Available online at www.sciencedirect.com
SCIENCE DIRECT
Cognitive Brain Research 17 (2003) 701–718
COGNITIVE BRAIN RESEARCH
www.elsevier.com/locate/cogbrainres

Research report

When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect

Jos J.A. van Berkum[a,b,c,*], Pienie Zwitserlood[d], Peter Hagoort[b,c], Colin M. Brown[b]

[a]*Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands*
[b]*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*
[c]*F.C. Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands*
[d]*Westfälische Wilhelms-Universität Münster, Germany*

Accepted 25 June 2003

Abstract

In two ERP experiments, we assessed the impact of discourse-level information on the processing of an unfolding spoken sentence. Subjects listened to sentences like *Jane told her brother that he was exceptionally quick/slow*, designed such that the alternative critical words were equally acceptable within the local sentence context. In Experiment 1, these sentences were embedded in a discourse that rendered one of the critical words anomalous (e.g. because Jane's brother had in fact done something very quickly). Relative to the coherent alternative, these discourse-anomalous words elicited a standard N400 effect that started at 150–200 ms after acoustic word onset. Furthermore, when the same sentences were heard in isolation in Experiment 2, the N400 effect disappeared. The results demonstrate that our listeners related the unfolding spoken words to the wider discourse extremely rapidly, after having heard the first two or three phonemes only, and in many cases well before the end of the word. In addition, the identical nature of discourse- and sentence-dependent N400 effects suggests that from the perspective of the word-elicited comprehension process indexed by the N400, the interpretive context delineated by a single unfolding sentence and a larger discourse is functionally identical.

Theme: Neural basis of behavior

Topic: Cognition

Keywords: Spoken language comprehension; Sentence interpretation in discourse context; Local and global meaning; Word recognition; N400; ERP

1. Introduction

How do people come to understand a sentence? One step towards comprehension is that the meanings of the individual words are related to each other in a way that respects the syntactic and, for spoken language, the prosodic structure of the sentence at hand. However, sentences are usually part of a wider discourse, and therefore need to be interpreted in the context of what has been said before. In the research reported below, we examine when and how listeners bring their knowledge about the prior discourse to bear on the interpretation of an unfolding spoken sentence.

Until the early 1990s, the dominant model of language comprehension was that of a stage-like process in which the reader or listener more or less sequentially worked through the various levels of language structure that linguistics had discovered (e.g. [16,19]): recognize the sublexical units, recognize the words, parse the sentence, interpret the sentence on its own terms, and finally compute its implications in terms of the wider discourse. Two things seemed to follow from this view. One is the local meaning hypothesis, the idea that people initially compute some sort of local, context-independent meaning of the sentence at hand before relating it to the prior discourse of which it is a part (e.g. [52]; see [23,24] and [8], for critical discussion). Second, and related, that discourse-associated processing should occur relatively

*Corresponding author. University of Amsterdam, Department of Psychology (PN) Roetersstraat 15, 1018 WB Amsterdam, The Netherlands. Tel.: +31-20-525-6921; fax: +31-20-639-1656.
E-mail address: berkum@psy.uva.nl (J.J.A. van Berkum).

0926-6410/03/$ – see front matter
doi:10.1016/S0926-6410(03)00196-4

'late' (on some relevant time scale), an idea that we will refer to as the late discourse hypothesis. One variant of this hypothesis, which can be traced back to early sentence processing models (e.g. [17]), is that incoming sentence input is related to the wider discourse only at the ends of major constituents. A perhaps more subtle variant is that sentential input and discourse-level representations do make contact at every word coming in, but that in terms of the various processes elicited by any given word, those that relate to discourse-level information occur relatively late, after the lexical, syntactic and sentence-level semantic aspects of the word at hand have been dealt with (see [18], [20] or [43], for discussion of this type of processing architecture).

In the last decade, this classic sequential-analysis model has been complemented by an important alternative view, often associated with a connectionist processing architecture (e.g. [42,58]), but not necessarily so (e.g. [31,34]; see also [43] for an early account along these lines). The essence of this alternative is that the comprehender is assumed to solve a constraint satisfaction problem at several 'levels' of language-related representation simultaneously, looking for the best interpretation of the incoming string of words not only in terms of their overt (and hence strongly constraining) orthographic or phonological form characteristics, but also—at the same time—in terms of their syntactic, semantic, and referential specifications (see [31,32] for an illuminating account). Because it is not easily formalized, discourse-level information is usually not taken into account by the existing models of this sort (but see [35]). However, the most natural prediction of such models is that discourse context can in principle affect the comprehension of an unfolding sentence extremely rapidly. In part, this follows from another assumption that is easily made within a simultaneous constraint satisfaction approach, which is that there is no fundamental processing distinction between the interpretation of incoming words within their local sentence context ('local meaning') and within the context of the wider discourse.

In two ERP experiments with written language materials [63], we recently obtained evidence that seemed to be at odds with the local meaning hypothesis as well as nontrivial variants of the late discourse hypothesis. Our starting point was that semantically anomalous words in single isolated sentences (e.g. *He spread the warm bread with* ***socks***) reliably elicit an N400 effect ([40], see [6,7,38,39,41] and [49] for reviews), which demonstrates that readers and listeners immediately relate the incoming words to a semantic representation of the preceding local sentence context. The question we addressed in our initial written-language experiments [63] was whether words that are fully compatible with the local sentence itself but anomalous with respect to the wider discourse would also elicit this effect. Native speakers of Dutch were asked to read (the Dutch equivalents of) sentences like *Jane told her brother that he was exceptionally quick/slow*, designed such that the alternative critical words were equally acceptable when the sentence was presented in isolation, but presented in the context of a larger discourse that rendered one of the critical words incoherent (e.g. because Jane's brother had in fact done something very quickly).

The results were very clear. Relative to a discourse-coherent counterpart (e.g. *quick*), discourse-anomalous words (e.g. *slow*) elicited a large standard N400 effect, i.e. a monophasic negative shift in the ERP that began at about 200–250 ms after word onset and peaked around 400 ms, with a centro-parietal maximum. As the effect revealed effects of discourse-level information within a mere 200–250 ms after word onset, regardless of whether the word at hand was in the middle or at the end of a sentence, we took this as evidence against the gist of the late discourse hypothesis. Furthermore, because the ERP effect of a discourse-dependent semantic anomaly emerged equally rapidly and was qualitatively identical to the standard N400 effect observed in response to a 'sentence-internal' semantic anomaly, this suggested, in contrast to what the local meaning hypothesis would lead one to expect, that from the perspective of the word-elicited comprehension process indexed by the N400, the interpretive context delineated by a single unfolding sentence and a larger discourse may well be functionally identical.

These findings replicated and extended the results of a pioneering written-language study conducted by St. George, Mannes, and Hoffman several years before [55]. St. George et al. asked their subjects to read a number of short stories supplied with or without a title. Although each story was locally coherent in that its individual sentences were interconnected and related to a single topic, it was very difficult to find out what that topic actually was if one had not been given the title as well. When the subjects read stories of each type in the ERP experiment, content words in stories without a title turned out to elicit larger N400s than the same words in stories that did have a title. Although the design of St. George et al. did not allow them to relate the N400 effect to specific critical words and to draw the associated timing inferences, their ERP results clearly demonstrated that the N400 is sensitive to discourse-level manipulations.

In the two ERP experiments reported below, we aimed to obtain discourse-dependent N400 effects in spoken language comprehension. One of our goals was to address an important concern that is frequently raised over the results of our 1999 written-language ERP studies (a concern that also holds for the St. George et al. written-language experiment). To avoid eye movement artifacts, we had presented the written sentences word by word at a fixed rate of 600 ms/word, with each word centered on a computer screen. The concern that is very often raised over this so-called serial visual presentation (SVP) procedure is that the results might hinge on the unnatural way in which language is being presented. In the present study, we directly address this concern. Subjects listened to fully

natural spoken versions of the 1999 written-language materials. If we obtain the same effect with natural speech input, then we clearly have more than an SVP-induced artefact to consider.

In addition to examining the robustness of earlier results, the present replication study also allowed us to have a closer look at when and how the processing of an unfolding spoken sentence is affected by discourse-level information. Spoken language is the archetypical form of language, and experiments with natural spoken language must therefore play a vital role in the study of language comprehension. More importantly, the use of spoken language actually has particular significance for the question we address, the timing of discourse-level involvement in sentence comprehension. The reason is, that in contrast to a written word presented all at once on a display, a spoken word itself takes time to unfold. The use of spoken language therefore allows us to begin to relate the timing of a discourse-dependent N400 effect to how the lexical signal itself unfolds in time.[1]

2. Experiment 1 (sentences in discourse)

2.1. Method

Apart from the use of a spoken version of the materials, the method of Experiment 1 is identical to that of its written-language predecessor [63].

[1]Following up on the 1999 written-language study [63], Salmon and Pratt [51] have recently also examined the ERP effects of sentence- and discourse-dependent anomalies with spoken-language stimuli. Unfortunately, the design of their study prevents us from accepting the data as informative. Salmon and Pratt had each of their subjects listen to a block of coherent and anomalous discourses, followed by a block of coherent and anomalous isolated sentences. However, subjects heard each of the four variants of a single item. Elaborating on the example item given in [51], a subject would thus hear (a) ''My computer system suddenly broke down. I'm glad I had backup disks. Fortunately, I didn't lose any **files**.'' and (b) ''My computer system suddenly broke down. I'm glad I had backup disks. Fortunately, I didn't lose any **friends.**'' in the first block (in this or the reverse order), later followed by (c) ''Fortunately, I didn't lose any **friends**'' and (d) ''Fortunately, I didn't lose any **rainbows**'' (in this or the reverse order) in the second block. Our main concern is with this repetition of materials, particularly since it is systematically confounded with critical factors. With the isolated sentences always following the stories, for instance, 'rainbows' in anomalous sentence (d) has never been heard before in this carrier phrase, whereas 'friends' in coherent sentence (c) has been presented in the same carrier phrase once before (in story b). Furthermore, when presented in the isolated sentences block, a carrier phrase such as ''Fortunately, I didn't lose any . . .'' may well have reminded listeners of the discourse context in which it had been presented twice before in the stories block, a repetition effect that would invalidate the story–sentence comparison. The fact that the ERP subjects were asked to also evaluate whether the last (i.e. ERP-critical) word of the story or sentence was semantically appropriate is not unlikely to have increased the impact of repetition.

2.1.1. Subjects

We recruited 24 right-handed native speakers of Dutch (19 female subjects, mean age 23, range 19–36 years) from our local subject pool. None had any neurological impairment, had experienced any neurological trauma, or had used neuroleptics. Also, none of them had participated in the written-language studies or its pretests.

2.1.2. Materials

The critical items were 80 short Dutch stories that described a wide variety of easily imaginable situations and events. Each story consisted of three sentences, of which the third, the carrier sentence, contained a critical word (CW). For each story, two CW alternatives were available: a discourse-coherent CW that was a good continuation of the earlier discourse, and a discourse-anomalous CW that did not continue the discourse in a semantically acceptable way. The difference in coherence between two CWs usually hinged on considerable inferencing about the discourse topic and the situation it described. An example follows below, with the coherent and anomalous CWs in boldface (coherent first), and the approximate English translation in brackets; the complete set of materials can be obtained from the first author.

(3) *Zoals afgesproken zou Jane om vijf uur 's ochtends haar zus en haar broertje wakker maken. Maar de zus had zich al gewassen, en het broertje had zich reeds aangekleed. Jane vertelde het broertje dat hij bijzonder* ***vlot*****/*****traag*** *was.*

(*As agreed upon, Jane was to wake her sister and her brother at five o'clock in the morning. But the sister had already washed herself, and the brother had even got dressed. Jane told the brother that he was exceptionally* ***quick*****/*****slow***.)

To avoid a sentential confound, each discourse-anomalous CW was chosen such that within the local carrier sentence it was roughly as acceptable as its discourse-coherent CW counterpart (see [63] for two pretests relevant to this criterion). Also, neither of the CWs was used in the preceding context. The discourse-coherent and -anomalous CWs were equated on word class and inflection (e.g. both were plural nouns) within each story, and matched across the entire set of stories on average orthographic length (7.7 and 7.5 letters, respectively, with no CW over 10 letters) and word frequency (46 and 42 occurrences per million, respectively, using the 42 million word CELEX written wordform corpus). Of the 80 carrier sentences, 35 had sentence-final CWs and 45 had sentence- (and clause-) medial CWs.

To assess the level of semantic constraint provided by our critical discourse contexts, we posthoc conducted a story completion ('cloze') test in which we truncated the 80 critical stories (as well as the 80 critical sentences used in Experiment 2) just before the critical word, and asked 24 native speakers of Dutch who had not participated in any of our EEG experiments to complete the text in a

natural way. As expected, 0% of the respondents continued our stories with the discourse-anomalous word (e.g. *traag [slow]* in the above example). Averaged across stories, the discourse-coherent word (e.g. *vlot [quick]*) was supplied by 18% of the respondents, with story-specific coherent word response percentages or 'cloze probabilities' ranging from 0 to 92%.

For the spoken-language study, the carrier sentences and their discourse contexts were recorded separately with a normal speaking rate and intonation, by the same female native speaker. A trained phonetician identified the acoustic onsets (and offsets) of the critical words in their carrier sentences. The average critical word lasted 558 ms (range 185–879 ms) in the discourse-anomalous CW condition, and 549 ms (range 190–874 ms) in the discourse-coherent CW condition.

Two different stimulus lists were used, each for half of the subjects. For the first list, 40 discourse-coherent and 40 discourse-anomalous critical story trials were pseudo-randomly mixed with 160 comparable filler story trials such that neither coherent nor anomalous CW trials occurred more than four times consecutively, and such that trials of each type were matched on average list position. The second list was derived from the first by replacing all discourse-coherent CWs by their anomalous counterparts and vice versa. Subjects always heard just one version of an item.

2.1.3. Procedure, EEG recording and analysis

After electrode application, subjects sat in a sound-attenuating booth and listened to the stimuli over headphones. They were told that EEG recording would only occur as they heard the last sentence of a story, and that during recording they should avoid all movement and fixate on an asterisk displayed on the screen before them. Subjects were asked to process each story for comprehension. No additional task demands were imposed.

Each trial consisted of a 300-ms auditory warning tone, followed by 700 ms of silence, the spoken discourse context, 1000 ms of silence, and the spoken carrier sentence. The 1000 ms separating the carrier sentence from its context did not perceptually break the story into two parts, and approximated natural pausing times measured between sentences within a context. To inform subjects when to fixate and sit still for EEG recording, an asterisk was displayed from 1000 ms before onset of the carrier sentence to 1000 ms after its offset. After a short practice, the trials were presented in five blocks of 15 min, separated by rest periods.

The EEG was recorded from 13 tin electrodes in an electrode cap, each referred to the left mastoid. Three electrodes were placed according to the international 10–20 system over midline sites at Fz, Cz, and Pz locations. Ten electrodes were placed laterally over symmetrical positions: left and right frontal (F7, F8), anterior temporal (LAT, RAT, halfway between F7–T3 and F8–T4, respectively), temporal (LT, RT, laterally to Cz, at 33% of the interaural distance), temporo-parietal (LTP, RTP, posterior to Cz by 13% of the nasion–inion distance, and laterally by 30% of the interaural distance each), and occipital (LO, RO, halfway between T5–O1 and T6–O2, respectively). Vertical eye movements and blinks were monitored via a supra- to suborbital bipolar montage. A right to left canthal bipolar montage was used to monitor for horizontal eye movements. Activity over the right mastoid bone was recorded on an additional channel to determine if there were differential contributions of the experimental variables to the two presumably neutral mastoid sites (no such differential effects were observed). The EEG and EOG recordings were amplified with Nihon Kohden AB-601G bioelectric amplifiers, using a high cutoff of 30 Hz and a time constant of 8 s. Impedances were kept below 3 kΩ for the EEG electrodes and below 5 kΩ for the EOG electrodes. The EEG and EOG signals were digitized on-line with a sampling frequency of 200 Hz, and screened off-line for eye movements, muscle artifacts, electrode drifting, and amplifier blocking in a critical window that ranged from 150 ms before to 1200 ms after acoustic onset of the critical word. Trials with such artifacts (6.0%) were rejected.

For each subject, average waveforms were computed across all remaining trials per condition after normalizing the waveforms of the individual trials on the basis of the 150 ms pre-CW baseline. Subsequent analyses of variance (ANOVAs) used mean amplitude values computed (for each subject and condition) in the standard N400 latency range of 300–500 ms after onset of the CW. Two supplementary latency ranges, 150–300 ms and 500–700 ms, were used to complement the standard latency range analysis whenever appropriate. Univariate F-tests with more than one degree of freedom in the numerator were adjusted by means of the Greenhouse–Geisser/Box's epsilon hat correction. All results were first evaluated in an omnibus ANOVA that crossed the coherence factor (anomalous, coherent) with a 13-level electrode factor. The scalp distribution of the coherence effect was subsequently explored in two separate ANOVAs, one with a three-level midline-electrode factor (Fz, Cz, Pz), and the other with a hemisphere (left, right) by lateral-electrode (F7/F8, LAT/RAT, LT/RT, LTP/RTP, LO/RO) design.

2.2. Results

Fig. 1 displays the grand average ERPs time-locked to the acoustic onset of discourse-anomalous and discourse-coherent spoken critical words. As is often the case with fully connected speech input, there are no clear exogenous components in these ERPs. However, there is a clear differential effect of discourse coherence. Relative to their discourse-coherent counterparts, discourse-anomalous CWs elicited a large and widely distributed negative deflection that emerged in the grand average at about 150–200 ms

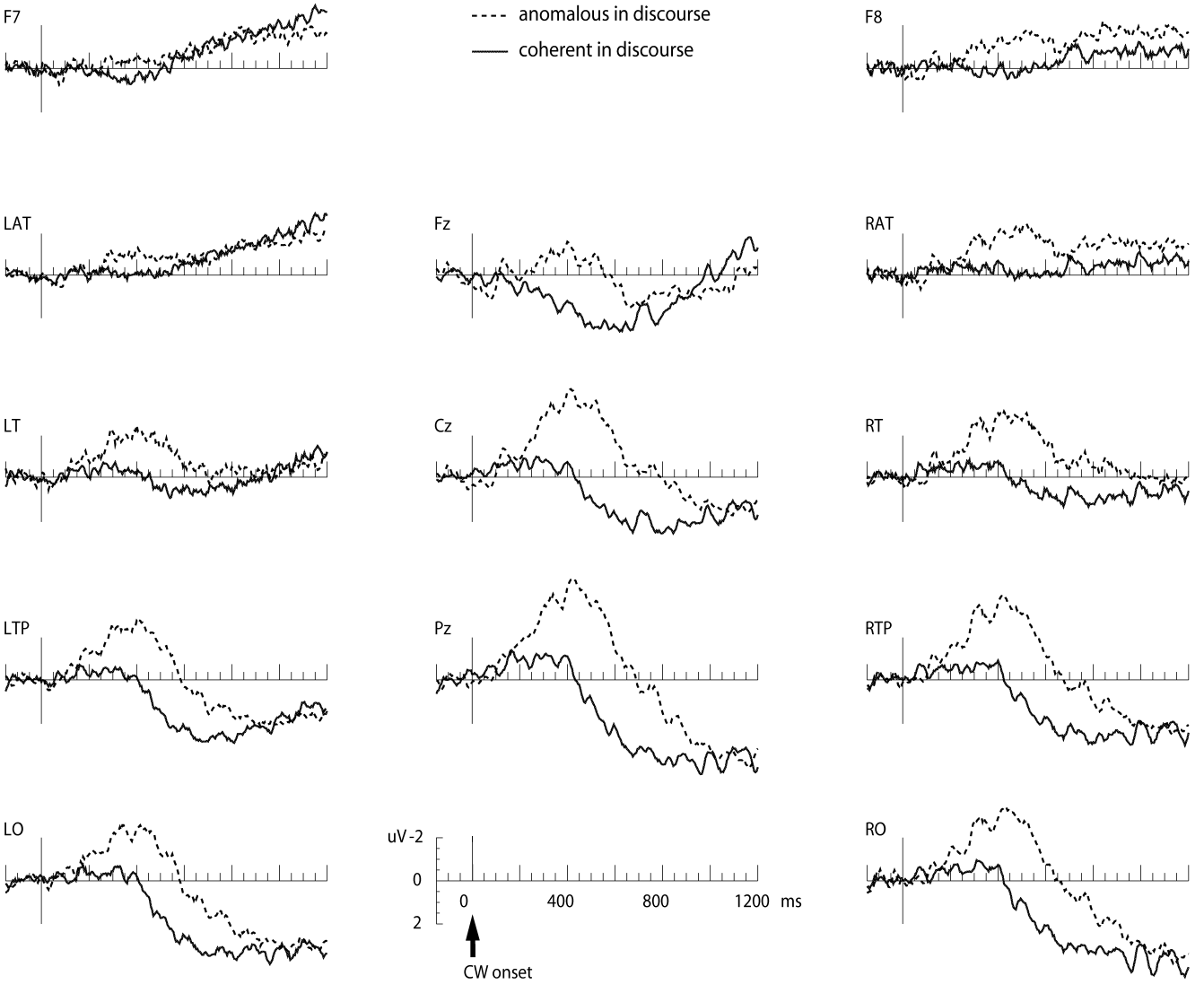

Fig. 1. Discourse-dependent semantic anomaly effects. Grand average ERPs elicited by spoken words that were semantically coherent (solid line) or anomalous (dotted line) with respect to the discourse context (Experiment 1). In this and all following figures, negativity is plotted upwards, and the acoustic onset of the critical word (CW) is at 0 ms. LAT, RAT = left, right anterior temporal, LT, RT = left, right temporal, LTP, RTP = left, right temporo-parietal, LO, RO = left, right occipital.

after their acoustic onset, peaked at about 400 ms, lasted for about 800–1000 ms, and reached its maximum over centro-parietal scalp sites. In view of these characteristics, it can without any doubt be qualified as a standard N400 effect.

Using mean amplitude in the standard 300–500 ms latency range used for testing N400 effects, the overall Coherence (2) × Electrode (13) analysis of variance (ANOVA) revealed a main effect of Coherence [$F(1,23) = 66.37$, MSe = 11.50, $P < 0.001$, corresponding to a 2.21 μV mean amplitude effect], as well as a Coherence × Electrode interaction [$F(12,276) = 7.37$, MSe = 1.28, $P = 0.001$]. An additional Coherence (2) × Electrode (3) ANOVA of the midline sites also revealed a Coherence main effect [$F(1,23) = 46.33$, MSe = 7.34, $P < 0.001$], but no clear Coherence × Electrode interaction [$F(2,46) = 2.92$, MSe = 1.31, $P = 0.086$]. The Coherence (2) × Hemisphere (2) × Electrode (5) ANOVA of the left- and right-lateral sites yielded a Coherence main effect [$F(1,23) = 68.99$, MSe = 6.64, $P < 0.001$] and a Coherence × (anterior-to-posterior) Electrode interaction [$F(4,92) = 9.17$, MSe = 1.67, $P = 0.003$], but no clear Coherence × Hemisphere interaction [$F(1,23) = 3.13$, MSe = 2.84, $P = 0.090$], nor a Coherence × Hemisphere × Electrode interaction [$F(4,92) = 0.71$, MSe = 0.18, $P = 0.522$].

We conducted onset analyses by testing the Coherence main effect in consecutive mean amplitude latency bins of 10 ms wide (e.g. 100–110 ms, 110–120 ms, etc.). The first significant main effect of Coherence was observed in the 200–210 ms latency range [$t(23) = -2.09$, $P = 0.047$ two-tailed], followed by a long and uninterrupted row of significant effects starting at 230–240 ms [$t(23) = -2.12$,

$P = 0.045$]. A marginally significant effect emerged at the intervening 210–220 ms interval [$t(23) = -1.98$, $P = 0.060$], although not at the 220–230 ms interval [$t(23) = -1.60$, $P = 0.124$].

The waveforms in Fig. 1 suggest that the N400 effect might be preceded by a smaller negative deflection in the 100–150 ms latency range. Comparable negativities have occasionally been reported to precede the N400 effect with spoken anomalous words [9,26,65]. In the ERPs computed across all 80 critical stories, however, there was no reliable main effect of Coherence in this latency range [$F(1,23) = 1.57$, MSe = 12.71, $P = 0.223$], nor any interaction with Electrode site [$F(12,276) = 0.37$, MSe = 0.86, $P = 0.815$]. We return to a relevant reanalysis of the data below.

2.3. CW position and CW length

The N400 effect of discourse coherence in Experiment 1 was obtained with a set of 80 items of which 35 had the critical word in sentence-final position, and 45 in sentence-(as well as clause-) medial position. As shown in the upper panel of Fig. 2 for Pz, a clear discourse-dependent N400 effect emerged for critical words in either position. A mean amplitude ANOVA in the 300–500 ms latency range revealed that Coherence did not significantly interact with Word Position [$F(1,23) = 1.07$, MSe = 30.12, $P = 0.312$]. Separate mean amplitude ANOVAs in the 300–500 ms latency range for each of the two CW types revealed a significant Coherence simple main effect both for sentence-final CWs [$F(1,23) = 22.86$, MSe = 43.48, $P < 0.001$] and for sentence-medial CWs [$F(1,23) = 39.66$, MSe = 13.93, $P < 0.001$]. Most relevant, the latter effect suggests that the rapid impact of discourse does not occur only at the end of a sentence or major clause, as part of some 'sentence-final wrap-up', but also occurs as these major units of language input are still unfolding.

The acoustic length of our critical words ranged from 185 to 897 ms across the entire set. In principle, the early onset of the average discourse-dependent N400 effect could thus solely reflect item-specific N400 effects elicited by a few very short critical words. To examine this, we divided the critical trials into those with a 'short' CW (up

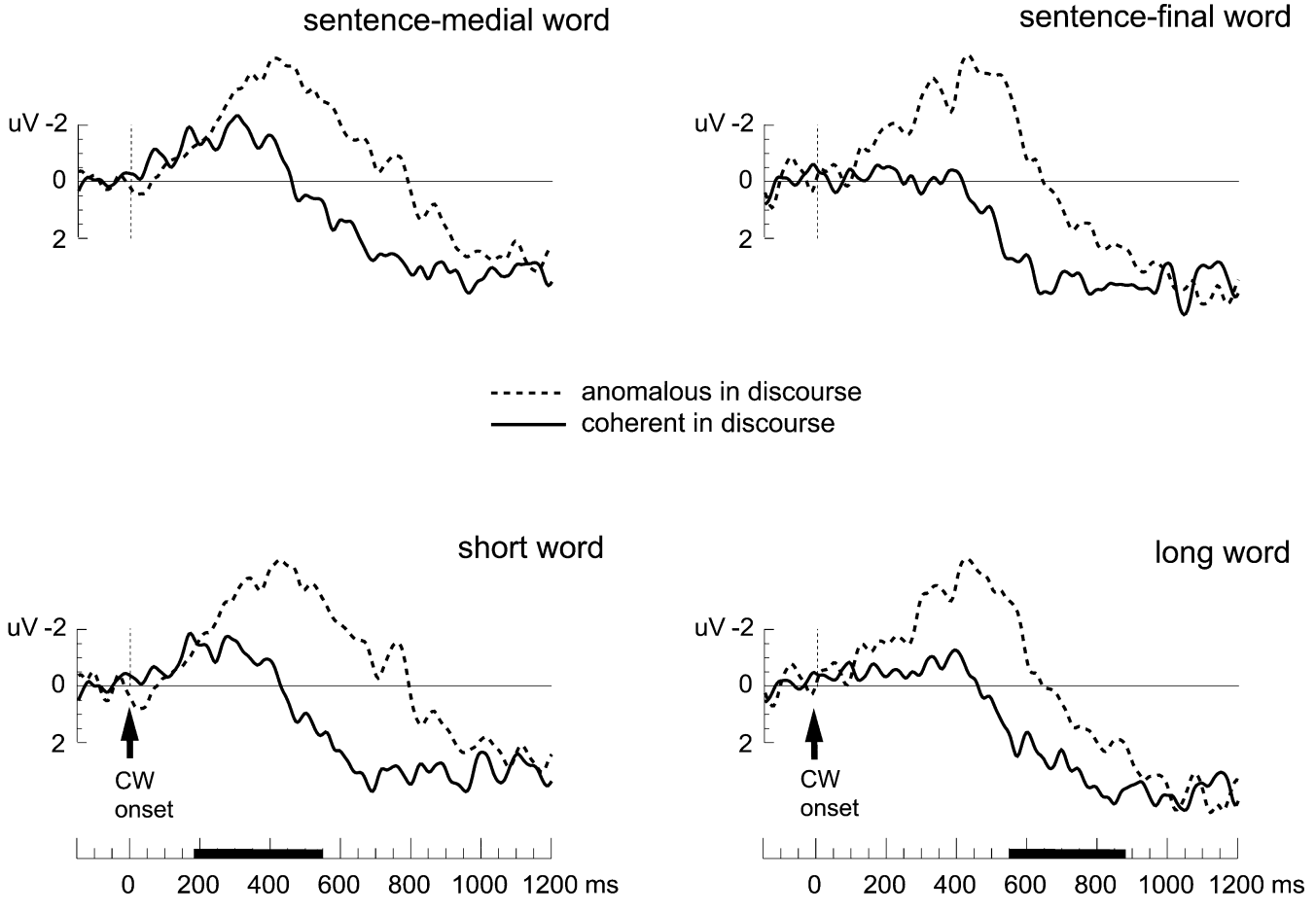

Fig. 2. Discourse-dependent semantic anomaly effects by word position and length. Grand average ERPs, at Pz, elicited by spoken words that were coherent (solid line) or anomalous in discourse (dotted line), for words in sentence-medial (top left) and sentence-final (top right) position, as well as for short words (bottom left; below 550 ms, and a mean duration of 451 ms) and long words (bottom right; over 550 ms, and a mean duration of 652 ms). Thick horizontal bars mark the range of acoustic word offsets for short and long critical words, respectively.

to 550 ms, average length 451 ms across 78 CWs) and those with a 'long' CW (over 550 ms, average length 652 ms across 82 CWs). As shown for Pz in the lower panel of Fig. 2, essentially the same N400 effect is obtained for both. A mean amplitude ANOVA in the 300–500 ms did not reveal a significant Coherence×Word Length interaction [$F(1,23)=0.02$, MSe=28.84, $P=0.887$], and simple effects tests revealed a significant discourse-dependent N400 effect for short words [Coherence: $F(1,23)=29.94$, MSe=23.85, $P<0.001$], as well as for critical words of at least 550 ms acoustic length [Coherence: $F(1,23)=23.52$, MSe=27.93, $P<0.001$]. As for the latter set of long words, *t*-tests in consecutive mean amplitude latency bins of 10 ms wide revealed an initial significant main effect of Coherence in the 200–210 ms latency range [$t(23)=-2.07$, $P=0.050$ two-tailed, and flanked by marginally significant effects in the immediately adjacent latency ranges], followed by a more sustained significant Coherence effect beginning in the 280–290 ms latency range [$t(23)=-3.52$, $P=0.002$].

2.4. Low- and higher-constraint stories

Although the 80 critical stories were on average only moderately constraining (with coherent words having an average cloze value of 18%), the item set did contain stories for which the coherent critical word was highly predictable (with item-specific cloze values of up to 92%; see Methods section). To rule out the possibility that the discourse-dependent N400 effect shown in Fig. 1 hinged on a subset of highly constraining stories only, we selectively re-averaged the EEG data for low-constraint critical stories. These were 33 stories for which the coherent word had a very low predictability, with no cloze probability above 5%, and a mean cloze value of only 1%. In these low-constraint items, the anomalous word hence did not disconfirm a strong expectation for the coherent word. As displayed in the left panel of Fig. 3, such discourse-anomalous words still elicited a large N400 effect [corresponding to a 1.9 μV mean amplitude effect in the 300–500 ms latency range; $F(1,23)=40.10$, MSe=14.71, $P<0.001$]. This reveals that the discourse-dependent N400 effect does not hinge on a disconfirmed strong lexical prediction.[2] Along the way, it also suggests that the effect at hand is not likely to derive from a difference in cloze probability between coherent and anomalous critical words (averages of 1 and 0% in this reanalysis, respectively).

The right panel of Fig. 3 displays the results for the remaining 47 stories. Because the coherent words that contribute to this complementary dataset have cloze probabilities ranging from 5 to 92% (mean cloze probability 30%), we label the set of stories involved as 'higher-constraint' only. In these 47 higher-constraint stories, discourse-anomalous words elicited a large N400 effect as well, corresponding to a 2.5 μV mean amplitude effect size in the 300–500 ms latency range [$F(1,23)=28.44$, MSe=34.18, $P<0.001$]. Although numerically somewhat larger, the latter N400 effect did not differ statistically from the N400 effect elicited in low-constraint stories [$F(1,23)=0.81$, MSe=29.44, $P=0.378$].

In higher-constraint stories, the discourse-dependent N400 effect is preceded by a short-lived negativity in the 100–150 ms latency range [$F(1,23)=4.92$, MSe=24.30, $P=0.037$], most prominent at Fz, Cz, and the right-sided electrode sites F8, RAT, RT, and RTP. The effect was not observed at discourse-incoherent words in low-constraint stories [$F(1,23)=0.61$, MSe=41.30, $P=0.442$]. Because of its selective occurrence, it is tempting to relate this early negativity to previously reported ERP effects of a mismatch between context-based phonological or acoustic expectations and the actual input, such as the MMN [47], the PMN [9], or the N200 effect [65]. However, in view of the relatively low signal-to-noise ratio involved in this posthoc reanalysis, we believe that such discussion must await a convincing replication (with materials dedicated to this issue).

2.5. Spoken versus written language

We computed average difference waveforms for the current spoken-language effect as well as for its written-language counterpart ([63], Experiment 1), each computed by subtracting the ERPs elicited by discourse-coherent CWs from those elicited by discourse-anomalous CWs. The two resulting net N400 effects—obtained with the same set of coherent and anomalous stories—are displayed in Fig. 4.

A joint Coherence (2)×Electrode (13)×Modality (2, spoken vs. written language input) ANOVA on mean amplitudes in the 300–500 ms latency range revealed a main effect of Coherence [$F(1,46)=88.87$, MSe=15.17, $P<0.001$], but no Coherence×Modality interaction [$F(1,46)=0.37$, MSe=15.17, $P=0.548$], nor a Coherence×Electrode×Modality interaction [$F(12,552)=0.47$, MSe=1.15, $P=0.730$]. Thus, within the standard latency range used for testing N400 effects, the discourse-dependent N400 effect obtained with fully connected speech input was statistically indistinguishable from the equivalent effect obtained with written-language serial visual presentation input in terms of its size as well as its scalp distribution.

An inspection of Fig. 4 clearly suggests that the spoken- and written-language N400 effects do differ in both an earlier and a later latency range, as well as in their approximate peak latency. Modality differences outside of

[2] In principle, some of the critical stories might be relatively predictive with respect to a word other than the two critical words used in the ERP experiment. However, because both of the critical words would then disconfirm the specific lexical prediction, this cannot explain the differential N400 effect at hand.

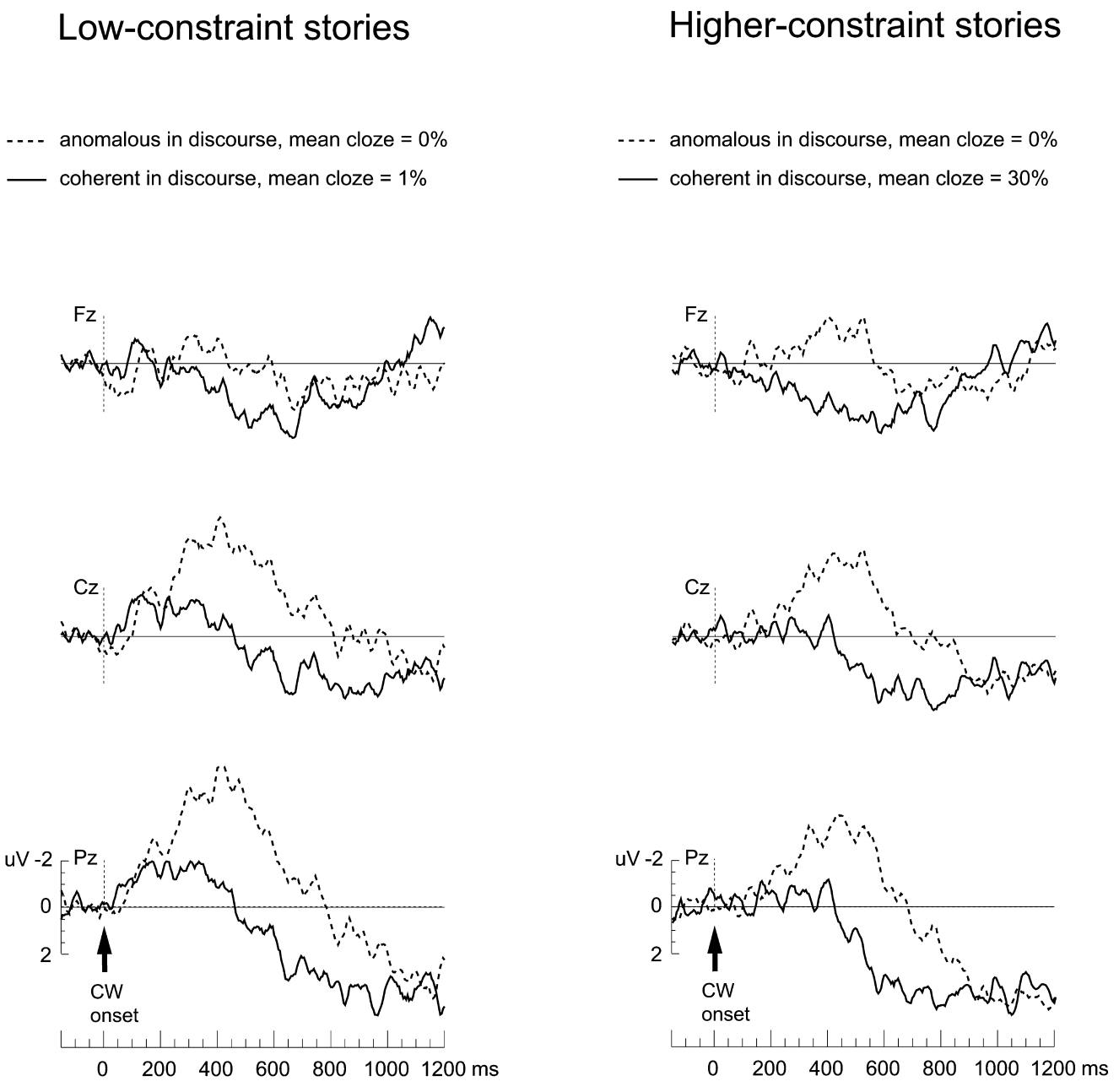

Fig. 3. Discourse-dependent semantic anomaly effects in low- and higher-constraint stories. Grand average ERPs elicited at three midline sites by spoken words that were semantically coherent (solid line) or anomalous (dotted line) with respect to a discourse context in which the coherent word had a very low cloze probability (left panel) or not (right panel). See text for further explanation.

the 300–500 ms latency range were evaluated in posthoc ANOVAs conducted on mean amplitude values in the 150–300 ms and 500–700 ms latency ranges. In the early 150–300 ms range, spoken-language presentation resulted in a Coherence main effect [$F(1,23)=7.99$, MSe = 9.42, $P=0.010$], but written-language presentation did not [$F(1,23)=0.19$, MSe = 8.36, $P=0.671$; data from [63]]—a combined analysis in this early latency range yielded a significant Coherence×Modality interaction [$F(1,46)=5.54$, MSe = 8.89, $P=0.023$]. A similar pattern was observed in the late 500–700 ms latency range. Here, spoken-language presentation again resulted in a Coherence main effect [$F(1,23)=37.82$, MSe = 20.22, $P<0.001$] whereas written-language presentation again did not [$F(1,23)=1.50$, MSe = 43.27, $P=0.233$; data from [63]]—the corresponding Coherence×Modality interaction was again significant [$F(1,46)=6.05$, MSe = 31.75, $P=0.018$].

As reported before, *t*-tests in consecutive mean amplitude latency bins of 10 ms wide revealed the earliest significant Coherence effect to emerge in the 200–210 ms latency range for spoken-language input. An equivalent analysis of the written-language data (not reported in [63])

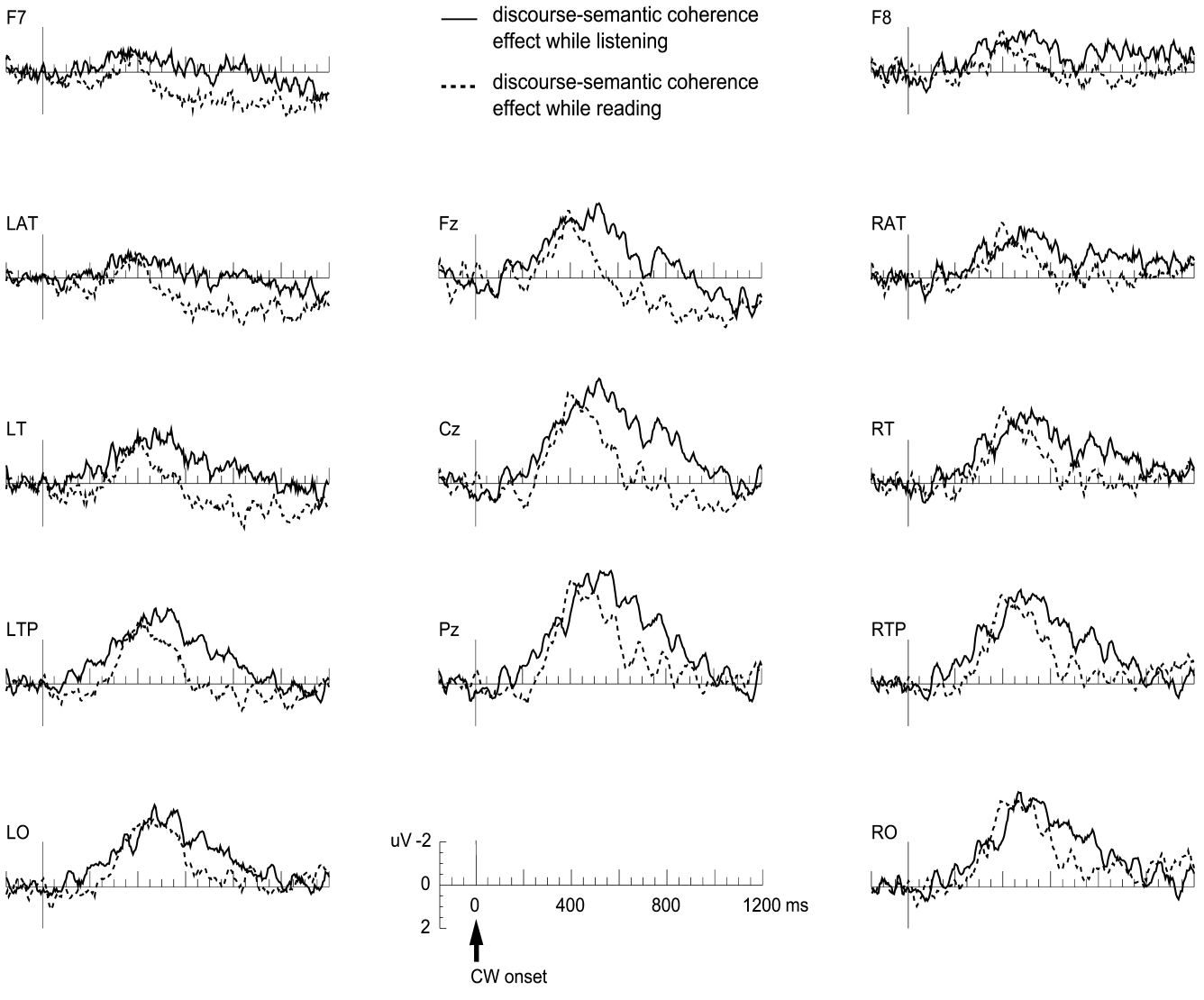

Fig. 4. Net discourse-dependent semantic anomaly effects in spoken and written language. Discourse-semantic coherence effects (anomalous–coherent difference waveforms) in spoken (solid line, Experiment 1) and written language comprehension (dotted line; data from [63], Experiment 1).

revealed the earliest significant Coherence effect to emerge in the 280–290 ms latency range [$t(23) = -2.20$, $P = 0.038$]. Although onset analyses with a 10-ms bin width have only limited power, the observed difference in when the earliest significant effect shows up does confirm what Fig. 4 suggests, namely that relative to their written-language counterparts, discourse-anomalous spoken words elicit an earlier processing effect in ERPs.

2.6. Discussion

In Experiment 1, subjects listened to short stories, of which the last sentence occasionally contained a word that was locally coherent but anomalous with respect to the wider discourse. Relative to a discourse-coherent control word, these discourse-anomalous words elicited a large N400 effect, i.e. a monophasic negative shift in the ERP that began at about 150–200 ms after acoustic word onset and peaked around 400 ms, with a centro-parietal maximum. Also, as with the written-language effect, a discourse-dependent N400 effect could be observed for spoken critical words in sentence-final as well as sentence-medial position. In spite of a radically different mode of language presentation, we have thus replicated all the critical written-language findings reported before in [63].

In addition, a posthoc inspection of the data revealed that a discourse-dependent ERP effect not only emerged very early for words of relatively short duration, but also for spoken words of relatively long duration (all over 550 ms long). Reanalysis of the data obtained with stories in which the coherent critical word had a very low cloze probability (all below 5%, and 1% on average) revealed that the differential ERP effect elicited by a discourse-anomalous word also did not hinge on the disconfirmation of a strong lexical expectation.

As revealed in Fig. 4, the discourse-dependent spoken-

language effect has an earlier onset, a later peak, and a later offset than the equivalent written-language effect. These differences may have arisen for a number of reasons. First, there is inevitable variance in the precise determination of where a spoken word begins. To the extent that a critical ERP effect is time-locked to word-onset (or to some word-specific delay relative to word onset, e.g. a uniqueness point), this variability in onset measurement will result in (additional) 'smearing', i.e. broadening of the grand average ERP. In addition, due to general and language-specific constraints on pronunciation, phonemes of the preceding word can contain subtle cues concerning the initial phonemes of the critical word. To the extent that listeners use these cues, analysis procedures that average the EEG signals relative to the onset of the critical word's very own first phoneme (like the procedure currently used) may in fact be slightly biased. A third issue to consider is that, although the critical words used in the spoken- and written-language study were identical, the temporal parameters of their presentation are of course radically different. In particular, whereas a written word is presented all at once, a spoken word needs time to unfold. Although this difference could be taken to suggest a later spoken-language effect across the board (and hence be inconsistent with the observed earlier onset of this effect), it is really not known how this difference plays out in comprehension. After all, our own ERP findings, as well as earlier behavioral results [43], show that listeners by no means need to have heard the complete word to start relating it to the interpretive context. In all, a wide variety of factors may underlie the somewhat different time-courses of the spoken- and written-language N400 effects.

Although onset latency differences in the order of 50–100 ms are non-negligible and as such merit attention, we actually find it remarkable how similar the ERP response of the comprehension system is, given that the average critical word is instantaneously visible in the written-language study, but takes about 550 ms to completely unfold in the present spoken-language experiment. The fact that we replicate the N400 effect reported in [63] with natural speech input demonstrates that the written-language effect did not depend on the use of relatively slow serial visual presentation. However, with the critical words unfolding in such radically different ways across the two studies, the remarkable similarity in the timing of these ERP effects not only validates an earlier SVP-based dataset, but also testifies to a rather surprisingly invariant system response across input modality.

The discourse-dependent interpretation of the obtained N400 effect relies on the assumption that across the set of 80 experimental items, the alternative critical words (e.g. *quick* and *slow*) are on average equally acceptable within the context provided by the local carrier sentence alone (e.g. *Jane told the brother that he was exceptionally* —). However, although the carrier sentences had been designed to achieve this, an empirical validation of this assumption is in order. In Experiment 2, we therefore presented the same critical words in just their local carrier sentence, i.e. without the prior discourse. If the N400 effect obtained in Experiment 1 really is a discourse-dependent effect, it should vanish if the discourse is taken away.

3. Experiment 2 (sentences in isolation)

3.1. Method

Apart from the use of a fully connected speech version of the materials, the method of Experiment 2 is also identical to that of its written-language predecessor ([63] Exp. 2).

3.1.1. Subjects

Experiment 2 was conducted with 16 right-handed native speakers of Dutch (14 female, mean age 22, range 19–27 years) recruited from the same subject pool. None had any neurological impairment, had experienced any neurological trauma, or used neuroleptics. Also, none of them had participated in the spoken main experiment or any of the preceding written-language studies.

3.1.2. Materials

In this control experiment, we presented all 80 critical carrier sentences of Experiment 1 in isolation, i.e. without their two-sentence discourse context. In the present context, we refer to these as formerly discourse-anomalous or formerly discourse-coherent carrier sentences, depending on their relation to the discourse context of Experiment 1. As these two sets of sentences had been designed to be equally coherent in isolation and were therefore predicted to elicit identical ERPs at the critical word, Experiment 2 also included a sensitivity check to make sure that the expected null result could not be attributed to, for instance, an inattentive group of subjects. The sensitivity check consisted of 80 Dutch sentences containing a local, sentence-dependent semantic anomaly (e.g. the Dutch equivalent of *Gloomily the men stood around the* ***pencil*** *of the president*), and 80 derived sentences in which the sentence-anomalous critical word was replaced by a sentence-coherent word (*Gloomily the men stood around the* ***grave*** *of the president*, see [63] for details). The average critical word lasted 443 ms (range 246–663 ms) in the sentence-anomalous CW condition, and 419 ms (range 200–735 ms) in the sentence-coherent CW condition. When completing truncated versions of these sentences (e.g. *Gloomily the men stood around the . . .*) in a posthoc cloze test, 0% of the respondents continued with the sentence-anomalous critical word (e.g. *pencil*), and 43% [range 0–100%] continued with the sentence-coherent critical word (e.g. *grave*).

Two different stimulus lists were used in Experiment 2, each for half of the subjects. For the first list, 40 formerly

discourse-anomalous carrier sentences, 40 formerly discourse-coherent carrier sentences, 40 sentences with a sentence-anomalous critical word, and 40 matched sentences with a sentence-coherent critical word were pseudo-randomly mixed with 280 filler sentences such that no sentence of any critical type occurred more than four times consecutively, and such that sentences of each type were matched on average list position. The second stimulus list was derived from the first by replacing all coherent (or formerly coherent) CWs by their anomalous (or formerly anomalous) counterparts and vice versa.

3.1.3. Procedure, EEG recording and analysis

The procedure and EEG handling was identical to that in Experiment 1, apart from the presentation of isolated sentences instead of mini-discourses. Each isolated-sentence trial began with a 300-ms warning beep, followed after 1200 ms of silence by a single spoken sentence. To help subjects avoid eye movements, a fixation asterisk was displayed on a computer screen from 1000 ms before sentence onset to 1000 ms after sentence offset. The next trial began 2500 ms after sentence offset. As in Experiment 1, the EEG and EOG signals were screened off-line for eye movements, muscle artifacts, electrode drifting, and amplifier blocking in a critical window that ranged from 150 ms before to 1200 ms after acoustic onset of the critical word. Trials containing such artifacts were rejected (6.7% of the carrier sentences, and 7.0% of the sentences with a locally anomalous or coherent word).

3.2. Results

Fig. 5 displays the grand average ERPs time-locked to the acoustic onset of the formerly discourse-anomalous and

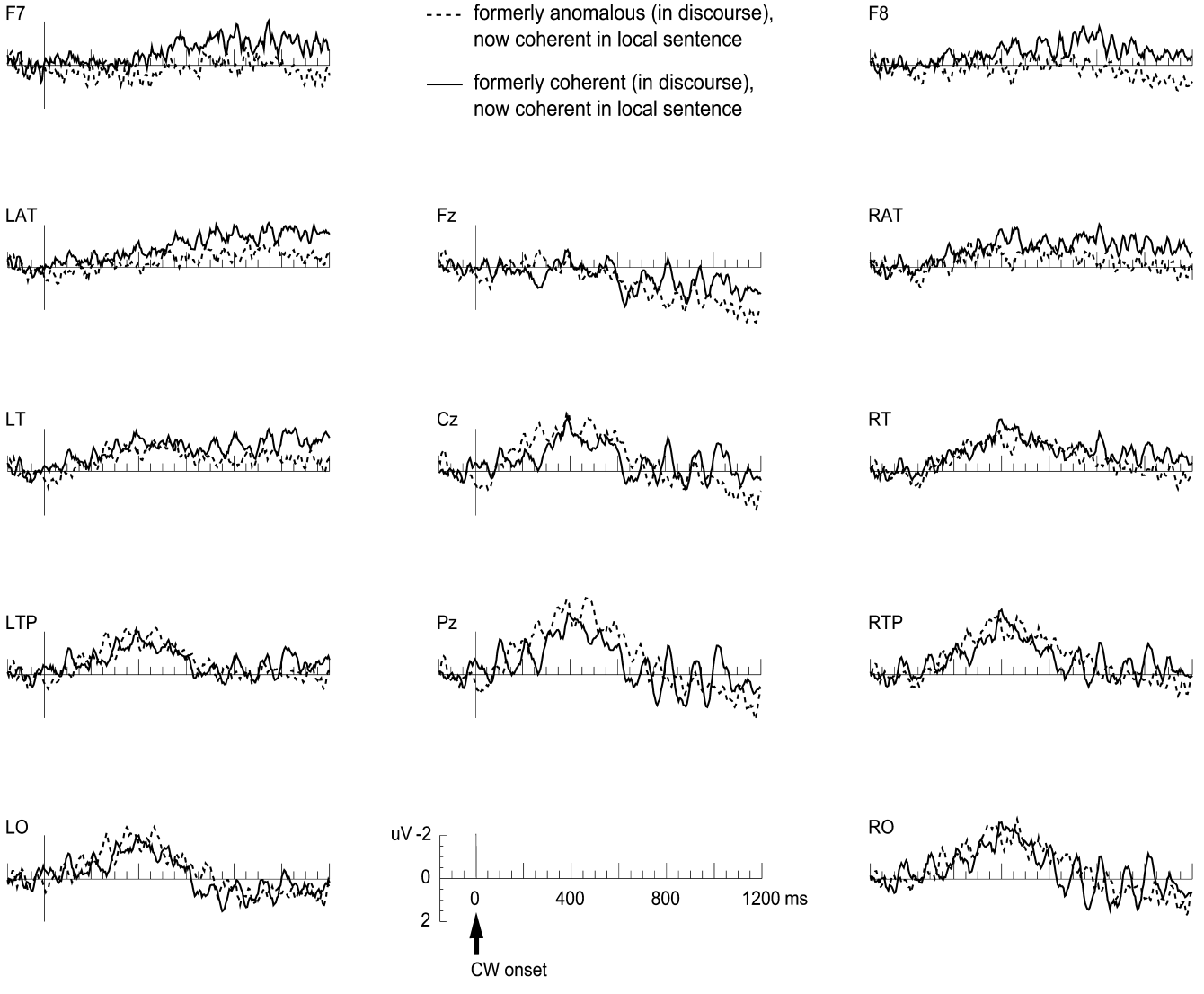

Fig. 5. Results for critical sentences without the biasing discourse. Grand average ERPs elicited by spoken words that were semantically coherent (solid line) or anomalous (dotted line) with respect to the original discourse context, but have now been presented in their local carrier sentence only (Experiment 2).

-coherent critical words, still embedded in their original carrier sentence, but now presented without the preceding discourse context. Again, there are no clear exogenous components in either of the waveforms. Also, both CW types elicit a clear N400 component in the average ERP, with a centro-parietal maximum, a slight right-over-left asymmetry, and a peak at about 400 ms after acoustic word onset. Critically, though, there is no obvious amplitude difference between the N400 elicited by formerly discourse-anomalous and formerly discourse-coherent CWs.

The overall Former-Coherence (2)×Electrode (13) ANOVA on mean amplitude in the 300–500 ms latency confirmed that there was no statistically significant difference [$F(1,15)=0.16$, MSe = 10.34, $P=0.699$], nor a significant interaction with Electrode [$F(12,180)=2.00$, MSe = 1.04, $P=0.146$]. There were no significant effects involving Former-Coherence in the ANOVA of the midline sites either (Former-Coherence: $F(1,15)=0.35$, MSe = 7.21, $P=0.564$; Former-Coherence×Electrode: $F(2,30)=1.09$, MSe = 1.52, $P=0.325$), nor in the ANOVA of the left- and right-lateral sites [Former-Coherence: $F(1,15)=0.94$, MSe = 5.72, $P=0.349$; Former-Coherence×Hemisphere: $F(1,15)=0.05$, MSe = 1.56, $P=0.832$; Former-Coherence×Electrode: $F(4,60)=3.15$, MSe = 1.21, $P=0.085$; Former-Coherence×Hemisphere×Electrode: $F(4,60)=0.46$, MSe = 0.13, $P=0.658$].

The absence of a differential N400 effect in Fig. 5 can clearly not be attributed to insensitivity of the present ERP experiment, for as shown in Fig. 6, the sentence-dependent semantic anomalies in this experiment elicited a solid N400 effect. Using mean amplitude in the standard 300–500 ms latency range, the overall Coherence (2)×Electrode (13) ANOVA revealed a main effect of Coherence [$F(1,15)=14.99$, MSe = 16.94, $P=0.002$, corre-

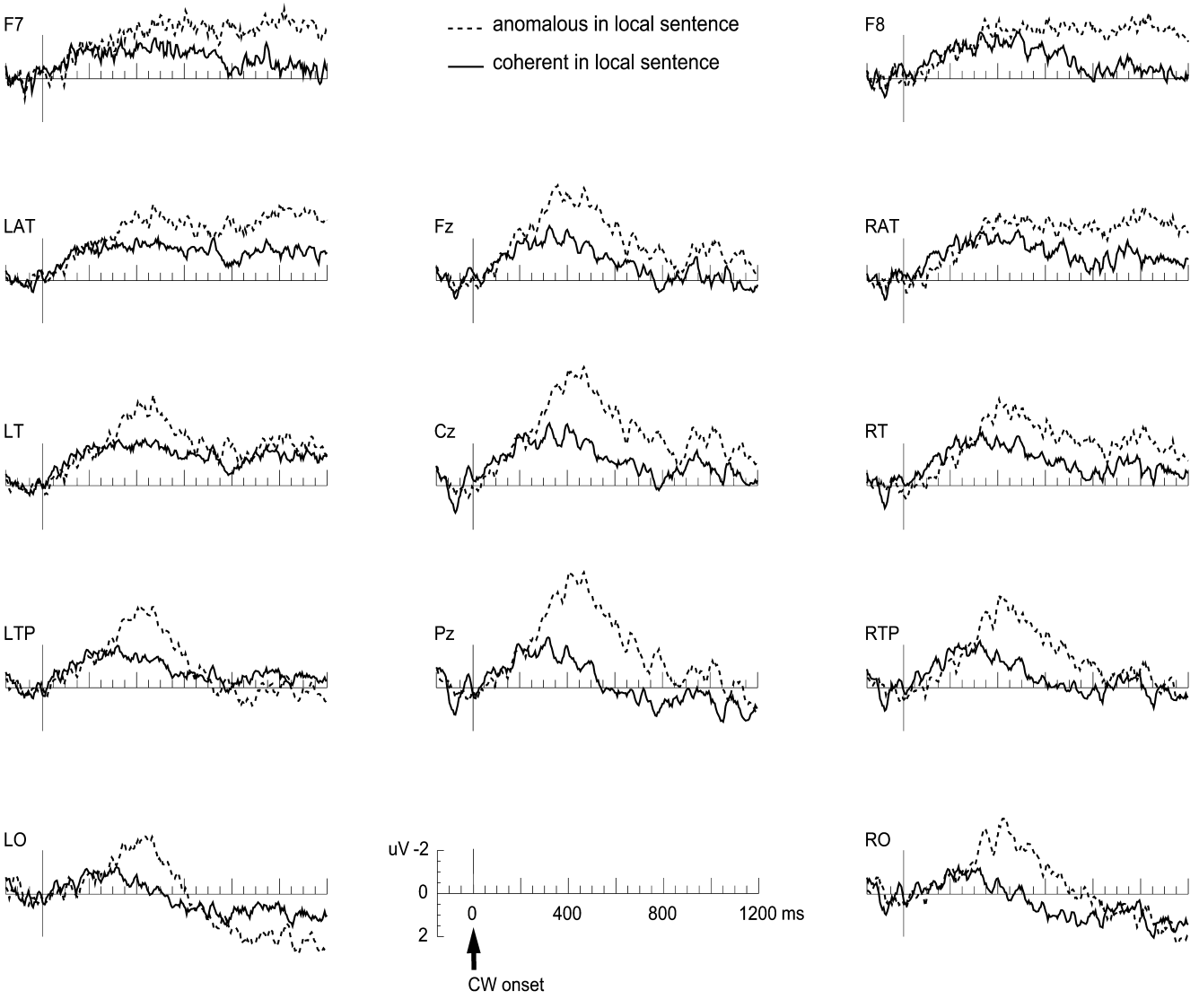

Fig. 6. Sentence-dependent semantic anomaly effects. Grand average ERPs elicited by spoken words that were semantically coherent (solid line) or anomalous (dotted line) with respect to the local sentence context (Experiment 2).

sponding to a 1.56 μV mean amplitude effect], as well as a Coherence×Electrode interaction [$F(12,180)=5.13$, MSe=0.89, $P=0.003$]. The Coherence (2)×Electrode (3) ANOVA of the midline sites also revealed a Coherence main effect [$F(1,15)=17.93$, MSe=8.10, $P=0.001$], but no clear Coherence×Electrode interaction [$F(2,30)=2.95$, MSe=0.91, $P=0.091$]. The Coherence (2)×Hemisphere (2)×Electrode (5) ANOVA of the left- and right-lateral sites yielded a Coherence main effect [$F(1,15)=12.98$, MSe=10.31, $P=0.003$] and a Coherence×(anterior-to-posterior) Electrode interaction [$F(4,60)=5.50$, MSe=0.97, $P=0.022$], but no Coherence×Hemisphere interaction [$F(1,15)=0.03$, MSe=2.45, $P=0.876$], nor a clear Coherence×Hemisphere×Electrode interaction [$F(4,60)=2.80$, MSe=0.27, $P=0.078$]. As in the written-language study, the scalp distributions of the discourse- and sentence-dependent N400 effects were statistically indistinguishable: Coherence (2)×Electrode (13)×Anomaly-dependence (2, discourse- vs. sentence-dependent) $F(12,456)=0.76$, MSe=1.13, $P=0.515$.

3.3. Discourse-embedded versus isolated critical sentences

We combined the data of the critical sentences that had been presented both in a discourse context in Experiment 1 and without their original context in Experiment 2 in a joint Coherence (2)×Electrode (13)×Context (2, with or without discourse) ANOVA on mean amplitudes in the 300–500 ms latency range. As suggested by the difference between Figs. 1 and 5, this revealed a substantial Coherence×Context interaction [$F(1,38)=30.85$, MSe=11.04, $P<0.001$], which did not depend on site (Coherence×Context×Electrode: $F(12,456)=0.85$, MSe=1.19, $P=0.465$).

3.4. Discussion

The result of Experiment 2 is unambiguous. When presented within the original carrier sentence but without the prior discourse, the ERP elicited by formerly discourse-anomalous critical words (e.g. *slow* in *Jane told the brother that he was exceptionally* ***slow***) no longer differed from the ERP elicited by formerly discourse-coherent critical words (e.g. *quick*). Also, the presence of a solid sentence-dependent N400 effect revealed that the experiment was sensitive to manipulations of semantic coherence. Together, these results confirm that the N400 effect obtained in Experiment 1 hinged on how the critical words related to the wider discourse, and not on unintended differences in their fit to the local sentence context.

4. General discussion

We conducted two ERP experiments to assess the impact of discourse-level information on the comprehension of an unfolding spoken sentence. Relative to coherent control words, spoken words that did not fit the wider discourse elicited a large standard N400 effect (Experiment 1). Furthermore, when we removed the discourse and presented the same words within their local carrier sentence context only (Experiment 2), the differential N400 effect disappeared, confirming that it indeed hinged on how the words at hand related to the prior discourse.

The pattern of results obtained with spoken language input is essentially identical to that obtained with written-language versions of the same materials [63]. In line with several other studies that elicited equivalent ERP effects with spoken versus serially presented written sentences (e.g. [15,26,27,37]), therefore, the earlier written-language ERP effect reported in [63] (as well as, by inference, the N400 effect reported in [55]) does not appear to depend on the use of a serial visual presentation procedure or a relatively slow rate of presentation. More interestingly, our findings show (a) that listeners very rapidly bring their knowledge about the discourse to bear on the processing of an unfolding spoken sentence, and (b) that the semantic comprehension process indexed by the N400 is indifferent to where the semantic constraints originally came from. We discuss each of these two implications in turn.

4.1. Early contact between discourse and the unfolding acoustic signal

The rapid discourse context effects observed with spoken-language input are inconsistent with the gist of the late discourse hypothesis, in two different ways (see [59] for a principled discussion in terms of 'word-elicited impulse responses'). First, as shown in Fig. 2, discourse-incoherent words elicit a clear N400 effect not only in sentence-final position, but also in sentence- (and clause-) medial position. This demonstrates that sentence processing is incremental 'all the way up', i.e. that, as a sentence unfolds, every individual word coming in is immediately related to the overall discourse (cf. [43]). Second, our results clearly contradict the notion that on any given word that does make contact with discourse information, the establishment of that contact occurs 'relatively late' (on some as yet to be determined relevant time scale). As revealed by Fig. 1, the processing consequences of a spoken word that does not fit the wider discourse begin to show up in ERPs at about 150–200 ms after its onset. A comparison to Fig. 6 reveals that this is by no means later than the processing consequences of a spoken word that does not fit the 'local' sentence context (see also [26,65,67]). Furthermore, the discourse-dependent semantic ERP effect emerges in the same latency range as where the first ERP effects of syntactic processing show up (sometimes referred to as early left anterior negativities or ELAN effects; see [21] and [28] for reviews). For various reasons, these ERP-based observations do not rule out the possibility of a principled delay in which the information associated with individual words makes contact with

discourse-level representations.[3] The comparison does suggest, however, that if such a delay exists, it is not a very large one. That is, discourse is not that late at all.

The early impact of discourse-level conceptual information observed here and in our previous written-language study [63] accords well with other reports of rapid discourse context effects in language comprehension (e.g. [1,2,5,10,13,14,22,30,43,46,50,53,55,58,60–62,64]). If readers or listeners encounter a singular definite noun for which the earlier discourse had introduced two equally suitable referents (e.g. *the girl* in a story about two girls), for example, the processing consequences of the referential ambiguity begin to emerge in the ERP at about 300 ms after onset of the head noun [60,62], with syntactic parsing decisions being affected at the very next word [60,61]. More generally, our demonstration of rapid discourse context effects clearly converges with demonstrations of the rapid impact of nonlinguistic context, obtained in so-called visual-world experiments (e.g. [3,4,11,12,57]).

Perhaps the most striking aspect of the present spoken-language result is what it tells us about the speed with which discourse information contacts an incoming spoken word. On average, the discourse-anomalous critical words in Experiment 1 took about 550 ms to unfold. In spite of this, Fig. 1 reveals that the spoken word input is making contact with discourse-level information by about 150–200 ms after the acoustic onset of these words. Also, this early onset does not hinge on just the shortest words in our materials, for as shown in the lower right panel of Fig. 2, a similarly early ERP effect emerges in the average waveforms computed over a subset of words of at least 550 ms duration (and a mean duration of 652 ms). Thus, in spoken-language comprehension, discourse does not only come into play 'very rapidly', but can do so long before a word has been fully heard. In fact, with an average phoneme length of about 80 ms across our set of anomalous critical words (average CW duration of 558 ms, divided by an average 6.7 phonemes per CW), the results suggest that the unfolding words in our study are related to discourse-level representations within the first two or three phonemes only.

Two ERP studies with sentence-semantic anomalies [67,66] have recently examined the onset of the N400 effect relative to the critical word's so-called isolation point, the point in the word's acoustic signal at which most listeners have heard enough of the word to be able to uniquely identify it (if presented in isolation). In both studies, the N400 effect actually began to emerge well before this isolation point. Basically, this suggests that listeners relate the incoming acoustic signal to the semantic context delineated by the unfolding sentence not just before they have heard the complete word, but before they can actually know exactly what the unfolding word itself is going to be. As such, the ERP evidence is consistent with earlier chronometric data obtained by Zwitserlood [69], which showed that sentence-semantic information differentially affected the activation level of word candidates before the acoustic input itself had uniquely specified a word.

In the psycholinguistic literature, very early contacts between lexical input and interpretive context are usually conceptualized as context effects in word recognition (e.g. [67,69,70]). Our rapid discourse-dependent ERP effect could also be conceptualized as such. Note that listeners will usually need less acoustic signal to detect that an unfolding word is not going to fit the context ('discrepancy point', cf. [67]) than they would need to identify the word in isolation (uniqueness or isolation point). In the example item shown before, for instance, the first three phonemes of the discourse-anomalous word *slow* (i.e, the acoustic equivalent of *slo_*) do not reliably identify the word as *slow*, and also allow for words like *slope*, *slogan*, *slow-motion*, and *slow-witted*. However, to the extent that none of the words that are still compatible with the acoustic input fits the semantic context (or at least does not do so straightforwardly), the word recognition system may at this point detect that something odd is about to show up.[4] Several models of spoken word recognition [44,45,48] allow one to formulate early context effects in exactly this way.

By definition, however, the discourse-dependent N400 effect hinges on lexical and higher-level interpretive processing. Functionally speaking, the immediate neural generator of this effect therefore need not participate in 'low-level' word recognition processes at all, but may instead play a role in the construction of a high-level conceptual interpretation for the unfolding message. This may sound unlikely, for the effects of discourse- (and sentence-) semantic anomaly emerge well before a word has been fully heard, and (as established for sentence-semantic context, [67,66]) possibly even before the word

[3]For one, the delay might be extremely small. Also, it is perhaps unlikely that the handful of ERP effects currently recognized as relevant to language comprehension (e.g. N400, P600/SPS, LAN) together reflect all of the functionally distinct processes that make up the language comprehension system.

[4]Because our materials had not been designed with a spoken-language replication in mind, they do not lend themselves to a useful ERP analysis relative to word isolation (or uniqueness) points. However, a posthoc analysis of our stories revealed that in approximately 45 of the 80 cases, it was impossible to come up with a contextually reasonable continuation after having heard just the first two phonemes of the anomalous critical word. Thus, as they unfolded, more than half of the anomalous critical words had already ruled out all contextually reasonable candidate words at either the first or the second phoneme. With an average phoneme length of approximately 80 ms, this suggests that the unfolding acoustic signal of over half of our anomalous words ruled out all contextually appropriate lexical alternatives somewhere in the first 80–160 ms from acoustic word onset. Although the latter is a coarse estimate only (and one that ignores the potential impact of earlier co-articulatory cues), it is interestingly compatible with the observed early onset of the discourse-dependent N400 effect.

has become unique. However, we cannot rule out that for an as yet undetermined subset (one, some, all) of the lexical candidates being activated by an unfolding but as yet non-unique spoken word, the associated semantic information is rapidly made available to higher-level interpretive processing. If this occurs, it implies that extremely early sentence- and discourse-dependent N400 effects can in principle also arise at the level of meaning construction.

This ambiguity in the interpretation of our ERP data echoes an unresolved issue in the wider N400 literature. Research on the N400 has led to a fairly general consensus that within the language domain this ERP component reflects some aspect(s) of the processes that relate the meaning of a particular word to a higher-order semantic interpretation of the unfolding message [7]. However, there is as yet no consensus on whether those processes should be viewed as part of a word recognition system that is somehow sensitive to higher-level semantic context, or as part of a sentence- and discourse-level semantic integration system that rapidly incorporates the semantics released by complete or unfolding words (nor whether this is really the right question to ask; see [31] [32] for an account in which the distinction seems to blur).

Whatever the ultimate answer, though, the present findings, as well as other spoken-language data [67,69], clearly suggest some form of continuous mapping between the unfolding acoustic signal on the one hand and the interpretive context on the other. Because discourse-anomalous words can elicit an N400 effect relative to coherent but unexpected alternatives (Fig. 3, left panel), the continuous mapping mechanism involved does not appear to operate by rapidly matching the incoming acoustic signal to an expectation-based phonological template prepared in advance for a single strongly expected coherent word. That is, the system is somehow able to very rapidly detect a mismatch between the unfolding acoustic signal and the discourse context even if the latter does not allow the listener to predict and simply check, at some phonological level, for the presence of a particular word.

4.2. Functional equivalence of sentence- and discourse-semantic context

Our ERP findings also bear on the precise relation between sentence- and discourse-dependent interpretation. According to the local meaning hypothesis (e.g. [52]), the language comprehension system initially computes some sort of local, context-independent meaning of the sentence at hand, before relating it to the prior discourse of which it is a part. Under this two-stage model of interpretation, one might expect the ERP effect of a discourse-dependent semantic anomaly to be later than or qualitatively different from the ERP effect of a sentence-dependent semantic anomaly. However, neither turns out to be the case. A discourse-dependent semantic anomaly elicits the exact same ERP effect as a sentence-dependent semantic anomaly, without any noticeable delay.

Research on the N400 effect has consistently shown that this ERP effect is not a simple reflection of semantic anomaly, but instead reflects some nontrivial aspect of normal language comprehension [7,41]. For example, another discourse-level EEG experiment has recently shown that a coherent spoken word that is relatively unpredictable in the discourse elicits a larger N400 than a much more predictable coherent word [64]. In the light of such findings, the observed equivalence of sentence- and discourse-dependent N400 effects we now report—although obtained via semantic anomalies—can be taken to suggest that it is the same normal comprehension process running into trouble in either context.

We propose a very simple explanation for this equivalence (cf. [63]). If within the language domain the N400 can be taken to reflect some aspect(s) of the processes that relate the meaning of a particular word to the interpretive context, the only additional assumption needed to account for our findings is that it apparently does not matter whether the interpretive context involved was provided by the first few words of a single unfolding sentence or, say, a 500-page novel. One way to achieve this is to invoke the notion of a ‘situation model’ (e.g. [25,33,35,68]), the comprehender’s mental representation of the state of affairs described in a text. If, as seems plausible, a situation model can be generated not only on the basis of a large piece of prior discourse, but also on the basis just the first few words of a single unfolding sentence, then the equivalence of sentence- and discourse-induced N400 effects can be taken to reflect something about the fit between an incoming word and a situation model generated by earlier (single- or multi-sentence) linguistic input.

A second, slightly different way to conceive of the interpretive context for an incoming word is to equate it with the common ground (e.g. [54,8]), the knowledge base that listeners and speakers mutually assume to have in common, and which provides the background against which all linguistic utterances are to be understood (and composed). Because the common ground for a linguistic exchange not only includes shared knowledge of the evolving discourse (including a mutually assumed situation model), but also, for instance, of the language and its conventions, the physical setting in which the current discourse takes place, the culture that the interlocutors are assumed to belong to, and the world in general, it constitutes an interpretive context that is very different from a strict situation model. However, as with a situation model, it is easy to see that a common ground context would in actual processing completely ‘absorb’ the difference between information that happens to be provided by an extensive preceding discourse and that happens to be provided by just the first few words of a single unfolding sentence. Whether the interpretive context involved in the processes generating the N400 is indeed to be equated with

either the situation model or the common ground remains to be established.

At this point, it is important to make clear that we do not argue that the sentence domain is irrelevant to interpretation. As illustrated by the fact that *I just saw him pick up your camera!* conveys something quite different from *I just saw him. Pick up your camera!,* the sentence is a primary device for partitioning and organizing a linguistic message (see [20] and [56] for thorough discussions). Also, the syntactic structure of an unfolding sentence imposes important local constraints both on how the message so far is to be understood and on how the next word is to be interpreted. What our findings can be taken to suggest, however, is that such local constraints immediately merge with the conceptual constraints imposed by the widest available interpretive domain. One way in which the system might achieve this is by immediately computing contextually enriched sentence meaning (which is then subsequently, but incrementally, used to augment the discourse representation with). Alternatively, sentence-level interpretive constraints might act as real-time constraints on how incoming words directly augment the discourse representation, without the computation of an explicit intermediate sentence-level semantic representation. Either way, what matters is that computed meaning is never context-free ('local meaning'), but is immediately formulated within the widest interpretive context available.

Finally, what do the present results tell us about the functional interpretation of the N400? As mentioned before, our data suggest that the comprehension process generating the N400 is sensitive to the degree of semantic friction between an incoming word and the relevant interpretive context, regardless of whether the latter was delineated by a larger piece of discourse or just the first part of an unfolding isolated sentence. Earlier work [36] had already shown that the N400 also does not seem to care all that much about whether the interpretive context for some word was an unfolding isolated sentence or a single prime word. At the other side of the contextual spectrum, recent work by Hagoort, Hald, and Petersson [29] has shown that the N400 does not care about whether the effective interpretive context involves relatively general 'semantic' knowledge (which renders a sentence like *The city of Rome is very **thin*** anomalous) or very specific 'pragmatic' knowledge about the real world (rendering *The city of Rome is very **new*** anomalous). This could be taken to suggest that in language comprehension, the interpretive context involved in the generation of N400 effects can be anything as long as it has meaning and as such provides a background to human communication.

5. Conclusion

The results of our ERP study support a number of conclusions. First, and in line with earlier written-language findings [55,63], the evidence suggests that the incoming words of an unfolding spoken sentence make contact with 'global' discourse-level semantic information in a way that is indistinguishable from how they make contact with 'local' sentence-level semantic information. The equivalent impact of sentence- and discourse-semantic contexts on the word-elicited N400 suggests that the process reflected by this ERP component basically doesn't care where the semantic context originally came from, and simply evaluates the incoming words relative to the widest interpretive domain available. Second, the process at hand also does not appear to depend much on whether the incoming word is a written one presented instantaneously or a spoken one taking half a second or more to unfold. For a system dealing with meaning, it would of course be unreasonable to expect radically different modes of operation for spoken and written language input. However, and perhaps counterintuitively, a discourse-dependent anomaly can be detected at least as fast with a spoken word unfolding gradually over time as with a written word displayed at once. Third and related, our findings reveal that in natural spoken-language comprehension, an unfolding word can be mapped onto discourse-level representations extremely rapidly, after only two to three phonemes, and in many cases well before the end of the word. When hearing speech in context, listeners apparently need very little to start tying the two together.

Acknowledgements

We thank Petra van Alphen, Ellen de Bruin, René de Bruin, Jelle van Dijk, Jesse Jansen, Valesca Kooijman, Marieke van der Linden, John Nagengast, Edith Sjoerdsma, Cathelijne Tesink, and Johan Weustink for their help. Supported by an NWO Innovation Impulse Vidi grant to JvB, a DFG grant to JvB, PZ and PH, and by NWO grant 400-56-384 to CB and PH.

References

[1] G.T.M. Altmann, Thematic role assignment in context, J. Mem. Lang. 41 (1999) 124–145.
[2] G.T.M. Altmann, A. Garnham, Y. Dennis, Avoiding the garden path: eye movements in context, J. Mem. Lang. 31 (1992) 685–712.
[3] G.T.M. Altmann, Y. Kamide, Incremental interpretation at verbs: restricting the domain of subsequent reference, Cognition 73 (1999) 247–264.
[4] J.E. Arnold, J.G. Eisenband, S. Brown-Schmidt, J.C. Trueswell, The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking, Cognition 76 (2000) B14–B26.
[5] M.A. Britt, The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases, J. Mem. Lang. 33 (1994) 251–283.
[6] C.M. Brown, P. Hagoort, On the electrophysiology of language comprehension: implications for the human language system, in:

M.W. Crocker, M. Pickering, C. CliftonJr. (Eds.), Architectures and Mechanisms for Language Processing, Cambridge University Press, Cambridge, 2000, pp. 213–237.
[7] C.M. Brown, P. Hagoort, M. Kutas, Postlexical integration processes in language comprehension: evidence from brain-imaging research, in: M.S. Gazzaniga (Ed.), The New Cognitive Neurosciences, MIT Press, Cambridge, MA, 2000, pp. 881–895.
[8] H.H. Clark, Using Language, Cambridge University Press, Cambridge, 1996.
[9] J.F. Connolly, N.A. Phillips, Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences, J. Cognit. Neurosci. 6 (1994) 256–266.
[10] S. Crain, M. Steedman, On not being led up the garden path: the use of context by the psychological parser, in: D.R. Dowty, L. Karttunen, A.M.N. Zwicky (Eds.), Natural Language Parsing, Cambridge University Press, Cambridge, 1985, pp. 320–358.
[11] D. Dahan, M.K. Tanenhaus, C.G. Chambers, Accent and reference resolution in spoken-language comprehension, J. Mem. Lang. 47 (2002) 292–314.
[12] K. Eberhard, M. Spivey-Knowlton, J. Sedivy, M. Tanenhaus, Eye movements as a window into real-time spoken language processing in natural contexts, J. Psycholing. Res. 24 (1995) 409–436.
[13] K.D. Federmeier, M. Kutas, A rose by any other name: long-term memory structure and sentence processing, J. Mem. Lang. 41 (1999) 469–495.
[14] K.D. Federmeier, M. Kutas, Right words and left words: electrophysiological evidence for hemispheric differences in meaning processing, Cogn. Brain Res. 8 (1999) 373–392.
[15] K.D. Federmeier, D.B. McLennan, E. De Ochoa, M.K. Kutas, The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an ERP study, Psychophysiology 39 (2) (2002) 133–146.
[16] J.A. Fodor, The Modularity of Mind, MIT Press, Cambridge, MA, 1983.
[17] J.A. Fodor, T.G. Bever, M.F. Garrett, The Psychology of Language, McGraw-Hill, New York, 1974.
[18] J.D. Fodor, W. Ni, S. Crain, D. Schankweiler, Tasks and timing in the perception of linguistic anomaly, J. Psycholing. Res. 25 (1) (1996) 25–57.
[19] K. Forster, Levels of processing and the structure of the language processor, in: W.E. Cooper, E.C. Walker (Eds.), Sentence Processing, Erlbaum, Hillsdale, NJ, 1979.
[20] L. Frazier, On Sentence Interpretation, Kluwer, Dordrecht, 1999.
[21] A.D. Friederici, The neurobiology of language comprehension, in: A.D. Friederici (Ed.), Language Comprehension: A Biological Perspective, Springer, Berlin, 1998, pp. 263–301.
[22] S. Garrod, M. Terras, The contribution of lexical and situational knowledge to resolving discourse roles: bonding and resolution, J. Mem. Lang. 42 (2000) 526–544.
[23] R.W. Gibbs, Literal meaning and psychological theory, Cognit. Sci. 8 (1984) 275–304.
[24] R.W. Gibbs, Interpreting what speakers say and implicate, Brain Lang. 68 (1999) 466–485.
[25] A.M. Glenberg, What memory is for, Behav. Brain Sci. 20 (1997) 1–55.
[26] P. Hagoort, C.M. Brown, ERP effects of listening to speech: semantic ERP effects, Neuropsychologia 38 (2000) 1518–1530.
[27] P. Hagoort, C.M. Brown, ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation, Neuropsychologia 38 (2000) 1531–1549.
[28] P. Hagoort, C.M. Brown, L. Osterhout, The neurocognition of syntactic processing, in: C.M. Brown, P. Hagoort (Eds.), The Neurocognition of Language, Oxford University Press, Oxford, 1999, pp. 273–316.
[29] P.H. Hagoort, L. Hald, K.-M. Petersson, Semantic vs. world knowledge integration during sentence comprehension, in: Presented 9th Annual Meeting of the Cognitive Neuroscience Society (CNS-2002), San Francisco, April 14–16, 2002.
[30] D.J. Hess, D.J. Foss, P. Carroll, Effects of global and local context on lexical processing during language comprehension, J. Exp. Psychol.: Gen. 124 (1) (1995) 62–82.
[31] R. Jackendoff, The representational structures of the language faculty and their interactions, in: C.M. Brown, P. Hagoort (Eds.), The Neurocognition of Language, Oxford University Press, Oxford, 1999, pp. 37–79.
[32] R. Jackendoff, Foundations of Language, Oxford University Press, New York, 2002.
[33] P.N. Johnson-Laird, Mental Models, Cambridge University Press, Cambridge, 1983.
[34] G.A.M. Kempen, Human grammatical coding (2001) Manuscript.
[35] W. Kintsch, Comprehension: A Paradigm for Cognition, Cambridge University Press, Cambridge, 1998.
[36] M. Kutas, In the company of other words: electrophysiological evidence for single-word and sentence context effects, Lang. Cognit. Proc. 8 (4) (1993) 533–572.
[37] M. Kutas, Views on how the electrical activity that the brain generates reflects the functions of different language structures, Psychophysiology 34 (1997) 383–398.
[38] M. Kutas, K.D. Federmeier, Electrophysiology reveals semantic memory use in language comprehension, Trends Cognit. Sci. 12 (2000) 463–470.
[39] M. Kutas, K.D. Federmeier, S. Coulson, J.W. King, T.F. Münte, Language, in: J.T. Cacioppo, L.G. Tassinary, G.G. Berntson (Eds.), Handbook of Psychophysiology, Cambridge University Press, Cambridge, 2000, pp. 576–601.
[40] M. Kutas, S.A. Hillyard, Reading senseless sentences: brain potentials reflect semantic incongruity, Science 207 (1980) 203205.
[41] M. Kutas, C.K. Van Petten, Psycholinguistics electrified: event-related brain potential investigations, in: M.A. Gernsbacher (Ed.), Handbook of Psycholinguistics, Academic Press, New York, 1994, pp. 83–143.
[42] M.C. MacDonald, N.J. Pearlmutter, M.S. Seidenberg, Lexical nature of syntactic ambiguity resolution, Psychol. Rev. 101 (4) (1994) 676703.
[43] W.D. Marslen-Wilson, L.K. Tyler, The temporal structure of spoken language understanding, Cognition 8 (1980) 1–71.
[44] W.D. Marslen-Wilson, A. Welsh, Processing interactions and lexical access during word recognition in continuous speech, Cognit. Psychol. 10 (1978) 29–63.
[45] J.L. McClelland, J.L. Elman, The TRACE model of speech perception, Cognit. Psychol. 18 (1996) 1–86.
[46] J.L. Myers, E.J. O'Brien, Accessing the discourse representation during reading, Discourse Proc. 26 (1998) 131–157.
[47] R. Näätänen, The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm), Psychophysiology 38 (1999) 1–21.
[48] D. Norris, SHORTLIST: A connectionist model of continuous speech recognition, Cognition 52 (1994) 189–234.
[49] L. Osterhout, P.J. Holcomb, Event-related potentials and language comprehension, in: M.D. Rugg, M.G.H. Coles (Eds.), Electrophysiology of Mind, Oxford University Press, Oxford, 1995, pp. 171–215.
[50] C.A. Perfetti, M.A. Britt, Where do propositions come from, in: C.A. Weaver, S. Mannes, C.R. Fletcher (Eds.), Discourse Comprehension, Erlbaum, Hillsdale, NJ, 1995, pp. 11–34.
[51] N. Salmon, H. Pratt, A comparison of sentence- and discourse-level semantic processing: an ERP study, Brain Lang. 83 (3) (2002) 367–383.
[52] J. Searle, Literal meaning, in: J. Searle (Ed.), Expression and Meaning, Cambridge University Press, Cambridge, 1979.
[53] M.J. Spivey-Knowlton, M.K. Tanenhaus, Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency, J. Exp. Psychol.: Learn. Mem. Cognit. 24 (1998) 1521–1543.

[54] R.C. Stalnaker, Assertion, in: P. Cole (Ed.), Syntax and Semantics 9: Pragmatics, Academic Press, New York, 1978, pp. 315–332.
[55] M. St. George, S. Mannes, J.E. Hoffman, Global semantic expectancy and language comprehension, J. Cognit. Neurosci. 6 (1) (1994) 70–83.
[56] D.J. Townsend, T.G. Bever, Sentence Comprehension: The Integration of Habits and Rules, MIT Press, Cambridge, MA, 2001.
[57] J.C. Sedivy, M.K. Tanenhaus, C.G. Chambers, G.N. Carlson, Achieving incremental semantic interpretation through contextual representation, Cognition 71 (1999) 109–147.
[58] M.K. Tanenhaus, C. Trueswell, Sentence comprehension, in: J.L. Miller, P.D. Eimas (Eds.), Speech, Language, and Communication, Academic Press, San Diego, 1995, pp. 217–262.
[59] J.J.A. Van Berkum, 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, ERP and Beyond. Psychol. Press. in press.
[60] J.J.A. Van Berkum, C.M. Brown, P. Hagoort, Early referential context effects in sentence processing: evidence from event-related brain potentials, J. Mem. Lang. 41 (1999) 147–182.
[61] J.J.A. Van Berkum, C.M. Brown, P. Hagoort, When does gender constrain parsing? Evidence from ERPs, J. Psycholing. Res. 28 (5) (1999) 555–571.
[62] J.J.A. Van Berkum, C.M. Brown, P. Hagoort, P. Zwitserlood, Event-related brain potentials reflect discourse-referential ambiguity in spoken-language comprehension, Psychophysiology 40 (2003) 235–248.
[63] J.J.A. Van Berkum, P. Hagoort, C.M. Brown, Semantic integration in sentences and discourse: evidence from the N400, J. Cognit. Neurosci. 11 (6) (1999) 657–671.
[64] J.J.A. Van Berkum, V. Kooijman, C.M. Brown, P. Zwitserlood, P. Hagoort, Do listeners use discourse-level information to predict upcoming words in an unfolding sentence? an ERP study, in: 9th Annual Meeting of the Cognitive Neuroscience Society (CNS-2002), San Francisco, April 14–16, 2002.
[65] D. Van den Brink, C.M. Brown, P. Hagoort, Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects, J. Cognit. Neurosci. 13 (7) (2001) 967–985.
[66] D. Van den Brink, C.M. Brown, P. Hagoort, The cascaded nature of lexical selection and integration is spoken words as revealed by N200 and N400 effects. (2003) submitted for publication.
[67] C. Van Petten, S. Coulson, S. Rubin, E. Plante, M. Parks, Time course of word identification and semantic integration in spoken language, J. Exp. Psychol.: Learn. Mem. Cognit. 25 (1999) 394–417.
[68] R.A. Zwaan, Situation models: the mental leap into imagined worlds, Curr. Directions Psychol. Sci. 8 (1) (1999) 15–18.
[69] P. Zwitserlood, The locus of effects of sentential-semantic context in spoken-word processing, Cognition 32 (1989) 25–64.
[70] P. Zwitserlood, Spoken words in sentence contexts, in: A.D. Friederici (Ed.), Language Comprehension: a Biological Perspective, Springer, Berlin, 1999.