

Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension

JOS J. A. VAN BERKUM,^{a,b,c} COLIN M. BROWN,^b PETER HAGOORT,^{b,c} AND PIENIE ZWITSERLOOD^d

^aDepartment of Psychology, University of Amsterdam, Amsterdam, The Netherlands

^bMax Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

^cF. C. Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands

^dWestfälische Wilhelms-Universität Münster, Germany

Abstract

In two experiments, we explored the use of event-related brain potentials to selectively track the processes that establish reference during spoken language comprehension. Subjects listened to stories in which a particular noun phrase like “the girl” either uniquely referred to a single referent mentioned in the earlier discourse, or ambiguously referred to two equally suitable referents. Referentially ambiguous nouns (“the girl” with two girls introduced in the discourse context) elicited a frontally dominant and sustained negative shift in brain potentials, emerging within 300–400 ms after acoustic noun onset. The early onset of this effect reveals that reference to a discourse entity can be established very rapidly. Its morphology and distribution suggest that at least some of the processing consequences of referential ambiguity may involve an increased demand on memory resources. Furthermore, because this referentially induced ERP effect is very different from that of well-known ERP effects associated with the semantic (N400) and syntactic (e.g., P600/SPS) aspects of language comprehension, it suggests that ERPs can be used to *selectively* keep track of three major processes involved in the comprehension of an unfolding piece of discourse.

Descriptors: Spoken language comprehension, Referential ERP effect, Sentence processing, Discourse context, Anaphoric processing, Brain imaging

For most of us, the subjective experience of reading or listening to our native language is one in which we immediately perceive *meaning*. However, underneath the apparently simple feat of language comprehension lies an extremely complex piece of neuronal machinery, the task of which is to analyze the input at a variety of levels of linguistic organization, while keeping up with input rates that average around 3–5 words per second. Imagine hearing a fairy tale end with “And so the little girl lived happily ever after.” To understand this sentence, listeners must recognize each individual word as it comes in, and retrieve its meaning and syntactic category (noun, verb, etc.) from among the tens of thousands of words stored in their mental dictionary. Listeners also need to work out the syntactic relations among the words (e.g., such that “the” is correctly taken to be the definite article for “little girl” rather than for “and so”). Most important of all, listeners need to construct a conceptual

interpretation of the sentence in this context, to find out what the speaker or writer had meant to convey. Among other things, this involves a *semantic analysis*, in which the meanings of larger constituents (e.g., “the little girl”) are constructed from the meanings of their individual words. It also involves finding out to what particular entity in the earlier discourse a phrase like “the little girl” actually refers to, a vital aspect of comprehensions usually designated as *referential analysis* (or *anaphoric processing*).

A major task of psycholinguistics is to find out how these syntactic, semantic, and referential analyses are orchestrated as language comprehension unfolds in time. Research on this topic has primarily focused on the relative timing and the informational dependency of these various analyses. An important unresolved question, for example, is whether the different types of analysis are conducted in some principled sequential order, with some theorists assigning a fundamental priority to syntactic analysis (e.g., Frazier & Clifton, 1996; Mitchell, Cuetos, Corley, & Brysbaert, 1995), and others instead arguing for a simultaneous evaluation of syntactic, semantic, and referential aspects of the input (e.g., Jackendoff, 1999; Kempen, 2001; MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995). A closely related question is whether the results of semantic and referential processing affect the initial syntactic analysis of the input (e.g., Altmann, 1988; Crain & Steedman, 1985) or not (e.g., Ferreira & Clifton, 1986).

We thank Petra van Alphen, Ellen de Bruin, René de Bruin, Jelle van Dijk, Valesca Kooijman, Bram de Kruijff, Martijn Morren, Mante Nieuwland, John Nagengast, Edith Sjoerdsma, Johan Weustink, and three anonymous reviewers for their help. Supported by a DFG (German Science Foundation) grant to PZ, and PH, by an NWO (Dutch Science Foundation) grant to CB and PH, and by an NWO Innovation Impulse Vidi grant to JvB.

Address reprint requests to: Dr. Jos J. A. van Berkum, University of Amsterdam, Department of Psychology (PN), Roetersstraat 15, 1018 WB Amsterdam, The Netherlands. E-mail: berkum@psy.uva.nl.

To address these processing issues, psycholinguists have turned to research paradigms that help them track language comprehension as it unfolds in synchrony with the input. One such “on-line” methodology that has proved to be particularly useful involves the registration of event-related brain potentials (ERPs) as subjects read or listen to language input. An important benefit of this brain measure is that the ERP signal can *selectively* reflect particular aspects of real-time language comprehension, with high temporal resolution. It is by now well established, for instance, that the processing consequences of a semantic integration problem and those of a syntactic integration problem show up in the ERP signal in qualitatively different ways. Semantic integration problems, such as caused by “socks” in “He buttered the warm bread with *socks* and jam,” invariably elicit an N400 effect (Kutas & Hillyard, 1980), a negative deflection in the ERP that starts at about 150–250 ms after word onset, peaks at about 400 ms, and is usually largest over centro-parietal areas of the scalp. Problems with syntactic integration, as in “The spoilt child *throw* the toy on the floor,” however, consistently elicit a so-called P600/SPS effect, a positive deflection that starts at about 500 ms after word onset, and is as such clearly different from the N400 effect (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992; for reviews, see Brown, Hagoort & Kutas, 2000; Hagoort, Brown, & Osterhout, 1999). In addition, syntactic processing problems have been observed to elicit a variety of earlier short-lived and frequently left-lateralized anterior negativities (LANs), again qualitatively distinct from the N400 effect (e.g., Friederici, Hahne, & Mecklinger, 1996; Kluender & Kutas, 1993; Neville, Nicol, Barss, Forster, & Garrett, 1991; see Brown, Hagoort, et al., 2000; Friederici, 1998; Hagoort et al., 1999; for reviews). Although the exact functional interpretation, the degree of language specificity, and the underlying neural generators of these various ERP effects are still under research, the data do suggest that *within the domain of language comprehension*, these effects dissociatively reflect certain aspects of semantic and syntactic processing, respectively. It is with such selective electrophysiological markers in hand that ERP researchers now help unravel the interplay of semantic and syntactic processes in written as well as spoken language comprehension.

In the present experiments, we explore the possibility that event-related brain potentials methodology can also be used to selectively track some of the *referential* aspects of language comprehension while people listen to a piece of discourse. This is relatively unexplored territory. Although ERPs have been used to address issues in referential processing before (e.g., Osterhout, Bersick, & McLaughlin, 1997; Osterhout & Mobley, 1995; Schmitt, Lamers, & Münte, 2002; Streb, Rösler, & Hennighausen, 1999; for a brief review, see Kutas, Federmeier, Coulson, King, & Münte, 2000), none of the studies that we know of have directly looked for an ERP signature of referential analysis in discourse-level spoken-language comprehension. The aim of the present work is to see whether event-related brain potentials can directly tap into some of the processes involved in establishing reference in spoken discourse, by means of an ERP signature that is—within the domain of language comprehension—distinct both from the standard semantics-related N400 effect and from syntax-related ERP effects such as the P600/SPS.

Preliminary evidence obtained in our laboratory with written language materials (Van Berkum, Brown, & Hagoort, 1999a) suggests that this may be a real possibility. To study the impact of referential factors on written sentence comprehension, we conducted an ERP experiment in which we manipulated the

number of candidate referents for a singular definite noun phrase (NP). In the experiment, subjects were asked to read several ministories, such as the one below (translated from Dutch, boldface added):

(1) David had told the boy and the girl to clean up their room before lunchtime. But the boy had stayed in bed all morning, and the girl had been on the phone all the time. David told **the girl** that had been on the phone to hang up.

Following earlier research on this topic (e.g., Crain & Steedman, 1985), the stories were varied such that they provided either a *single* unique referent for the NP “the girl” in the last sentence, as in (1), or *two* equally eligible referents, as in (2):

(2) David had told the two girls to clean up their room before lunchtime. But one of the girls had stayed in bed all morning, and the other had been on the phone all the time. David told **the girl** that had been on the phone to hang up.

This simple discourse manipulation turned out to have a clear and reliable effect in the ERP waveforms elicited by a written noun: relative to its referentially successful counterpart (e.g., “the girl” in a one-referent context), a referentially ambiguous noun (e.g., “the girl” in a two-referent context) elicited a widely distributed negative deflection, emerging at about 300 ms after noun onset. The effect was largest at anterior sites, where it was also particularly sustained (see Van Berkum, Brown, et al., 1999a, their Figures 1a and 1b).

As discussed by Van Berkum, Brown, and Hagoort (1999a), this ERP effect elicited by referential ambiguity has potentially important implications. First of all, the rapid divergence of ERP waveforms elicited by referentially ambiguous and unambiguous nouns by itself suggested that language users can very rapidly determine whether a singular definite noun has a unique referent in the earlier discourse or not, within at most some 300 ms. Secondly, because the reference-related ERP effect was qualitatively very different from both the semantics-related N400 effect and the syntax-related P600/SPS effect, it suggested that the processing consequences of at least one type of referential problem are both interestingly different *and* separately observable in ERPs. The comparison to these other types of ERP effects was facilitated by the fact that in the same written-language study, with the same subjects, we also obtained syntax-related P600/SPS effects (Van Berkum, Brown, et al., 1999a, Van Berkum, Brown, & Hagoort, 1999b) as well as discourse-induced semantics-related N400 effects (Van Berkum, Hagoort, & Brown, 1999).

This referentially induced sustained negative shift in ERPs might help us gain a deeper understanding of the mechanisms by which reference is established and maintained in language comprehension. As discussed by Van Berkum, Brown, et al. (1999a), the ERP effect at hand was very similar to sustained frontal ERP effects observed under conditions of increased working memory load in language processing (e.g., Fiebach, Schlesewsky, & Friederici, 2001; Friederici et al., 1996; King & Kutas, 1995; Kluender & Kutas, 1993; Kutas, 1997; Müller, King & Kutas, 1997; Münte, Schiltz, & Kutas, 1998; Rösler, Heil, & Röder, 1997; Rösler, Pechmann, Streb, Röder, & Hennighausen, 1998; Vos, Gunter, Kolk, & Mulder, 2001) as well as in nonlinguistic processing tasks (e.g., Donaldson & Rugg, 1999; Rösler, Heil, & Glowalla, 1993; Rugg & Allen,

2000). This suggested that the referentially induced frontal negativity might itself reflect an increased demand on memory resources (an issue we return to in the General Discussion), and need thus not be *specific* to referential ambiguity in language comprehension. Note, however, that (as suggested by the results of Van Berkum, Brown, et al., 1999a) this would not diminish the utility of the effect to—within the domain of language comprehension—*selectively* index aspects of referential, rather than semantic or syntactic processing. As such, a referentially induced sustained negative shift can also help us gain a deeper understanding of how and when the mechanisms by which reference is established interact with the semantic and syntactic analysis of an unfolding sentence.

The main goal of the present experiments is to examine the ERP effect of referential ambiguity in *spoken* language comprehension. Spoken language is the dominant form of language. With the exception of sign languages for the deaf, all the world's languages naturally evolved into spoken form, and the corresponding writing systems, if present at all, are all relatively recent cultural inventions. Also, whereas children usually effortlessly acquire language from the speech they hear around them, learning to read is a cumbersome process that requires explicit school-based teaching. Furthermore, to the extent that the human species is endowed with a biological faculty or “instinct” for language (e.g., Pinker, 1994), it will be a system that has evolved to subserve communication by means of spoken (and possibly gestural)—but not written—language. All this does, of course, not imply that discourse- and other high-level aspects of language comprehension system should *always* preferably be studied with spoken-language input. It does imply, however, that spoken language plays a vital role in the study of language comprehension.

A secondary reason for conducting the present spoken-language research is that it allows us to address a concern that is sometimes raised over the use of serial visual presentation (SVP). In the Van Berkum, Brown, et al. (1999a) experiments, we presented our materials at a rate of 600 ms/word. This is a common procedure for written-language ERP studies that try to cover new ground in sentence comprehension research. However, a 600 ms/word input rate is about twice as slow as the average natural reading rate (~250 ms/word; Rayner, 1998). In principle, the referentially induced ERP effect in our written-language study might thus have arisen because of an unrepresentative input rate, perhaps because subjects had “spare time” to look for and exploit certain regularities in the many critical ministories presented to them. In spoken-language Experiment 1, we directly address this concern by presenting the same set of ministories as fully connected natural speech recorded with a normal speaking rate (~2–3 words/s; Levelt, 1989). In Experiment 2, we again use spoken language, but we now present only a small subset of the original critical stories, amid a large number of unrelated filler stories. In both experiments, the issue is whether referentially ambiguous spoken nouns elicit a differential ERP effect relative to referentially unique nouns, and if so, to what extent this spoken-language effect resembles the effect observed with written language.

EXPERIMENT 1

Method

Apart from the use of fully spoken materials and different subjects, the method of this experiment is identical to its written-language predecessor (Van Berkum, Brown, et al., 1999a).

Participants

We recruited 24 right-handed native speakers of Dutch (19 female participants, 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.

Materials

Every participant listened to 240 three-sentence ministories, of which the final sentence contained a critical singular NP. Each story-final target sentence was embedded in a discourse that introduced either one or two referents for this critical singular NP, and differed in no other way (see examples 1 and 2). The full set of Dutch materials is available from the first author.

In formulating the discourse contexts, we took great care to avoid phrasing or content that foregrounded one candidate referent at the expense of the other one (for details, see Van Berkum, Brown, et al., 1999a). We also made sure that the use of the critical noun in the story-final sentence was equally felicitous in the one- and two-referent conditions. Furthermore, to rule out a lexical priming confound, this critical noun was used equally often in the preceding one- and the two-referent contexts.

To explore a different research issue (see Van Berkum, Brown, et al., 1999a, 1999b; Brown, Van Berkum, & Hagoort, 2000), a critical NP was always in object position, and was followed by a relative clause or a sentential complement. All sentential continuations were such that referential ambiguity was preserved for at least two more words after the critical noun.¹ Furthermore, 40 of the 240 story-final sentences also contained a semantically odd word in the remainder of the sentence, at least four words after the critical noun. The semantic anomaly hinged on the semantic relation between this sentence and its discourse context, and was, for current purposes, primarily expected to distract the participants' attention from aspects of the design at hand (for details and the associated ERP effects, see Van Berkum, Zwitserlood, Hagoort, & Brown, in press; see also Figure 2C). All of the abovementioned factors were counterbalanced with respect to the referential manipulation.

The critical sentences and their discourse contexts were recorded with a normal speaking rate and intonation in two separate sessions, by the same female native speaker. The discourse context recordings were those of the preceding written-language experiment (where only the critical last sentence of every story had, in fact, been presented visually). The one-referent context had been recorded first for half of the items, and second for the other half, and recordings with referent-biasing prosody were redone. To enable later EEG averaging, a trained phonetician identified the acoustic onsets (and offsets) of the critical nouns in these sentences. The acoustic duration of these nouns ranged from 169 to 781 ms, with an average of 448 ms.

Procedure, EEG Recording, and Analysis

After electrode application, participants 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. Participants were asked to process each story for comprehension. No additional task demands were imposed.

¹Although the present experiments aimed to explore the impact of emerging referential ambiguity, we did not systematically manipulate (and cannot, therefore, have a clean look at ERP effects associated with) the moment at which referential ambiguity was *resolved*.

We used two different trial sequences, identical to those used in the written-language study. Each trial consisted of a 300-ms auditory warning tone, followed by 700 ms of silence, the spoken discourse context, 1,000 ms of silence, and the spoken final sentence. An informal pretest (as well as later remarks of our participants) indicated that the 1,000 ms separating the final sentence from its context did not perceptually break the story into two parts. To inform participants when to fixate and sit still for EEG recording, an asterisk was displayed from 1,000 ms before onset of the target sentence to 1,000 ms after its offset. After a short practice, the trials were presented in five blocks of 15 min, separated by rest periods.

The EEG recording parameters were as in the written language study (for details, see Van Berkum, Brown, et al., 1999a). 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 2,250 ms after the acoustic onset of the critical noun. Trials containing such artifacts were rejected (12.2%).

Average waveforms in this critical time window were computed for each participant in each referential condition, after normalizing the waveforms of the individual trials relative to a 150-ms prestimulus baseline interval preceding the critical word. Subsequent analyses of variance (ANOVAs) used mean amplitude values computed for each participant in a latency window of 300–600 ms after onset of the critical noun, the interval that had also been used in the written-language study. 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 included a 13-level electrode factor orthogonal to the referential context factor. The scalp distribution of various ERP effects 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.

Results

Figure 1 displays, for each electrode, the grand average event-related brain potentials time-locked to the acoustic onset of critical singular nouns presented in a one-referent or a two-referent discourse context. Inherent to the use of fully connected speech input, there are no clear exogenous ERP components. The waveforms do however clearly reveal a rapid effect of referential context: Relative to their referentially unique counterparts, referentially ambiguous nouns elicit a negative deflection in the average ERP that begins at about 300–400 ms after their acoustic onset, and does not have an obvious hemispheric asymmetry. At posterior sites, the effect terminates at about 600 ms poststimulus. At anterior sites, the negativity is a more sustained one. In view of these characteristics, the effect of referential ambiguity can clearly not be qualified as a regular N400 effect (an issue to which we return below).

Statistical analysis supports the above observations. Using mean amplitude in the 300–600-ms latency range after onset of the critical noun, an omnibus ANOVA with referential context (one- or two-referent) and electrode (13 sites) yielded a significant effect of referential context, $F(1,23) = 7.54$, $MSE = 5.75$,

$p = .011$. On average, mean amplitudes were $0.5 \mu V$ more negative for nouns presented in a two-referent context than for the same nouns presented in a one-referent context. In the 300–600-ms latency range, referential context did not significantly interact with electrode site, $F(12,276) = 0.25$, $MSE = 0.34$, $p = .851$, nor—in subanalyses—with any of the more specific topographical factors, all F s < 1 . A separate trend analysis over four successive blocks of 60 trials each revealed that the magnitude of the referential context effect did not evolve as the session progressed (mean two-referent – one-referent effect sizes were $0.47 \mu V$ for block 1, $0.55 \mu V$ for block 2, $0.59 \mu V$ for block 3, and $0.40 \mu V$ for block 4; Context \times Block $F(3,69) = 0.07$, $MSE = 16.58$, $p = .970$).

A supplementary ANOVA on mean amplitudes in the 600–900-ms latency range confirmed that the later part of the effect had a more limited anterior distribution. There was no significant main effect of referential context, $F(1,23) = 3.17$, $MSE = 4.65$, $p = .088$, but referential context clearly interacted with electrode, $F(12,276) = 4.58$, $MSE = 0.29$, $p = .006$. Topographical subanalyses revealed significant interactions of referential context with a three-level midline electrode factor, $F(2,46) = 11.01$, $MSE = 0.27$, $p = .001$, and a five-level left/right-averaged lateral electrode factor, $F(4,92) = 5.47$, $MSE = 0.41$, $p = .019$. But referential context did not interact with hemisphere, $F(1,23) = 0.07$, $MSE = 0.37$, $p = .798$, nor with Hemisphere \times Five-Level Lateral Electrode, $F(4,92) = 0.27$, $MSE = 0.09$, $p = .722$. Electrode-specific analyses revealed significant referential context effects at Fz, F7, F8, LAT, and RAT only.

A second supplementary mean amplitude ANOVA revealed that the referentially induced ERP effect was not significantly affected by whether the noun was followed by a complement clause or a relative clause. Referential context did not interact with subsequent clause type, either in the 300–600 ms latency range (Referential Context \times Subsequent Clause Type: $F(1,23) = 1.60$, $MSE = 15.97$, $p = .219$; Referential Context \times Electrode \times Subsequent Clause Type: $F(12,276) = 0.30$, $MSE = 1.16$, $p = .845$) or in the 600–900 ms latency range (Referential Context \times Subsequent Clause Type: $F(1,23) = 1.25$, $MSE = 20.84$, $p = .276$; Referential Context \times Electrode \times Subsequent Clause Type: $F(12,276) = 1.96$, $MSE = 1.39$, $p = .134$).

To statistically evaluate the onset of the referentially induced negative shift, we conducted one-tailed repeated measures t tests over the mean amplitudes in the one- and two-referent context in a series of latency “bins” of 50-ms width, each bin shifted 10 ms later in time relative to the preceding one (so 0–50 ms, 10–60 ms, 20–70 ms, etc.). Taking the onset of at least five consecutive latency ranges as the onset of the effect, the ERPs to referentially ambiguous and unambiguous nouns started to diverge significantly in the 310–360 ms latency range ($t(23) = 1.88$, $SD = 1.18$, $p = .037$; with a more agnostic two-tailed criterion, the effect emerged in the 320–370 ms latency range).

Discussion

Relative to their referentially unique counterparts, referentially ambiguous spoken nouns elicited a widely distributed and frontally sustained negativity, emerging at about 300–400 ms after their acoustic onset. Despite a radically different mode of language presentation, this spoken-language effect is similar to the effect obtained with written-language materials presented at a rate of 600 ms/word (Van Berkum, Brown, et al., 1999a, Figure 1).

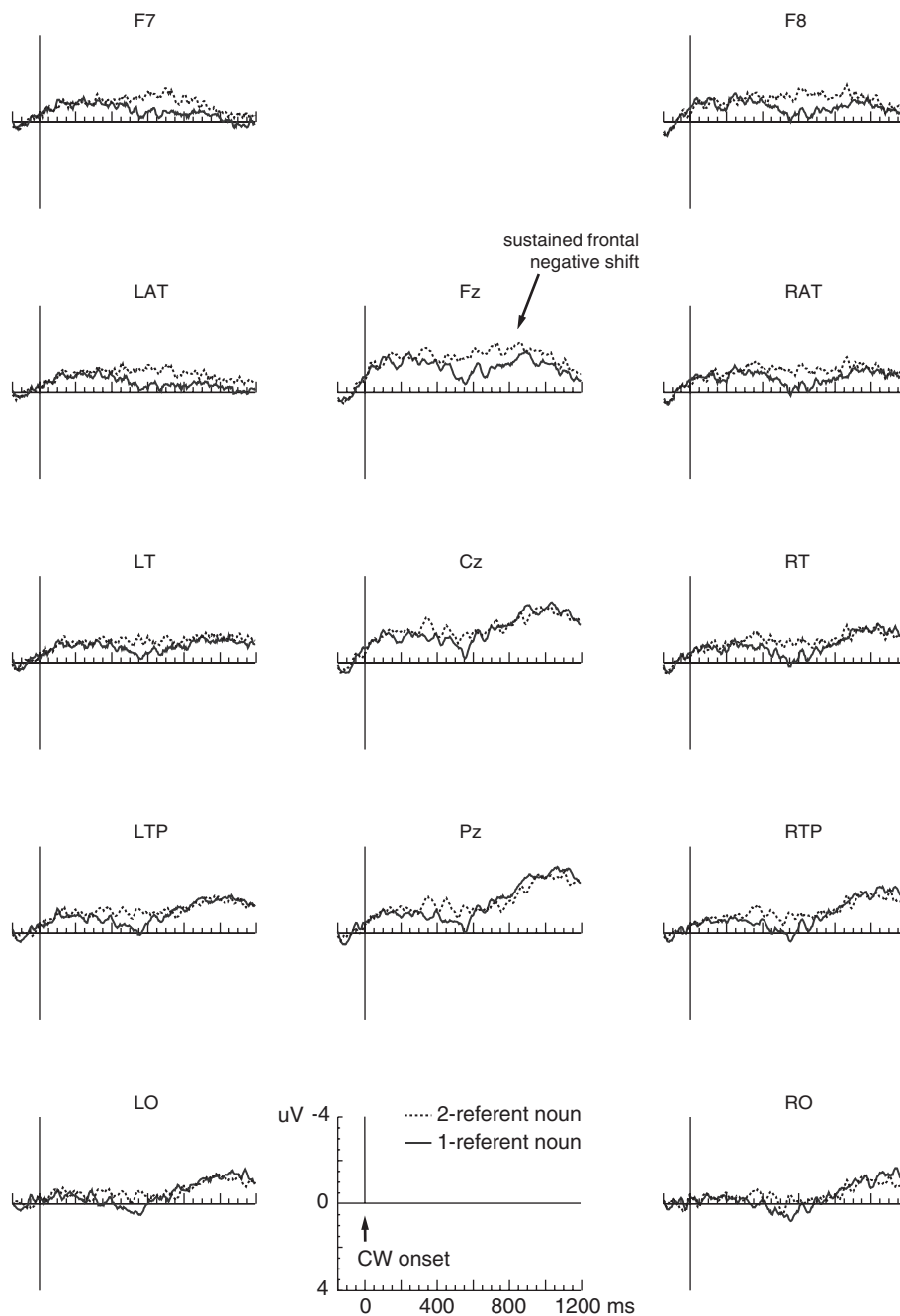


Figure 1. Grand average ERP waveforms elicited by spoken singular nouns presented in a one-referent context and a two-referent context in Experiment 1. In this and all following figures, acoustic onset of the noun (CW) is at 0 ms, and negative polarity is plotted upwards.

For ease of comparison, Figure 2A redisplayes the ERP waveforms previously obtained with written language next to the waveforms now obtained with spoken-language versions of the same materials (Figure 2B), for the same three midline sites, and at the same voltage and time scales. The waveforms themselves are very different across modality, for whereas the abrupt visual onset of a written word at 0 ms (as well as that of its successor 600 ms later) elicits a clear N1-P2 complex, a spoken word embedded in a continuous stream of overlapping speech sounds does not. However, the *differential effect* of referential ambiguity is essentially identical across input modality, sharing overall

morphology (a sustained negative shift), scalp distribution (an anterior maximum), magnitude (below $1 \mu\text{V}$, i.e., a very small effect), and approximate onset (at about 300–400 ms).² In line

²When we combined the spoken-language data of Experiment 1 and the written-language data of its predecessor in a supplementary mean amplitude ANOVA, the referential context effect did not significantly differ as a function of input modality, either in the 300–600 ms latency range (Referential Context \times Modality: $F(1,46) = 0.71$, $MSE = 6.58$, $p = .403$; Referential Context \times Electrode \times Modality: $F(12,552) = 0.48$, $MSE = 0.33$, $p = .700$) or in the 600–900 ms latency range (Referential Context \times Modality: $F(1,46) = 1.63$, $MSE = 6.74$, $p = .209$; Referential

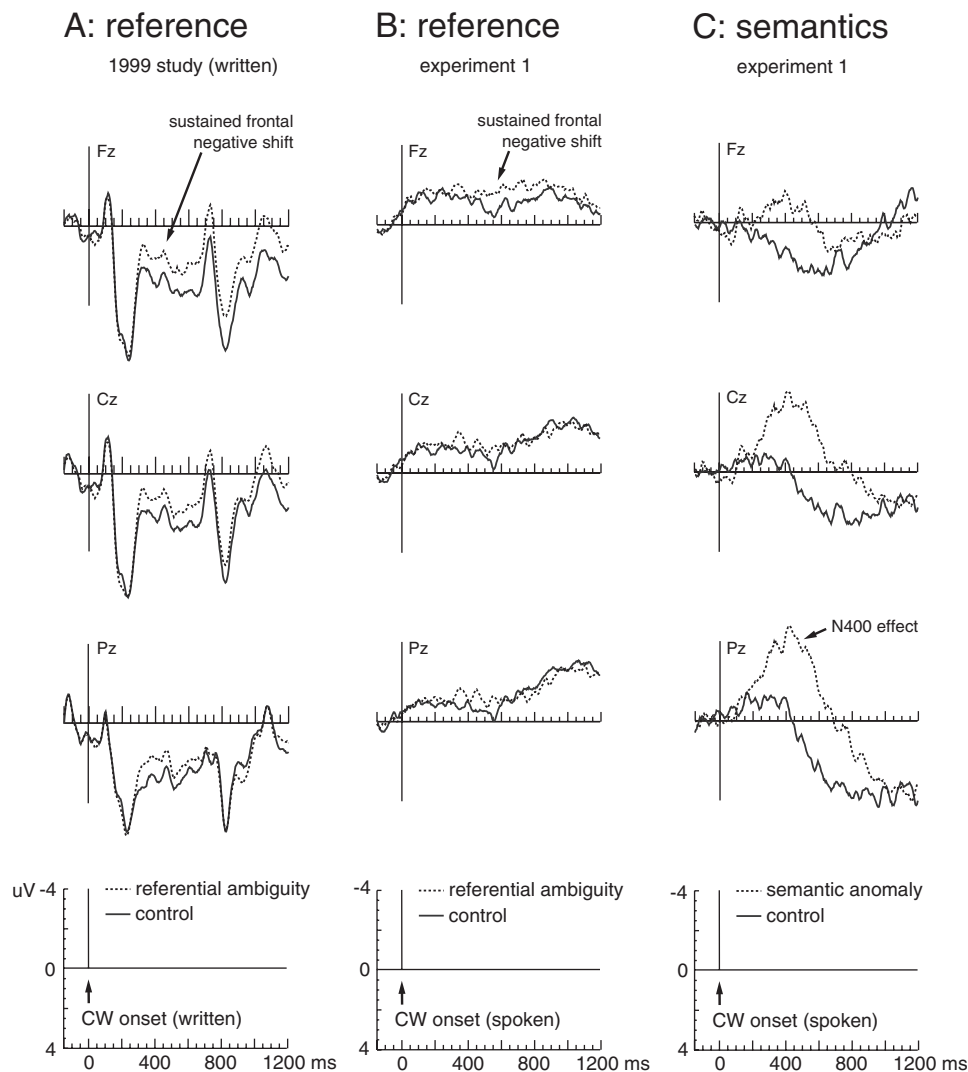


Figure 2. Referentially and semantically induced ERP effects in written and spoken language. A: Grand average ERP waveforms elicited by referentially unambiguous and referentially ambiguous *written* nouns, data from Van Berkum, Brown, and Hagoort (1999a). B: Grand average ERP waveforms elicited by referentially unambiguous and referentially ambiguous *spoken* nouns in Experiment 1. C: Grand average ERP waveforms elicited by semantically coherent and semantically anomalous spoken words in Experiment 1 (for details, see Van Berkum, Zwitserlood, Hagoort, & Brown, in press).

with several other studies that elicited equivalent ERP effects with spoken and serially presented written sentences (e.g., Van Berkum, Zwitserlood, et al., in press; Brown, van Berkum, et al., 2000; Federmeier, McLennan, De Ochoa, & Kutas, 2002; Hagoort & Brown, 2000a, 2000b; Kutas, 1993, 1997; Osterhout & Holcomb, 1993; Van den Brink, Brown, & Hagoort, 2001; Van Petten, Coulson, Rubin, Plante, & Parks, 1999), the discourse-referential ERP effect observed in the Van Berkum, Brown, et al. (1999a) study therefore does not seem to depend on the use of a Serial Visual Presentation procedure or a relatively slow input rate.

Context \times Electrode \times Modality: $F(12,552) = 2.38$, $MSE = 0.30$, $p = .062$). The latter trend towards a significant three-way interaction of referential context, electrode, and modality in the 600–900 ms latency range reflects a (numerically) slightly more left-dominant distribution of the written-language effect, relative to the spoken-language effect, in this late interval.

With SVP-based artifacts out of the way, the difference emerging between ERPs elicited by referentially ambiguous and referentially successful nouns can be taken to show that within about 300–400 ms after noun onset, the processing system has at some level determined whether a singular definite spoken noun has a single unique referent in the earlier discourse or not. Furthermore, although discourse-induced referential ambiguity could in principle be viewed as a problem with conceptual interpretation, the observed ERP effect is clearly not simply a standard *N400* effect, the effect that is usually associated with interpretive problems. Using discourse-dependent *semantic anomalies* such as “John ate a sandwich with *salami*” in a discourse that depicted John as a strict vegetarian, we did in fact also obtain a standard (but discourse-dependent) *N400* effect in the same experiment. Figure 2C displays this discourse-semantic *N400* effect for three midline sites (for details, see Van Berkum, Zwitserlood, et al., in press). As can be seen in Figure 2C, the *N400* effect emerged at about 150–200 ms after acoustic word

onset and clearly peaked at about 400 ms, with a maximum at centro-parietal sites. These are the typical features of a language-induced N400 effect (cf. Kutas & Van Petten, 1994). Also, as can be seen in Figure 2B, what we observe with a discourse-dependent referential problem is a very different set of features. Not only is the referentially induced ERP effect observed in Experiment 1 much smaller, it also has a much later onset (at about the time where the N400 effect reaches its peak), an anterior maximum, and a very different morphology (lacking a clear peak). In the same group of subjects listening to the same stories in the same experiment, problems with reference and problems with the semantics thus elicit *qualitatively* very different ERP effects. We return to this in the general discussion.

Before we can argue that the ERP effect elicited by referentially ambiguous nouns indeed reflects some aspect of naturally occurring referential processing in language comprehension, however, we need to address a potential concern over the design of Experiment 1, inherited from its written-language precursor. Due to the constraints imposed by a different question addressed in these experiments (see Van Berkum, Brown, et al., 1999a), half of the 240 story trials in Experiment 1 contained a referentially ambiguous critical noun. In principle, this may have led subjects to increasingly focus their attention on referential aspects of the study.

Postsession interviews conducted for Experiment 1 revealed that by the end of the session, many of our subjects had indeed noticed the occurrence of referentially ambiguous NPs. We believe that to a large extent this is an unavoidable consequence of the fact that the explicit and successful identification of referents is central to language comprehension. If subjects aim for comprehension, and if referential ambiguity causes coherence to break down (as in our materials), people simply notice the ambiguity as an interpretive problem. However, the trend analysis over four successive trial blocks revealed that the magnitude of the referential context effect did not evolve as the experimental session progressed. This suggests that, although referential ambiguity may itself be salient when it occurs, the ERP effect at hand is probably not associated with an evolving strategic expectation. The next experiment provides additional evidence for this claim.

EXPERIMENT 2

In Experiment 2, each participant encountered only 30 critical trials with a referentially ambiguous noun, amid 30 matched control trials and 180 unrelated trials. The issue was whether we would nevertheless still obtain an effect of referential ambiguity in ERPs.

Method

Subjects

For Experiment 2, 24 right-handed native speakers of Dutch (18 female participants, mean age 22, range 18–28 years) were recruited from our local subject pool. None had any neurological impairment, had experienced any neurological trauma, or had used neuroleptics. Also, none had participated in Experiment 1 or the earlier written-language study (Van Berkum, Brown, et al., 1999a).

Materials

Critical items were 60 of the 240 stories used in Experiment 1, each in a one- and a two-referent version. In all 60 stories, the critical noun was followed by a restrictive relative clause (as in examples 1 and 2). Because they had to be mixed with a large set of new stories (see below), all 60 recycled stories were recorded again for Experiment 2, by a (different) female native speaker, using the same counterbalanced recording procedure as in Experiment 1. A trained phonetician identified the acoustic onsets (and offsets) of the critical nouns in these sentences. Critical noun duration ranged from 172 to 678 ms, with an average of 443 ms. The 60 critical stories were mixed with 180 stories that did not involve referential ambiguity. Of the latter, 120 addressed a different issue (for further details, see Van Berkum, Kooijman, Brown, Zwitserlood, & Hagoort, 2002). The remaining 60 stories were true fillers. In the ERP experiment, every participant saw 30 critical stories in the one-referent condition, and 30 in the two-referent condition, amid the 180 other stories. These trials were pseudo-randomly mixed such that the largest sequence of consecutive critical trials was four, and the largest sequence of critical trials in the same referential condition was two. The resulting randomization was divided in five blocks of 48 trials, each of which started with two additional filler trials. A 20-trial comparable practice sequence preceded the first experimental block. A second trial list was derived from the first by rotating referential conditions over items. In the experiment, half of the participants were tested with the first list, the other half with the second list.

Procedure, EEG Recording and Analysis

The test situation, instructions, and stimulus presentation parameters were identical to those of Experiment 1. The EEG was recorded from 29 silver-chloride electrodes mounted in an elastic cap, each referred to the left mastoid. Five electrodes were placed over the standard 10% system midline sites Fz, FCz, Cz, Pz, and Oz. Nine pairs of electrodes were placed over the standard lateral sites AF3/AF4, F3/F4, F7/F8, FC3/FC4, FT7/FT8, C3/C4, CP3/CP4, P3/P4, and PO7/PO8. Three additional pairs were placed laterally over symmetrical nonstandard positions: (a) a temporal pair (LT and RT) placed laterally to Cz, at 33% of the interaural distance, (b) a temporo-parietal pair (LTP and RTP) placed 30% of the interaural distance lateral and 13% of the nasion-inion distance posterior to Cz, and (c) a parietal pair midway between LTP/RTP and PO7/PO8 (LP and RP). Vertical eye movements 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 presumably neutral mastoid site. No such differential effects were observed.

The EEG and EOG recordings were amplified with a NeuroScan SynAmp Model 5083 EEG amplifier, using a hi-cut of 30 Hz and a time constant of 8 s (0.02 Hz). Electrode impedances were kept below 3 k Ω for the EEG recording and below 5 k Ω for the EOG recording. The EEG and EOG signals were digitized on-line with a sampling frequency of 500 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 2,850 ms after the acoustic onset of the critical noun. Trials containing such artifacts were rejected (20.2%).

Averaging and subsequent statistical analysis procedures were as in Experiment 1, but because of a different electrode configuration, effects were tested in somewhat different ANOVA designs. Two omnibus ANOVAs were conducted first, one crossing referential context (one or two referents) with a simple five-level midline-electrode factor (Fz, FCz, Cz, Pz, Oz), and one that fully crossed the referential context factor with a hemisphere (left/right) by anteriority (anterior/posterior) factor. The latter analysis effectively defined four quadrants: (1) left-anterior, involving AF3, F3, F7, FC3, and FT7, (2) right-anterior, involving AF4, F4, F8, FC4, and FT8, (3) left-posterior, involving LTP, CP3, LP, P3, and PO7, and (4) right-posterior, involving CP4, RTP, P4, RP, and PO8. As the configuration of electrodes within each of these quadrants is different, the quadrant ANOVA was conducted over mean amplitude values that were averaged per quadrant across the five electrodes that it contained. If necessary, these two omnibus tests were followed by more specific ANOVA's.

Results

Figure 3 displays, for each electrode, the grand average event-related brain potentials time-locked to the acoustic onset of critical singular nouns presented in a one-referent or a two-referent discourse context. Again no clear exogenous ERP components can be distinguished in the average waveforms. But, just as in Experiment 1, referentially ambiguous nouns elicit a frontally dominant negative deflection in the average ERP relative to their referentially unique counterparts. Also, this negativity is again a sustained one at anterior sites only.

Due to the substantially lower number of critical trials involved (at most 30 per condition per subject, instead of 120 trials in Experiment 1), the ERP waveforms in this experiment are more noisy than those of Experiment 1, making it more difficult to identify the onset of the sustained negative deflection in Figure 3. Possibly because of this lower signal-to-noise ratio, the ERP elicited by referentially ambiguous nouns in Experiment 2 displays a transient and relatively sharply peaked negativity already at about 100–150 ms after noun onset. The frontally dominant sustained negative shift, however, emerges at approximately the same time (~300–400 ms) as in Experiment 1.

Using mean amplitude in the 300–600-ms latency range after onset of the critical noun, the ANOVA with referential context (one- or two-referent) and midline electrode (five sites) did not yield a significant main effect of referential context, $F(1,23) = 2.90$, $MSE = 8.23$, $p = .102$, nor an interaction of context with electrode, $F(4,92) = 2.29$, $MSE = 1.36$, $p = .125$. In the associated mean quadrant ANOVA, however, the main effect of referential context was significant, $F(1,23) = 8.73$, $MSE = 3.24$, $p = .007$. In the 300–600-ms latency range at hand, the referential context effect was not significantly modulated by hemisphere, $F(1,23) = 2.52$, $MSE = 0.74$, $p = .126$, anteriority, $F(1,23) = 1.30$, $MSE = 2.34$, $p = .265$, or Hemisphere \times Anteriority, $F(1,23) = 1.84$, $MSE = 0.25$, $p = .188$. Looking at each of the quadrants separately, referential context elicited a significant simple main effect in the left-anterior quadrant (corresponding to a two-referent – one-referent effect size of $0.92 \mu\text{V}$, $F(1,23) = 7.87$, $MSE = 6.46$, $p = .010$), the right-anterior quadrant ($1.12 \mu\text{V}$, $F(1,23) = 12.89$, $MSE = 5.82$, $p = .002$), and the right-posterior quadrant ($0.81 \mu\text{V}$, $F(1,23) = 4.29$, $MSE = 9.18$, $p = .050$). No significant referential effect was obtained in the

left-posterior quadrant ($0.22 \mu\text{V}$, $F(1,23) = 0.25$, $MSE = 11.40$, $p = .620$). In none of the four quadrants did referential context interact with the five-level electrode factor (LA: $p = .221$; RA: $p = .514$; LP: $p = .830$; RP: $p = .554$).

Following up on the supplementary ANOVA conducted for Experiment 1, we also conducted these midline and quadrant analyses on mean amplitude values in the consecutive 600–900-ms latency range. The ANOVA with referential context (one- or two-referent) and midline electrode (five sites) did not yield a significant main effect of referential context, $F(1,23) = 0.50$, $MSE = 15.50$, $p = .487$, but in this later latency range referential context did interact with electrode site, $F(4,92) = 13.88$, $MSE = 1.38$, $p < .001$. In the mean quadrant ANOVA over the 600–900-ms latency range, the main effect of referential context was also not significant, $F(1,23) = 3.49$, $MSE = 5.10$, $p = .075$. Instead, this factor significantly interacted with anteriority, $F(1,23) = 11.00$, $MSE = 3.13$, $p = .003$, and Hemisphere \times Anteriority, $F(1,23) = 5.26$, $MSE = 0.15$, $p = .031$, although not with hemisphere, $F(1,23) = 0.29$, $MSE = 0.87$, $p = .597$. Looking at each of the quadrants separately in the 600–900-ms latency range, referential context elicited a significant simple main effect in the left-anterior quadrant (two-referent – one-referent effect size of $1.51 \mu\text{V}$, $F(1,23) = 13.79$, $MSE = 9.91$, $p = .001$), in the right-anterior quadrant ($1.40 \mu\text{V}$, $F(1,23) = 12.39$, $MSE = 9.49$, $p = .002$), but not in the left-posterior ($-0.44 \mu\text{V}$, $F(1,23) = 0.87$, $MSE = 13.16$, $p = .360$) or right-posterior ($-0.04 \mu\text{V}$, $F(1,23) = 0.01$, $MSE = 13.64$, $p = .936$) quadrants. In none of the four quadrants did referential context interact with the five-level electrode factor (LA: $p = .183$; RA: $p = .124$; LP: $p = .221$; RP: $p = .247$). This supplementary analysis confirms that the referentially induced negative shift is a sustained one at anterior sites only.

As for the transient and relatively sharply peaked negativity at about 100–150 ms after noun onset, an ad hoc ANOVA over mean amplitudes in the 100–150-ms latency range revealed a main effect in the midline ANOVA, $F(1,23) = 4.28$, $MSE = 12.94$, $p = .050$, as well as in the mean quadrant ANOVA, $F(1,23) = 6.97$, $MSE = 5.20$, $p = .015$. In the latter analysis, the negativity did not significantly interact with hemisphere, $F(1,23) = 2.68$, $MSE = 0.89$, $p = .116$, anteriority, $F(1,23) = 0.01$, $MSE = 1.87$, $p = .919$, or the two together, $F(1,23) = 0.01$, $MSE = 0.20$, $p = .906$.

Discussion

As opposed to Experiment 1, in which 120 of the 240 story trials (50%) contained a referentially ambiguous noun, only 30 of 240 story trials (12.5%) were referentially ambiguous in Experiment 2. As an unavoidable consequence, the waveforms in Figure 3 are somewhat noisier than those displayed in Figure 1. In spite of this, however, we see a comparable phenomenon: Relative to referentially successful control nouns, referentially ambiguous spoken nouns again very rapidly elicit a sustained frontal negative shift. Also, as in Experiment 1, the effect of a discourse-induced referential problem is again qualitatively different from that of a discourse-induced semantic problem (see Figure 2C for the latter).

As can be seen in Figures 1 and 3, there are nonnegligible differences between the present referentially induced frontal negative shift and the one obtained in Experiment 1. However, if the ERP effect obtained in Experiment 1 had hinged on strategic

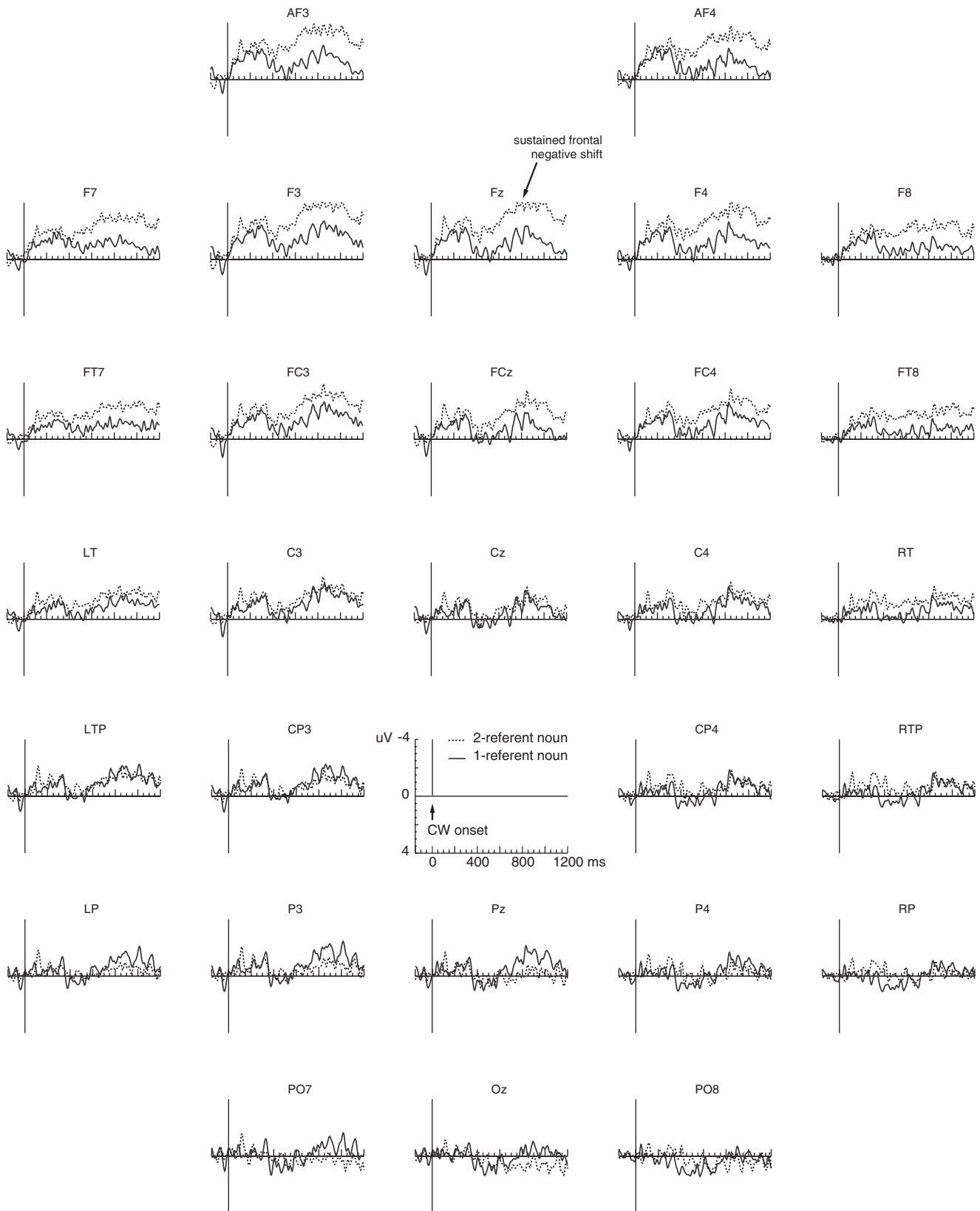


Figure 3. Grand average ERP waveforms elicited by spoken singular nouns presented in a one-referent context and a two-referent context in Experiment 2.

processing induced by a high proportion of referentially ambiguous trials, one would have expected to find a smaller and/or later effect of referential ambiguity in Experiment 2, and possibly even no effect at all. This is clearly not the case. Instead, the present ERP effect appears to be larger and earlier in onset.³ Furthermore, despite the differences, the overall finding is again a frontally dominant, sustained negative shift. This suggests that the referential ERP effect observed in Experiment 1 does not hinge on a high proportion of referentially ambiguous trials.

General Discussion

In two experiments, we explored the use of event-related brain potentials to selectively track the processes that establish reference during spoken language comprehension. In Experiment 1, referentially ambiguous spoken nouns elicited a negative shift that emerged in the ERPs at about 300–400 ms after acoustic onset, and that, although widely distributed, was most prominent and most sustained at anterior recording sites. As such, the present spoken-language ERP effect is identical to the effect we elicited with referentially ambiguous *written* nouns (Van Berkum, Brown, et al., 1999a). Moving beyond this strict cross-modality replication, Experiment 2 revealed that the presence of a frontally sustained negative shift does not critically depend on a high proportion of referentially ambiguous trials.

The most direct implication of these spoken-language findings is that we are dealing with a referentially induced ERP effect that is essentially independent of the methodological particulars of the original written-language experiment. In the original Van Berkum, Brown, et al. (1999a) written-language study, 50% of the stories contained a referentially ambiguous noun phrase, and the words in critical sentences were presented serially at a fixed and relatively slow pace. Our present results suggest that neither feature of the original study was critical to obtain a referentially induced negative shift.

More interestingly, our findings show (a) that very shortly after acoustic word onset (for the materials used in our study), the processing system has at some level determined whether a singular definite noun has a single unique referent in the earlier discourse or not, and (b) that we can use ERPs to *selectively* keep track of at least some aspect of referential processing as a sentence is heard (or read) in discourse. We discuss both of these implications in turn.

Rapid Discourse-Referential Processing

The higher-level processes associated with establishing reference, particularly those that require consulting one's model of the prior discourse, are frequently assumed to be rather slow, as compared to the lower-level syntactic and sentence-semantic aspects of comprehension. However, in Experiment 1 (see Figure 1), our listeners needed at most only 300–400 ms to detect, at some level, a difference between, say, “the girl” in a discourse that had introduced a single girl or that had introduced *two* girls. Also, the results of Experiment 2 suggest that referential ambiguity can perhaps be detected even earlier than that. Given that all known sentence-semantic and -syntactic ERP effects emerge within

some 500 ms after critical word onset (for review, see Brown, Hagoort, et al., 2000), this discourse-referential ERP effect is thus clearly “in the same ballpark.”

The critical nouns in Experiments 1 and 2 were singular Dutch nouns whose plural form usually involves the addition of a plural-marking suffix to a singular stem, as in “meisje/meisjes” (“girl/girls”). Given that noun number is marked as a suffix, one might reasonably ask how rapidly referential ambiguity can have been detected for our critical words *in principle*. After all, if subjects can only find out that a plural suffix is missing after having heard the complete stem, and with critical nouns having an average singular stem length of about 440 ms, would not the ERP effect need to emerge somewhat later than we now see it emerge?

Note, however, that there is great variation in the duration of the critical nouns used in our study, with some nouns taking less than 200 ms to unfold. Such short nouns may have contributed to an early onset of the effect in the grand average ERP waveforms even if one would have to wait for the end of the word. Furthermore, as the exact acoustic realization of a Dutch word stem usually varies with whether a suffix is present or not, listeners very often do *not* need to wait that long. For instance, 14 of the 240 critical nouns in Experiment 1 would have a different *stem vowel* in their plural form (e.g., “lid/leden,” “member/members”), and many more of them have different stem-final consonants in the plural. Next to such salient changes, the addition of a suffix also affects the noun stem in more subtle ways, partly because of the smearing (“co-articulation”) of adjacent speech sounds, and partly because of other systematic effects (e.g., stem shortening when a suffix is added; Jongman, 1998). Listeners have been shown to be sensitive to these changes (Salverda, Dahan, & McQueen, 2002). Finally, in sufficiently well-delimited referential domains, the onset of a de-accented noun may actually allow the listener to begin to identify a plausible referent well before the noun has become acoustically unique (Dahan, Tanenhaus, & Chambers, 2002). In all, these observations suggest that our subjects will frequently have been able to detect that a noun is in singular form well before the end of this noun.

We do not take our results to suggest that referential ambiguity is *always* detected within some 300–400 ms. There is good evidence, for example, that such detection depends on the degree to which each of the candidate referents is in focus (for reviews, see Garnham, 2001; Myers and O'Brien, 1998), on the degree to which comprehension of the text actually requires a referential expression to be resolved (Levine, Guzmán, & Klin, 2000), and on whether the noun is de-accented or not (Dahan et al., 2002). Finally, as will be clear from the previous discussion, the moment at which an unfolding noun “betrays” its number depends on a wide variety of factors, including where the language at hand codes a noun for its number (e.g., suffix, prefix, or other), the duration of the spoken noun at hand, and the way a noun's stem changes with pluralization. These “acoustic” sources of variability in when the critical information becomes available may, because of the resulting latency jitter, well account for the fact that none of the presently reported ERP waveforms to spoken input displays a crisp onset of the frontal negative shift. The fact that a *written* word does elicit a negative shift with a relatively crisp onset (best seen in van Berkum, Brown, et al., 1999a, Figure 1b) is consistent with such an account. In all, the currently observed early onset of the referential ERP effect should primarily be taken as an existence proof, revealing that

³The signal-to-noise ratio of the ERPs at hand, averaged across at most 30 trials per condition, is substantially lower than that of the ERPs obtained in Experiment 1, which were averaged across some 120 trials per condition. We therefore refrain from a statistical comparison of effects across the two experiments.

discourse-dependent referential ambiguity *can* be detected within that short amount of time.⁴

Apart from its rapid emergence, the referentially induced ERP effect observed with spoken language is also “immediate” in suggesting that reference is established *incrementally*, that is, at each relevant word coming in. In Dutch, a referentially ambiguous noun phrase can always be extended by a post-nominal modifier that supplies additional information, such as a relative clause (“the girl that was waiting”). In principle, the comprehension system might thus delay its attempt to establish reference until later sentential input signals that the noun phrase is unequivocally complete (cf. Perfetti & Britt, 1995). What our present and earlier (van Berkum, Brown, et al., 1999a) findings suggest is that under the conditions tested in these studies, the comprehension system does *not* do this, and instead initiates sufficient referential processing at the head noun to at least determine, within some 300–400 ms after noun onset, whether it is referentially unique or not.

Our ERP evidence for incremental and very rapid referential processing is consistent with delays observed in the reading of referentially ambiguous words (e.g., Corbett, 1984; for an overview, see Garnham, 2001, or Myers & O’Brien, 1998).⁵

⁴Although referentially ambiguous spoken nouns elicited frontally dominant sustained negative shifts in Experiment 1 as well as in Experiment 2, the effects were not completely identical (compare Figure 3 to Figure 1). Several differences between the experiments may account for this. One is that the critical items used in Experiment 2 were only a subset of those used in Experiment 1. In addition, they were recorded anew for Experiment 2. On average, the two sets of critical nouns will thus have somewhat different acoustic realizations, possibly including differences (in mean and variability) in how much of the noun was needed to detect its singular number. In addition, there will be accidental differences in the degree of referential ambiguity generated by specific texts (as well as their specific spoken recording). A final obvious difference is that whereas 50% of the stories used in Experiment 1 contained a referentially ambiguous noun, only 12.5% of the stories used in Experiment 2 contained such nouns. Note once again, however, that if the ERP effect obtained in Experiment 1 had hinged on increased strategic processing induced by a higher proportion of referentially ambiguous trials, one would have expected to find a smaller and/or later effect of referential ambiguity in Experiment 2, and possibly even no effect at all. Instead, what the waveforms suggest is a bigger and somewhat earlier effect. Furthermore, if the presence of a higher proportion of ambiguous trials in Experiment 1 would matter, one might expect to see their impact develop throughout the four blocks of Experiment 1. This was not the case either. In all, we cannot as yet identify the exact cause of the differences between the results of Experiments 1 and 2. What remains is that in both experiments referentially ambiguous nouns elicited a frontally sustained negative shift, and no clear N400 effect.

⁵Such convergence is not a necessity, for in a (noncumulative) word-by-word self-paced reading experiment recently conducted in our laboratory with a subset of the same materials, referentially ambiguous nouns did *not* slow down reading. In the experiment, 24 subjects read 20 stories with a referentially unique noun and 20 stories with a referentially ambiguous noun, amid 80 unrelated stories. Mean reading times in the one- and two-referent conditions were 253 and 250 ms on the critical singular noun, $F = 0.39$, $MSE = 574$, $p = .539$, 253 and 248 ms on the next word, $F = 1.32$, $MSE = 295$, $p = .262$, and 233 and 235 ms on the word after that, $F = 0.32$, $MSE = 320$, $p = .578$. The design did not allow us to look for a (“spilled-over”) reading time effect on subsequent words. In the same experiment, however, semantically coherent words that were somewhat less predictable given the wider discourse did result in slower reading than highly predictable words (343 vs. 299 ms respectively, $F = 11.30$, $MSE = 4142$, $p = .003$); the absence of a referential ambiguity effect can therefore not be due to an accidentally inattentive group of subjects. Under the conditions tested here, brain potentials thus appear to be more sensitive to a referential processing problem than self-paced reading times.

Our findings also converge with head-mounted eye-tracking results obtained in so-called visual-world experiments (e.g., Altmann & Kamide, 1999; Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Dahan et al., 2002; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Salverda et al., 2002; Sedivy, Tanenhaus, Chambers, & Carlson, 1999). In a typical visual-world experiment, the subject is repeatedly instructed to select and move a specific object amid an array of objects, and his or her eye fixation behavior is recorded as the spoken instruction unfolds. Such experiments have shown that listeners begin to fixate the intended referent of a critical spoken noun long before the noun has been fully heard (for recent findings and review, see Dahan et al., 2002). The convergence of these results with our own suggests that at least one basic aspect of referent identification is insensitive to whether the candidate referents are actually “out there” in the visual scene or represented as entities in one’s mental model of the discourse. Related to this, it apparently also does not matter whether language users need the output of referential processing to physically grasp or point to an object in front of them (the typical task in visual-world experiments) or whether they can sit back and just listen to a short story, without any additional response task at all. What this suggests is that we are looking at a relatively robust, task-independent aspect of the language comprehension system.

Selective Tracking of Referential Processes

In addition to timing, there is also information in the *nature* of the referentially induced ERP effect, a frontally dominant and sustained negative shift. First of all, the effect is qualitatively different from both a P600/SPS effect and an N400 effect. Our material had been designed to elicit not only discourse-level referential ERP effects, but also discourse-induced N400 effects associated with semantic problems (Van Berkum, Hagoort, et al., 1999; Van Berkum, Zwitserlood, et al., in press), as well as P600/SPS effects associated with problems in syntactic analysis (Van Berkum, Brown, et al., 1999a; Brown, van Berkum, et al., 2000). Neither of these effects, elicited in the same group of subjects, resembles the frontally sustained negativity induced by discourse-dependent referential ambiguity. The syntactic P600/SPS effects were, as their name suggests, positive deflections. Discourse-dependent semantic anomalies did elicit a negative deflection, but as illustrated for spoken language input in Figure 2C, this was a classical N400 effect with a well-articulated peak at about 400 ms after word onset and a clear centro-parietal maximum over the scalp (for details, see Van Berkum, Hagoort, et al., 1999; Van Berkum, Zwitserlood, et al., in press).

It is of particular interest to note that the ERP effect of referential ambiguity is not simply a standard N400 effect. Within the domain of language processing, the N400 component is commonly taken to reflect some aspect of the processes that relate the meaning of a particular word to a higher-order conceptual interpretation of the unfolding message, with words that are more difficult to relate to this context eliciting larger N400 components (for a recent review, see Brown, Hagoort, et al., 2000). As a referentially ambiguous expression might be more difficult to integrate into a high-level conceptual interpretation of the unfolding utterance than a referentially successful one, referential ambiguity could, in principle, also have elicited a standard centro-parietally distributed N400 effect.

In view of the relatively transient posterior negativity obtained in Experiment 1 at about 300–600 ms after word onset, we cannot rule out the possibility that the referentially induced

sustained frontal negativity is sometimes accompanied by a small N400 effect. However, the fact that referentially ambiguous nouns do not *necessarily* elicit an N400 effect—as becomes evident from the data of Experiment 2—seems to imply that the N400 component reflects only *some* of the dimensions along which an incoming word can be said to fit the conceptual context. Furthermore, the fact that problems with the “referential fit” of a word to its conceptual context robustly elicits a qualitatively different frontal ERP effect suggests that the processing of a referentially problematic interpretation, such as hearing “he saw the girl” with two girls in the context, is in at least some ways qualitatively different from that of a semantically problematic interpretation, such as when hearing “he sailed the girl.”

Note that it is, in fact, not immediately obvious why this should be so, because semantics (sense) and reference are inextricably intertwined aspects of language. One could easily argue, for instance, that hearing “the girl” with two referents in context is problematic because the semantics of this NP is not specific enough. Also, the semantic problem with something like “he sailed the girl” ultimately depends on reference, in that there would be nothing wrong with it in a context in which we know “the girl” to refer to a beloved boat instead of a human being. Nevertheless, what our ERP results suggest is that in on-line language comprehension the two types of problems have at least partially distinct processing implications.⁶

Although within the context of this work we sometimes refer to the observed negativity elicited by referentially ambiguous nouns as “the referentially induced negative shift” or “the referentially induced ERP effect,” it is important to reemphasize that we do *not* claim that this ERP effect is uniquely associated with referential ambiguity in language comprehension. As mentioned in the Introduction, the frontally sustained negative shift reported for written and now also spoken referentially ambiguous nouns is very similar to sustained ERP effects observed under conditions of increased memory load in language comprehension (e.g., Fiebach et al., 2001; Friederici et al., 1996; King & Kutas, 1995; Kluender & Kutas, 1993; Kutas, 1997; Müller et al., 1997; Münte et al., 1998; Rösler et al., 1997, 1998; Vos et al., 2001) as well as in nonlinguistic processing tasks (e.g., Donaldson & Rugg, 1999; Rösler et al., 1993; Rugg & Allen, 2000).

It is not difficult to imagine why referential ambiguity might be a memory-demanding situation. For one thing, referential ambiguity may trigger additional retrieval from episodic discourse memory, associated with a search for less obvious clues that might help to infer the most plausible referent (Myers

& O’Brien, 1998). Alternatively, referential ambiguity may require the system to actively maintain two candidate fillers for an unresolved single referential slot in working memory (for an account of referentially induced working memory load in sentence comprehension, see Gibson, 1998). The latter would explain why the ERP effect of referential ambiguity resembles the ERP effect elicited by various other types of expressions that impose a higher load on working memory, such as (a) object-relative clauses like “The fireman *who the cop speedily rescued* sued the city over working conditions” (King & Kutas, 1995; Kutas, 1997; Müller et al., 1997) in which the reader or listener must deal with “the cop” without yet knowing what to do with “the fireman,” (b) temporal expressions like “*Before* the psychologist submitted the article, the journal changed its policy” (Münte et al., 1998) in which the information supplied in the first phrase does not describe what actually happened first, or (c) expressions like “The *pitcher* fell down and broke/cursed” that contain a lexically ambiguous word (Hagoort & Brown, 1994). Our current data do not allow us to discriminate between an exclusively language-specific interpretation—limited to referential ambiguity and perhaps other referential problems—on the one hand and any of these more generic memory-related hypotheses on the other.⁷

In the domain of language processing, memory-related sustained frontal negativities are sometimes referred to as left anterior negativities or LAN effects, because of their frequent left-anterior maximum over the scalp (e.g., Kluender & Kutas, 1993). Some particularly early left anterior negativities have however also been claimed to directly reflect aspects of early syntactic processing (e.g., Friederici et al., 1996; for a review, see Friederici, 1998). It is as yet unclear to what extent the early and later LAN effects all reflect the same set of neuronal generators (for recent discussions, see Brown & Hagoort, 2000; Brown, Hagoort, et al., 2000; Friederici, 1998; and Hagoort et al., 1999). The exact relationship between the LAN family of effects and our (bilaterally and globally distributed) referentially induced sustained negative shift thus remains to be established. In view of its selective emergence (in Experiment 2 only), we also as yet refrain from speculating on the nature of the transient negativity visible around 100–150 ms in Figure 3.

The current findings support a number of conclusions. First, discourse-level referential ambiguity can elicit very rapid processing effects in comprehension, within only a few hundred milliseconds after onset of the referring noun. Second, the observed effects of referential ambiguity are independent of whether the referring expressions are presented visually at 600 ms/word or as fully connected natural speech. This implies that whatever is reflected by our referentially induced ERP effect is a relatively robust ingredient of normal language comprehension. Third, the fact that referentially ambiguous nouns elicit a frontally sustained negative deflection suggests that at least some of the processing consequences of referential ambiguity may involve an increased demand on memory resources. Finally, our discourse-induced referential ERP effect is qualitatively very different from both discourse-induced semantic N400 effects and discourse-induced syntactic P600/SPS effects, in spoken- as well

⁶A potentially useful distinction made in the anaphoric processing literature is between identifying the appropriate referent of some referring expression (“the girl”), and integrating what is already known about this referent with what is now supplied with the anaphoric expression (“David told...”). In terms of this so-called *bonding* versus *resolution* distinction (Garrod & Sanford, 1994; Garrod & Terras, 2000; Sanford, 1985), one speculative possibility worthwhile exploring is that, with referring expressions, the sustained frontal negative shift may reflect processing consequences associated with a *bonding* problem, whereas N400 effects might instead reflect interpretive problems that arise during semantic *resolution*. This obviously raises the issue why the onset of the sustained frontal negative shift appears to be later than the onset of the standard N400 effect (typically around 150–250 ms after noun onset). Note, however, that it need not be the case that coindexation processes are necessarily completed *before*, and fully independent of, integration processes. Also, the referentially induced ERP effect may well reflect the *downstream consequences* of a bonding problem (e.g., renewed memory search, increased working memory load), rather than the problem itself.

⁷Ongoing research in our laboratory (partly reported by Van Berkum, 1999) suggests that a referentially ambiguous pronoun, as in “David shot at John as *he* jumped over the wall,” also elicits a sustained frontal negativity. The present ERP effect thus need not reflect a process that is specific to associating *nouns* with their referents.

as written-language comprehension. This indicates that the processing implications of a difficulty with the referential analysis of language input differ from those associated with the semantic and the syntactic analysis. Moreover, and precisely because of

this dissociation, our work suggests that event-related brain potentials can be used to *selectively* keep track of three major processes involved in the comprehension of an unfolding piece of discourse.

REFERENCES

- Altmann, G. T. M. (1988). Ambiguity, parsing strategies, and computational models. *Language and Cognitive Processes*, 3(2), 73–97.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264.
- Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, 76, B14–B26.
- Brown, C. M., & Hagoort, P. (2000). On the electrophysiology of language comprehension: Implications for the human language system. In M. W. Crocker, M. Pickering, & C. Clifton, Jr. (Eds.), *Architectures and mechanisms for language processing* (pp. 213–237). Cambridge, UK: Cambridge University Press.
- Brown, C. M., Hagoort, P., & Kutas, M. (2000). Postlexical integration processes in language comprehension: Evidence from brain-imaging research. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 881–895). Cambridge, MA: MIT Press.
- Brown, C. M., Van Berkum, J. J. A., & Hagoort, P. (2000). Discourse before gender: An event-related brain potential study on the interplay of semantic and syntactic information during spoken-language understanding. *Journal of Psycholinguistic Research*, 29, 53–68.
- Crain, S., & Steedman, M. (1985). 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* (pp. 320–358). Cambridge, UK: Cambridge University Press.
- Corbett, A. T. (1984). Pronominal adjectives and the disambiguation of anaphoric nouns. *Journal of Verbal Learning and Verbal Behavior*, 23, 683–695.
- Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, 47, 292–314.
- Donaldson, D. I., & Rugg, M. D. (1999). Event-related potential studies of associative recognition and recall: Electrophysiological evidence for context dependent retrieval processes. *Cognitive Brain Research*, 8, 1–16.
- Eberhard, K., Spivey-Knowlton, M., Sedivy, J., & Tanenhaus, M. (1995). Eye movements as a window into real-time spoken language processing in natural contexts. *Journal of Psycholinguistic Research*, 24, 409–436.
- Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. K. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, 39, 133–146.
- Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348–368.
- Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research*, 30, 321–338.
- Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.
- Friederici, A. D. (1998). The neurobiology of language comprehension. In A. D. Friederici (Ed.), *Language comprehension: A biological perspective* (pp. 263–301). Berlin: Springer.
- Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1219–1248.
- Garnham, A. (2001). Mental models and the interpretation of anaphora. Hove, UK: Psychology Press.
- Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language*, 42, 526–544.
- Garrod, S. C., & Sanford, A. J. (1994). Resolving sentences in discourse context: How discourse representation affects language understand- ing. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 675–698). New York: Academic Press.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Hagoort, P., & Brown, C. M. (1994). Brain responses to lexical–ambiguity resolution and parsing. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45–80). Hillsdale, NJ: Erlbaum.
- Hagoort, P., & Brown, C. M. (2000a). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia*, 38, 1518–1530.
- Hagoort, P., & Brown, C. M. (2000b). 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, 1531–1549.
- Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439–483.
- Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 273–316). Oxford: Oxford University Press.
- Jackendoff, R. (1999). The representational structures of the language faculty and their interactions. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 37–79). Oxford: Oxford University Press.
- Jongman, A. (1998). Effects of vowel length and syllabic structure on segment duration in Dutch. *Journal of Phonetics*, 26, 207–222.
- Kempen, G. A. M. (2001). *Human grammatical coding*. Manuscript.
- King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7, 376–395.
- Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5, 196–214.
- Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, 8, 533–572.
- Kutas, M. (1997). Views on how the electrical activity that the brain generates reflects the functions of different language structures. *Psychophysiology*, 34, 383–398.
- Kutas, M., Federmeier, K. D., Coulson, S., King, J. W., & Münte, T. F. (2000). Language. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 576–601). Cambridge, UK: Cambridge University Press.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). New York: Academic Press.
- Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT Press.
- Levine, W. H., Guzmán, A. E., & Klin, C. M. (2000). When anaphor resolution fails. *Journal of Memory and Language*, 43, 594–617.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, 24, 469–488.
- Müller, H. M., King, J. W., & Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research*, 5, 193–203.
- Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, 395, 71–74.

- Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, 26, 131–157.
- Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3, 151–165.
- Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory and Cognition*, 25, 273–285.
- Osterhout, L., & Hagoort, P. H. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King, and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes*, 14, 1–14.
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785–806.
- Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes*, 8, 413–437.
- Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, 34, 739–773.
- Perfetti, C. A., & Britt, M. A. (1995). Where do propositions come from? In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension* (pp. 11–34). Hillsdale, NJ: Erlbaum.
- Pinker, S. (1994). *The language instinct*. London/New York: Penguin.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Rösler, F., Heil, M., & Glowalla, U. (1993). Monitoring retrieval from long-term memory by slow event-related brain potentials. *Psychophysiology*, 30, 170–182.
- Rösler, F., Heil, M., & Röder, B. (1997). Slow negative brain potentials as reflections of specific modular resources of cognition. *Biological Psychology*, 45, 109–141.
- Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language*, 38, 150–176.
- Rugg, M. D., & Allan, K. (2000). Memory retrieval: An electrophysiological perspective. In M. D. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 805–816). Cambridge, MA: MIT Press.
- Salverda, A. P., Dahan, D., & McQueen, J. M. (2002). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. Manuscript submitted for publication.
- Sanford, A. J. (1985). Aspects of pronoun interpretation: Evaluation of search formulations of inference. In G. Rickheit & H. Strohner (Eds.), *Inferences in text processing* (pp. 183–204). Amsterdam: North-Holland.
- Schmitt, B. M., Lamers, M. J. A., & Münte, T. F. (in press). Electrophysiological estimates of biological and syntactic gender violation during pronoun. *Cognitive Brain Research*.
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71, 109–147.
- Streb, J., Rösler, F., & Hennighausen, E. (1999). Event-related responses to pronoun and proper name anaphors in parallel and nonparallel discourse structures. *Brain and Language*, 70, 273–286.
- Tanenhaus, M. K., & Trueswell, C. (1995). Sentence comprehension. In J. L. Miller, & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 217–262). San Diego: Academic Press.
- Van Berkum, J. J. A. (1999). Establishing reference in language comprehension. In V. Van Geenhoven, & W. Warner (Eds.), *Max Planck Institute for Psycholinguistics Annual Report 1999* (pp. 35–37). Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics.
- Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999a). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, 41, 147–182.
- Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999b). When does gender constrain parsing? Evidence from ERPs. *Journal of Psycholinguistic Research*, 28, 555–571.
- Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11, 657–671.
- Van, Berkum J. J. A., Kooijman, V., Brown, C. M., Zwitserlood, P., & Hagoort, P. (2002). *Do listeners use discourse-level information to predict upcoming words in an unfolding sentence? An ERP study*. Presented at the 9th Annual Meeting of the Cognitive Neuroscience Society (CNS-2002), San Francisco, April 14–16.
- Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (in press). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*.
- Van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 Effects. *Journal of Cognitive Neuroscience*, 13, 967–985.
- Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 394–417.
- Vos, S. H., Gunter, T. C., Kolk, H. H. J., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology*, 38, 41–63.

(RECEIVED May 30, 2002; ACCEPTED August 15, 2002)