THE ELECTROPHYSIOLOGY OF DISCOURSE AND CONVERSATION

Jos J.A. van Berkum

Department of Psychology, University of Amsterdam, The Netherlands
F.C. Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands

To appear in: M. Spivey, M. Joanisse, & K. McRae (Eds), *The Cambridge Handbook of Psycholinguistics* Cambridge: Cambridge University Press.

Address for correspondence: Jos J.A. van Berkum
                             University of Amsterdam
                             Department of Psychology (PN)
                             Roetersstraat 15
                             1018 WB Amsterdam
                             The Netherlands
                             e-mail: berkum@psy.uva.nl
                             phone: +31-20-525-6921 (secr. -6840)
                             fax: +31-20-639-1656

1. Introduction

What's happening in the brains of two people having a lively 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 neuronal systems engaged in highly complex, high-speed interaction. 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 collaborative enterprise (Clark, 1996), in which interlocutors continuously keep track of each others state of mind (e.g., beliefs, intentions), 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, and some focus is clearly essential. Fortunately, others have already recently taken up the task to review the implications of patient research (Brownell & Friedman, 2001; Marr, 2004), as well as those of fMRI and PET (Ferstl, in press). The aim of the present chapter is to review the EEG research on discourse and conversation, and to see what electrophysiology tells us about the systems involved. EEG studies conducted on discourse-level language use have thus far practically all focused on the *comprehension of multi-sentence text*, and as such provide a convenient operational definition of the domain. However, an important theme developed in this review is that when it comes to 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?

For most of us, 'neuroimaging' has come to stand for functional magnetic resonance imaging (fMRI) or positron emission tomography (PET). To be sure, the images produced by fMRI and PET are aesthetically appealing and very informative. Furthermore, because they are 3-dimensional, they strongly resonate with our desire to look into our working brains. However, in the end, these images are images of blood, and although neurons need blood to keep working, the real work for which the brain has become famous – information processing or computation – is carried by electricity (and, at synapses, electro-chemical transduction). As it happens, this is precisely what the EEG taps in to.

The EEG or electro-encephalogram is a record of tiny voltage fluctuations over time, picked up via electrodes located at the scalp. Roughly speaking, these scalp-recorded 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 in the brain, the EEG is usually a bit of a mess. However, researchers have discovered various ways to extract useful information from this signal. In the most common method, the EEG is recorded to *repetitions* of some critical event *A* (say, a word), and the associated bits of EEG are subsequently averaged to create the event-related brain potential or ERP elicited by that event. If the latter systematically differs from the ERP to some other event *B* (say, a beep), the differential ERP effect is evidence that, at that very moment in time, the brain is sensitive to as least some of the differences between *A* and *B*.

Now, to a psycholinguist, all of this probably doesn't sound too exciting so far. However, in the late 1970's, ERP 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 deflection in the ERP 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, as in "He spread the warm bread with BUTTER", 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 less easy to integrate into the sentence-semantic context (see Kutas & Van Petten, 1994, for a review).

The significance of this phenomenon was enhanced by the end of the 80's, when syntactically anomalous or unexpected words were found to elicit a very different ERP effect, the so-called P600/SPS effect (Osterhout & Holcomb, 1992; Hagoort, Brown, & Groothusen, 1993). The discovery of two very different ERP 'signatures' not only raised interesting theoretical questions about the architecture of the sentence comprehension system and the types of representations it computed, but also implied that ERPs could be used as a tool 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 several other language-relevant ERP phenomena have been discovered since (see Osterhout, this volume, for review). With a small repertoire of ERP effects in hand, sentence comprehension researchers are now successfully examining the processes involved.

So far, 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 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. Because *at least* 30-40 trials are needed per condition to obtain a relatively clean ERP, factorial sentence-level EEG experiments require the presentation of many lengthy trials, as well as sometimes months of work to create the materials. Another problem is that within each of these lengthy trials, people are not supposed to move their eyes, head, or body. With a longer fragment of text or conversation in each trial, all this is only going to get worse.

Fortunately, it turns out that, in spite of such constraints, it is not impossible to use EEG to track discourse-level comprehension (Van Berkum, 2004). Moreover, because EEG allows one to selectively keep track of various sub-processes with high temporal resolution, without the need to impose an additional task, the additional work sometimes really pays off. In the below, I review what can be learned from the handful of discourse-level EEG studies conducted over the past decade. I begin by examining studies that used the N400 to address the construction of discourse-level meaning (section 3), then turn to experiments 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 we face in combining electrode caps with conversation (section 7). Finally, I pull out some major implications of the research reviewed (section 8), and return to the issue of what discourse-level processing really is (section 9).

3. Making sense of discourse

Although prose passages had in fact already featured in one of the early landmark ERP experiments on how people make sense of language (Kutas & Hillyard, 1983), 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 its individual sentences were interconnected and related to a single topic, it was rather difficult to find out what that topic actually was if the story did not have a title (in this case, *Procedure for washing clothes*). In the ERP experiment, the content words in stories without a title turned out to elicit larger N400 deflections than the same words in stories with a title. This clearly indicated that the N400 is not only sensitive to local lexical or sentential context, but also to global discourse-level context (see Van Petten, 1995, for indirect evidence pointing the same way).

Building on this initial exploration, Van Berkum and colleagues (Van Berkum, Hagoort & Brown, 1999c; Van Berkum, Zwitserlood, Brown & Hagoort, 2003b) 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, as in:

2. When he arrived at the animal testing facility, the lab assistant was stopped dead by a boy and a girl. The girl had chained herself to the fence, and the boy was standing right next to her. The assistant warned the girl that she would be arrested / nauseous soon.

Whereas St. George et al. had ensured attention by informing their subjects about a later recall task, participants in the Van Berkum et al. experiments were simply asked to read or listen for comprehension, without any additional demands. Nevertheless, words that were not supported by the wider discourse (e.g., "nauseous") elicited a much larger N400 than discourse-supported words (e.g., "arrested"), 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 words as well as clause-medial words, these findings suggested that *every* incoming word is immediately related to the wider discourse. Furthermore, with spoken words (Van Berkum et al., 2003b), the effect of discourse-level fit emerged as early as 150 ms after acoustic word onset, i.e., only some *2-3 phonemes into the word*. This suggested that spoken words are actually related to the wider discourse extremely rapidly, well before they have been fully pronounced, and possibly even before they have become acoustically unique. Importantly, a post-hoc split of the data revealed that the early impact of discourse-level information did *not* depend on whether or not subjects had been expecting the discourse-coherent word (see also Camblin, 2005). That is, even in essentially non-predictable, open-ended situations, an unfolding spoken word and the meaning(s) associated with it is mapped onto discourse-level representations this quickly. 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. This indicated that discourse- and sentence-dependent semantic constraints are brought to bear on comprehension as part of the same unified interpretation process (an issue to which we return below).

The relevance of identical sentence- and discourse-dependent anomaly effects would of course be somewhat limited if the commonality simply reflected some common error detection process, activated by two otherwise very different comprehension processes. However, it has long been know that the word-elicited N400 is no simple anomaly detector, but a reliable index of the ease with which lexical meaning is integrated into the wider sentential context (e.g., Kutas & Van Petten, 1994; Coulson & Federmeier, in press). In line with this, Otten and Van Berkum

(2005b) recently showed that relative to highly expected words in discourse (such as "sword" in example 3), words that are merely somewhat less expected (such as "lance") also elicit an N400 effect. In all, evidence from the N400 consistently indicates that words are related to the wider discourse very rapidly, and in a way that is no different from how they are related to local sentence-level context. This accords well with models of language comprehension that do not make a distinction between the computation of sentence- and discourse-level meaning.

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

Two recent observations have corroborated this perspective. First, note that although the various discourse-dependent N400 effects reported so far reflect something about the ease with which an incoming word is related to 'the wider discourse', it is not immediately clear from those studies what the critical context really is. Van Berkum et al. (1999c; 2003b) assumed that some kind of message-level representation (e.g. a situation model) was at work. However, in some of the stories in the Van Berkum et al. experiments, the context also contained words that might prime the discourse-supported word at a more superficial level. This can be illustrated with example (3), in which "sword" is not only strongly supported by the specific message conveyed by the story so far, but may also receive support from words like "dragon", "knight", and "threatened", via some form of convergent lexical priming, or – somewhat more refined – via the scenario suggested by these words. To examine these possibilities, Otten and Van Berkum (2005b) modified their highly constraining stories like in (3) such that, although the lexical support for a word like "sword" was still there, the message-level representation no longer strongly supported either critical word:

4. The benevolent sorcerer saw that the dragon threatened the brave knight.
   He quickly reached for a… (sword/lance)

In stories like (4), the large N400 effect obtained with words like "lance" as compared to "sword" in (3) largely collapsed, indicating that the latter had to a large extent critically hinged on message-level fit. However, words like "lance" did still elicit a small left-dominant N400-like effect relative to "sword". This remaining effect suggests that the mere 'bag of words' in prior discourse can also provide more or less support for the current one, regardless of whether these words are carefully arranged to convey a strongly constraining message, as in (3), or not, as in (4). Whether this more shallow form of contextual support hinges on 'simple' intra-lexical associations like "knight"- "sword", or on the scenario suggested by words like "dragon", "knight", and "threatened", remains to be explored (and in the end, one may not be so different from the other). What the results clearly indicate, though, is that prior discourse can provide support for specific words in two very different ways (see Camblin, 2005, for a related result).

In a second critical extension of the initial findings, Nieuwland and Van Berkum (2005b) have recently examined whether the wider discourse can also make *locally elicited N400 effects go away*. In the first experiment, subjects listened to stories like (5) in which two human beings were engaged in social interaction. In the critical condition, the noun denoting one of the persons (e.g., "sailor") was systematically replaced by one denoting an inanimate object (e.g., "yacht").

5. Once upon a time a psychotherapist was consulted by a sailor (yacht) with emotional problems. The sailor (yacht) confided her that everything in life had gone wrong and started crying. The psychotherapist consoled the sailor (yacht) by stating that everybody experiences these kinds of trouble every now and then. But the sailor (yacht) doubted whether to continue outlining his problems to her. The psychotherapist advised the sailor (yacht) to be honest not only with her, but especially with himself. At that moment the sailor (yacht) cried out that he was absolutely terrified of water.

We know that in the absence of a wider discourse, animacy violations such as in "a psychotherapist was consulted by a yacht" elicit a large N400 effect at the inanimate noun. However, we also expected a narrative in which particular inanimates come alive to gradually and implicitly define a cartoon-like 'genre', i.e. an interpretive context in which it is not uncommon for inanimate objects to have animate properties. Based on our earlier evidence for the equivalence of sentence- and discourse-level context, we predicted that this emerging 'pragmatic' context, once in place, should be able to overrule a local 'semantic' anomaly, and make the N400 effect go away. And this is indeed what we found. Whereas story-initial animacy violations elicited a sizeable N400 effect, later such violations (in the 3rd and 5th sentence) no longer did so. A follow-up experiment (Nieuwland & Van Berkum, 2005b, Experiment 2) showed that this was not due to a lexical repetition or story position confound, and revealed that a sufficiently strong discourse context can even make listeners prefer the animacy-violating expression "the peanut was in love" over the locally correct and canonical expression "the peanut was salted". In all, these results provide strong evidence that constraints from wider discourse are brought to bear on comprehension in the same unified interpretation process as 'classic' local semantic features such as animacy.

The studies discussed so far have all used ERPs to examine when and how discourse-level meaning is brought to bear on language comprehension. Obviously, however, what's involved in making sense of discourse is more than just a fight between local and global semantic constraints. To work out how the two sentences in "The deadline for my chapter was 10 days ago. If I don't finish it soon, the editor will start bugging me." hang together, for example, people need to make specific 'bridging inferences', so that they come to understand that the editor referred to in the second sentence is not just anybody, but presumably responsible for the book that is supposed to contain the chapter alluded to in the first. How such inferences are made is still under intense debate (Zwaan & Singer, 2003). Another hotly debated issue is to what extent people routinely make so-called elaborative inferences, i.e., ideas that are strongly implied by the text but that are not strictly required to make sense of it (e.g., that, if the chapter is still not done within a month, worse things might happen).

The word-elicited N400 has been shown to be highly sensitive to the level of contextual support from prior text, even in situations where people are merely asked to read or listen for comprehension (Van Berkum et al., 1999c, 2003b), and it is therefore a very attractive tool with which to examine the mechanisms underlying various types of text-based inferences. Again, St. George, Mannes, and Hoffman (1997) were the first to do so, in an experiment that compared low- and high-span readers on the degree to which they made bridging and elaborative inferences in 4-sentence texts. Unfortunately, with only a very limited number of texts per condition (15), the resulting ERP waveforms 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, St. George et al. (1997) resorted to comparing the average ERPs to *all* words in the critical sentence, a procedure that must inevitably have attenuated critical differences between conditions.

In a very recent ERP study on inferences, (Kuperberg, Caplan, Eddy, Cotton & Holcomb, 2005) avoided both of these problems. Participants read sentences like (6)

6. The next morning John had many bruises.

in the context of an explicitly supportive discourse context (7), an inferentially supportive discourse context (8) or a non-supportive discourse context (9):

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

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. In fact, the latter context was as effective in attenuating the N400 as an explicitly supportive one. This reveals that readers can rapidly connect the sentences of an unfolding discourse via additional inferences (but see Yang, Perfetti and Schmalhofer, 2005, for evidence that this need not always be the case)

      EEG researchers have also begun to study the comprehension of various 'figurative' uses of language, such as metaphors and jokes. For most people, jokes like "I still miss my ex-wife, but I'm improving my aim" require considerable inferencing, and so do metaphoric expressions like "Power is a strong intoxicant" (both examples from Seana Coulson). However, this does not necessarily set the processing of such passages apart from the processing of non-figurative ones. Recent analyses of the semantics-pragmatics distinction (e.g., Sperber & Wilson, 1995; Clark, 1996; Kempson, 2001) strongly suggest that inferencing is at the heart of *all* non-trivial language comprehension, an issue to which we will return later. Compatible with this perspective, ERP studies on figurative language (see Coulson, 2004 for review) suggest that the processing difficulty associated with such language reflects the complexity of such inferences, rather than a radically different mode of comprehension.

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 meaning. These examples reveal that establishing reference in discourse is a non-trivial job. To examine the nature and time course of such discourse-level referential analysis, Van Berkum, Brown and Hagoort (1999a) presented their EEG subjects with short written passages that contained a referentially unambiguous or ambiguous expression in the last sentence, such as "the friend" in (10) and (11) respectively.

10. Just as the elderly hippie had lit up a joint, he got a visit from a friend and a nephew. Even though his friend had had quite a few drinks already, and the nephew had just smoked quite a lot of pot already, they insisted on smoking along. The hippie warned the friend that there would be some problems soon.

11. Just as the elderly hippie had lit up a joint, he got a visit from two friends. Even though one of his friends had had quite a few drinks already, and the other one had just smoked quite a lot of pot already, they insisted on smoking along. The hippie warned the friend that there would be some problems soon.

      Because a referentially ambiguous noun may well be more difficult to integrate into an unfolding representation of discourse-level meaning, one might expect it to elicit a centro-parietally distributed N400 effect (compared to a referentially unambiguous control). However, referentially ambiguous nouns elicited a frontally dominant sustained negative shift instead, emerging at about 300 ms after noun onset. The early emergence of this effect indicated that people can very rapidly work out whether a singular definite NP is referentially ambiguous in discourse or not. Moreover, the fact that we did not obtain an N400 effect here revealed that the processing implications associated with difficulties in making sense and establishing reference are at least partially distinct. Comparable findings have been obtained with referentially ambiguous nouns in *spoken* discourse (Van Berkum, Brown, Hagoort, & Zwitserlood, 2003a), as well as with referentially ambiguous pronouns in sentences like "David shot at John as he jumped over the fence" (Van Berkum, Zwitserlood, Bastiaansen, Brown, & Hagoort, 2004).

Although the early onset of the discourse-induced frontal negativity show that readers and listeners try to establish reference in discourse very rapidly, the effect may arise at various different levels of the comprehension system By using critical stories in which, say, one of the two mentioned friends had just left the scene so that only one eligible discourse entity remained, Nieuwland, Otten, and Van Berkum (2005) have recently shown that the referentially induced frontal ERP effect reflects deep, situation-model ambiguity, rather than the more superficial ambiguity associated with the fact that two different friends had been mentioned in the discourse. Thus, reference is established very rapidly at the deepest level, with nouns contacting situation-model discourse entities within only a few hundred milliseconds. Furthermore, the Nieuwland et al. findings suggest that referential ambiguity at the level that is most relevant for comprehension, the situation model, can be tracked in ERPs relatively selectively.

Anderson and Holcomb (2005) recently examined the processing of sentence-initial definite and indefinite noun phrases (e.g., "The/A stone…") in discourse contexts that provided a suitable candidate for definite reference ("Tommy threw a stone/rock towards the pond"). Literally repeated critical nouns elicited a smaller N400 than nouns that were synonyms, and both in turn elicited a smaller N400 than newly introduced nouns. Furthermore, these effects were not modulated by whether the article suggested a previously introduced ("The…") or new ("A…") discourse entity. These findings suggest that the N400 repetition and semantic priming effects observed in word lists and single sentences also occur across a sentence boundary, and do so independently of the referential status of the critical NP. However, why the latter did not also have an effect on the N400 remains to be established. There is good evidence, for example, that in terms of N400 amplitude, the discourse-dependent referential infelicity of repeating a proper name can totally outweigh the benefits of lexical repetition (Swaab, Camblin & Gordon, 2004).

Evidence from ERPs also suggests that discourse-referential factors can rapidly influence syntactic parsing. In the abovementioned Van Berkum et al. (1999a) experiment, for example, discourse-induced referential ambiguity not only elicited a frontal negativity at the ambiguous noun, but also had an immediate impact on how the parser analysed a subsequent local *syntactic* ambiguity. For instance, if "the friend" in "The hippie warned the friend that…" was referentially ambiguous, the parser was more inclined to take the syntactically ambiguous word "that" as the beginning of a referentially restrictive *relative clause* ("…that had had quite a few drinks") than as the beginning of a complement clause ("…that there would be some problems soon"). In fact, discourse-dependent referential ambiguity could even briefly lure the parser into pursuing an analysis that was formally blocked by a local syntactic gender agreement constraint (see Van Berkum, Brown, & Hagoort 1999b, for detailed discussion). Using two-sentence question-answer sequences, Bornkessel, Schlesewsky and Friederici (2003) have recently also obtained evidence that discourse-referential factors can sometimes outweigh locally unambiguous syntactic cues.

5. Anticipating upcoming discourse

In the field of sentence comprehension research, behavioral and ERP work had already shown that the processing of a single isolated sentence is to a large extent incremental, with readers and listeners trying to relate every incoming word to a partial syntactic and semantic analysis of the sentence as it unfolds. The ERP findings reviewed in the preceding two sections reveal that such incremental processing actually holds 'all the way up', in that readers and listeners also immediately relate the syntactic, semantic and referential implications of incoming words to their knowledge of the wider discourse.

As with any amazingly fast brain response, it makes sense to ask to what extent the speed with which incoming words are taken 'all the way up' depends 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, on the basis of information supplied by any other level.

So far, only a few EEG studies have directly addressed discourse-based anticipation, with a focus on the prediction of upcoming words. Some of the first N400 experiments on semantic 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. In line with this, words that are highly predictable from the wider discourse context have been found to elicit a smaller N400 than less expected words (Otten & Van Berkum, 2005b). However, although such findings are compatible with the hypothesis that readers and listeners anticipate discourse-predictable words as the discourse unfolds in real time, they are not definitive evidence for the latter. The reason is that the processing benefits indexed by a reduced N400 may also reflect the ease with which unfolding discourse-predictable words are integrated into the wider context *as part of the processes that are driven by the unfolding word itself.*

To avoid the ambiguity in interpreting 'predictability effects' on the expected word itself, Van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005) probed for traces of lexical anticipation *before* the discourse-predictable word would come along. In the ERP experiment, participants listened to (Dutch) mini-stories such as

> 12. The burglar had no trouble locating the secret family safe. Of course, it was situated behind a…

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 would at this point first hear a gender-inflected adjective whose syntactic gender either agreed with the anticipated noun, as in (13)

> 13. The burglar had no trouble locating the secret family safe. Of course, it was situated behind a **big**$_{NEU}$ but rather unobtrusive painting$_{NEU}$

or did not agree with this expected noun, as in (14)

> 14. The burglar had no trouble locating the secret family safe. Of course, it was situated behind a ***big***$_{COM}$ but rather unobtrusive bookcase$_{COM}$.

Relative to the gender-congruent prenominal adjective in (13), the gender-incongruent adjective in (14) 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, it suggested that discourse-level information can indeed lead people to anticipate specific upcoming words 'on-line', as a local sentence 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. In follow-up research, Otten and Van Berkum (2005a) examined whether these predictions were being driven by a true message-level representation of the discourse (as had been assumed by Van Berkum et al., 2005a), or whether they instead involved some form of convergent or scenario-based lexical

priming. The ERP results actually suggest that *both* sources of constraint allow people to anticipate specific upcoming words. We briefly return to this issue in section 8.

The ERP evidence for discourse-based anticipation reviewed so far suggests 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, 2005b; Van Berkum, Otten, Jansen, Koornneef, De Boer, Huizinga, & Junge, 2005b) and, in a sentence paradigm, upcoming phonological word forms (DeLong, Urbach & Kutas, 2005), as well as evidence compatible with the discourse-based anticipation of the semantic features of upcoming words (Federmeier & Kutas, 1999a, 1999b).

6. Getting the speaker into the picture

In EEG research with spoken language input, the necessity of a speaker is often seen as an unavoidable complication of one's experiment, the price to be paid for having naturally unfolding input instead of the unnatural serial visual presentation of words in the center of a screen. In line with this perspective, speakers are usually chosen on how well they can produce sentences or stories without drawing attention to themselves (by, say, hesitations, or a highly salient accent). However, pragmatic analyses of language meaning (e.g., Clark, 1996; Kempson, 2001; Sperber & Wilson, 1995; Wilson & Sperber, 2004) have compellingly shown that listeners can only really make sense of language if they take the speaker and his or her perspective into account. This suggests that to address important issues in comprehension, we need to bring the speaker into the picture, rather than tucking him or her completely out of sight.

One immediate reminder of the speaker's existence and state of mind is when he or she hesitates in delivering a message. In a refreshing ERP study, MacGregor, Corley, and Donaldson (2005) recently examined the impact of such hesitations on the listener. In the experiment, ERPs elicited by highly and less expected words in standard, fluently delivered utterances such as in (15) were compared to ERPs elicited by the same words after a brief dysfluency (16).

15. Everyone's got bad habits, and mine is biting my NAILS/TONGUE.
16. Everyone's got bad habits, and mine is biting my… er… NAILS/TONGUE.

In a fluently delivered utterance, highly expected words elicited a much reduced N400 relative to their less expected counterparts, in line with prior demonstrations of sentence- and discourse-dependent predictability effects. After a disfluency, however, the size of this N400 effect was substantially reduced. Because MacGregor et al. assumed that the standard predictability effect reflects true lexical prediction, they concluded that in the face of disfluency, listeners are less likely to predict what speakers are about to say. Such an account makes sense, but because the critical assumption need not be correct (see section 5), the evidence is as yet not compelling. Still, something interesting is clearly going on here. More generally, the study is a good example of how one might address at least some issues relevant to conversation without actually having one.

Van Berkum, Van den Brink, Tesink, Kos, Müller, & Hagoort (2005c) recently also used EEG to examine the role of speaker-related information in language comprehension. The goal of the experiment, which involved 21 different speakers, was to examine when and how listeners make use of inferences about the identity of the speaker to make sense of what they say. 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), inferences about the speaker are directly relevant to interpretation. There is a debate, however, on how and precisely when such knowledge is used. In the EEG experiment, people heard utterances whose content did or did not match probabilistic inferences

supported by the speaker's voice, as in (17) delivered with a male of female voice, (18) delivered with the voice of a young kid or an adult, and (19) delivered with a lower- or upper-class accent.

17. I always rent movies with lots of *violence* in it.
18. On Saturday I spent the whole afternoon playing *marbles* on the street.
19. 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. The fact that such words elicited a 'standard' N400 effect indicates that voice-based speaker inferences are brought to bear on language comprehension as rapidly as 'sentence-internal' lexical constraints, and do so as part of the same unified utterance interpretation process.

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 (see section 2, and Van Berkum, 2004) will surely also have caused them to stay away from conversation. On top of this, 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. So, how to make sure that at least one of the interlocutors is at some point presented with a critical event, the EEG response to which we want to measure? As in recent eye tracking research on conversation (see the papers in Trueswell & Tanenhaus, 2005, for examples), probably the best way to do this is to have the participant interact with a confederate of the experimenter, in a collaborative setting that is relatively easy to script and manipulate (such as the referential communication task). Unfortunately, fitting many such lengthy conversational trials into an EEG experiment is a considerable challenge. On top of this, in conversation people tend to speak and move. They make gestures to accompany what they're saying, change posture to express how they feel, and may point to or manipulate objects in the scene. They also make eye movements to visually scan the scene and their interlocutor, and to regulate conversational turn-taking. Because movements cause large recording artifacts, all this is bad news for an EEG experiment. 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.

In view of the daunting task of getting EEG to work in two-person conversation, 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. Magne and colleagues (2005) have recently exploited this paradigm to examine the role of prosody in discourse context. In the ERP experiment, 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 (20) versus (21):

20. Q: Did he give his fiancée a ring or a bracelet?   A: He gave a RING to his fiancée.
21. 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. 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.

Unfortunately, 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 secondary task like this carries the risk that language-related ERP effects become confounded with task-related ERP effects (see also Van Berkum, 2004). Magne and colleagues are certainly not alone in using a secondary task to ensure attention, for in single sentence ERP research, the use of such tasks is widely accepted standard practice. However, the creative elbow room allowed for by having more than one sentence per trial can be used to engage the subject's attention in a somewhat more natural way. To me, it seems that unless there are specific theoretical reasons to check for comprehension (e.g., to sort EEG trials on whether participants 'got the point'), discourse-level EEG researchers should take advantage of that opportunity whenever they can, and stop running the unnecessary risks involved in a secondary task.

The above discussion points to another important motivation to try to combine electrode caps with conversation. Most of the EEG research on language comprehension falls squarely within the *language-as-product* tradition (Clark, 1996), in which readers and listeners are faced with bits of text they essentially don't care about. Of course, to the extent that the processes under study do their job largely regardless of what the linguistic input is for – as is commonly and probably 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 would be good to be able 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 many more such examples). Such paradigms would extend the generality of results obtained with snippets of text, and would allow us to address specific issues associated with action-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?

In the below, I summarize what the EEG studies reviewed so far have to say about the architecture of discourse-level language use. Of course, the neurocognition of discourse and conversation has also been probed in behavioral work with patients and/or hemifield presentation techniques (see Brownell & Friedman, 2001; Marr, 2004; Beeman & Chiarello, 1998), as well as in more recent studies using fMRI and PET (see Ferstl, in press). Any grand theory of how people comprehend discourse and conduct a conversation should adequately reflect what's been found with *all* of these methods.

(1) Rapid incremental processing all the way up. First of all, the evidence from EEG unequivocally reveals that 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,

---

[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.

Jackendoff, 2002), such incremental processing can betray itself in various ways, some of which have been summarized in figure 1. 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 group of subjects: a rapid ERP effect of discourse-level referential ambiguity, a downstream P600/SPS 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. Note that, for the materials in this study, 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. Somehow, unfolding words really make surprisingly 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*.

*** insert figure 1 about here ***

Of course, the extant EEG data should not be taken to suggest that constraints from prior discourse should *always* have an immediately visible impact on processing. The speed with which a conflict between such constraints and a local one will be detected presumably depends on many things, including the relative strength and availability of the various cues at hand, and the amount of computation involved in working out their implications. What the EEG data should be taken to suggest is that constraints from prior discourse *can* have an immediately visible impact on processing, i.e., that there are no *principled* delays on discourse-dependent computations built into architecture of the comprehension system.

(2) Discourse supports 'presonance'. The evidence discussed in section 5 suggests that readers and listeners go beyond incremental processing, and can actually use discourse-level information to make predictions about upcoming communication, as the current utterance is unfolding. As discussed more extensively in Van Berkum et al. (2005a), such predictions need not involve costly conscious deliberation. Instead, they may well virtually come for free, as the result of automatic pattern completion at multiple, highly interrelated levels of representational structure. To capture the idea that much of prediction in natural language comprehension may well involve the automatic retrieval of bits of temporally or sequentially organized memory structure, rather than some form of strategic deliberation, Zwaan (2005) recently coined the term 'presonance' (a variant of 'resonance', the mechanism by which information stored in memory is assumed to be automatically retrieved). What the EEG records suggest is that such presonance can be about the semantic, syntactic, and phonological features of specific upcoming words (Federmeier & Kutas, 1999a,b; Van Berkum et al., 2005a, Otten & Van Berkum, 2005b; DeLong et al., 2005), about whom or what will be referred to soon (Nieuwland & Van Berkum, 2005b, Van Berkum et al., 2005b), and about upcoming syntactic structure (Bornkessel et al., 2003; Van Berkum et al., 1999a). Moreover, results of a recent ERP study on discourse-based lexical anticipation (Otten & Van Berkum, 2005b) can be taken to suggest that the anticipation of upcoming words draws on a message-level representation of the discourse as well as on some more shallow form of priming. In all, it seems that people can extrapolate unfolding discourse at a variety of levels in a relatively cost-free way.

(3) Discourse can overrule local constraints. The evidence reviewed also clearly attests to the power of discourse-level representations relative to more locally imposed constraints. For instance, discourse-level constraints can actually briefly overrule locally unambiguous syntactic information (Van Berkum et al., 1999a,b, see also Bornkessel et al., 2003;). Likewise, a suitable cartoon-like wider discourse can completely overrule a severe local animacy violation, and do so

rapidly enough to prevent the latter from eliciting a differential ERP effect (Nieuwland & Van Berkum, 2005b). Although a peanut in love is certainly an odd thing to encounter, a suitable discourse context can apparently take all of the *communicative* surprise away.

(4) No context-free sentence-internal interpretation. We have seen that regardless of whether words are anomalous (or somewhat less expected) in the context of a wider discourse or in the context of a single sentence, they elicit the same N400 effect at the same time. This suggests that words are immediately related to the widest interpretive context available, and that it does not matter whether this context was provided by the first few words of a single unfolding sentence or, say, a 500-page novel. Furthermore, the interpretive context to which the meaning of an incoming word is immediately related also contains such things as the discourse genre (i.e., whether we are in a cartoon-like story or not; Nieuwland & Van Berkum, 2005b), whether delivery is fluent or not (MacGregor et al., 2005), and whatever inferences can reasonably be inferred about the speaker from his or her voice (Van Berkum et al., 2005c). In addition, classic sentence-level 'semantic' constraints such as animacy effortlessly merge with global discourse-level constraints, such that problems with the former can be totally pre-empted by the latter (Nieuwland & Van Berkum, 2005b). All this is difficult to reconcile with the idea that language users initially compute a local, sentence-internal and context-independent meaning ("what is said") before they determine what the sentence really means in context ("what is meant"). That is, the interpreting brain does not seem to honor the classic distinction between 'semantics' and 'pragmatics', at least, not in the sense of a principled 2-step analysis.

(5) Shallow vs. deep interpretation. Although language interpretation does not appear to be a two-step affair in the sense of context-free followed by context-dependent analysis, the extant EEG evidence does suggest that interpretation involves a quick-and-dirty initial analysis followed (or partially overlapped) by more precise computations. One piece of evidence pointing in this direction is that in the Otten & Van Berkum (2005b) study, 'prime control stories' elicited a left-dominant N400-like attenuation for nouns (e.g.,"sword") that were supported by other words in the context ("dragon", "knight", "threatened"), regardless of the specific message they conveyed. Furthermore, in a separate experiment that probed for the prediction of those nouns by means of a gender-inflected prenominal adjective (Otten & Van Berkum, 2005a), readers were found to anticipate specific upcoming words at two different levels as well, again involving the precise message conveyed by the context as well as the 'bag of words' used to convey this with. Finally, we have recently obtained ERP results (Nieuwland & Van Berkum, 2005a) that suggest that, if a totally discourse-anomalous word is highly scenario-relevant as well as semantically related to the expected discourse-coherent word, the interpretive system *initially* seems to think that all is well (in line with behavioral evidence for semantic illusions), and then very rapidly thereafter discovers that something is quite wrong. Again, this points to some form of shallow heuristics-based analysis, which can later be overruled by the outcome of a more precise analysis.

Of course, the idea that language interpretation may sequentially or in part simultaneously involve a shallow and a more precise analysis has been incorporated in several comprehension models, including Scenario Mapping and Focus theory (Sanford & Garrod, 1998), the Construction-Integration model (Kintsch, 1998), and the LAST model (Townsend & Bever, 2001). Also note that in (until now relatively separate) work on hemispheric specialization in language comprehension, it has been proposed that whereas left-hemisphere language areas might be best at extracting a precise message-level representation and at rapidly suppressing apparently irrelevant information, the right hemisphere might take a somewhat more fuzzy approach, based on associatively related or scenario-relevant information (see, e.g., the chapters in Beeman & Chiarello, 1998; Long, Baynes & Pratt, in press). I believe it will be very useful to try to more precisely relate the various language-relevant ERP phenomena to 2-process models of this sort

(see Coulson, 2004, Federmeier & Kutas, 1999b, and Nieuwland & Van Berkum, 2005a, for initial efforts in this direction).

9. Discourse-unit and discourse-level processing

To get this review started in a manageable way, I provisionally defined discourse-level language comprehension as the processes involved in digesting units of language bigger than a single sentence, i.e., a fragment of text or conversation. In a unit-oriented scheme of things, a discourse, the biggest chunk, is made up of two or more sentences, which in turn consist of words (or, perhaps better, morphemes), which in spoken language contain phonemes, et cetera. With relatively separate research communities for the study of lexical processing, sentence processing, and text or conversational processing, research in psycholinguistics has partly clustered along these lines, a division of labor that is inevitably also reflected in the structure of this handbook. However, with respect to the processes involved in the brain of a language user, a unit-oriented division of labor is not necessarily the most helpful one. To be sure, listeners process phonemes, words, sentences, and discourse. However, they also process incoming language at the level of its phonological, syntactic, and conceptual structure. These levels of analysis are partly orthogonal to the units-based organization. A word, for instance, can be seen as an idiosyncratic bundle of bits of phonological, syntactic, and conceptual information, which in a sentence collectively and incrementally contribute to phonological, syntactic, and conceptual analysis (Jackendoff, 2002).

Because syntax is about sentence structure, syntactic analysis neatly falls within the scope of a single unit of language, the sentence. However, *conceptual* analysis – interpretation – has *two different* units of language as its domain: a sentence, and a discourse. And here is the crux: as summarized in the previous section, *nothing* in the ERP evidence reviewed here supports the idea that the processes involved in *discourse-unit* conceptual interpretation are any different from those involved in the conceptual interpretation of a single sentence. Instead, all the evidence suggests that with every word coming in, people immediately compute contextualized *discourse-level* meaning, by projecting the conceptual implications of the word at hand up to the level of what the discourse is about. If the linguistic context is just the first few words of a single sentence, it is those few words, together with who is speaking, and in what situation, that define what the discourse is about.

The traditional idea that there might be important qualitative (and timing) differences between 'sentence-unit' and 'discourse-unit' interpretation, originally supported by the classic linguistic division between semantics and pragmatics, is also increasingly losing its support in the field of pragmatics itself. In the traditional framework, discourse-unit processing was predicted to be slow and laborious processing, because of all the presumed additional inferential work that would be needed to relate the different sentences to each other. However, modern pragmatic analyses reveal that such inferences are also required to make sense of the words in a single sentence (e.g., Clark, 1996; Kempson, 2001; Sperber & Wilson, 1995). This straightforwardly explains why differences in the 'contextual fit' of a critical word elicit the same type of N400 effect regardless of whether the context is defined by the first few words of a single sentence or a larger piece of discourse. In fact, what the word-elicited N400 might be revealing to us is *the amount of inferential work* required to determine how the meaning(s) afforded by the word at hand can be construed to make sense in the textual and communicative context at hand, possibly along the lines of mechanisms suggested by Relevance Theory (Sperber & Wilson, 1995; Wilson & Sperber, 2004; see Coulson, 2002, for a similar suggestion).

The fact that recent pragmatic analyses and the EEG data reviewed here collectively downplay the relevance of the discourse-sentence split for the interpreting brain has important implications for how we define research on the neurocognition of discourse and conversation. For one, it suggests that there is nothing particularly special about the inferences used to connect

sentences in text. Furthermore, and directly related, it implies that to study the many interpretive phenomena associated with discourse and conversation, one need not always go beyond the single sentence, which is good news for the cognitive neuroscientist. A third 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 the processing of unrelated single sentences (as is typically assumed in fMRI research on text comprehension; see Ferstl, in press, for review). Of course, the associated subtraction may well reveal various systems that are being taxed more by a piece of discourse than by a bunch of single sentences. It may also reveal some of the processes that by necessity require a large piece of text (such as those that are sensitive to 'text structure', story grammar, et cetera). However, if much of the really interesting discourse-conceptual stuff, such as drawing all kinds of bridging inferences to work out what the sentence really means, 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.

Ironically, then, what the psychophysiology of discourse-*unit* comprehension suggests is that with respect to interpretation, the brain doesn't really care about the discourse-sentence unit split. To the language user, discourse-level processing is simply language-driven conceptual processing, regardless of whether it occurs in a single sentence or a longer discourse. And intuitively, this makes sense. Does it really matter, for example, whether the targeted entity of a free referential pronoun like "he" has been introduced in the previous sentence or in the current one? And is the 'sentence-internal' interpretive context set up by "The US policy towards post-1959 Cuba has been extremely…" really so different from that set up by some prior piece of discourse? Of course, a well-designed prior text on the recent history of Cuba may *align* different readers or listeners on whether they prefer "intelligent", "humane", "aggressive" or "stupid". However, in the absence of such prior discourse, comprehension will still be guided by a vast interpretive context, potentially including one's prior knowledge of Cuban history, one's general view on the quality of US foreign policy, and whether the speaker is, say, Fidel Castro or Condoleeza Rice.

[text body: 9630 words]

Figure captions:

Figure 1. Three types of discourse-induced ERP effects during written language comprehension. From left to right: A sustained frontal negative shift to a discourse-induced *referential* problem ("friend" is referentially ambiguous in the 2-referent context), a P600/SPS effect to a discourse-induced *syntactic* problem ("there" rules out the provisional relative-clause analysis pursued at "that" in the 2-referent context), and an N400 effect to a discourse-induced *semantic* problem ("fascists" does not fit the wider story context); see text for further explanation. The example item is shown here in several variants (1- and 2-referent contexts, coherent/anomalous ending), but any one subject saw just a single variant. Data were obtained in a single ERP experiment, aspects of which were described in Van Berkum et al., 1999a, and Van Berkum et al., 1999c.

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.

Beeman, M., & Chiarello, C. (1998) (Eds.), Right hemisphere language comprehension: Perspectives from cognitive neuroscience. Mahwah, NJ: Lawrence Erlbaum Associates.

Bornkessel, I., Schlesewsky, M., Friederici, A.D. (2003). Contextual information modulates initial processes of syntactic integration: the role of inter- versus intrasentential predictions. Journal: Journal of Experimental Psychology: Learning, Memory, and Cognition, 29(5), 871-882.

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. (2005). Manuscript in preparation (dissertation).

Clark, H.H. (1996). Using language. Cambridge: Cambridge University Press.

Coulson, S. (2004). Electrophysiology and pragmatic language comprehension. In I. Noveck and D. Sperber (Eds.), Experimental Pragmatics (pp. 187-206). San Diego: Palgrave Macmillan.

Coulson, S. & Federmeier, K.D. (in press). Words in context: ERPs and the lexical/postlexical distinction. Journal of Psycholinguistic Research

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.

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

Federmeier, K.D. & Kutas, M. (1999b). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. Cognitive Brain Research, 8, 373-392.

Ferstl, E.C. (in press). The functional neuronatomy of text comprehension: What's the story so far? In F. Schmalhofer & C. A. Perfetti (Eds.), Higher-level language processes in the brain: Inference and comprehension processes, Mahwah, NJ: Lawrence 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.

Jackendoff, R. (2002). Foundations of language. New York: Oxford University Press.

Kempson, R. (2001). Pragmatics: Language and communication. In Aronoff, M. & Rees-Miller, J. (Eds.), Handbook of linguistics. Malden, Massachusetts: Blackwell.

Kintsch, W. (1998). Comprehension: A paradigm for cognition. Cambridge: Cambridge University Press.

Kuperberg, G., Caplan, D., Eddy, M., Cotton, J., & Holcomb, P. (2004). Electrophysiological correlates of processing causal relationships between sentences. Journal of Cognitive Neuroscience, supplement.

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., & Van Petten, C.K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M.A. Gernsbacher (Ed.), Handbook of Psycholinguistics (pp. 83-143). New York: Academic Press.

Long, D.L., Baynes, K., & Pratt, C. (in press). Sentence and discourse representation in the two hemispheres. In F. Schmalhofer & C. A. Perfetti (Eds.), Higher-level language processes in the brain: Inference and comprehension processes, Mahwah, NJ: Lawrence Erlbaum.

MacGregor, L.J., Corley, M., & Donaldson, D.I. (2005, September). It's the way that you say it: Everyday disfluency in speech has immediate and lasting effects on the listener. Paper presented at AMLaP'05, Ghent, Belgium, September 5-7 2005.

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.

Nieuwland, M.S., Otten, M., & Van Berkum, J.J.A. (2005). The 'Problem of 2' revisited: Tracking discourse-level referential processing with ERPs. 9th International Conference for Cognitive Neuroscience (ICON-2005), Havana, September 5-10.

Nieuwland, M.S. & Van Berkum, J.J.A. (2005a). 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. (2005b). When peanuts fall in love: N400 evidence for the power of discourse. Journal of Cognitive Neuroscience, accepted pending minor revisions.

Osterhout, L., & Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. Journal of Memory and Language, 31, 785-806.

Otten, M., & Van Berkum, J.J.A., (2005a). Discourse based lexical anticipation during language processing: Prediction or Priming? 18th Annual CUNY Sentence Processing Conference (CUNY-2005), Tucson, March 31 - April 2.

Otten, M., & Van Berkum, J.J.A., (2005b). The influence of message-based predictability and lexical association on the N400 effect. Annual meeting of the Cognitive Neuroscience Society (CNS-2005), New York, April 9-12.

Sanford, A.J., & Garrod, S. (1998). The role of scenario mapping in text comprehension. Discourse Processes, 26, 159-190.

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.

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.

Townsend, D.J., & Bever, T.G. (2001). Sentence comprehension: The integration of habits and rules. Cambridge, MA: MIT Press.

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., 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., & Zwitserlood, P. (2003a). Event-related brain potentials reflect discourse-referential ambiguity in spoken-language comprehension. Psychophysiology, 40, 235-248.

Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005a). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. Journal of Experimental Psychology: Learning, Memory, & Cognition, 31(3), 443-467.

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

Van Berkum, J.J.A., Otten, M., Jansen, J., Koornneef, A., De Boer, S., Huizinga, F., & Junge, C. (2005b). Why does "David praised Linda because he was proud" feel odd? ERP evidence for the predictive use of verb-based interpersonal bias. 9th International Conference for Cognitive Neuroscience (ICON-2005), Havana, September 5-10.

Van Berkum, J.J.A., Van den Brink, D., Tesink, C., Kos, M., Müller, O., & Hagoort, P. (2005c). That's odd for you to say! An EEG study on how voice-inferred speaker characteristics affect language comprehension. 9th International Conference for Cognitive Neuroscience (ICON-2005), Havana, September 5-10.

Van Berkum, J.J.A., Zwitserlood, P., Bastiaansen, M.C.M., Brown, C.M. & Hagoort, P. (2004). So who's "he" anyway? Differential ERP and ERSP effects of referential success, ambiguity and failure during spoken language comprehension. Annual meeting of the Cognitive Neuroscience Society (CNS-2004), San Francisco, April 18-20.

Van Berkum, J.J.A., Zwitserlood, P., Brown, C.M., & Hagoort, P., (2003b) 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. (1995). Words and sentences: Event-related brain potential measures. Psychophysiology, 32, 511-525.

Wilson, D., & Sperber, D. (2004). Relevance Theory. In G. Ward and L. Horn (eds.) Handbook of Pragmatics. Oxford: Blackwell, 607-632.

Yang, C.L., Perfetti, C.A., & Schmalhofer, F. (2005). ERPs expose integration processes in text comprehension. Submitted for publication.

Zwaan, R.A. & Kashak, M.P. (2005). Language comprehension as a means of "re-situating" oneself. To appear in Robbins, P., & Aydede, M. (Eds.), Cambridge handbook of situated cognition. Cambridge University Press.

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.

Just as the elderly hippie had lit up a joint, he got a visit from a friend and a nephew (two friends).
Even though his friend (one of his friends) had had quite a few drinks already, and the nephew
(the other one) had just smoked quite a lot of pot already, they insisted on smoking along. The
hippie warned the **friend** *that* **there** *would be some* **problems/fascists** *soon.*
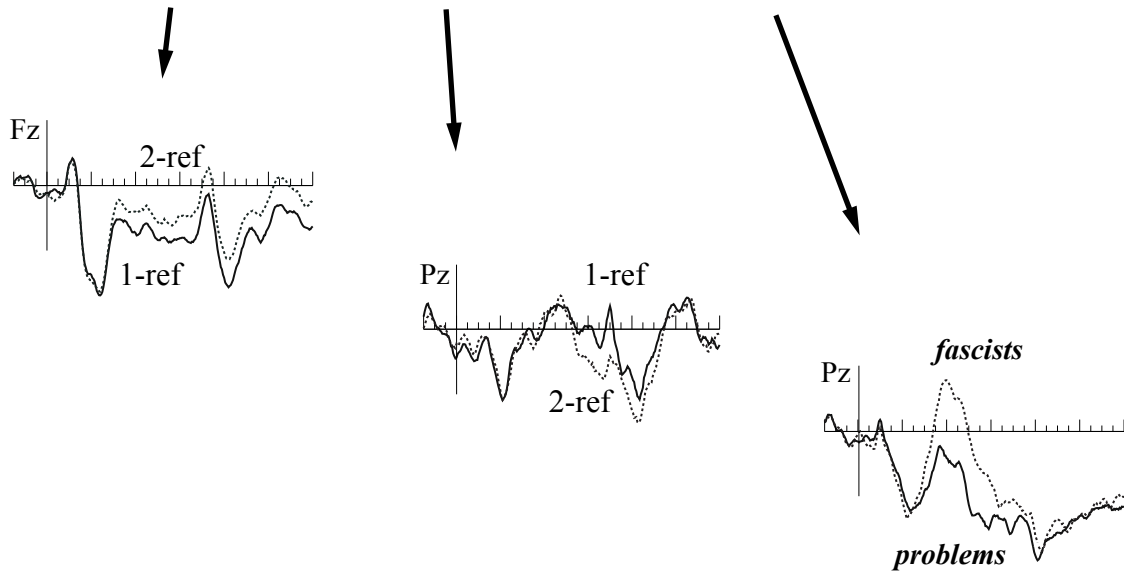

Figure 1. Three types of discourse-induced ERP effects during written language comprehension.
From left to right: A sustained frontal negative shift to a discourse-induced referential problem
("friend" is referentially ambiguous in the 2-referent context), a P600/SPS effect to a discourse-induced
syntactic problem ("there" rules out the provisional relative-clause analysis pursued at "that" in the
2-referent context), and an N400 effect to a discourse-induced semantic problem ("fascists" does not
fit the wider story context); see text for further explanation. The example item is shown here in several
variants (1- and 2-referent contexts, coherent/anomalous ending), but any one subject saw just a
single variant. Data were obtained in a single ERP experiment, aspects of which were described in
Van Berkum et al., 1999a, and Van Berkum et al., 1999c.