

THE ELECTROPHYSIOLOGY OF DISCOURSE AND CONVERSATION

Jos J.A. Van Berkum

Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands

Donders Institute for Brain, Cognition and Behaviour:
Centre for Cognitive Neuroimaging
Nijmegen, The Netherlands

October 1, 2008 (vanberkum-handbookofpsycholinguistics-revision.doc/pdf)

To appear in:

M. Spivey, M. Joannis, & K. McRae (Eds), *The Cambridge Handbook of Psycholinguistics*,
Cambridge: Cambridge University Press.

Address for correspondence:

Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands
www.josvanberkum.nl, jos.vanberkum@mpi.nl, +31-24-3521309 (911), fax: +31-24-3521213

1. Introduction

What's happening in the brains of two people having a conversation? One reasonable guess is that in the fMRI scanner we'd see most of their brains light up. Another is that their EEG will be a total mess, reflecting dozens of interacting neuronal systems. Conversation recruits *all* of the basic language systems reviewed in this book. It also heavily taxes cognitive systems more likely to be found in handbooks of memory, attention and control, or social cognition (Brownell & Friedman, 2001). With most conversations going beyond the single utterance, for instance, they place a heavy load on episodic memory, as well as on the systems that allow us to reallocate cognitive resources to meet the demands of a dynamically changing situation. Furthermore, conversation is a deeply social and collaborative enterprise (Clark, 1996; this volume), in which interlocutors have to keep track of each others state of mind and coordinate on such things as taking turns, establishing common ground, and the goals of the conversation.

For somebody invited to review the neurocognition of discourse and conversation in a handful of pages, this doesn't sound too good. Fortunately, others have already recently surveyed the implications of patient research (Brownell & Friedman, 2001; Marr, 2004), as well as those of fMRI and PET (Ferstl, 2007; Ferstl, Neumann, Bogler & von Cramon, 2008). Here, I review the EEG research on discourse and conversation to date, to see what electrophysiology tells us about the systems involved. EEG studies on discourse and conversation have thus far practically all focused on the *comprehension of multi-sentence text*, and as such they provide a convenient operational definition of the domain. However, an important theme developed in this review is that for the processes that extract meaning from language, the classic separation between a single sentence and a whole bunch of them may not be all that relevant.

2. Why bother with EEG?

The term 'neuroimaging' has come to stand for functional magnetic resonance imaging (fMRI) or positron emission tomography (PET). Images obtained with fMRI or PET can be extremely informative, and they strongly resonate with our desire to look into our working brains. However, in the end, these are images of blood, and, for that matter, slowly responding blood. Neurons do need it, but the real work for which the brain has become famous – computation – is carried by electricity (and, at synapses, electro-chemical transduction). This is precisely what the EEG taps in to. The EEG or electro-encephalogram is a record of tiny voltage fluctuations over time, usually recorded at the scalp. Roughly speaking, these fluctuations arise when large populations of similarly oriented neurons are simultaneously (de-) activated. Because a great many such ensembles are at work at any one time, the EEG is usually a bit of a mess. However, there are ways to extract useful information from this signal. In the most common method, the EEG is recorded to many instances of some critical event (say, a word), and the associated bits of EEG are then averaged to create an event-related brain potential or ERP.

Now, to a psycholinguist, this probably doesn't sound too exciting so far. However, in the late 1970's, researchers discovered that words that were semantically anomalous in their local sentence context, as in "He spread the warm bread with socks", elicited a negative ERP deflection that peaked at about 400 ms after the offending word (Kutas & Hillyard, 1980). Because this so-called N400 effect was not elicited by a typographic anomaly, it was taken to reflect some aspect of how words are related to their semantic context. Follow-up experiments soon confirmed this, and made clear that N400 effects actually reflected graded modulations of an underlying N400 *component*, elicited by every content word, with an amplitude that increased to the extent that the word was harder to relate to its sentence-semantic context (see Kutas & Federmeier, 2000; Kutas, Van Petten, & Kluender, 2006, for review).

The significance of all this was enhanced when syntactically anomalous or unexpected words were found to elicit a very different ERP effect, the so-called P600 effect (Osterhout & Holcomb, 1992; Hagoort, Brown, & Groothusen, 1993; see Osterhout, Kim & Kuperberg, this volume; Kuperberg, 2007, for review). The discovery of two very different ERP ‘signatures’ associated with sentence-level language comprehension not only raised interesting theoretical questions about the architecture of the comprehension system and the types of representations it computed, but also implied that ERPs could be used to *selectively* keep track of specific aspects of comprehension, as a sentence unfolded in time. Of course, a process as complex as this was bound to generate more than just two ERP effects, and many other language-relevant ERP phenomena have been discovered since (cf. Connolly, this volume; Coulson, this volume, Osterhout et al., this volume). With a steadily growing repertoire of ERP effects in hand, researchers are now examining the processes involved in comprehending single sentences.

Until quite recently, however, psycholinguists have been reluctant to use EEG to examine the comprehension of larger units of language. One reason is that psycholinguistic ERP research is for historical reasons strongly rooted in the sentence processing community. This means that (at the time of writing) most of the people with EEG expertise and easy access to EEG labs have sentence processing issues in mind, whereas those most interested in discourse and conversation are short of expertise and labs. Furthermore, combining EEG with single sentences is already difficult enough as it is, and with a text or conversation in each trial, things are just getting worse.

As testified by the work reviewed here, though, it is not impossible to use EEG to track discourse comprehension (see Van Berkum, 2004, for a methodological tutorial). Moreover, the additional work does sometimes really pay off. Below, I take stock of the discourse-oriented EEG studies conducted so far. I begin with studies that used the N400 to address the construction of discourse-level meaning (section 3), then turn to research on how people find out to what or whom certain expressions refer (section 4), and subsequently review the evidence for discourse-based anticipation (section 5). After this review of exclusively text-oriented research, I briefly look at initial attempts to draw the speaker into the picture (section 6), and discuss the challenge of combining electrode caps with conversation (section 7). Finally, I extract several general implications (section 8), and end with where I think the field should be heading (section 9). For related research on figurative processing, I refer to Coulson (this volume).

3. Making sense of discourse

3.1 Initial observations

Although prose passages had in fact already featured in several of the early landmark ERP experiments on language comprehension (Kutas & Hillyard, 1983; Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991), ERP research specifically aimed at *discourse*-level comprehension took off in 1994, with a pioneering study by St. George, Mannes, and Hoffman (1994). Building on classic work in cognitive psychology, St. George et al. asked their subjects to read short stories with or without a title, such as:

1. The procedure is actually quite simple. First you arrange things into different groups depending on their makeup. Of course, one pile may be sufficient depending on how much there is to do. If you have to go somewhere else due to lack of facilities that is the next step, otherwise you are pretty well set. It is important not to overdo any particular endeavor. That is, it is better to do too few things at once than too many. In the shorter run this may not seem important, but complications from doing too many can easily arise. A mistake can be expensive as well. The manipulation of the appropriate mechanisms should be self-explanatory, and we need not dwell on it here. At first the whole procedure will seem complicated. Soon, however, it will become just another facet of life. It is difficult to foresee any end to the necessity of this task in the immediate future, but then one can never tell.

Whereas each story was locally coherent in that individual sentences were interconnected and related to a single topic, it was difficult to find out what that topic actually was without a story title (e.g., *Procedure for washing clothes*). Content words in stories without a title elicited larger N400 deflections than the same words in titled stories, clearly indicating that the N400 is not only sensitive to local lexical or sentential context.

In follow-up research, Van Berkum and colleagues (Van Berkum, Hagoort & Brown, 1999; Van Berkum, Zwitserlood, Brown & Hagoort, 2003) examined the brain's response to words that were equally acceptable in their local carrier sentence but differed radically in how well they fit the wider discourse (e.g., "The dean told the lecturer that there was ample reason to promote him" in a context where the lecturer had just been found out to be a fraud). Words that were not supported by the wider discourse elicited a much larger N400 than discourse-supported words, in readers and listeners alike. Furthermore, this N400 effect collapsed when the same carrier sentences were presented in isolation. Because the discourse-dependent N400 effect emerged for clause-final as well as clause-medial words, these findings suggested that *every* incoming word is immediately related to the wider discourse. Furthermore, with spoken words, the discourse-dependent N400 effect had its onset as early as 150 ms after acoustic word onset, i.e., only some *2-3 phonemes into the word* (such early onsets are quite common for spoken-language N400 effects). Spoken words are thus related to the wider discourse extremely rapidly, well before they have been fully pronounced, and possibly even before they have become acoustically unique. Finally, the timing, shape and scalp distribution of the N400 effect elicited by discourse-dependent anomalies did not differ from that of the 'classic' sentence-dependent N400 effect, indicating that discourse- and sentence-dependent semantic constraints have a comparable impact on some aspect of semantic processing.

Three additional observations ruled out relatively trivial accounts of discourse-dependent N400 effects. First, like its sentence-dependent counterpart, it does not depend on anomaly but can also be elicited by somewhat less expected coherent words like "lance" in (2) (Otten & Van Berkum, 2007; see also Burkhard, 2006, 2007; Ditman & Kuperberg, 2007; St. George, Mannes & Hoffman, 1997; Van Berkum, Brown, Zwitserlood, Kooijman & Hagoort, 2005).

2. The brave knight saw that the dragon threatened the benevolent sorcerer.
He quickly reached for a **sword** / **lance** and ... (story continues)

Thus, in contrast to a recent suggestion (Geurts, in press), we can take these ERP modulations to reflect something about the normal computations involved in understanding sentences in context. Second, discourse-dependent N400 effects also emerge at story positions where no *particular* word is highly expected (Van Berkum, Zwitserlood, Brown & Hagoort, 2003; Camblin, Gordon, & Swaab, 2007), and thus do not critically depend on disconfirming a specific lexical prediction. Third, experiments that controlled for potential confounding effects from lexical priming have shown that at least *some* discourse-dependent N400 effects hinge on the precise message-level representation established for the discourse at that point. For example, whereas "lance" elicited a large standard N400 effect relative to "sword" in (2), it did not do so in (3), where the message was altered while keeping the same prime words (Otten & Van Berkum, 2007; see also Ditman & Kuperberg, 2007).

3. The benevolent sorcerer saw that the dragon threatened the brave knight.
He quickly reached for a **sword** / **lance** and ... (story continues)

In all, the N400 findings so far consistently indicate that words are related to a message-level representation of the wider discourse very rapidly, in a way that is no different from how they are related to local sentence-level context.

3.2 What is going on here?

Because these are ‘message-level’ discourse effects, it is tempting to take them to *directly* reflect some aspect of high-level conceptual structure building, such as (contextually modulated) sentence-level semantic composition (e.g., Jackendoff, 2007) or the incremental expansion of a discourse model (e.g., along the lines of Discourse Representation Theory; Kamp & Reyle, 1993; Geurts, 1999). This *integration account* of the N400 is implicitly or explicitly behind much discourse-level N400 research (including e.g., Van Berkum, Hagoort & Brown, 1999), and discourse-dependent N400 effects are sometimes taken as strong evidence *for* the integration view. However, these effects are also fully compatible with a *memory retrieval account* (Kutas & Federmeier, 2000; Kutas et al., 2006), in which “N400 amplitude is a general index of the ease or difficulty of retrieving stored conceptual knowledge associated with a word (or other meaningful stimuli), which is dependent on both the stored representation itself, and the retrieval cues provided by the preceding context” (Kutas et al., 2006, p. 669). After all, prior discourse can make for a fine retrieval cue.

These alternative accounts of discourse-level N400 effects are difficult to disentangle, and, according to some (e.g., Coulson & Federmeier, *in press*), may be two sides of the same coin. However, as argued in detail elsewhere (Van Berkum, *in press*), a long-term memory retrieval account of discourse-dependent N400 effects allows these findings to be much more parsimoniously unified with a wide range of other N400 phenomena. For example, N400 effects can also be obtained by semantically odd combinations that are non-linguistic, such as an unexpected turn of events in a film clip (e.g., a guy in front of the bathroom mirror putting shaving cream on his chin, and then picking up a rolling pin; Sitnikova, Kuperberg & Holcomb, 2003) or a cartoon-like picture sequence (West & Holcomb, 2002). Furthermore, lexical or lexico-conceptual factors such as repetition, word frequency, associative or semantic word-word priming, the number of orthographically similar words, and concreteness also modulate the N400 (see Kutas et al., 2006 for review; or Van Berkum, *in press*, for references). Although the impact of such factors is often attenuated in richer contexts, the fact that they modulate the N400 speaks against an account in which the N400 *directly* reflects sophisticated conceptual structure building. Although the N400 might be sensitive to the latter, it clearly reflects a more general mechanism.

A central idea in the so-called memory-based text processing tradition (e.g., Gerrig & McKoon, 1998; Gerrig & O’Brien, 2005) is *readiness*, the timely availability of plausibly relevant information. Firmly based in general models of human memory (e.g., Ratcliff & McKoon, 1988; Anderson, 1990), research in this tradition has shown that as we read through a text, potentially relevant additional information waxes and wanes without cost, as a function of how our content-addressable long-term memory passively ‘resonates’ to currently active representations (Myers & O’Brien, 1998). What is being anticipated here is not necessarily upcoming communication – our long-term memory is simply trying to predict information that might soon be needed (cf. Anderson, 1990). However, to the extent that such cost-free memory-based anticipation is also sensitive to complex *message-level* cues, this automatic resonance may well be a critical ingredient in how discourse-dependent N400 effects emerge. In the St. George et al. (1994) study, for instance, the presence of a story title can help retrieve the relevant meaning of content words in the story. In the Otten & Van Berkum (2007) study, the precise message-level representation of “The brave knight saw that the dragon threatened the benevolent sorcerer. He quickly reached for a...” should facilitate retrieval from memory of a relevant meaning for “sword”, more so than for “lance”. And in a story about a swindling lecturer (Van Berkum, Hagoort & Brown, 1999), the exact meaning of “The dean told the lecturer that there was ample reason to...”, presumably increase the readiness of concepts involving grave consequences (being reprimanded or fired), at the expense of happier concepts (e.g., being praised or promoted).

Like the integration account, the memory retrieval account of N400 effects is as yet far from explicit. It also has one distinct and experimentally nasty drawback: If a message-level

representation of the prior discourse is but one of a mixed bag of memory cues that can facilitate the retrieval of this word's meaning from conceptual long-term memory, next to, say, associatively and semantically related prime words, scenario-based world knowledge activated the preceding text, and aspects of the visual and auditory scene (Kutas & Federmeier, 2000; Van Berkum, *in press*), scientific prediction doesn't get any easier. After all, the net N400 amplitudes observed are then a function of some unknown (and possibly weighed) balance of high- and lower-level retrieval cues. But then again, perhaps nature just isn't any more accommodating. Below, we'll encounter several cases where exactly this mixing of higher- and lower-level cues seems to be at hand.

3.3 Inference

The word-elicited N400 is clearly sensitive to the level of contextual support from prior text. In principle, this makes it an attractive tool with which to study the mechanisms below various types of text-based inferences. Again, St. George, Mannes, and Hoffman (1997) were the first to do so, in an experiment on bridging and elaborative inferences in text. Unfortunately, with only few texts per condition (15), the ERPs were relatively noisy and hence difficult to interpret. Also, because not every item had been designed to elicit an N400 effect at a single critical word, ERPs were pooled across *all* words in the critical sentence, a procedure that must inevitably have attenuated critical differences between conditions.

Two more recent studies on inferences avoided both of these problems. In the first (Ditman & Kuperberg, 2007), participants read sentences like (7) in the context of an explicitly supportive discourse context (4), an inferentially supportive discourse context (5) or a non-supportive discourse context (6):

4. Mark and John were having an argument. Mark began to hit John hard.
5. Mark and John were having an argument. Mark got more and more upset.
6. Mark and John were gambling at the casino. They won every game of blackjack.

7. The next morning John had many **bruises**.

As might be expected, a critical word like "bruises" elicited a smaller N400 in an explicitly supportive discourse context than in a non-supportive one. However, such words also elicited a smaller N400 effect in the inferentially supportive discourse context, indicating that readers can rapidly work out how the sentences of an unfolding discourse are connected.

Results from the second study (Yang, Perfetti, & Schmalhofer, 2005) seem to be at odds with this conclusion. Here, participants read sentences like (12) in the context of a referentially explicit discourse context (8), a referentially paraphrased context (9), an inferentially supportive context (10) or a non-supportive context (11):

8. After being dropped from the plane, the bomb hit the ground and exploded.
9. After being dropped from the plane, the bomb hit the ground and blew up.
10. After being dropped from the plane, the bomb hit the ground.
11. Once the bomb was stored safely on the ground, the plane dropped off its passengers and left.

12. The **explosion** was quickly reported to the commander.

The referentially explicit and referentially paraphrased contexts attenuated the N400 on critical words (e.g. "explosion"), relative to the non-supportive context. But inferentially supportive context did *not*, suggesting that the critical word was as unexpected there as it was in the non-supportive control context. This could be taken to reflect that readers were not making the needed inference (i.e., a bomb hitting the ground will explode), at least not sufficiently rapidly.

Although it is not immediately clear how to reconcile these findings, the Ditman and Kuperberg (2007) study does provide an existence proof that inferences *can* be made sufficiently rapidly to support the processing of a subsequent critical word in text. Research discussed in sections 4 and 5 below (e.g., Burkhardt, 2006; Burkhardt & Roehm, 2007; Van Berkum et al., 2005; Otten & Van Berkum, 2007) point in the same direction. In line with the memory retrieval perspective outlined in section 3.2, as well as with behavioral evidence for automatic memory based inferences (Gerrig & O'Brien, 2005), I suspect that these inference-dependent N400 attenuations are the result of automatic 'anticipatory' processing before the critical word comes along, afforded by how our long-term memory works (see also section 5)

3.4 World knowledge, counterfactuals and negation

Several recent studies use the N400 to explore the interaction between a dynamically configured discourse context and stable, default background knowledge about the world. In a first study (Nieuwland & Van Berkum, 2006a), a cartoon-like discourse context was pitted against default knowledge about what animate and inanimate entities tend to be like in the real world:

13. A woman saw a dancing peanut who had a big smile on his face. The peanut was singing about a girl he had just met. And judging from the song, the peanut was totally crazy about her. The woman thought it was really cute to see the peanut singing and dancing like that. The peanut was **salted** / **in love**, and by the sound of it, this was definitely mutual. He was seeing a little almond.

In texts such as (13), discourse-supported but animacy-violating predicates ("in love") elicited a smaller N400 than animacy-respecting but discourse-inappropriate alternatives ("salted"), showing that in this case, discourse-contextual fit completely overruled local animacy requirements in driving the N400. A recent follow-up (Filik & Leuthold, 2008) confirmed these results, and showed that such fiction-dependent overrides of default world knowledge can be brought about with a single sentence (e.g., "The Incredible Hulk was annoyed there was a lot of traffic in his way. He picked up the **lorry** and..."). Under some models, such results might be very surprising. However, I suspect that even the first few words of a single sentence ("The Incredible Hulk picked up the **lorry**") can do the trick. Note that sentences are *always* interpreted against a background. In the absence of specific prior information, the interpretive background defaults to our conceptual knowledge of the real world. A prior discourse context simply 'customizes' the interpretive background. Either way, it will always be the case that some bits of information are more expected than others (cf. section 3.2).

This is not to say that well-known real-world constraints cannot 'leak through' and have their own, independent effect on N400 amplitude, over and above what might be expected on the basis of a dynamically configured message-level representation of the discourse (e.g., see Federmeier & Kutas, 1999; Hald, Steenbeek-Planting & Hagoort, 2007; Ferguson, Sanford & Leuthold, in press). In fact, if a message-level representation is but one of a mixed bag of memory cues that can facilitate the retrieval of the next word's meaning from memory (Kutas & Federmeier, 2000; Van Berkum, in press), this is exactly what one can expect. As discussed before, the net balance between message-level and other cues should then depend on the relative strengths of the cues. And indeed, lexical primes such as "arms" and "legs" or "moon" and "stars", which instantiate default world knowledge in a simple associative or category-dependent way, seem to have a bigger effect on the N400 in moderately constraining contexts (Camblin, Gordon & Swaab, 2007) than in strongly constraining ones (Van Petten, 1993).

The balance between the dynamically configured message and simpler word- or scenario-based memory cues may also be relevant to a recent ERP study on negation in counterfactuals (Ferguson, Sanford & Leuthold, in press).

14. If cats were not carnivores, they would be cheaper for owners to look after.
Families could feed their cat a bowl of **carrots** / **fish** and listen to it purr happily.

In stories such as (14), discourse-supported words like “carrots” elicited a larger N400 than words like “fish”, a result that was taken to index the failure of discourse context to immediately affect language comprehension. However, in items like these, the precise message-level representation faces rather tough competition from other cues that reinforce real-world contingencies, such as convergent or scenario-mediated priming from nearby words like “feed”, “cat”, and possibly “bowl”. In addition, the effectiveness of the message-level cue depends on how well (and how rapidly) readers have been able to incorporate the negator. At odds with early N400 observations (Fischler, Bloom, Childers, Roucos & Perry, 1983; Kounios & Holcomb, 1992), everyday language experience suggests that people can in principle interpret negations very effectively (“I’m afraid your paper is not suitable for publication in Science”), and recent evidence (Nieuwland & Kuperberg, in press) suggests that what may matter is how pragmatically felicitous the negation is. With negation ‘coming out of the blue’ (“if cats were not carnivores”, “a robin is not a bird”), it is probably much more difficult to rapidly construct an adequate message-level representation.

In all, it seems that in pragmatically felicitous narratives, the N400 is highly sensitive to how things fit what is being talked about right now, even if it happens to be some imaginary world or negated microworld. Whether real-world knowledge ‘leaks through’ in the amplitude of a word-elicited N400 may well depend on how effective the discourse is as a cue to long-term memory, relative to the (possibly convergent and/or scenario-mediated) word cues that provide access to default knowledge.

3.5 Semantic illusions and information structure

Behavioral research has shown that people sometimes overlook quite severe semantic anomalies in language, even when asked to look for them (see Sanford & Sturt, 2002, for review). For example, after having been told about a plane crash right on the border between France and Italy, many people miss the anomaly in “The authorities were trying to decide where to bury the survivors” (Barton & Sanford, 1983). What seems critical here is that (a) the context activates a strong scenario, (b) the impostor word – “survivors” – fits the scenario, (c) the impostor word is semantically related to the acceptable word – “victims” – it replaces, and (d) the impostor word is not marked as particularly noteworthy by linguistic focus, such as prosodic stress, italics, or cleft-constructions (“It was the survivors that...”).

The ERP studies reviewed before all indicate that semantically unexpected or anomalous words immediately generate an N400 effect. However, when embedded in spoken stories that obeyed the abovementioned illusion-inducing conditions, such as (15), the severe anomaly at “suitcase” did *not* elicit an N400 effect (Nieuwland & Van Berkum, 2005; see Bohan, 2008, for a similar finding).

15. A tourist wanted to bring his huge suitcase onto the airplane. However, because the suitcase was so heavy, the woman behind the check-in counter decided to charge the tourist extra. In response, the tourist opened his suitcase and threw some stuff out. So now the suitcase of the resourceful tourist weighed less than the maximum twenty kilos. Next, the woman told the **tourist** / **suitcase** that ... (story continues).

The precise interpretation of this remains to be established, and again depend in part on what the N400 can be assumed to reflect. On the integration assumption, the absence of an N400 effect to anomalous words suggests that the comprehension system suffers from a temporary semantic illusion (Nieuwland & Van Berkum, 2005; cf. Hoeks, Stowe & Doedens, 2004, for a similar result in sentences). However, the memory retrieval account of the N400 pursued in this review may open up other possibilities. For the Nieuwland & Van Berkum study, for example, a relevant

factor might be that “suitcase” has been mentioned several times before and is highly salient, eliminating the need for long-term memory retrieval. For both studies, the impostor word’s strong association to the scenario as well as to the coherent word may also facilitate retrieval.

The behavioral literature on semantic illusions has revealed that they are more likely to occur in situations where the word is deaccented or otherwise marked as business-as-usual (e.g., Sanford, Sanford, Molle & Emmott, 2006). This predicts that if the anomaly is brought into focus, an N400 effect should resurface, and studies investigating this are currently underway. Also, there is initial evidence that words brought into focus by accentuation elicit a larger N400 component than a deaccented word, irrespective of anomaly (Li, Hagoort & Yang, 2008). The story for how such phenomena mesh with the memory retrieval account is far from complete (but see Van Berkum, in press, for an attempt in terms of attention-controlled intensified memory retrieval, along lines suggested by Sanford & Garrod, 2005). But I suspect that a closer examination of these surprising phenomena will help us make important progress in unravelling the architecture of interpretation and its link to the N400.

In all, I have argued (see Van Berkum, in press, for details) that the many discourse-dependent modulations of a word-elicited N400 probably do *not directly* reflect the ease or difficulty with which the word’s meaning is integrated in a structured (e.g. Jackendoffian or DRT-style) conceptual representation of the unfolding message. In the face of many other N400 phenomena, a more parsimonious account is that the N400 directly reflects whatever sense-making is achieved through a more primitive resonance-based ‘analysis-by-retrieval’ in long-term memory (i.e., pattern recognition and pattern completion). Message-dependent N400 effects can then simply come about because the structured conceptual representations involved – if sufficiently rapidly computed – can act as retrieval cues (or ‘primes’) to semantic long-term memory. Of course, parsimony only takes you that far. But if the proposal holds up, more direct ERP reflections of message-level structure building must be sought elsewhere.

4. Establishing reference in discourse

In “The deadline for my chapter was 10 days ago. If I don’t finish it soon, the editor will start bugging me.”, determining the referent of “it” seems quite trivial. However, in “The deadline for my chapter was 10 days ago. If I keep on ignoring it, the editor will start bugging me.”, things are not so clear, for “it” can now refer to the chapter or the deadline. Moreover, upon closer examination, in the original example the neuter pronoun “it” can formally also refer to either, and the ambiguity is removed by the meaning of other words in its vicinity. These examples reveal that establishing reference in discourse is a non-trivial job.

4.1 Ambiguous referential expressions

In an ERP study that examined the nature and time course of such discourse-level referential analysis (Van Berkum, Brown & Hagoort, 1999a), participants were presented with short passages that contained a referentially unambiguous or ambiguous expression in the last sentence, such as “the lecturer” in (16).

16. In dismay, the faculty dean called the lecturer and the professor (the two lecturers) to his office. This was because the lecturer (one of the lecturers) had committed plagiarism, and the professor (the other one) had faked some of his research data. The dean told the **lecturer** that there was ample reason to sack him.

Relative to their referentially successful counterparts, referentially ambiguous expressions elicited a sustained negative ERP shift right at the ambiguous written noun, emerging at about 300-400 milliseconds over the front of the head. A similar effect was obtained with spoken nouns (Van

Berkum, Brown, Hagoort, & Zwitserlood, 2003), as well as with spoken and written pronouns in single sentences (“David shot at Michael as he jumped over the fence”; Nieuwland & Van Berkum, 2006b; see Van Berkum, Koornneef, Otten, & Nieuwland, 2007, for review, and Hammer et al., 2005, for a possible exception). Also, the mechanism involved is relatively smart: in a context in which, say, one of the two lecturers has just fled from the building, “The dean told the lecturer” does not elicit the effect, even though the episodic memory of the discourse does contain *two* lecturers (Nieuwland, Otten, & Van Berkum, 2007).

Thus, based on the unfolding sentence as well as prior text, people rapidly work out which are the conceptually *suitable* referents. Also, because the effect of referential ambiguity is largest over the front of the head, clearly different from the N400 effect, we can infer that different aspects of interpretation are handled by at least partly different networks in the brain (see Nieuwland, Petersson & Van Berkum, 2007, for convergent evidence from fMRI). What the network at hand is doing exactly remains to be established, but plausible options include computationally intensive inferencing as people are trying to resolve the ambiguity and, possibly related, a higher load on frontal working memory systems (cf. Müller, King & Kutas, 1997; Münte, Schiltz & Kutas, 1998; Pyllkänen, 2008; Streb, Rösler & Hennighausen, 1999; Baggio, Van Lambalgen & Hagoort, in press; see Van Berkum, in press; Van Berkum et al., 2007; Nieuwland & Van Berkum, 2008; for discussion)

Interestingly, if people do not know who or what is being referred to, this can immediately affect their word-by-word *syntactic* analysis of the sentence (Van Berkum, Brown & Hagoort, 1999a). For instance, when “the lecturer” in “The dean told the lecturer that...” is referentially ambiguous, people are lured into analyzing the subsequent word “that” as starting a relative clause that will tell them which lecturer was meant (e.g., “...the lecturer that had committed plagiarism”) rather than as starting some other syntactic structure (e.g., “...that there was ample reason to sack him”). If the sentence then continues with the latter after all, we see a P600 effect, an effect that is reliably elicited by words at which the currently pursued syntactic analysis runs into problems (Osterhout et al., this volume). Thus, referential factors can lead readers to momentarily perceive a syntactic error when there is none (see Van Berkum et al., 2007, for a comparable phenomenon in single sentences; and Bornkessel, Schlesewsky & Friederici, 2003 for related findings with question-answer pairs).

Even more strikingly, the conceptual ‘pull’ of referential ambiguity can sometimes briefly lure people into pursuing a syntactic analysis that is *ungrammatical*. In a Dutch version of “The dean told the lecturer that...”, for example, referential ambiguity can lead people to momentarily read “that” as a relative pronoun *even if* a relative pronoun of this form does not agree in gender marking with the gender of the preceding noun (Van Berkum, Brown & Hagoort, 1999b). Such observations are important, because they suggest that when interpreting language, people don’t just slavishly follow the syntax. Other recent ERP findings also suggest that good sense can sometimes outweigh good syntax (Kuperberg, 2007). Against so-called ‘syntax-first’ or ‘syntactocentric’ models of processing, this indicates that comprehenders are trying to deal with multiple levels of linguistic structure simultaneously, without giving one of them absolute priority.

In a recent ERP study on how referential processing interacts with semantic composition (Nieuwland & Van Berkum, in press), nouns embedded in a short story could be semantically anomalous, referentially ambiguous, or both (e.g., “The dean ate the lecturer” in a story about two equally salient lecturers). These doubly problematic nouns elicited a clear N400 effect, indicating a problem with how the word’s meaning fit the story context. However, they did not elicit the sustained frontal negativity typically observed with referentially ambiguous nouns, presumably indicating that people did not engage in additional work to disambiguate reference. This is in line with a standard story in which sense-dependent processing precedes – and sometimes preempts – referential computations (but see Nieuwland & Van Berkum, in press, for a catch involving P600 effects in a subset of the participants).

4.2 Unambiguous referential expressions

Referential ambiguity is not the only way to tap into aspects of establishing reference. Several studies have compared the effects of using a repeated, inferrable, or new nominal NP in text. In one of these (Anderson & Holcomb, 2005), participants read mini-stories like:

17. In dismay, Kathy sat nervously in the *cab/taxi* on her way to the airport.
The/A **cab** came very close to hitting a car.

Relative to newly introduced nouns (e.g., “cab” in the first sentence), synonyms in the second sentence (“cab” after “taxi”) elicited a smaller N400, and literally repeated nouns (e.g., “cab” after “cab”) elicited the smallest N400. Although now crossing a sentence boundary, these effects echo other N400 repetition and semantic priming effects, and are easily explained as a result of facilitated long-term memory retrieval (literally repeated NPs do sometimes elicit a *larger* N400 than a newly introduced NPs but this has been shown to depend on whether the discourse status of the antecedent makes repeated reference felicitous or not; Camblin, Ledoux, Boudewyn, Gordon & Swaab, 2007; Ledoux, Gordon, Camblin & Swaab, 2007; Swaab, Camblin & Gordon, 2004).

The N400 effects in the Anderson & Holcomb study did not depend on whether the article suggested a previously introduced (“The...”) or new (“A...”) discourse entity, a null result that was speculatively explained as resulting from weak information structure effects of definiteness being washed out by much stronger effects of repetition and (synonym-)semantic priming in this particular set of materials. Such an explanation is in line with the memory retrieval account of the N400 discussed before. However, and this is important, this should *not* be taken as evidence that readers entirely failed to pick up on the semantic and pragmatic difference between “The cab” and “A cab” – on the memory retrieval account, all one can infer here is that the distinction did not differentially affect memory retrieval at the subsequent noun. The stronger inference that the distinction was somehow not represented in conceptual structure only follows from the stronger assumption that the N400 directly reflects core semantic and pragmatic integration operations.

A second study on repeated, inferrable, and new nominal NPs (Burkhardt, 2006; see also Burkhardt, 2007; Burkhardt & Roehm, 2007) examined ERPs to the NP in the second sentence in:

18. Tobias was having a chat with Nina. He said that the **conductor** was very impressive.
19. Tobias visited a *concert* in Berlin. He said that the **conductor** was very impressive.
20. Tobias visited a *conductor* in Berlin. He said that the **conductor** was very impressive.

Relative to newly introduced nouns (e.g., “conductor” in (18)), repeated nouns (“conductor” in (20)) elicited a smaller N400. ‘Bridged’ nouns that were inferentially related to the antecedent (e.g., “conductor” after “concert” in (19)) also elicited a smaller N400, of equal amplitude as that to the repeated noun. The latter is interesting, as it seems to suggest that in coherent discourse, words like “concert” prepare the reader for “conductor” to the same extent as “conductor” itself does.

Relative to repeated nouns, new and bridged nouns also both elicited an enhanced positivity between 600 and 900 ms, which was interpreted as a P600 effect indexing anaphoric integration cost due to the establishment of an independent, new discourse referent (see Burkhardt, 2007, for a related result). Whether this positivity is indeed an instance of the P600 effect elicited by unexpected structure remains to be established. However, the findings do suggest that late positivities can sometimes index discourse-referential operations. A possibly related finding involving bare quantifiers (Kaan, Dallas & Barkley, 2007) was obtained in a very repetitive and possibly quite transparent design, and should therefore be interpreted with caution.

Two ERP studies have examined the implications of changing the referential status of pronouns across the sentence boundary. One of them (Streb, Hennighausen & Rösler, 2004) systematically varied the distance between a pronoun and its antecedent:

21. Beate has a small “boardy home” for animals.
Everywhere in the house are animals.
Tom is an old cat.
Today **it** scratched the door of the woman. [near]
22. The weather is beautiful today.
Gerhart is an experienced mountaineer.
Anna wants to go on a walking tour.
Then **he** shows the ascent to the tourist. [intermediate]
23. *Lisa* strolls across a bazaar.
Peter sells gems to tourists.
The gems are cut excellently.
Then **she** will buy a diamond from the trader. [far]

In line with the notion that pronouns tend to prefer highly salient antecedents (Garnham, 19??), more distant pronouns elicited larger N400 components. Because pronouns have relatively shallow semantic content, this could be taken as a challenge to the memory retrieval account of the N400. However, they do have *some* meaning, and retrieving this can still be more difficult. Also, a conceptual mismatch between context and word might also lead to intensified retrieval of information related to the *context*. Why such mismatch results in an N400 effect here and in a P600 effect in other studies (e.g., Osterhout & Mobley, 1995; Hammer, Jansma, Lamers & Münte, 2005; Van Berkum et al., 2007) is not yet clear.

Elevated N400 components have also been observed to text-embedded singular definite pronouns “he” or “she” as compared to plural “they” (Filik, Sanford & Leuthold, 2008):

24. The in-flight meal I got was more impressive than usual.
In fact, **she** / **they** courteously presented the food as well.
25. The in-flight meal I got from the stewardess was more impressive than usual.
In fact, **she** / **they** courteously presented the food as well.

Because “they” can more easily be used without an explicit antecedent (e.g., “They’re digging up the road again”), this N400 effect has been proposed to reflect a difference in the referential demands structurally imposed by “he” or “she”, as compared to “they”. However, note that it is not the case that “they” can *always* leave its referent relatively unspecified (e.g., “At the UN meeting, the Chinese felt increasingly uncomfortable because of the human rights issues continuously raised by other delegates. So they left.”). Important, the pronoun *itself* does not signal which situation is at hand. This N400 effect may thus be in need of another explanation.

In the same study, singular definite pronouns without explicit antecedent, like “she” in (24), also elicited a frontocentral positivity that peaked around 750 ms, relative to all other three conditions. Analogous to the late positivity to new and bridged nominals (Burkhardt, 2006; 2007), this – more anterior – late positivity was taken to reflect the discourse complexity involved in inferring and adding an additional entity to the discourse model. Of course, this convergence does not imply that *all* referentially induced late positivities reflect discourse-model expansion. For example, the P600 effect to the pronoun in sentences like “David praised Linda because **he**...” (where verb meaning sets up a referential expectation; Van Berkum et al., 2007) is difficult to explain in terms of the need to set up a new discourse entity, as in this case, the required entity is already there. Whether this means that we’re dealing with a single late positivity with an as yet ill-understood functional interpretation, or with two distinct positivities that happen to look the same, remains to be established (and may require other techniques, e.g., MEG).

So what does all this tell us? Although the diverse ERP findings on establishing reference in discourse resist a simple interpretation, several general observations can be made. First, these findings testify to the speed with which comprehenders try to bind incoming words to entities in the discourse or situation model. Furthermore, although contingent upon some form of compositional semantics, referential processing is not fully constrained by the syntax. Third, whereas the processing consequences of referential ambiguity consistently show up as sustained frontal negativities, the various computations that occur in the absence of such ambiguity show up in several very different ERP effects. The late positivity reported by Burkhardt (2006) and Filik et al. (2008) may genuinely reflect the establishing of a new discourse entity. Other late positivities (Van Berkum, Brown & Hagoort, 1999b; Van Berkum et al., 2007) are more plausibly explained by a strong referential attractor leading the system to at least momentarily put the blame on syntax. And many of the N400 effects to referring nouns and pronouns (Anderson & Holcomb, 2005; Burkhardt, 2006; possibly Streb et al., 2004) may well reflect simple (word-priming) or complex (information-structural) modulations of the ease with which the critical word's meaning is retrieved from memory, making these effects rather indirect – possibly sometimes even poor – indices of establishing a referential dependency. In all, there is clearly no unique ERP signature of referential processing, and the signatures that we get need not all *directly* reflect the processes that link an expression to its discourse-model referent.

5. Anticipation in discourse

To what extent does the speed with which readers and listeners relate the syntactic, semantic and referential implications of incoming words to their knowledge of the wider discourse depend on expectations? This is a difficult question to answer, and we won't get very far unless we ask an additional question: expectations of *what*? After all, language users are continuously analyzing the unfolding discourse at many different levels (such as phonology, syntax, semantics, and reference). Furthermore, listeners and readers not only keep a record of *what is being talked about*, i.e., the 'situation model' (Kintsch, 1998; Zwaan & Singer, 2003) or 'situational representation' (Clark, 1996), but they also keep track of *how the communicative enterprise itself is getting along*, encompassing the 'textbase' (Kintsch, 1998) or 'discourse record' (Clark, 1996) as well as, for example, inferences about what the speaker may or may not know, and about why this conversation is being held in the first place. In discourse, people might in principle develop expectations at *any* of these levels, based on information supplied by any other level.

Some of the first N400 experiments on interpretation in language comprehension were already framed as revealing the effects of word expectancy, in a way that was not deemed to be limited to single sentences (e.g., Kutas & Hillyard, 1984). The attenuated N400 elicited by discourse-predictable words (e.g., "sword" in example (2)) relative to coherent but somewhat less predictable control words ("lance"; Otten & Van Berkum, 2007) is compatible with the idea that readers and listeners actually anticipate such predictable words as the discourse unfolds. More ERP evidence for discourse-based anticipation of features has come from studies (e.g., Federmeier & Kutas, 1999; see Federmeier, 2007, for review) in which highly predictive stories were continued with the expected critical word, a semantically related but anomalous alternative, or an unrelated anomalous alternative, as in (26):

26. They wanted to make the hotel look more like a tropical resort.
So, along the driveway, they planted rows of **palms** / **pin**es / **tulips**.

Relative to unrelated anomalous words like "tulips", expected words like "palms" elicited an attenuated N400. However, anomalous words like "pin

plausible interpretation is that people were preactivating “palms” to such an extent that semantically related concepts were also preactivated. Note that the attenuated N400 to anomalous “pines” relative to equally anomalous “tulips” is not easy to account for if the N400 directly indexes high-level conceptual integration. But it is readily interpreted under a memory-based account of the N400; in fact, one can see this as another case of undesirable real-world knowledge leaking through (cf. section 3.4).

Stronger evidence that people predict upcoming words was obtained in an ERP study that probed for traces of lexical anticipation *before* the discourse-predictable word came along (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). Participants listened to (Dutch) mini-stories such as (27), which in a paper-and-pencil cloze test were predominantly completed with one particular critical noun (in this case, “painting”, the Dutch translation of which is a neuter-gender word). To test whether such discourse-based lexical prediction would also occur ‘on-line’ as part of real-time language comprehension, the EEG participants at this point first heard a gender-inflected adjective, whose syntactic gender either agreed with the anticipated noun, as in (28), or did not agree with this expected noun, as in (29).

27. The burglar had no trouble locating the secret family safe.

Of course, it was situated behind a...

28. ... **big**_{NEU} but rather unobtrusive painting_{NEU}

29. ... **big**_{COM} but rather unobtrusive bookcase_{COM}.

Relative to the gender-congruent prenominal adjective in (28), the gender-incongruent adjective in (29) elicited a small but reliable ERP effect right at the inflection. Because this prediction effect hinged on the idiosyncratic (hence memorized) syntactic gender of an expected but not yet presented noun, and because the effect collapsed when the prior context sentence was removed, it suggested that discourse-level information can indeed lead people to anticipate specific upcoming words as the text unfolds. In addition, the fact that such prediction could be probed via syntactic gender agreement suggested that the syntactic properties of those anticipated ‘ghost’ words can immediately begin to interact with locally unfolding syntactic constraints, such as the gender inflection on a prenominal adjective. Follow-up experiments have revealed that, at least when probed in this gender-sensitive paradigm, such discourse-based lexical predictions cannot be reduced to the effects of simple or convergent lexical priming (Otten & Van Berkum, in press; Otten, Nieuwland & Van Berkum, 2007). Also, comparable prediction has been observed in constraining single sentences (e.g., Wicha, Moreno & Kutas, 2004).

These findings suggest that people predict specific upcoming words as syntactic entities (‘lemmas’), i.e. a representation of the word in terms of its syntactic category and the associated syntactic features, (e.g., gender). However, consistent with the idea that opportunities for discourse-based prediction exist at many levels of analysis, there is also ERP evidence for the anticipation of upcoming referents (Nieuwland & Van Berkum, 2006a; Van Berkum et al., 2007) and, in a sentence paradigm, upcoming phonological word forms (DeLong, Urbach & Kutas, 2005). The mechanisms behind all this anticipation remain to be established. One currently debated idea is that comprehenders need to run a sophisticated forward model at various levels of analysis, as part of their basic comprehension strategy (see e.g., Pickering & Garrod, 2007). However, as argued elsewhere (Van Berkum, in press), much of it may actually come for free as a function of how human memory works (Gerrig & O'Brien, 2005).

6. Getting the speaker into the picture

In EEG research with spoken language materials, the need for a speaker is often seen as an unavoidable complication, the price to be paid for having naturally unfolding spoken input instead of word-by-word presentation on the screen. In line with this, speakers are usually chosen

on how well they produce text without drawing any attention to themselves (by, say, hesitations, or a highly salient accent). However, pragmatic analyses of language meaning (e.g., Carston, 2002; Clark, 1996; Grice, 1957; Kempson, 2001; Levinson, 1983; Sperber & Wilson, 1995; Wilson & Sperber, 2004) reveal that listeners can only really make sense of language if they take the speaker and his or her perspective into account. Thus, to understand the neurocognition of interpretation in full, we need to bring the speaker into the picture.

One immediate reminder of the speaker's existence is when he or she hesitates in delivering a message. In a refreshing ERP study (Corley, MacGregor & Donaldson, 2007), the impact of such hesitations on the listener have recently been examined with materials such as:

30. Everyone's got bad habits, and mine is biting my NAILS/TONGUE.
31. Everyone's got bad habits, and mine is biting my... er... NAILS/TONGUE.

In a fluently delivered utterance like (30), highly expected words elicited a much reduced N400 relative to their less expected counterparts, in line with other observations of predictability effects (e.g., Hagoort & Brown, 1994; Federmeier & Kutas, 1999; Otten & Van Berkum, 2007). But after a disfluency such as in (31), predictable words were much less effective in attenuating the N400. On an N400 integration account, this suggests that hesitation can interfere with integrating the meaning of a subsequent word. On a memory retrieval account, it suggests that implicit text-based anticipation is sensitive to speaker performance, possibly mediated by implicit knowledge about when (and when not) to expect fluent delivery. Either way, a marked delivery matters. More generally, the study is a good example of how to address at least some issues relevant to conversation without actually having one.

Another ERP study (Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008), recently examined when and how listeners make use of voice-based inferences about who the speaker is. As revealed by the fact that "I need to see my gynecologist" is really odd for a man to say (unless he's Arnold Schwarzenegger in this absurd Hollywood product), inferences about the speaker are directly relevant to interpretation. There is debate, however, on how and precisely when such knowledge is used (cf. Lattner & Friederici, 2003). In the EEG experiment, people heard utterances whose content did or did not match probabilistic inferences supported by the speaker's voice, as in (32) delivered with a male or female voice, (33) with the voice of a young kid or an adult, and (34) delivered with a stereotypically 'common' or 'upper-class' accent.

32. I always rent movies with lots of **violence** in it.
33. On Saturday I spent the whole afternoon playing **marbles** on the street.
34. I have a big **tattoo** on my back

Speaker-inconsistent critical words elicited a small but reliable N400 effect, beginning at 200-300 milliseconds from acoustic word onset. This reveals that listeners rapidly classify speakers on the basis of their voices, and anchor the utterance to the speaker immediately, somehow using their knowledge of the speaker's probable sex, age, and social stratum to evaluate the plausibility of what is being asserted. Again, the most parsimonious account for this is that, after for example having heard "I have a big" in a stereotypically 'upper-class' accent (instead of a more 'common' accent), listeners find it more difficult to retrieve the meaning of "tattoo" from long-term memory.

Finally, recent work on concurrent speech-gesture comprehension (Özyürek, Willems, Kita & Hagoort, 2007) has shown that a mismatch between a spoken word (e.g., "He slips on the roof and **rolls** down") and a simultaneous iconic co-speech gesture (e.g., finger-gestured *walking*) also elicits an increased N400. Because such gestures probably have little to no context-independent meaning stored in long-term memory, this result is sometimes taken as strong evidence for an integration account of the N400. However, because mismatching gestures can hinder the retrieval of a relevant meaning for concurrently presented words, this result is also

easily accommodated in a memory retrieval account of that component. Either way, the findings testify to the speed with which extra-sentential information picked up from the speaker can affect the processes involved in making sense of an utterance.

7. How about conversation?

Whereas the electrophysiology of text comprehension is now gradually picking up speed, the electrophysiology of conversation is as yet nonexistent. To some extent, this simply echoes the fact that most psycholinguistic theories and experiments are, implicitly or explicitly, about monologue. Furthermore, the various reasons that have caused many EEG experimenters to think twice before entering the domain of text comprehension will surely also have caused them to stay away from conversation. In addition, real two-person conversation brings with it an entirely new set of methodological problems. One of them is that in contrast to text, conversation cannot be fully pre-designed. As in eye tracking research on conversation (see Trueswell & Tanenhaus, 2005), probably the best way to make sure that EEG participants encounter the critical input at the right time is to have them interact with a confederate of the experimenter, in a collaborative setting that can be scripted in relatively natural ways. Unfortunately, fitting many lengthy conversational trials into an EEG experiment is not easy. Also, people tend to speak, gesture, and otherwise move a lot in conversation, which gives rise to large EEG artifacts. In all, we're currently facing the enormous challenge to come up with conversational settings that minimize movement, supply lots of critical trials, and are yet sufficiently natural to be worth our while. Whether a Virtual Reality setup can help attenuate some of these problems is currently being explored.

Some researchers have recently turned to other ways to get a handle on conversationally relevant phenomena. One, as we have seen, is to modify the classic monologue paradigm in ways that bring the speaker into the picture. Another promising avenue is to use an 'overhearing' paradigm, in which the participant listens to pre-designed conversational exchanges (e.g., Magne, Astésano, Lacheret-Dujour, Morel, Alter & Besson, 2005; Toepel, Pannekamp & Alter, 2007). In an ERP experiment on the role of prosody in discourse context (Magne et al., 2005), participants listened to relatively natural question-answer exchanges between male and female speakers, designed such that a contrastive accent in the answer did or did not match expectations raised by the question, as in (35) versus (36):

- | | |
|---|--|
| 35. Q: Did he give his fiancée a ring or a bracelet? | A: He gave a RING to his fiancée. |
| 36. Q: Did he give a ring to his fiancée or his sister? | A: He gave a RING to his fiancée. |

Surprisingly, whether a noun did or did not carry a contrastive accent by itself did not affect the ERPs, in spite of the fact that accent substantially altered the physical realization of the spoken noun (contra Li et al., 2008). The only thing that mattered was whether the prosody matched expectations raised by the question: relative to pragmatically congruous counterparts, the pragmatically incongruous NPs consistently elicited very rapid differential ERP effects. However, the results of this innovative experiment may have been compromised by the fact that at every trial, participants were asked to decide whether the intonation of the answer was coherent or not in relation to the question. Because secondary tasks recruit their own brain activity, and because the instruction to only respond after the linguistic input does not guarantee that the task-related *evaluation* will also be delayed that long, an additional task like this carries the risk that language-related ERP effects become confounded with task-related ERP effects (see also Van Berkum, 2004).

The possibility of task-related confounds should not in general deter people from studying language comprehension (and production) in a task environment. Most EEG research to date falls

squarely within the *language-as-product* tradition (Clark, 1996; 1997), in which readers and listeners are faced with bits of text they essentially don't care about (but see Van Berkum, Holleman, Nieuwland, Otten & Murre, submitted, for an exception). Of course, to the extent that the processes under study do their job in the same way regardless of what the linguistic input is for – as is commonly and perhaps reasonably assumed for syntactic parsing – this isn't too much of a problem. However, with *interpretation*, things are not so clear. Again, some processes may well turn out to do their job regardless of whether the input serves a bigger purpose. Also, if the stimuli themselves are sufficiently engaging, the experiment will be relevant to what language users do when they pick up a novel, or, say, a tabloid in a commuter train. Nevertheless, it seems critical to also record EEG in experiments where language truly matters to *getting things done* (*language-as-action*; Clark, 1996), such as building a Lego construction or baking a cake (see Trueswell & Tanenhaus, 2005, for examples). Such paradigms can help examine the generality of results obtained with snippets of text, as well as address specific issues associated with goal-oriented conversation. Moreover, and particularly relevant here, they would have the enormous advantage of getting *naturally motivated* language comprehension for free.¹

8. So, what kind of system is this?

Stepping back from all the details, we can discern several general operating principles at work in discourse comprehension.

(1) *Rapidly contextualized processing*. First of all, the words of an unfolding spoken or written sentence are very rapidly related to what the wider discourse is about, as soon as they come in. Because words encode information at various levels of linguistic analysis (phonology, syntax, meaning, and reference, Jackendoff, 2002), such discourse-sensitive incremental processing can show up in various ways, some of which have been summarized in figure 1.

*** insert figure 1 about here ***

Illustrated here is, from left to right, the harvest of a *single* EEG experiment, designed to obtain three different discourse-relevant effects in the same participants: a rapid ERP effect of discourse-level referential ambiguity, a downstream P600 garden-path effect associated with the discourse-dependent modulation of how the parser analyzes a subsequent syntactic ambiguity, and the N400 effect elicited by a discourse-dependent semantic anomaly. The impact of discourse-level factors does not only show up at the first relevant word in each of the three cases, but also shows up extremely rapidly *at* that particular word. Furthermore, studies that contrasted discourse-level factors with convergent lexical priming or other more superficial mechanisms confirm that what is at work here is more than just a lexical priming or repetition confound. Unfolding words make very rapid contact with a deep representation of the wider discourse, at the level of *what's being talked about*, as well as *how*, and *by whom*.

(2) *Discourse-based 'presonance' at multiple levels*. The evidence discussed in section 5 suggests that readers and listeners go beyond incremental processing, and can use discourse-level information to make predictions about what information might be needed soon, as well as about upcoming communication. Such predictions need not involve costly conscious deliberation. Instead, many of them may virtually come for free, as a result of how information in our content-addressable long-term memory resonates with low- and high-level aspects of currently processed information. This idea of predictive resonance has nicely been captured in the term 'presonance', coined by Zwaan and Kashak (2005). What the EEG records suggest is that such presonance can be driven by simple primes in prior discourse as well as by much more complex message-level

¹ Note that what matters is not so much whether the collaborative task at hand is a natural one, but whether the relevant utterances are a natural part of whatever is needed to get the job done.

representations of that prior discourse. Furthermore, the targets of anticipation can involve such things as the semantic, syntactic, and phonological features of specific upcoming words, whom or what will be referred to soon, and upcoming syntactic structure. In all, it seems that people can extrapolate unfolding discourse at a variety of levels in a relatively cost-free way. Of course, that doesn't make resonance (or more sophisticated forward modelling) a complete model of discourse comprehension – much of the work there *must* be about building precisely structured and dynamically evolving representations of what is said and meant, as well as of the situation being talked about. However, such complex representations do also drive simple long-term memory retrieval, and thereby prepare the system to a greater or lesser extent for the next bit of linguistic information coming up.

(3) *Simultaneous multiple constraint satisfaction.* In line with evidence from other methodologies, the ERP data are not in line with a classic feedforward model of comprehension in which different levels of language analysis – phonology/orthography, syntax, compositional sentence semantics, reference and other discourse-level pragmatics – only constrain each other in an ordered left-to-right fashion. Syntax does constrain interpretation, and specific combinations of word meanings do help define a referential target. But at the same time, the evidence that discourse-level constraints can briefly override locally unambiguous syntactic information, and that a suitable cartoon-like wider discourse can completely override a severe real-world anomaly, attests to the power that discourse-level representations can have over more local 'lower-level' constraints. In all, the picture painted here is that of a system that is trying to simultaneously satisfy constraints at multiple levels of representation (Trueswell & Tanenhaus, 2005; Jackendoff, 2002; 2007), without giving strict priority to any one level.

(4) *Discourse is not so special.* Nothing in the ERP evidence reviewed here supports the idea that the processes involved in discourse-dependent conceptual interpretation are any different from those involved in 'sentence-internal' conceptual interpretation. For example, discourse- and sentence-dependent N400 effects are identical, both for anomalies, for more subtle manipulations of contextual fit, and, as a special case of absence, for 'semantic illusions'. The ERP effect of referential ambiguity reported in section 4 can be elicited in texts and single sentences alike. And people predict upcoming words in discourse and single sentences alike. The sentence domain is of course highly relevant for syntax, as well as for the compositional semantics involved in determining *sentence* meaning. However, for much of the interpretation beyond that, the sentence boundary is irrelevant. Interpretation is fundamentally contextualized, and it doesn't seem to matter to the brain whether this context is defined by, say, the first few words of a single unfolding sentence, a 500-page Russian novel, or who the speaker is.

9. Where to go next?

9.1 From discourse-unit to discourse-level processing

To get this review started, I provisionally defined discourse comprehension as the processes involved in digesting *units of language* bigger than a single sentence, i.e., a fragment of text or conversation. Within neurocognitive research on language comprehension, this reflects current practice in thinking about discourse. But 'discourse' has an important second meaning, that of a *level of mental representation and processing*. This is evident in psycholinguistic research on situation models (e.g., Garnham, 1999; Johnson-Laird, 1983; Kinsch, 1998; Zwaan, 2004; Zwaan & Radvansky, 1998; see Clark, this volume), and in linguistic research on discourse-level representations (e.g., DRT; Kamp & Reyle, 1993; Geurts & Beaver, 2007). Discourse-level representation and processing is in principle orthogonal to whether the exchange consists of a single sentence or more, because even a single isolated sentence is analyzed at the discourse level (hence explaining the similarity of sentence- and discourse-dependent ERP effects). It just

happens to be the case that prior discourse context is often a convenient way to modulate discourse-level processing.

This has three important implications. First, we need not *always* go beyond the single sentence to study interpretive phenomena associated with discourse and conversation. Research on, say, discourse connectives or turn-taking obviously requires a bigger unit. But research on establishing reference or updating situation models can in principle be conducted with texts, conversations, and single sentences alike – good news for the cognitive neuroscientist. An important second implication, and the other side of the coin, is that “the neural substrate of discourse and conversation” cannot simply be isolated by comparing the processing of connected text to that of single sentences (a relatively standard working assumption in fMRI research on text comprehension). The comparison may well reveal various systems that are being taxed *more* by a long piece of discourse, as well as some of the processes that by necessity require text (e.g., those sensitive to discourse connectives). But if much of the really interesting discourse-level stuff occurs in *every* sentence regardless of whether or not it is embedded in a larger text, the subtraction is going to partial out the very stuff that one is after. The third and related implication is that in developing new research on the electrophysiology of discourse and conversation, researchers may want to think twice about what they are really after. Studying the neurocognition of discourse and conversation by examining language comprehension embedded in a big and rich *unit* of language use looks immediately relevant (and for a reason discussed below, it is). At the same time, I think the deeper thing we should be after is the neurocognition of discourse-level processing, as formalized in pragmatics and some corners of semantics (cf. Van Berkum, in press). If a big unit helps, that’s fine. But in this particular enterprise, size is not the issue.

9.2 Other directions

One of the strengths of ERPs is that their multidimensional nature (polarity, distribution, shape, and timing) can inform us about the identity of the processes engaged. However, this is difficult to benefit from when, instead of having two crisp ERP effects (say, N400 and P600) with clearly separable descriptions and well-understood functional significance, our inferential benchmark in the literature is a large set of many rather similar and many very different ERP effects, with multiple competing functional accounts for each. Unfortunately, it looks like we are steadily moving towards the latter situation. This doesn’t mean that the electrophysiology of discourse and conversation is doomed. It does suggest that whenever one can design an ERP experiments such that the logic does not solely rely on the *nature* of the effect, this is the safer way to go (Van Berkum, 2004). It also points to the growing importance of linking specific ERP effects to their neural sources, by whatever available means (constrained source modelling, using evidence from hemifield studies, MEG, and fMRI, etc.)

Another way out of the increasing mess of ERP effects is to make a serious start with relating ERP research to detailed linguistic (e.g., DRT-style) and computational (Graesser, McNamara & Rus, this volume) models of discourse-level interpretation, and to derive further studies from that (see Baggio et al., 2008, for a good example). Neuroimagers are sometimes inclined to think that, now that we can ‘look under the hood of the car’, we can pin down the architecture of comprehension by simply running more EEG, fMRI and TMS experiments. I think this is a mistake. Language interpretation is complex business, and we can only seriously move forward if we ground our neural research in more precise theories that also take the relevant semantic, pragmatic, and behavioral findings into account.

Finally, we must take EEG, MEG, and other neuroimaging measures a little further into the arena of real language use (Clark, 1996). As mentioned before, not all discourse-level issues necessarily require a big chunk of language to be investigated. However, language is first and foremost an instrument by which people coordinate with each other (Clark, this volume). Thus, as with any system designed for a purpose, it is not unlikely that at least some of the features of our

language comprehension and production system will be customized to operate in this particular cognitive and social niche. Ignoring that niche in our neurocognitive research would be a capital mistake. In the words of Small and Nussbaum (2004, p. 301; see also Clark, 1997; Kutas, 2006), it is high time to “examine cortical activity during language behavior that most closely matches conditions of evolution: language use by people at a time and place, aiming to understand and to be understood, fulfilling a purpose.”

NOTES

Address correspondence to Jos van Berkum, Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands (www.josvanberkum.nl, jos.vanberkum@mpi.nl)

Thanks to Michael Spivey, Mante Nieuwland, Marte Otten and Petra van Alphen for comments, and my colleagues in the MPI Neurobiology of Language group for discussion. Writing of this chapter was in part supported by the Dutch Science Foundation (*Vidi* grant 016.008.021).

FIGURE CAPTION

Figure 1. From left to right: A sustained frontal negative shift or ‘Nref effect’ to a discourse-induced referential problem (in the 2-referent version of the story, “lecturer” is ambiguous), a P600 effect to a discourse-induced syntactic problem (the provisional relative-clause analysis temporarily pursued at “that” in the 2-referent context is subsequently ruled out by “there”), and an N400 effect to a discourse-induced semantic problem (“promote” does not fit the wider story context). The example item is shown here in several variants (1- and 2-referent contexts, coherent/anomalous ending), but any one participant saw just a single version. ERP waveforms are time-locked to the presentation of the written critical word (0 msec) and are shown for 1200 msec each. Negative voltage is up. All data were obtained in a single reading experiment.

REFERENCES

- Anderson, J.E., & Holcomb, P.J. (2005). An electrophysiological investigation of the effects of coreference on word repetition and synonymy. *Brain and Language*, *94*, 200-216.
- Anderson, J.R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Erlbaum.
- Baggio, G., van Lambalgen, M., & Hagoort, P. (in press). Computing and re-computing discourse models: an ERP study. *Journal of Memory and Language*.
- Barton, S. B. & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, *21*, 477-487.
- Bohan, J. (2008). *Depth of Processing and Semantic Anomalies*. Unpublished dissertation.
- Bornkessel, I., Schlesewsky, M., Friederici, A.D. (2003). Contextual information modulates initial processes of syntactic integration: the role of inter- versus intrasentential predictions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 871-882.
- Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, *98*, 2, 159-168.
- Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *NeuroReport*, *18*(17), 1851-1854.
- Burkhardt, P. & Roehm, D. (2007). Differential effects of saliency: An event-related brain potential study. *Neuroscience Letters*, *413*(2), 115-120.
- Brownell, H., & Friedman, O. (2001). *Discourse ability in patients with unilateral left and right hemisphere brain damage*. In R.S. Berndt (Ed.). *Language and aphasia* (Handbook of neuropsychology, 2nd edition, volume 3), pp. 189-204. Amsterdam: Elsevier.
- Camblin, C.C., Gordon, P.C., Swaab, T.Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*, *56*, 103-128.
- Camblin, C.C., Ledoux, K., Boudewijn, M., Gordon, P.C., & Swaab, T.Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVPs. *Brain Research*, *1146*, 172-184.
- Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
- Clark, H.H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H.H. (1997). Dogmas of understanding. *Discourse Processes*, *23*, 567-598.
- Clark, H.H. (this volume).
- Connolly, J.F. (this volume).
- Coulson, S. (this volume).
- Coulson, S. & Federmeier, K.D. (in press). Words in context: ERPs and the lexical/postlexical distinction. *Journal of Psycholinguistic Research*.
- Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*, 658-668.
- Delong K.A., Urbach T.P., Kutas M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117-21.

- Ditman, T. & Kuperberg, G.R. (2007). The time course of building discourse coherence in schizophrenia: an ERP investigation. *Psychophysiology*, *44*, 991-1001.
- Federmeier, K.D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*, 491-505.
- Federmeier, K.D. & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469-495.
- Ferguson, H.J., Sanford, A.J., Leuthold, H. (in press). Eye-movements and ERPs reveal the time-course of processing negation and remitting counterfactual worlds. *Brain Research*.
- Ferstl, E. C. (2007). *The functional neuroanatomy of text comprehension: What's the story so far?* In F. Schmalhofer & C. Perfetti (Eds.), *Higher Level Language Processes in the Brain: Inference and Comprehension Processes* (pp. 53-102). Mahwah, NJ, Lawrence Erlbaum.
- Ferstl, E. C., Neumann, J., Bogler, C. & von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, *29*, 581-593.
- Filik R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: Evidence from the N400. *Psychophysiology*, *45*, 554-558..
- Filik R., Sanford A. J. & Leuthold H. (2008). Processing pronouns without antecedents: Evidence from event-related brain potentials *Journal of Cognitive Neuroscience* *20*(7),1315-1326.
- Fischler, I., Bloom, P.A., Childers, D.G., Roucos, S., & Perry, N.W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*(4), 400-409.
- Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove, UK: Psychology Press.
- Gerrig, R.J., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes*, *26*, 67-86.
- Gerrig, R. J., & O'Brien, E. J. (2005). The scope of memory-based processing. *Discourse Processes*, *39*, 225-242.
- Geurts, B. (1999). *Presuppositions and pronouns*. Oxford: Elsevier.
- Geurts, B. (in press). Neurocognition of language: good, bad, and bogus. *Language and Cognitive Processes*.
- Geurts, B. and D. Beaver (2007). Discourse representation theory. *Stanford encyclopedia of philosophy*.
- Graesser, A.C., McNamara, D.S., & Rus, V. (this volume).
- Grice, P. (1957). Meaning, *The Philosophical Review*, *66*, 377-88.
- Hagoort, P., & Brown, C.M. (1994). *Brain responses to lexical ambiguity resolution and parsing*. In C. Clifton Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45-80). Hillsdale, NJ: Erlbaum.
- Hagoort, P., Brown, C.M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439-483.
- Hagoort, P., Hald, L., Bastiaansen, M. & Petersson, K. M. (2004) Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438-441.
- Hald, L.A, Steenbeek-Planting, E.G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research*, *1146*, 210-218.
- Hammer, A., Jansma, B.M., Lamers, M., & Münte, T.F. (2005). Pronominal reference in sentences about persons or things: An electrophysiological approach. *Journal of Cognitive Neuroscience*, *17*, 227-239.

- Hoeks, J.C.J., Stowe, L.A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 18, 59-73.
- Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press.
- Jackendoff, R. (2007). A Parallel Architecture perspective on language processing. *Brain Research*, 1146, 2-22.
- Johnson-Laird, P.N. (1983). *Mental Models*. Harvard University Press.
- Kaan, E., Dallas, A.C., & Barkley, C.M. (2007). Dealing with new discourse referents: An ERP study. *Brain Research*, 1146, 199-209.
- Kamp, H., & Reyle, U. (1993). *From discourse to logic*. Dordrecht: Kluwer.
- Kempson, R. (2001). *Pragmatics: Language and communication*. In M. Aronoff & J. Rees-Miller (Eds.), *Handbook of linguistics*. Malden, Massachusetts: Blackwell.
- Kintsch, W., (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kounios, J., & Holcomb, P.J., (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General*, 121, 459-479.
- Kuperberg, G.R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23-49.
- Kutas, M. (2006). One lesson learned: frame language processing - literal and figurative - as a human brain function. *Metaphor and Symbol*, 21(4), 285-325.
- Kutas, M. & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4, 463-470.
- Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-205.
- Kutas, M., & Hillyard, S.A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and 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., Van Petten, C., & Kluender, R. (2006). *Psycholinguistics electrified II: 1994-2005*. In M. Traxler & M.A. Gernsbacher (Eds.), *Handbook of Psycholinguistics, 2nd Edition* (pp 659-724). New York: Elsevier.
- Lattner, S. & Friederici, A.D. (2003). Talker's voice and gender stereotype in human auditory sentence processing – evidence from event-related brain potentials. *Neuroscience Letters*, 339, 191-194.
- Ledoux, K., Gordon, P.C., Camblin, C.C., & Swaab, T.Y. (2007). Coreference and lexical repetition: Neural mechanisms of discourse integration. *Memory & Cognition*, 35, 801-815.
- Levinson, S.C. (1983). *Pragmatics*. Cambridge University Press.
- Li, X., Hagoort, P., & Yang, Y. (2008). Event-related Potential Evidence on the Influence of Accentuation in Spoken Discourse Comprehension in Chinese. *Journal of Cognitive Neuroscience*, 20, 906-915.
- Magne, C., Astésano, C., Lacheret-Dujour, A., Morel, M., Alter, K. & Besson, M. (2005). On-line Processing of "Pop-Out" words in spoken french dialogues. *Journal of Cognitive Neuroscience*, 17, 740-756.
- Marr, R.A. (2004). The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia*, 42, 1414-1434.

- Müller, H.M., King, J.W., & Kutas, M. (1997) Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research*, 5, 193-203.
- Münte, T.F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, 395, 71-74.
- Myers, J.L. & O'Brien, E.J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, 26, 131-157.
- Nieuwland, M.S. & Kuperberg, G.R. (in press). When the truth isn't too hard to handle: An Event-Related Potential study on the Pragmatics of Negation. *Psychological Science*.
- Nieuwland, M.S., Otten, M. & Van Berkum, J.J.A. (2007). Who are you talking about? Tracking discourse-level referential processing with ERPs. *Journal of Cognitive Neuroscience*, 19, 228-236.
- Nieuwland, M.S., Petersson, K.M. & Van Berkum, J.J.A. (2007). On sense and reference: Examining the functional neuroanatomy of referential processing. *NeuroImage*, 37, 993-1004.
- Nieuwland, M.S. & Van Berkum, J.J.A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24, 691-701.
- Nieuwland, M.S. & Van Berkum, J.J.A. (2006a). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18, 1098-1111.
- Nieuwland, M.S. & Van Berkum, J.J.A. (2006b). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, 1118, 155-167.
- Nieuwland, M.S. & Van Berkum, J.J.A. (2008). The neurocognition of referential ambiguity in language comprehension. *Language and Linguistics Compass*, 2(4), 603-630
- Nieuwland, M.S. & Van Berkum, J.J.A. (in press). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language*.
- Osterhout, L., & Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.
- Osterhout, L., Kim, A., & Kuperberg, G. (this volume).
- Osterhout, L. & Mobley, L.A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, 34, 739-773.
- Otten, M., Nieuwland, M.S. & Van Berkum, J.J.A. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, 8(89).
- Otten, M. & Van Berkum, J.J.A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research*, 1153, 166-177.
- Otten, M. & Van Berkum, J.J.A. (in press). Discourse-based anticipation during language processing: Prediction or priming? *Discourse Processes*.
- Özyürek, A., Willems, R. M., Kita, S. & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, 4, 605-616.
- Pickering, M.J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11, 105-110.
- Pylkkänen, L. (2008). Mismatching Meanings in Brain and Behavior. *Language and Linguistics Compass* 2(4), 712-738.
- Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, 95, 285-408.

- Sanford, A.J. & Garrod, S.C. (2005). Memory-based approaches and beyond. *Discourse Processes*, 39, 205-234.
- Sanford, A.J.S., Sanford, A.J., Molle, J. & Emmott, C. (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes*, 42 (2), 109-130.
- Sanford, A.J. & Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences*, 6, 382-386.
- Sitnikova, T., Kuperberg, G.R., & Holcomb, P.J. (2003). Semantic integration in videos of real-world events: an electrophysiological investigation. *Psychophysiology*, 40, 160-164.
- Small, S. L., & Nusbaum, H. C. (2004). On the neurobiological investigation of language understanding in context. *Brain & Language*, 89, 300-311.
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition*. Oxford: Blackwell
- St. George, M., Mannes, S., & Hoffman, J.E. (1994). Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience*, 6(1), 70-83.
- St. George, M., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience*, 9(6), 776-787.
- Streb, J., Rösler, F., & Hennighausen, E. (1999). Event-related responses to pronoun and proper name anaphors in parallel and nonparallel discourse structures. *Brain and Language*, 70, 273-286.
- Streb, J., Rösler, F., & Hennighausen, E. (2004). Different anaphoric expressions are investigated by event-related brain potentials. *Journal of Psycholinguistic Research*, 33, 175-201.
- Swaab, T.Y., Camblin, C.C., & Gordon, P.C. (2004). Electrophysiological evidence of reversed lexical repetition effects in language processing. *Journal of Cognitive Neuroscience*, 16, 715-726.
- Toepel, U., Pannekamp, A. & Alter, K. (2007). Catching the news: Processing strategies in listening to dialogs as measured by ERPs. *Behavioral and Brain Functions*, 3(53).
- Trueswell, J., & Tanenhaus, M. (2005) (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-action traditions*. Cambridge, MA: MIT Press.
- Van Berkum, J.J.A. (2004). *Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things?* In M. Carreiras & C. Clifton Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 229-270). New York: Psychology Press.
- Van Berkum, J.J.A. (in press). *The neuropragmatics of 'simple' utterance comprehension: An ERP review*. To appear in U. Sauerland & K. Yatsushiro (Eds.), *Semantic and pragmatics: From experiment to theory*. Palgrave.
- Van Berkum, J.J.A., Brown, C.M., & Hagoort, P. (1999a). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, 41, 147-182.
- Van Berkum, J.J.A., Brown, C.M., & Hagoort, P. (1999b). When does gender constrain parsing? Evidence from ERPs. *Journal of Psycholinguistic Research*, 28(5), 555-571.
- Van Berkum, J.J.A., Brown, C.M., Hagoort, P., & Zwitterlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken-language comprehension. *Psychophysiology*, 40, 235-248.
- Van Berkum, J.J.A., Brown, C.M., Zwitterlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 31, 443-467.
- Van Berkum, J.J.A., Hagoort, P., & Brown, C.M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6), 657-671.

- Van Berkum, J.J.A., Holleman, B., Nieuwland, M.S., Otten, M., & Murre, J. (submitted). *Right or wrong? The brain's fast response to moral statements.*
- Van Berkum, J.J.A., Koornneef, A.W., Otten, M., & Nieuwland, M.S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146*, 158-171.
- Van Berkum, J.J.A., Van den Brink, D., Tesink, C.M.J.Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20(4)*, 580-591.
- Van Berkum, J.J.A., Zwitserlood, P., Brown, C.M., & Hagoort, P., (2003) When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17*, 701-718.
- 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., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience, 3*, 131-150.
- West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research, 13*, 363-375.
- Wicha, N. Y. 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.
- Wilson, D., & Sperber, D. (2004). *Relevance Theory*. In G. Ward & L. Horn (eds.) *Handbook of Pragmatics*. Oxford: Blackwell, 607-632.
- Yang, C-L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33(1)*, 55-89.
- Zwaan, R.A., & Kaschak, M.P. (in press). Language comprehension as a means of “re-situating” oneself. In P. Robbins & M. Aydede (Eds.). *The Cambridge Handbook of Situated Cognition*. Cambridge, UK: Cambridge University Press.
- Zwaan, R.A., & Radvansky, G.A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162-185.
- Zwaan, R.A., & Singer, M. (2003). *Text comprehension*. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 83-121). Mahwah, NJ: Lawrence Erlbaum.

Discourse-dependent referential, syntactic, and semantic ERP effects

In dismay, the faculty dean called the lecturer and the professor (*the two lecturers*) to his office. This was because the lecturer (*one of the lecturers*) had committed plagiarism, and the professor (*the other one*) had faked some of his research data. The dean told the **lecturer** that **there** was ample reason to **sack/promote** him.

