., 1985. The influence of sentence constraint on the scope of facilitation for upcoming words. Journal of Memory and Language 24, 232–252.
Sedivy, J.C., Tanenhaus, M.K., Chambers, C.G., Carlson, G.N., 1999. Achieving incremental semantic interpretation through contextual representation. Cognition 71, 109–147.
Seidenberg, M.S., Tanenhaus, J.K., Lieman, J.M., Bienkowski, M., 1982. Automatic access of the meanings of ambiguous words in context: some limitations of knowledge-based processing. Cognitive Psychology 14, 489–537.
Senkfor, A.J., Van Petten, C., 1998. Who said what? An event-related potential investigation of source and item memory. Journal of Experimental Psychology. Learning, Memory, and Cognition 24, 1005–1025.
Spencer, K.M., Dien, J., Donchin, E., 1999. A componential analysis of the ERP elicited by novel events using a dense electrode array. Psychophysiology 36, 409–414.
Spencer, K.M., Dien, J., Donchin, E., 2001. Spatiotemporal analysis of the late ERP responses to deviant stimuli. Psychophysiology 38, 343–358.
Squires, N.K., Squires, K.C., Hillyard, S.A., 1975. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. Electroencephalography and Clinical Neurophysiology 38, 387–401.
Squires, K.C., Wickens, C., Squires, N.K., Donchin, E., 1976. The effect of stimulus sequence on the waveform of the cortical event-related potential. Science 193, 1142–1146.
Squires, K.C., Petuchowsky, S., Wickens, C., Donchin, E., 1977. The effects of stimulus sequence on even related potentials: a comparison of visual and auditory sequences. Perception & Psychophysics 22, 31–40.
Stanovich, K.E., West, R.F., 1981. The effect of sentence context on word recognition: tests of a two-process theory. Journal of Experimental Psychology. Human Perception and Performance 7, 658–672.
Stanovich, K.E., West, R.F., 1983. On priming by a sentence context. Journal of Experimental Psychology. General 112, 1–36.
Sutton, S., Tueting, P., Zubin, J., John, E.R., 1967. Information delivery and the sensory evoked potential. Science 155, 1436–1439.
Swaab, T., Brown, C., Hagoort, P., 1997. Spoken sentence comprehension in aphasia: event-related potential evidence for a lexical integration deficit. Journal of Cognitive Neuroscience 9, 39–66.
Tabossi, P., 1991. Understanding words in context. In: Simpson, G.B. (Ed.), Understanding Word and Sentence. : Advances in Psychology, 77. Elsevier, Amsterdam, pp. 1–22.
Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., Sedivy, J.C., 1995. Integration of visual and linguistic information in spoken language comprehension. Science 268, 1632–1634.
Thornhill, D.E., Van Petten, C. (submitted for publication). Lexical versus conceptual specificity of two ERP components elicited by plausible but unpredictable sentence completions.
Traxler, M.J., Foss, D.J., 2000. Effects of sentence constraint on priming in natural language comprehension. Journal of Experimental Psychology. Learning, Memory, and Cognition 26, 1266–1282.
Tulving, E., Gold, C., 1963. Stimulus information and contextual information as determinants of tachistoscopic recognition of words. Journal of Experimental Psychology 66, 319–327.
Tyler, L.K., Wessels, J., 1983. Quantifying contextual contributions to word recognition processes. Perception & Psychophysics 34, 409–420.
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, 657–671.
van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. Journal of Experimental Psychology. Learning, Memory, and Cognition 31, 443–467.
van den Brink, D., Brown, C.M., Hagoort, P., 2001. Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. Journal of Cognitive Neuroscience 13, 967–985.
van den Brink, D., Brown, C.M., Hagoort, P., 2006. The cascaded nature of lexical selection and integration in auditory sentence processing. Journal of Experimental Psychology. Learning, Memory, and Cognition 32, 364–372.

van Herten, M., Kolk, H.H., Chwilla, D.J., 2005. An ERP study of P600 effects elicited by semantic anomalies. Brain Research. Cognitive Brain Research 22, 241–255.
van Herten, M., Chwilla, D.J., Kolk, H.H., 2006. When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. Journal of Cognitive Neuroscience 18, 1181–1197.
Van Petten, C., 1993. A comparison of lexical and sentence-level context effects and their temporal parameters. Language and Cognitive Processes 8, 485–532.
Van Petten, C., 1995. Words and sentences: event-related brain potential measures. Psychophysiology 32, 511–525.
Van Petten, C., Kutas, M., 1990. Interactions between sentence context and word frequency in event-related brain potentials. Memory & Cognition 18, 380–393.
Van Petten, C., Kutas, M., 1991a. Influences of semantic and syntactic context on open and closed class words. Memory & Cognition 19, 95–112.
Van Petten, C., Kutas, M., 1991b. Electrophysiological evidence for the flexibility of lexical processing. In: Simpson, G.B. (Ed.), Understanding Word and Sentence. : Advances in Psychology, 77. Elsevier, Amsterdam, pp. 129–174.
Van Petten, C., Luka, B.J., 2006. Neural localization of semantic context effects in electromagnetic and hemodynamic studies. Brain and Language 97, 279–293.
Van Petten, C., Senkfor, A.J., 1996. Memory for words and novel visual patterns: repetition, recognition, and encoding effects in the event-related brain potential. Psychophysiology 33, 491–506.
Van Petten, C., Coulson, S., Rubin, S., Plante, E., Parks, M., 1999. Timecourse of word identification and semantic integration in spoken language. Journal of Experimental Psychology. Learning, Memory, and Cognition 25, 394–417.
Vissers, C.Th.W.M., Chwilla, D.J., Kolk, H.H.J., 2006. Monitoring in language perception: the effect of misspellings of words in highly constrained sentences. Brain Research 1106, 150–163.
West, R.F., Stanovich, K.E., 1982. Source of inhibition in experiments on the effect of sentence context on word recognition. Journal of Experimental Psychology. Learning, Memory, and Cognition 8, 395–399.
Wicha, N.Y., Moreno, E.M., Kutas, M., 2004. Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. Journal of Cognitive Neuroscience 16, 1272–1288.
Wlotko, E.W., Federmeier, K.D., 2007. Finding the right word: hemispheric asymmetries in the use of sentence context information. Neuropsychologia 45, 3001–3014.
Woodward, S.H., Ford, J.M., Hammett, S.C., 1993. N4 to spoken sentences in young and older subjects. Electroencephalography and Clinical Neurophysiology 87, 306–320.