Discourse-based lexical anticipation : the nature and contextual basis of predictions in language comprehension

Otten, M.

Publication date
2008

Chapter 6

What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse dependent N400 effect.

Chapter 2

A discourse context provides a reader with a great deal of information that can provide constraints for further language processing, at several different levels. In this experiment we used event-related potentials (ERPs) to explore whether discourse-generated contextual constraints are based on the precise message of the discourse or, more 'loosely', on the scenario suggested by one or more content words in the text. Participants read constraining stories whose precise message rendered a particular word highly predictable ("The manager thought that the board of directors should assemble to discuss the issue. He planned a…meeting") as well as non-constraining control stories that were only biasing in virtue of the scenario suggested by some of the words ("The manager thought that the board of directors need not assemble to discuss the issue. He planned a…"). Coherent words that were inconsistent with the message-level expectation raised in a constraining discourse, for example "session" instead of "meeting", elicited a classic centroparietal N400 effect. However, when the same words were only inconsistent with the scenario loosely suggested by earlier words in the text, they elicited a different negativity around 400 ms, with a more anterior, left-lateralized maximum. The fact that the discourse-dependent N400 effect cannot be reduced to scenario-mediated priming reveals that it reflects the rapid use of precise message-level constraints in comprehension. At the same time, the left-lateralized negativity in non-constraining stories suggests that, at least in the absence of strong message-level constraints, scenario-mediated priming does also rapidly affect comprehension.

Introduction

Over the last two decades, event related potentials (ERPs) have considerably advanced our understanding of the cognitive processes underlying language comprehension. The N400, an ERP component that is particularly sensitive to semantic processing, has played a major role in this. The N400 has been used to study the comprehension of written, spoken and signed language (see Kutas, Van Petten and Kluender (2006) for review), as well as the breakdown of language or of related cognitive functions in for example aphasia (Hagoort, Brown, & Swaab, 1996; Swaab, Brown, & Hagoort, 1997), schizophrenia (Sitnikova, Salisbury, Kuperberg, & Holcomb, 2002), and Alzheimer's disease (Ford et al., 1996). The N400 component was discovered by Kutas and Hillyard (Kutas & Hillyard, 1980), who found that a sentence-final word that was incongruent with the preceding linguistic context evoked a larger negativity than a congruent ending. Later experiments showed that the N400 was more than a semantic anomaly detector. In particular, relative to highly expected words, semantically coherent but unexpected words also evoked a larger N400, albeit not as large as anomalies (Kutas & Hillyard, 1984). Based on these and other findings, Kutas and colleagues (2006) have recently argued that the amplitude of the N400 reflects the degree to which "context aids in the interpretation of a potentially meaningful stimulus". In line with this, language researchers generally agree that the word-elicited N400 indexes how well the meaning of a word fits the constraints set by the context (Brown, van Berkum, & Hagoort, 2000; Chwilla, Brown, & Hagoort, 1995; Coulson & Federmeier, in press; Friederici, 1995; Hagoort et al., 2004; Osterhout & Holcomb, 1995; Van Berkum, Brown et al., 2003).

But what are those constraints? Following up on the pioneering N400 research of St. George, Mannes and Hoffman (1994), ERP experiments with text-level manipulations suggest that the N400 is not only sensitive to constraints provided by a single word prime or an unfolding single sentence, but is also highly sensitive to what the wider discourse is about (Brown et al., 2000; Federmeier & Kutas, 1999b; Nieuwland & Van Berkum, 2006b; St George et al., 1997; Van Berkum, Brown et al., 2003; Van Berkum et al., 2005). In stories such as (1), for example, the discourse-supported word "movie" elicited a much smaller N400 than the discourse-inappropriate word "book" (Van Berkum et al., 1999; 2003). Such discourse-dependent N400 effects have been interpreted as evidence that the language comprehension system immediately evaluates the current word against a precise message-level representation of what has been said so far.

*(1) David and Sabrina had been thinking about what they were going to do that evening, and eventually decided to go to the cinema. They hoped the **movie/ book** would be fun, but it turned out to be quite boring.*

However, an important alternative interpretation has as yet not been ruled out. Models of text comprehension and memory suggest that in addition to contributing to a precise message-level representation of the discourse, the words in a text can also provide semantic constraints in a much less precise way, via the activation of related information stored in long term memory (Kintsch, 1988; McKoon & Ratcliff, 1992; Sanford, 1990). In the above story, for instance, the mere presence of the word "cinema" could activate a going-to-the-movies scenario, which includes seeing a film. The attenuation of the N400 in (1) might thus also come about because the word "movie" is relevant to the scenario suggested by one or several words in the preceding text.

The difference between a message-level and scenario-mediated account for discourse-dependent N400 effects may not be obvious at first. After all, our understanding of what has been said so far, the precise message, will in part depend on our default knowledge about what things tend to go together in the world (as captured in scenarios, scripts, etcetera). However, consider what would happen if we change the precise message of the discourse, as in (2). Although the going-to-the-movies-scenario is still implied by the words in the context, the actual message of the story does not really support either "movie" or "book".

*(2) David and Sabrina had been thinking about what they were going to do that evening, and eventually decided not to go to the cinema. They hoped the **movie/ book** would be out on dvd soon, and went to the pub.*

In the ERP study reported below, we try to disentangle the effects of message- and scenario-level constraints by exploiting the possibility to change the message of a story while leaving the scenario-relevant words in that story intact. The goal is to examine whether the discourse-dependent N400 effect hinges on constraints provided by the exact message-level representation of the prior text (as assumed in Van Berkum et al., 1999; 2003), or whether it can perhaps be accounted for – entirely or in part – by scenario-mediated lexical priming.

Behavioral experiments have shown that this scenario-mediated priming does play a role in comprehension, and can do so even when the scenario-generated information is irrelevant to, or at odds with, the actual message (Duffy et al., 1989; Garrod & Terras, 2000; O'Seaghdha, 1997). Garrod and Terras (2000), for example, showed that the word "pen" is initially just as

effectively integrated when presented in a sentence following "The teacher wrote a letter" as it is after the sentence "The teacher wrote the exercise on the blackboard". Only in regression path analysis and second pass reading times a significant difference was observed between the appropriate and inappropriate contextual message. This indicates that participants did not at first notice the message-level incongruence of "pen", presumably because "pen" is strongly associated with the default scenario activated by the verb ("to write"). Results like these imply that contextually activated scenarios can prime scenario-related concepts, leading to facilitated processing of these concepts and the words that denote them. Furthermore, they show that scenario-related facilitation is in some cases initially stronger than the support provided by the actual message of the discourse

To account for such results, models of text comprehension (Kintsch, 1988; 1998; Sanford & Garrod, 1981; 1998) usually include an initial stage in which all potentially relevant information is retrieved in a way that is highly sensitive to the set of words and concepts in the text, regardless of the precise message of the text. For example, the scenario-mapping and focus theory of Sanford and Garrod (1981; 1998) proposes that the word currently read or heard (e.g. "movie") is initially evaluated in terms of a 'quick & dirty' match to prior words in the text and the scenario suggested by those words, before it is mapped more carefully onto the precise message conveyed by that text. Kintsch (1988; 1998) has made similar proposals. The dominant model for the relatively shallow initial retrieval process involved in these accounts is the resonance model (Cook et al., 1998; Myers & O' Brien, 1998; Myers et al., 1994). According to this model, individual concepts from the linguistic input send out a signal to long term memory. Concepts in memory then resonate as a function of their relatedness to the input, based on the overlap between the semantic and contextual features of the concepts involved. Eventually, those concepts that have the highest level of activation enter working memory. The resonance process is assumed to be fast-acting and autonomous (or "dumb" (Myers & O' Brien, 1998)). As a result, activated information in long-term memory can be irrelevant to the specific meaning conveyed by the complete discourse, and may in fact even be incongruent with that message.

In a recent ERP experiment Hoeks, Stowe and Doedens (2004) directly compared the effects of message-level constraint and scenario fit on the N400 by contrasting the ERP evoked by a scenario-related verb in a highly constraining sentence like "The javelin was by the athletes thrown" (approximate translation from Dutch) and in a much less constraining sentence like "The javelin has the athletes thrown". In spite of the difference in

message, both sentences induce the same athletics-scenario due to the presence of related words ("javelin", "athletes") in the context. Surprisingly, the N400 to scenario-related verbs was independent of whether the actual message of the context supported or prohibited this verb. This result suggests that the N400 is sensitive to the effects of scenario fit, and that these scenario effects can under some conditions fully determine the amplitude of this component, at least in the early stages of comprehension.

If the N400 is indeed as sensitive to scenario fit as the results of Hoeks et al. (2004) suggest, this has important consequences for the functional interpretation of the N400, but also for the interpretation of experiments in which the N400 is used as a measure of message-level semantic integration or contextual facilitation. More specifically, if the N400 elicited by words in text is (also) dependent on scenario-based constraints, then differential N400 effects cannot be automatically taken as an indication that the words at hand are differentially integrated with a message-level representation of the text. As already discussed, this ambiguity affects the interpretation of discourse-dependent N400 effects, and we focus our experiment on this type of effect. However, consistent with the account given for discourse-dependent N400 effects in Van Berkum et al. (1999; 2003), the ambiguity also affects the interpretation of sentence-dependent N400 effects, such as for those elicited by, e.g., "He mailed the letter without a stamp" or "She locked the valuables in the safe" (Kutas & Hillyard, 1984). As with longer texts, it is not clear whether attenuated N400 effects in sentences like these reflect support from the precise message, or from the general scenario suggested by, say, "mailed" and "letter".

In our experiment, we examined whether the discourse-induced N400 effect results from a mismatch between the incoming word and the specific constraints set by the precise message of the discourse, or whether this N400 effect reflects scenario-based fit. We roughly followed the logic embodied in examples (1) and (2), avoiding semantic anomalies, since semantically congruent words that differ in sentence- or discourse-based predictability can also elicit a large differential N400 effect (DeLong et al., 2005; Hagoort & Brown, 1994; Kutas & Hillyard, 1984; Van Berkum et al., 2005).

Table 6.1 shows the two types of stories used in the experiment. In the so-called message- and scenario-biased (or MS-biased) story, the text leading up to the critical word has a highly constraining message, such that when people are asked to complete the story in a cloze test they predominantly converge on "meeting", and very rarely come up with the equally congruent word "session". This predictability critically hinges on the prior discourse

Table 6.1. An approximate English translation of one of the stories shown to the participants, in all four conditions. The critical nouns are printed in boldface.

1. Message & Scenario biased Discourse	
Bias-Consistent Noun	*Bias-Inconsistent Noun*
The manager thought that the board of directors should assemble to discuss the issue. He planned a **meeting** where the staff members involved would be present as well.	The manager thought that the board of directors should assemble to discuss the issue. He planned a **session** where the staff members involved would be present as well.
2. Scenario biased Discourse	
Bias-Consistent Noun	*Bias-Inconsistent Noun*
The manager thought that the board of directors need not assemble to discuss the issue. He planned a **meeting** where the staff members involved would be present and nobody else.	The manager thought that the board of directors need not assemble to discuss the issue. He planned a **session** where the staff members involved would be present and nobody else.

supplied by the preceding sentence (Van Berkum et al., 2005). Next, because the discourse-dependent attenuation of the N400 for a word like "meeting" might also be caused by its support from the meeting-scenario suggested by the individual words in the MS-biased story quoted in Table 6.1 like "manager", "board", "directors", "assemble" and "discuss", we created so-called scenario-biased (or S-biased) control stories. In these stories, we changed the precise message of the first sentence such that neither of the two critical words (nor any other word) was particularly predictable. At the same time, we made sure to preserve the scenario-relevant content words of their MS-biased counterparts (e.g., "manager", "board", "directors", "assemble" and "discuss"), so that these S-biased control stories would remind people of the same scenario.

If the discourse-dependent N400 effect is solely a result of the mismatch between the incoming word and a precise message-level representation of the discourse, then the N400 effect elicited by MS-biased stories should not be elicited in S-biased stories. Alternatively, if the discourse-dependent N400 effect hinges solely on fit to the scenario induced by a set of content words, the N400 effect elicited by MS-biased stories should be comparable to the one elicited in S-biased stories. Of course, to the extent that the precise message conveyed by the discourse and the scenario induced by a set of words in that

discourse both affect the N400, both types of stories should evoke an N400effect. However, in this case the N400 effect for S-biased stories, where only the scenario supports the bias-consistent word, should be smaller than the N400 effect observed in MS-biased stories, in which the bias-consistent word is supported by both the scenario and the message of the story.

Methods

Participants

36 right-handed native speakers of Dutch (27 female participants, mean age 21, range 18-26 years) took part in the experiment, as part of a course requirement. None had any neurological impairment, had experienced neurological trauma, or used neuroleptics. None of the participants had participated in the pre-tests conducted during the material construction phase.

Materials

The stimuli in this experiment were 160 mini-stories of two sentences, with the first sentence establishing the discourse context, followed by the local carrier sentence containing the critical word. For each item a *message- and scenario-biased* (or *MS-biased*) discourse as well as a *scenario-biased* (or *S-biased*) control discourse was created, both containing the same scenario-relevant words (see Table 6.1 and the Appendix 3 for examples of the stimulus materials). All stories were designed to suggest a specific message-predictable word right after the indefinite article in the target sentence (the second sentence) in the MS-biased condition, but not in the S-biased condition. In addition, S-biased stories suggested a scenario that favoured the bias-consistent critical word over the bias-inconsistent one. Across the MS- and S-biased conditions, we refer to the message/scenario-predictable words as *bias-consistent* words, and to the coherent but essentially unpredictable control word as *bias-inconsistent* words.

The level of *message-level* constraint for each of the two conditions of each critical story was determined in a pencil-and-paper cloze test, prior to the EEG-experiment. In this pretest, we showed the MS-biased and S-biased mini-stories up to (and thus not including) the critical word to 66 participants, and asked them to complete the story with the first thing that came to mind. The two versions of the items were divided over separate lists, so that a participant never saw an item in more than one context condition. For each item the cloze value (the proportion of participants who filled in the critical word) for the bias-consistent word and the bias-inconsistent control word was calculated, in

both the MS-biased and the S-biased condition. Only those items were selected in which the cloze value for the *bias-consistent* word was higher than .50 in the MS-biased version of the story and lower than .30 in the S-biased version, with a difference between these two values of at least .25. In the resulting itemset, bias-consistent words had a mean cloze value of .77 (sd = .13) across all MS-biased stories and .18 (sd = .15) across all S-biased stories. The mean cloze value for corresponding bias-inconsistent control words was .04 in both the MS-biased stories (sd = .06) and the S-biased stories (sd = .07).

Bias-consistent and bias-inconsistent words were matched on average length and frequency: The mean length of the discourse-predictable and control word was respectively 6.2 (sd = 2.2) characters and 6.8 characters (sd = 2.5), and the mean frequency for discourse-predictable and control words was respectively 30.5 (sd = 52.6) and 30.3 (sd = 65.6) per 1 million, as stated in the Celex database. All words preceding the critical noun and the two words that followed the critical noun in the second sentence were identical for each of the four conditions, and as such did not differ in length or frequency. After the two identical words that followed the critical noun, the remainder of the story sometimes varied between MS- and S-biased stories, to avoid coherence breaks at the message level.

The 160 items (40 for each of the four conditions shown in Table 6.1) were pseudo-randomly mixed with 80 filler items addressing an unrelated issue (Van Berkum et al., 2007). By rotating the conditions in this list, three more lists of stimuli were created. Each of the four lists contained all 160 experimental stimuli, 80 stories in the MS-biased discourse version and 80 with a S-biased discourse. Half of the 80 MS-biased items and half of the 80 S-biased items contained the bias-consistent word, while the remaining 40 ended with a bias-inconsistent word. Each participant was shown one of these four lists of stimuli, so that one participant saw all stories, but never in more than one condition.

Procedure, EEG recording and Analysis

Each participant saw 240 stories, 160 of which were critical for the current issue. Participants were asked to read for comprehension and were not required to perform any other task. The electroencephalogram (EEG) was recorded with 30 electrodes (FP1, FP2, F9, F7, F3, Fz, F4, F8, F10, FT9, FC5, FC2, FC6, FC1, FT10, T7, C3, Cz, C4, T8, CP5, CP1, Cp2, Cp6, P7, P3, Pz, P4, P8 and Oz) mounted in an elastic cap, each referenced to the left mastoid. Blinks and vertical eye-movements were registered by placing an electrode under the left eye, initially referenced to the left mastoid, but later

rereferenced to an electrode above the left eye (Fp1). Electrode impedance was kept below 5 kOhms during the experiment. The EEG was amplified, band-pass filtered at 0.03 Hz-100 Hz and sampled with a frequency of 500 Hz.

During the comprehension task the participants sat in a comfortable chair in a normally lit room. The stimuli were presented in black 36 point courier new font on a white background on a fast TFT display (Iiyama TXA 3834 MT) positioned approximately 80 cm away from the participant. Before each trial, a fixation cross was shown in the center of the screen for 2.5 s. Participants were asked to avoid blinks and eye movements when the words were presented on screen, and were encouraged to blink when the fixation cross was shown. To signal the start of each trial to the participant a beep sounded 1 s before the onset of the first word.

The stories were then presented word for word. To make the visual presentation more natural, words were presented using a Variable Serial Visual Presentation (VSVP) procedure, in which the presentation time of each non-critical word varied with its length. Non-critical word duration consisted of a standard offset of 187 ms plus and additional 27 ms per letter (with an upper bound of 10 letters for each word). In the present experiment, durations varied from 214 ms for a one-letter word to 450 ms for words consisting of ten or more letters. The interword interval was always 106 ms. The presentation of clause-final words preceding a comma was prolonged with an additional 200 ms. In addition, presentation time for sentence-final words was extended with an extra 293 ms, followed by a 1 s pause until the next sentence began. These various parameters were based on natural reading times (Haberlandt & Graesser, 1985; Legge et al., 1997), a subjective assessment of the naturalness of the resulting presentation, and technical constraints imposed by the video refresh rate. Note that to the extent that critical words, or words close to the critical word, differ in average length, the above procedure will induce unintended shifts in the ERP waveforms (particularly the exogenous deflections associated with visual word onset and offset). To avoid spurious ERP effects due to these shifts, words whose exogenous components fall in the critical EEG epoch (or baseline) should therefore be equated across condition on their presentation time. In the present study, the critical noun and the three words that followed were presented with a fixed duration of 346 ms, based on the average critical word length across all 240 stories in the experiment (6 characters). Participants did not notice the alternation between completely variable and semi-fixed word duration presentation within a single story.

The data were re-referenced off-line to the average of right and left mastoids. Blinks and eye movements were removed from the data using a

procedure based on Independent Component Analysis (ICA) as described by Jung and colleagues. (Jung, Makeig, Humphries et al., 2000; Jung, Makeig, Westerfield et al., 2000). After that the data were segmented, timelocked to the onset of the critical word, from 500 ms before critical word onset until 1200 ms after critical word onset. Segments in which the signal exceeded (-)100 µV and those containing linear drift that was not related to the onset of the critical word and exceeded (-) 40 µV were eliminated off-line. Due to the presence of artifacts, for each of the four conditions about 7% of the trials were deleted. The remaining trials were normalized by subtracting the mean amplitude in a 200 ms pre-stimulus interval. For each participant the trials were then averaged for each of the four conditions, timelocked to the onset of the word.

Analyses of Variance (ANOVAs) were conducted, using mean amplitude values computed for each participant and condition in the 300-500 ms N400 window for each electrode. Univariate F tests with more than one degree of freedom in the numerator were adjusted by means of the Greenhouse-Geisser or Huynh–Feldt correction where appropriate. Uncorrected degrees of freedom and corrected *P*-values are reported in the Results. The results were evaluated in an overall ANOVA with the factors Context (MS-biased and S-biased) and Consistency (bias-consistent word and bias-inconsistent word). To evaluate differences in scalp distribution of the observed effect, an ANOVA with the factors and Context, Consistency and Electrode was conducted after the data were scaled by vector-length (McCarthy & Wood, 1985). Each participant's amplitude value was divided by the square root of the sum of the squared amplitudes over all electrodes for each condition of that participant (i.e. the vector-length for each condition). The nature of significant interactions with Electrode was assessed in a quadrant ANOVA crossing Context and Consistency with a Hemisphere (left/right) and Anteriority (anterior/posterior) factor. Four quadrants were defined: (1) left-anterior, comprising FP1, F3, F7, F9, FC1, FC5 and FT9; (2) right-anterior, comprising FP2, F4, F8, F10, FC2, FC6 and FT10; (3) left-posterior, comprising C3, T7, CP1, CP5, P3 and P7; (4) right-posterior, comprising C4, T8, CP2, CP6, P4 and P8.

Results

Figure 6.1 shows for each electrode the ERPs evoked by the bias-consistent (solid line) and bias-inconsistent (dotted line) word in the MS-biased condition. Figure 6.2 shows the ERPs for the same words in the S-biased condition. As can be seen in Figure 6.1, semantically coherent words that are not what would be expected given the highly constraining message of the discourse so far elicit a clear N400 effect, relative to coherent expected words. However, Figure 6.2 reveals that the same bias-inconsistent words also evoke a negativity in the N400 latency range when the discourse is only biasing through a related scenario, whilst the actual message is not constraining. Figure 6.3 shows the difference waves (ERP for inconsistent words – ERP for consistent words) for the MS-biased (solid line) and the S-biased discourse (dotted line), together with the scalp topographies of the bias inconsistency effect between 300 and 500 ms for the two context conditions. What can be

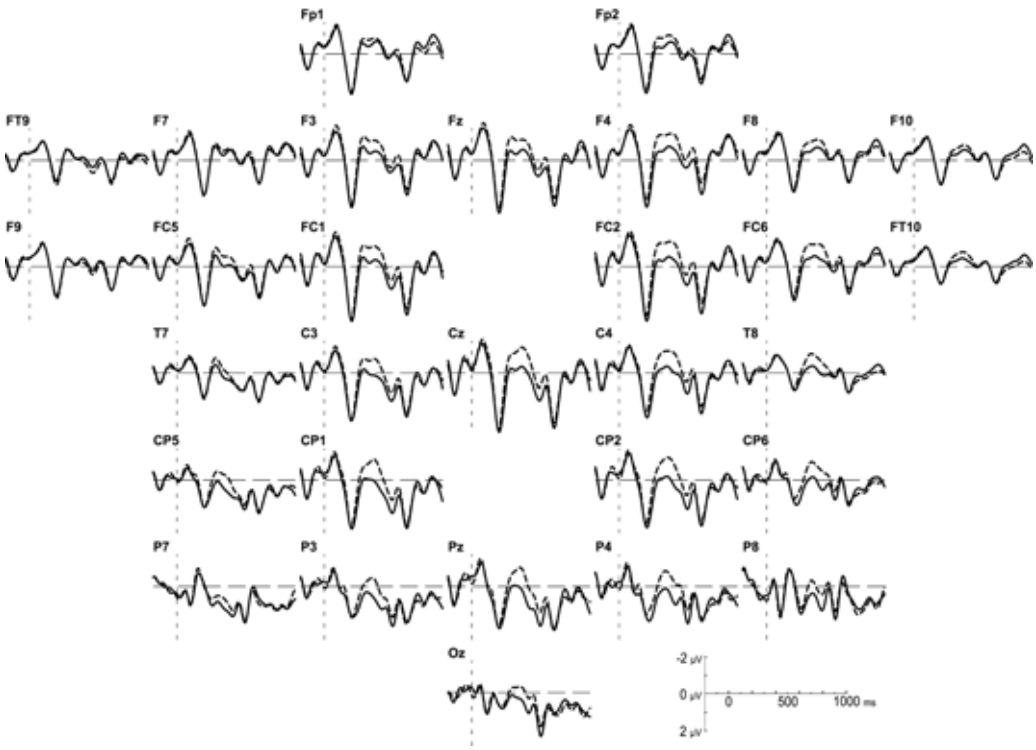

Figure 6.1 *Grand average ERPs elicited by the critical nouns in a MS biased discourse. The solid waveforms, represent the response to bias-consistent target nouns; the dotted waveforms represent responses to bias-inconsistent target nouns.*

seen here is that, whereas the bias inconsistency effect in MS-biased stories has a typical N400 distribution (a centro-parietal and slightly right-lateralized maximum), the effect in S-biased stories has a very different scalp distribution, with a more anterior, left-central maximum.

The overall Context x Consistency ANOVA on mean amplitudes in the 300-500 ms latency range reveals a significant main effect of Consistency ($F(1,35) = 11,7$; $p = .002$), with inconsistent words evoking a larger negativity than consistent words, and a main effect of Context ($F(1,35) = 5.4$, $p = .03$), which is a result of a larger negativity for words presented in a MS-biased discourse than a S-biased discourse, regardless of the nature of the word (consistent or inconsistent).

In spite of the difference in overall effect size visible in Figure 6.3, the Consistency x Context interaction in the overall ANOVA was not significant

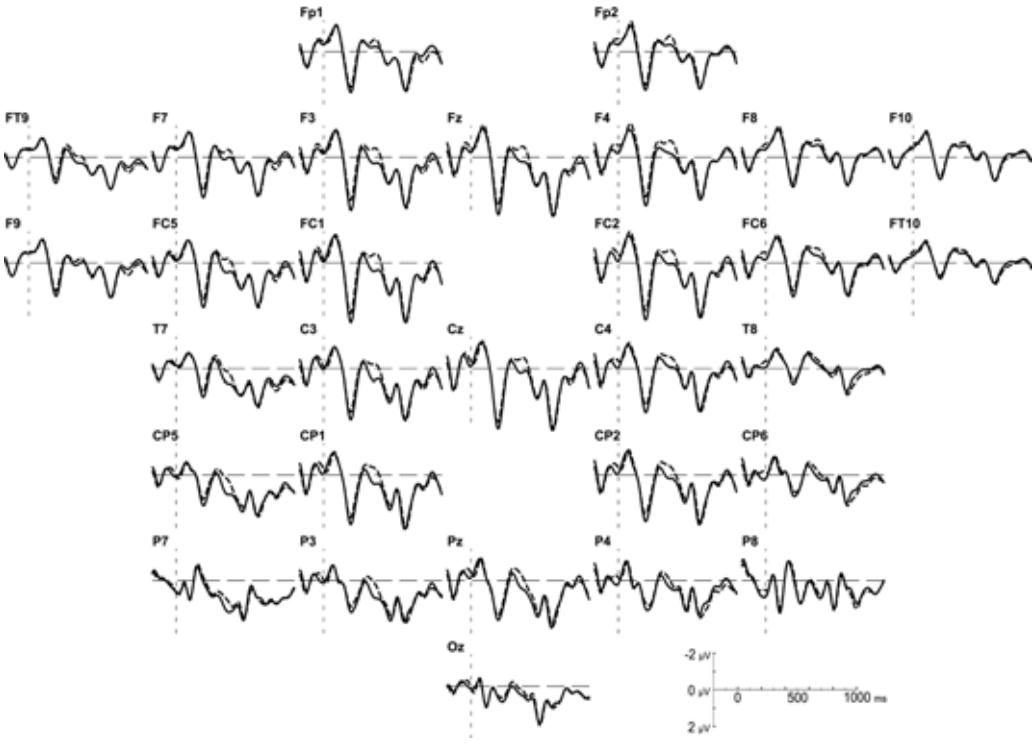

Figure 6.2 *Grand average ERPs elicited by the critical nouns in an S biased discourse. The solid waveforms, represent the response to bias-consistent target nouns; the dotted waveforms represent responses to bias-inconsistent target nouns.*

(F(1,35) = 1.6, p = .22). However, the difference in effect distribution was statistically corroborated, after scaling the data, by a reliable Consistency x Context x Electrode interaction (F(29,1015) = 2.4, p = .04) in the overall ANOVA, as well as a reliable Consistency x Context x Hemisphere interaction (F(1,35) = 8.0, p = .008) in the quadrant ANOVA. Follow-up tests on the unscaled data indicated that whereas the two consistency effects did not reliably differ over the left hemisphere (F(1,35) = 0.4, p = .54), they did differ

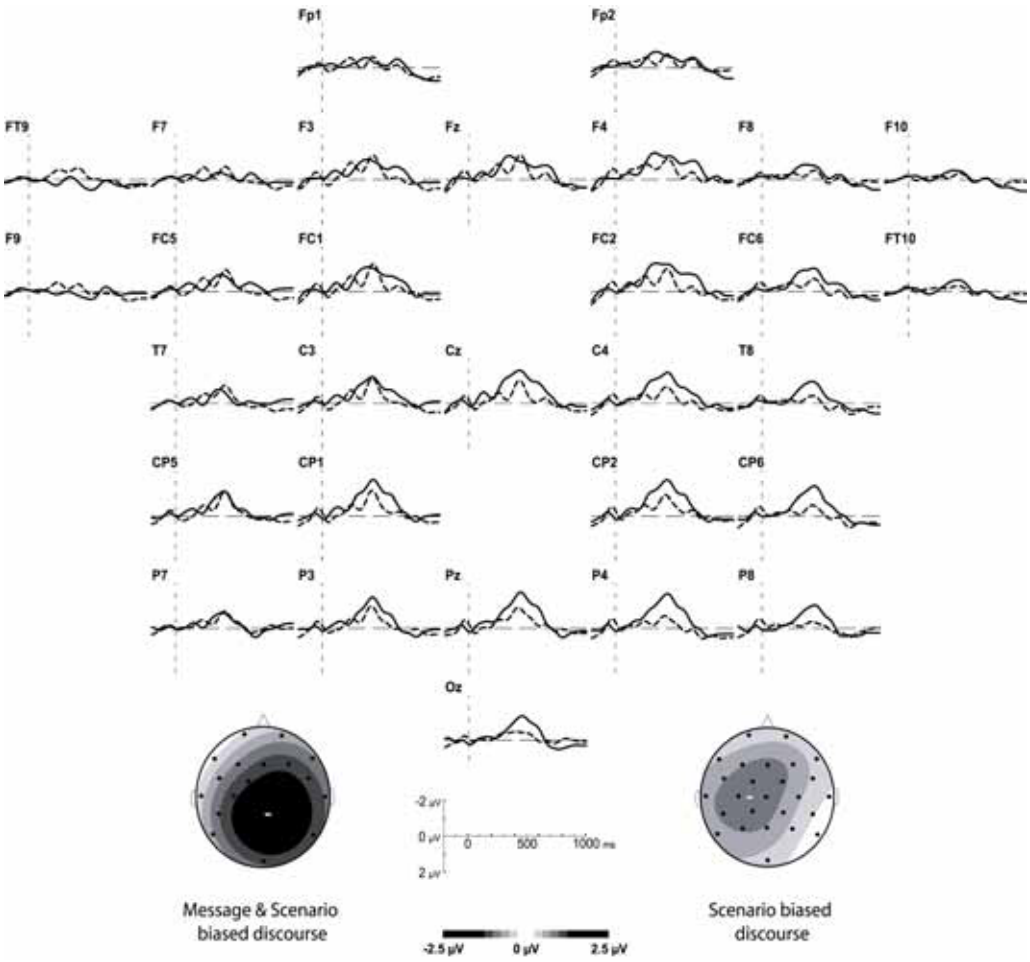

Figure 6.3 *Difference waves for Message & Scenario and Scenario biased Discourses. The difference waves depicted here result from subtracting the ERP elicited by bias-consistent targets from the ERP evoked by bias-inconsistent targets. Solid waveforms represent the difference wave for the MS biased discourse; dotted waveforms represent the difference wave for the Scenario biased discourse. Also shown are the scalp distributions for the effect of Bias-Consistency in each of the two conditions.*

over the right hemisphere ($F(1,35) = 4.6$, $p = .04$). Further simple main effects analyses revealed a reliable negativity over the left hemisphere in the MS-biased condition ($F(1,35) = 5.9$, $p = .020$) as well as the S-biased condition ($F(1,35) = 5.6$, $p = .023$), but a reliable negativity over the right hemisphere in the MS-biased condition only ($F(1,35) = 10.7$, $p = .002$), and not in the S-biased condition ($F(1,35) = 1.8$, $p = .19$).

Discussion

We examined whether word-elicited N400 effects within a coherent discourse are solely determined by constraints set by the precise message of the discourse or by the contextually activated scenario. In an MS-biased discourse, where both the message of the discourse *and* the invoked scenario supported a specific noun, critical words that were inconsistent with these biases indeed elicited a larger N400 than the completely consistent critical word. However, in the S-biased discourse, where the consistent word was only supported by the scenario suggested by one or more content words in the text, bias-inconsistent words also elicited an increased negativity around 400 ms, now with a more anterior, left-lateralized maximum. The fact that the discourse-dependent N400 effect cannot be reduced to scenario-mediated priming reveals that it reflects the rapid use of precise message-level constraints in comprehension. At the same time, the presence of residual left-dominant negativity in S-biased stories suggests that, at least in the absence of strong message-level constraints, scenario-mediated priming does also rapidly affect comprehension. We discuss both implications in turn, and then discuss their possible relation.

A message-based discourse-level N400 effect

In MS-biased stories designed to generate specific word expectations, coherent words that did not meet those expectations elicited a classic N400 effect, peaking at 400 ms, and with a centroparietal, slightly right-lateralized maximum. As suggested by Figure 6.3, this effect cannot be explained by scenario-mediated (or other word-based 'non-message') priming, for when the message-level constraint was removed while all potentially scenario-relevant words were still present (as in S-biased control stories), the same critical words no longer elicited a classic N400 effect. Although critical words in S-biased control stories also gave rise to a negativity, its very different scalp distribution implies that the N400 effect observed in MS-biased stories at least partly reflects other processes than those induced by scenario consistency. Since the two types of stories only systematically differ in the precise message

and its degree of predictability, the N400 effect observed in MS-biased stories must reflect, at least in part, the impact of message-level constraints.

This result supports earlier claims (Brown et al., 2000; Nieuwland & Van Berkum, 2006b; St George et al., 1994; Van Berkum, Brown et al., 2003) that comprehenders immediately relate the meaning of every incoming word to a precise message-level representation of the wider discourse so far. That is, comprehension is incremental all the way up to the level that matters most: what the story or conversation is about. Note that earlier reports on discourse-dependent N400 effects invariably relied on semantic anomalies (the only earlier study in which discourse-dependent cloze differences also elicited an N400 effect was aimed at a different phenomenon, and did not cleanly manipulate N400-relevant cloze probabilities; Van Berkum et al., 2005). In the present study a discourse-dependent N400 was elicited by acceptable but unexpected words, which were thus not anomalous. This is entirely consistent with the fact that subtle cloze manipulations can also induce N400 effects in coherent single sentences (DeLong et al., 2005; Hagoort & Brown, 1994; Kutas & Hillyard, 1984).

The effect of scenario fit

As just discussed, the data show that the discourse dependent N400 effect cannot be reduced to mere differences in scenario fit: the actual message of the discourse is taken into account when evaluating new information. However, when the message of the discourse was not particularly constraining, as is the case in the S-biased discourse, bias-inconsistent words still evoked a differential negativity in the N400 time window. The difference in scalp distribution for the two effects after normalization shows that the negativity in the S-biased condition is not just a smaller version of the standard N400 effect present in the MS-biased discourse. As a result, the negativity in the S-biased condition cannot be uniquely attributed to the small remaining difference in cloze value between the scenario-consistent and the scenario-inconsistent word.

But how then should we interpret this unexpected effect related to scenario bias? The timing and polarity of the effect are in line with a modulation of the classic N400, but whereas the distribution of the latter is consistently described to have a centro-parietal maximum (Curran, Tucker, Kutas, & Posner, 1993; Johnson & Hamm, 2000; Kutas & Van Petten, 1994; Osterhout & Holcomb, 1995), the current effect has a more anterior left-lateralized maximum. This distribution resembles the scalp topography of the left anterior negativity or LAN (Coulson et al., 1998; Friederici, Hahne, &

Mecklinger, 1996; Gunter, Stowe, & Mulder, 1997 ; Osterhout & Mobley, 1995). LAN components, however, are related to morphosyntactic violations, whereas the critical manipulation in the present design is purely lexical-semantic. A more appealing explanation for this might therefore be that the observed scenario-related negativity recruits some (but clearly not all) of the neuronal generators that also underlie the classic N400 effect. This is in line with the observed similarities in timing and polarity for the MS- and S-related effects. Such an account would suggest that scenario-based priming lies at the basis of the classic N400 as well, but that it is not the only factor that determines the fit of incoming semantic information. [5]

Apart from this, the mere fact that scenario-based differences in support for particular words elicit a left-lateralized central negativity, instead of a classic N400 effect, may have an interesting additional implication for language research on (or with) the N400. Although researchers agree that the word-elicited N400 indexes how well the meaning of a word fits the constraints set by the context, there is disagreement over whether N400 context effects reflect contextual modulations of the ease of lexical information retrieval (Federmeier, Segal, Lombrozo, & Kutas, 2000), whether they instead directly reflect the compositional processes involved in making sense of language (i.e., semantic integration or 'unification', (Brown et al., 2000; Chwilla et al., 1995; Hagoort et al., 2004; Van Berkum, Brown et al., 2003)), or whether they perhaps necessarily reflect both because the two aspects cannot be meaningfully separated (Coulson & Federmeier, in press). It seems that only a perspective that includes message-level semantic integration as a critical component can account for our current findings. An account in which the word-elicited N400 purely indexes lexical retrieval (and contextual modulations thereof) cannot easily explain why, in our study, differences in scenario-based (or other word-based priming) support for a critical word do not modulate the classic, centroparietal N400.

One thing that is important to note is that it is difficult to separate scenario fit from simple "intra-lexical" word-word priming. In the example

[5] The fact that violations of message- and scenario-level constraints engage (at least partially) nonoverlapping neuronal systems might also be related to findings that the two hemispheres are differentially sensitive to the different levels of contextual constraint, as is apparent from divided visual field studies (Atchley, Burgess, & Keeney, 1999; Beeman, Friedman, Grafman, Perez, & et al., 1994; Chiarello, Liu, Quan, & Shears, 2000; Faust, 1998; Faust & Gernsbacher, 1996) and ERP research (Coulson, Federmeier, Van Petten, & Kutas, 2005).

story in Table 6.1, the scenario-biased context also contains words that in isolation would be a moderate to strong prime for the consistent word. Since it has often been shown that individually presented primes do not exert a facilitative effect over intervening words or longer SOA's (> 700 ms) (Masson, 1995; Neely, 1977, 1991; Ratcliff & McKoon, 1988, 1995), it seems unlikely that the N400 effect observed in natural language utterance (where related words are usually separated by time and intervening words) is dependent on strict single-word priming. Kutas and colleagues have shown that within a sentential context, incongruent words that are related to the contextually expected target also evoke an attenuated N400 (Federmeier & Kutas, 1999a, 1999b; Kutas & Hillyard, 1984). This was taken to imply that N400 amplitude serves as an index of relative priming by words present in context. However, taking into account the fuzzy boundary between word-word priming and scenario-based priming, this effect could just as well be ascribed to scenario-based priming. Previous research therefore does not support a conclusion that all effects of scenario-based priming are actually effects of simple lexical priming, or vice-versa. These two types of priming could very well be instantiations of the same underlying process. Both processes require that information in working memory activates dormant information in long term memory through previously established connections. In case of scenario priming, the information that generates the inference, and the knowledge that is in turn activated by that inference, is more elaborate than in word-word priming, but the underlying mechanisms of storage and retrieval seem closely related.

The role of background information in language comprehension

Our experiment suggests that both message- and scenario-based constraints determine the early integration of words within a discourse. This is consistent with a core assumption in several models of language comprehension, such as the Construction-Integration model (Kintsch, 1988; 1998), or the Scenario-Mapping and Focus theory (Sanford & Garrod, 1998). The assumption is that during comprehension, words are mapped onto prior discourse in two different ways: a relatively crude mapping onto potentially relevant scenarios and other static 'default' knowledge structures in long-term memory (e.g., via an automatic resonance process (Myers & O' Brien, 1998; Myers et al., 1994)), and a more precise mapping onto the actual meaning dynamically constructed for the discourse.

Our findings do not support the idea that scenario- and message-level information is (always) used in *consecutive stages* of integration, as suggested

by Sanford and Garrod (1998). Instead, our results suggest that people can use scenario- and message-related constraints equally rapidly. This implication of our findings must be treated with some caution, for we do not know whether the N400 evoked in the MS-condition is actually a result of an addition of a message-based N400 and the scenario-based negativity observed in the S-biased condition, or a 'pure' N400 effect elicited solely by message-based constraints. Determining this would require 'message-only' biased discourses without scenario-based or other lower-level forms of contextual support (which appear very difficult, perhaps even impossible, to create). Thus, our data do not allow us to infer whether message- and scenario-level information can actually be used *simultaneously*, or whether strong message-level constraint simply overrides the scenario-based information. What we can infer is that strong message-level expectations can be brought to bear on processing as rapidly as scenario-based constraints in the absence of strong message-level expectations.

Other ERP studies that focused on the interaction between contextual constraint and scenario-based (or word-word) priming have found mixed results. Evidence for the prevalence of message-level congruity over lexical association (Van Petten et al., 1999) has been found, as well as additivity of these two factors (Van Petten, 1993). In addition, as discussed in the introduction, Hoeks et al (2004) found that scenario-related information can overrule the actual message of the discourse. In a recent review Ledoux et al. (2006) state that the exact interaction of message-level information and lexical association depends on the amount of constraint imposed by the message. In a constraining context, effects of lexical association are overridden by message-level effects, whereas in a less constraining discourse the effects of lexical association are present. The present results suggest that for scenario and message information this could also very well be the case. The data presented by Hoeks et al. (2004) furthermore suggest that when the scenario constraint is particularly high, the effect of message-level information is temporarily diminished. Recent N400 evidence that listeners momentarily fail to notice a coherence break when the anomalous word is scenario-relevant (Nieuwland & Van Berkum, 2005) also points in that direction. Taken together, the extant ERP data suggest that both types of information, scenario-based (or based on lexical association) and message-based, are taken into account when processing new information, and that the amount of influence each type of information exerts seems to depend on the relative amount of constraint it receives from the context.

Conclusion

Our first conclusion is that discourse-based N400 effects cannot completely be attributed to scenario fit (or other lower-level priming mechanisms). The actual message of the broader discourse plays a significant role in the early processing of incoming information. Second, message-level fit does not seem to be the only factor influencing early processing of words in coherent text. When the message of the discourse does not generate strong expectations, a clear differential effect of scenario fit shows up, as a left-dominant negativity around the same time as the discourse-based N400 effect. Thus, both the actual message of the context and the scenario suggested by the 'bag of words' in the context can affect early processing in the N400 latency range, as people read a coherent text. Third, the different scalp distributions of the ERP effects observed in the scenario-biased and the message- & scenario-biased condition suggest that different constellations of neural generators are involved when incoming information mismatches constraints based on message-level information as opposed to when the constraints are based on induced scenario's. It is clear, from our own experience, that readers nearly always extract the correct message from a discourse and our results show that message-dependent processing begins extremely rapidly. At the same time, however, basic scenarios play their automatic part in the early stages of language processing, irrespective of that message.