

## Early Referential Context Effects in Sentence Processing: Evidence from Event-Related Brain Potentials

Jos J. A. van Berkum, Colin M. Brown, and Peter Hagoort

*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

An event-related brain potentials experiment was carried out to examine the interplay of referential and structural factors during sentence processing in discourse. Subjects read (Dutch) sentences beginning like “David told the girl that . . .” in short story contexts that had introduced either one or two referents for a critical singular noun phrase (“the girl”). The waveforms showed that within 280 ms after onset of the critical noun the reader had already determined whether the noun phrase had a unique referent in earlier discourse. Furthermore, this referential information was immediately used in parsing the rest of the sentence, which was briefly ambiguous between a complement clause (“ . . . that there would be some visitors”) and a relative clause (“ . . . that had been on the phone to hang up”). A consistent pattern of P600/SPS effects elicited by various subsequent disambiguations revealed that a two-referent discourse context had led the parser to initially pursue the relative-clause alternative to a larger extent than a one-referent context. Together, the results suggest that during the processing of sentences in discourse, structural and referential sources of information interact on a word-by-word basis. © 1999 Academic Press

*Key Words:* discourse context; referential ambiguity; parsing; syntactic ambiguity resolution; ERP.

When we read a book or listen to speech in our native language, we usually have a sense of immediate understanding, of recognizing and interpreting every word as soon as we see or hear it. Psycholinguistic experiments have to a large extent confirmed this intuition, by showing that as a sentence unfolds over time, every new word is related to the local sentence context within only a few hundred milliseconds, both in terms of its syntactic features (“parsing”) and in

terms of its semantics. To extract the syntactic and semantic structure of a given sentence is, however, only part of what it means to comprehend. Sentences almost invariably occur in discourse and can only be properly understood in the context of what has been said before (Clark, 1996). In a coherent text or conversation, the definite NPs in an utterance like *They told the girl that the house was gone*, for example, are all likely to refer to entities introduced before. Without identifying those referents, what is said cannot be related to what is already known, and sentence meaning is left in mid-air.

The present study examined two closely related aspects of processing sentences in discourse. First, we examined the speed and incrementality of the mechanisms involved in referent identification—do these also deliver their output within only a few hundred milliseconds after a relevant word or do they for some reason substantially lag behind the processes that recover syntactic and semantic structure? Second, can the results of referential processing—if delivered quickly enough—guide the further analysis of sentence *structure*, i.e., parsing?

We thank Karin Remmerswaal and John Haasen for help in running the experiment and the subsequent ERP analyses; Edith Sjoerdsma for lending her voice to the stimuli; Herbert Baumann, René de Bruin, Inge Doehring, John Nagen-gast, and Johan Weustink for technical assistance; and Randi Martin as well as several anonymous reviewers for helpful comments. Part of this work has been reported at the Fourth Annual Meeting of the Cognitive Neuroscience Society, Boston, March 23–25, 1997; the Sixth Winterconference of the Dutch Psychonomics Society, December 19–20, 1997; and the 11th Annual CUNY Conference on Human Sentence Processing, New Brunswick, March 19–21, 1998. The research was supported in part by Grant 400-56-384 from the Netherlands Organization for Scientific Research.

Address correspondence and reprint requests to Jos J.A. van Berkum, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. E-mail: [jos.van.berkum@mpi.nl](mailto:jos.van.berkum@mpi.nl). Fax: +31-24-3521213.

## THE TIME COURSE OF REFERENTIAL PROCESSING

The time course of referential processing has received considerable attention in research on anaphor resolution (see Garrod & Sanford, 1994, for an overview). Some findings suggest that readers begin to associate anaphoric definite nominal NPs with their linguistic antecedents very rapidly, while the head noun is being read (e.g., Dell, McKoon, & Ratcliff, 1983; Duffy & Rayner, 1990; Garnham, Oakhill, & Cain, 1997; Garrod, O'Brien, Morris, & Rayner, 1990). There are also indications, however, that the referential processing of such NPs may be delayed by one or several words (e.g., Garrod, Freudenthal, & Boyle, 1993; Greene, McKoon, & Ratcliff, 1992).

Consistent with the evidence for delays, one theory (Britt, Gabrys, & Perfetti, 1993; Perfetti & Britt, 1995) predicts that the sentence processor will only begin to look for a definite NP's referent when the word *following the noun* indicates that the NP can be closed. In a similar vein, it has been suggested that since definite NPs are by no means always used anaphorically, the sentence processor should not commit itself to a particular prior referent too early (Garrod, 1994; Garrod & Sanford, 1994).

We addressed this issue by asking people to read sentences such as *David told the girl that there would be some visitors* in a discourse context that introduced either a single referent for the critical singular definite article noun phrase (e.g., a single girl) or two. If people try to establish reference incrementally and if the results of this process become available rapidly enough, the critical noun might elicit differential processing effects as a function of whether the discourse supplies one or two eligible referents.

## REFERENTIAL CONTEXT EFFECTS IN PARSING

Whereas the evidence for incremental referential processing is still mixed, there is good evidence for incremental parsing, i.e., for the immediate syntactic assignment of every incoming word (see Mitchell, 1994, and Tanen-

haus & Trueswell, 1995, for surveys). In particular, if an incoming word is ambiguous because it can be assigned to the current partial phrase structure in more than one way, the parser provisionally commits itself, at least to some degree, to just one specific assignment (see Frazier & Clifton, 1996; and above surveys). For example, if a sentence fragment like *David told the girl that . . .* is presented in isolation, people initially tend to take the locally ambiguous word *that* to be the head of a subsequent complement clause, as in *David told the girl that there would be some visitors*, even though it could also be the head of a relative clause, as in *David told the girl that had been on the phone to hang up* (Altmann, Garnham, & Dennis, 1992; van Berkum, Brown & Hagoort, 1998).

An important unresolved issue in the parsing literature is whether such initial parsing preferences can change if the ambiguous sentence is presented in discourse context. Crain and Steedman (1985) pointed out, for example, that the use of a restrictive relative clause is much more felicitous in a discourse context that leaves the preceding NP referentially ambiguous (e.g., by having provided two equally plausible referents for *the girl*) than in a context that by itself already clearly suggests a single unique referent.

According to various context-sensitive theories of parsing, the parser can make immediate use of such information (e.g., Altmann, 1988; Crain & Steedman, 1985; Gibson, 1998; Lewis, 1993; Ni, Crain, & Schankweiler, 1996; Spivey-Knowlton & Tanenhaus, 1998; Kempen, 1998). According to so-called syntax-first theories of parsing, however, referential information is initially ignored, because—depending on the specific theory—the parser is architecturally constrained (cf. Fodor, 1983) to always first try the syntactic structure that is simpler (Ferreira & Clifton, 1986; Frazier, 1987; Frazier & Rayner, 1982; Rayner & Pollatsek, 1989; De Vincenzi & Job, 1995), encountered more often (Brysbart & Mitchell, 1996; Mitchell, Cuertos, Corley, & Brysbart, 1995), or the better option for other syntax-related reasons (e.g., Konieczny, Hemforth, & Voelker, 1994). Hybrid theories also exist, allowing for context-sensitive pars-

ing of some types of ambiguity, but assuming syntax-first parsing for others (e.g., Britt et al., 1993; Frazier & Clifton, 1996).

The empirical evidence is mixed (see Mitchell, 1994 and Spivey-Knowlton & Tanenhaus, 1994 for reviews). Following up on Crain and Steedman (1985), many psycholinguists have attempted to uncover referential context effects in parsing (e.g., Altmann, 1988; Altmann et al., 1992; Altmann, Van Nice, Garnham, & Henstra, 1998; Altmann & Steedman, 1988; Boland, 1997; Britt, 1994; Britt, Perfetti, Garrod, & Rayner, 1992; Clifton & Ferreira, 1989; Crain, Ni, Schankweiler, Conway, & Braze, 1996; Ferreira & Clifton, 1986; Konieczny, Hemforth, & Voelker, 1994; Mitchell, Corley, & Garnham, 1992; Murray & Liversedge, 1994; Ni et al., 1996; Rayner, Garrod, & Perfetti, 1992; Spivey-Knowlton & Sedivy, 1995; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Spivey-Knowlton & Tanenhaus, 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1996; Van Nice, Garnham, & Altmann, 1997). For many of the studies that revealed apparently early referential context effects, however, the exact interpretation is complicated by the fact that the parser's commitments were probed relatively late in the sentence and thus do not necessarily reflect the way the parser *initially* resolved the ambiguity (cf. Mitchell et al., 1992). And the studies that did not reveal *any* effects of referential context throughout the processing of the critical sentence are in principle open to a methodological concern about whether the particular contexts used in the experiment were sufficiently biasing (cf. Spivey-Knowlton & Tanenhaus, 1994).

In view of these concerns, a maximally informative experimental design is one in which parsing commitments are probed at the earliest possible word, and—should no context effects emerge at that critical position—one in which the referential context manipulation can be shown to have an effect at a later probe position. In a self-paced reading experiment designed like this, Mitchell et al. (1992) obtained a pattern of results that seemed to support a syntax-first account of parsing. Mitchell et al. presented target sentences with a complement/relative-

clause ambiguity (e.g., *The politician told the woman that . . .*) in one- and two-referent discourse contexts and probed the parser's commitments by disambiguating either immediately after the word *that* or several words downstream. Referential context had a marginal effect at the late probe position, suggesting that the parser was by then pursuing the context-supported analysis, i.e., a complement clause in one-referent context and a relative clause in two-referent context. But it had no effect immediately after *that*, leading Mitchell et al. to conclude that the parser was initially insensitive to discourse context.

The syntax-first account of these results has not gone unchallenged, in part because of the way Mitchell et al. probed the early commitments of the parser (Altmann, Garnham, & Henstra, 1994). Also, because the *late* referential effect was only marginally significant, the Mitchell et al. contexts were perhaps not sufficiently effective to elicit early effects (Spivey-Knowlton & Tanenhaus, 1994). Related to this, it is possible that the detection of early referential context effects requires a more sensitive on-line measure than self-paced reading latencies.

The goal of the present experiment was to look for early referential context effects in parsing with a maximally sensitive design. First, we used event-related brain potential (ERP) measures, which, as is discussed below, have proven to be highly—and quite selectively—sensitive to modulations of the parsing process. Second, we designed our Dutch sentence materials to probe parsing commitments immediately after the first word in the ambiguous region (and even earlier, as will be seen). Before we turn to ERPs, we first explain the critical features of our materials.

Like English, the grammar of Dutch leaves room for a local complement/relative-clause ambiguity, as illustrated by example sentences (1) and (2) in the lower half of Table 1. Note that the local ambiguity arises at *dat* and is resolved by the very next word, where the expletive pronoun *er* disambiguates toward a complement clause and the auxiliary *had* disambiguates toward a restrictive relative clause

TABLE 1

Example Item Paradigm with Early Expletive/Auxiliary Disambiguation (see Method Section)

### One-Referent context

*David had de jongen en het meisje (de vrouw) gezegd hun kamer voor de lunch op te ruimen. Maar de jongen had de hele ochtend liggen slapen, en het meisje (de vrouw) had voortdurend zitten bellen.*

David had told the boy and the girl (the woman) to clean up their room before lunch time. But the boy had stayed in bed all morning, and the girl (the woman) had been on the phone all the time.

### Two-Referent context

*David had de twee meisjes (vrouwen) gezegd hun kamer voor de lunch op te ruimen. Maar het ene meisje (de ene vrouw) had de hele ochtend liggen slapen, en het andere (de andere) had voortdurend zitten bellen.*

David had told the two girls (women) to clean up their room before lunch time. But one of the girls (women) had stayed in bed all morning, and the other had been on the phone all the time.

#### (1) Early-complement target

<i>David vertelde</i>	<i>het meisje</i>	<i>dat</i>	<b><i>er</i></b>	<i>visite kwam.</i>
David told	the girl <sub>NEU</sub>	that <sub>COMPL</sub>	<b>there</b>	would be some visitors.

#### (2) Early-relative target

<i>David vertelde</i>	<i>het meisje</i>	<i>dat</i>	<b><i>had</i></b>	<i>zitten bellen</i>	<i>op te hangen.</i>
David told	the girl <sub>NEU</sub>	that <sub>RELPR(NEU)</sub>	<b>had</b>	been phoning	to hang up.

#### (3) Immediate-complement target

<i>David vertelde</i>	<i>de vrouw</i>	<b><i>dat</i></b>	<i>er</i>	<i>visite kwam.</i>
David told	the woman <sub>COM</sub>	<b>that<sub>COMPL</sub></b>	there	would be some visitors.

#### (4) Immediate-relative target

<i>David vertelde</i>	<i>de vrouw</i>	<b><i>die</i></b>	<i>had</i>	<i>zitten bellen</i>	<i>op te hangen.</i>
David told	the woman <sub>COM</sub>	<b>that<sub>RELPR(COM)</sub></b>	had	been phoning	to hang up.

*Note.* All disambiguating words are in bold. The two context stories have the critical referents denoted by a neuter gender noun (for early-disambiguation targets) with the alternative common gender noun (for immediate-disambiguation targets) in brackets. COMPL = complementizer; RELPR(NEU) = relative pronoun for neuter gender nouns; RELPR(COM) = relative pronoun for common gender nouns. Translations are approximate, to preserve overall fluency and meaning.

(in written Dutch, nonrestrictive relative clauses require a comma after the noun).

In the experiment, we presented these sentences in one- and two-referent context stories such as illustrated in the upper half of Table 1 (cf. Crain & Steedman, 1985). Syntax-first and context-sensitive parsing accounts make different predictions for the impact of such discourse contexts. Under a syntax-first account, the referential status of a preceding NP will not have any implications for how the parser initially handles the structural ambiguity arising at *dat*, for in both cases the parser will simply prefer the less complex (or more frequent) complement-clause analysis. In context-sensitive accounts, in contrast, the referential status of the critical NP does matter. Whereas a referentially successful NP will lead the parser to initially pursue the more felicitous complement-clause analysis of *dat*, a referentially ambiguous NP

will lure it, at least to a larger extent, into pursuing the more felicitous relative-clause analysis. A complement-clause disambiguation should thus generate more parsing problems in a “garden-pathing” two-referent context than in a supportive one-referent context, whereas a relative-clause disambiguation should generate the opposite result, i.e., more parsing problems in the now “garden-pathing” one-referent context than in the now supportive two-referent context. Note that we phrase these as *relative* effects because to the extent that referential context is not the only factor that determines parsing preferences, the use of, for instance, a two-referent context will not necessarily result in an *absolute* preference for the relative clause over the complement clause.

Sentences like (1) and (2) probe the commitment of the parser at the same position that Mitchell et al. (1992) took to be the earliest

possible one. To probe for referential effects even earlier than that, though, we exploited a feature of Dutch that is not available in English. Compare, in Table 1, sentences (3) and (4) to (1) and (2). Note, first of all, that the Dutch complement/relative-clause ambiguity resembles the English ambiguity only for nouns with grammatically neuter gender (so-called “het-words” like *het meisje*, in (1) and (2), marked by the subscript NEU) because only the *neuter* form of the relative pronoun is identical to the complementizer *dat*. As shown in (3) and (4), so-called common gender nouns (or “de-words,” like *de vrouw*, marked by the subscript COM) require the relative pronoun form *die* instead. Linguistically speaking, the use of a common gender noun thus completely eliminates the local complement/relative-clause ambiguity.

From a (particular) *processing* point of view, however, the ambiguity may still arise in a sentence like (3)—if only very briefly. We actually began our design of the experiment with these so-called immediate-complement sentences because we had the intuition that when reading something like *David vertelde de vrouw dat . . .* in a discourse with two women in focus, the use of *dat* “feels” as if a gender agreement error has been made at that point. This informal observation suggested that the word *dat* was somehow briefly taken as a relative pronoun (and subsequently again rejected as such on the grounds of its incongruent gender).

To make sense of this observation, we envisaged the following chain of events. (1) When processed as part of *David vertelde de vrouw dat . . .*, the lexically ambiguous wordform *dat* makes available two morpholexical entries: the generic complementizer *dat*<sub>COMPL</sub> and the relative pronoun *dat*<sub>RELPR(NEU)</sub>. (2) The parser initially ignores gender agreement and therefore considers two candidate analyses, namely [<sub>s</sub>David vertelde [<sub>NP</sub> de vrouw] [<sub>CC</sub> *dat*<sub>COMPL</sub> . . .]] as well as [<sub>s</sub>David vertelde [<sub>NP</sub> de vrouw [<sub>RC</sub> *dat*<sub>RELPR</sub> . . .]] . . .]. (3) In a two-referent discourse context, the parser provisionally commits itself to the relative-clause analysis of *dat*, at least to a larger extent than in a one-referent context. (4) Before processing the next

word, the preferred analysis is checked on gender agreement. (5) To the extent that a two-referent context has biased the parser to pursue the relative-clause analysis, the resulting gender agreement violation forces it to abandon its preferred analysis again.

Note that although the *referential* prediction is located in the third phase only, the account also assumes that structural ambiguity can be generated and resolved by the very same word. If we take the latter as a working assumption, a sentence like (3), in Table 1, provides us with, for all intents and purposes, an *immediate* probe of initial parsing commitments. Obviously, this assumption should be reexamined in the light of the results. Also, for unambiguous sentences like (4), the above account does not hold because *die* cannot be a complementizer in Dutch and therefore does not generate the complement/relative-clause ambiguity at any point in time. These so-called immediate-relative sentences will serve a variety of control purposes.

#### PREDICTIONS FOR EVENT-RELATED BRAIN POTENTIALS

To study the time course of basic referential processes and their impact on subsequent parsing in discourse, we made use of event-related brain potential (ERP) methodology. In the domain of language processing, two different brain responses have been particularly informative so far: the N400 (Kutas & Hillyard, 1980) and the P600/SPS (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992; the P600/SPS label combines the two earlier alternative names for this brain response, “P600”, and “Syntactic Positive Shift”). The N400 is a distinct negative deflection in the ERP waveform, typically onsetting around 200 ms and peaking around 400 ms after word onset, and commonly associated with semantic aspects of word and sentence processing (see Brown & Hagoort, in press; Kutas & Van Petten, 1994; Osterhout & Holcomb, 1995; and Osterhout, McLaughlin, & Bersick, 1997, for overviews). The P600/SPS is a positive deflection in the waveform with a typical onset around 500 ms after word onset and a duration of several hundred milliseconds, thought to be associated, in

the language domain, with aspects of *syntactic* processing (Hagoort et al., 1993; Osterhout & Holcomb, 1992; Osterhout & Hagoort, 1999). The P600/SPS has been elicited by a wide variety of local grammatical violations (e.g., of phrase structure, verb subcategorization, number agreement, and gender agreement), in Dutch, English, and German (e.g., Ainsworth-Darnell, Shulman, & Boland, 1998; Friederici, Hahne, & Mecklinger, 1996; Hagoort & Brown, 1994, 1997, 1998; Hagoort et al., 1993; Neville et al., 1991; Osterhout & Holcomb, 1992, 1993; Osterhout, McKinnon, Bersick, & Corey, 1996; Osterhout & Mobley, 1995). Importantly, the P600/SPS is not only elicited by words that unrecoverably *violate* the syntax (e.g., *\*The children throws the toy on the floor*), but also by words that signal that the *preferred* syntactic analysis can no longer be maintained (e.g., Osterhout, Holcomb, & Swinney, 1994; Brown, Hagoort, & Vonk, 1997).

The latter regularity allowed us to formulate highly specific ERP predictions regarding discourse context effects in parsing. First, to the extent that a two-referent context immediately biases the parser to pursue a relative-clause analysis, early-complement disambiguation [*er* in sentence (1)] should yield a P600/SPS in this context, as compared to a one-referent context. And under the working assumption discussed before, immediate-complement disambiguation [*dat* in sentence (3)] should do so too. Conversely, to the extent that a one-referent context biases the parser to pursue a complement-clause analysis, early-relative disambiguation [*had* in sentence (2)] should yield a P600/SPS in *this* context, as compared to a two-referent context. Note that we predict P600/SPS effects here (and not, for instance, N400 effects) precisely because we infer from earlier work that, within the language domain, the P600/SPS is a selective and reliable index of syntactic garden-pathing. Furthermore, note that for present purposes it does not matter whether the P600/SPS *specifically* reflects initial detection, subsequent diagnosis, or repair—all we need to know here is that it more generally reflects a syntactic dead end.

As for a syntax-first parser, it will always initially pursue the complement-clause analysis. Early-complement and immediate-complement

disambiguation should therefore allow it to maintain its analysis in both discourse contexts, yielding no differential P600/SPS effect across contexts. Also, early-relative disambiguation should force it to abandon its preferred analysis, again in both discourse contexts, and should thus not yield a differential P600/SPS effect across contexts either. In all, a context-contingent S600/SPS should not occur in any of these three critical sentence types.<sup>1</sup>

With respect to the time course of basic referential processes at the preceding critical noun, the existing literature points toward at least two possible effects in the ERP record. First, to the extent that referential ambiguity makes it more difficult to integrate the meaning of a definite NP into a higher-level semantic representation of the discourse, one might expect to see a larger N400 component in response to a referentially ambiguous noun than in response to a referentially successful noun (see Brown & Hagoort, in press and Osterhout & Holcomb, 1995 for an interpretation of N400 amplitude effects in terms of meaning integration; see St. George, Mannes, & Hoffman, 1994; 1997, and van Berkum, Hagoort, & Brown, in press for N400 effects in discourse). Second, referential ambiguity may be associated with a more extensive use of memory resources, which has been observed to elicit slow negative shifts in the EEG (Kutas & King, 1996; Rösler, Heil, & Glowalla, 1993).

## METHOD

### *Subjects*

The experiment was conducted with 24 native speakers of Dutch (20 female subjects,

<sup>1</sup> An event-related brain potentials experiment recently conducted in our lab with the same target sentence materials (van Berkum, Brown, & Hagoort, 1998) showed that in the absence of a discourse context, the relative clause disambiguation exemplified in sentence (2) of Table 1 elicited a significant P600/SPS as compared to the complement-clause disambiguation exemplified in sentence (1). This extends earlier reports of a default complement clause preference in the complement/relative clause ambiguity (e.g., Altmann et al., 1992) and at the same time shows that ERPs are sensitive enough to reveal such a preference in the current experiment should no context effects emerge.

TABLE 2

Example Item Paradigm with Early Nominative/Reflexive Pronoun Disambiguation

**One-Referent context**

*De aardige reus werd onderweg vergezeld door een elfje (een fee) en een kabouter. Het elfje (de fee) had zich vastgeklampt aan zijn bovenarm, de kabouter had zich genesteld in een comfortabele broekzak.*

On the road, the gentle giant was accompanied by an elf (fairy) and a goblin. The elf (fairy) had clung [itself] to his upper arm, the goblin had ensconced itself in a comfortable trouser-pocket.

**Two-Referent context**

*De aardige reus werd onderweg vergezeld door twee elfjes (feeën). Het ene elfje (de ene fee) had zich vastgeklampt aan zijn bovenarm, het andere (de andere) had zich genesteld in een comfortabele broekzak.*

On the road, the gentle giant was accompanied by two elfs (fairies). One of the elfs (fairies) had clung [itself] to his upper arm, the other had ensconced itself in a comfortable trouser-pocket.

**(1) Early-complement target**

<i>De reus waarschuwde</i>	<i>het elfje</i>	<i>dat</i>	<i>ze</i>	<i>niet moest vallen.</i>
The giant warned	the elf <sub>NEU</sub>	that <sub>COMPL</sub>	she	shouldn't fall off.

**(2) Early-relative target**

<i>De reus waarschuwde</i>	<i>het elfje</i>	<i>dat</i>	<i>zich</i>	<i>had vastgeklampt</i>	<i>niet te vallen.</i>
The giant warned	the elf <sub>NEU</sub>	that <sub>RELPR(NEU)</sub>	itself	had clung	not to fall.

**(3) Immediate-complement target**

<i>De reus waarschuwde</i>	<i>de fee</i>	<i>dat</i>	<i>ze</i>	<i>niet moest vallen.</i>
The giant warned	the fairy <sub>COM</sub>	that <sub>COMPL</sub>	she	shouldn't fall off.

**(4) Immediate-relative target**

<i>De reus waarschuwde</i>	<i>de fee</i>	<i>die</i>	<i>zich</i>	<i>had vastgeklampt</i>	<i>niet te vallen.</i>
The giant warned	the fairy <sub>COM</sub>	that <sub>RELPR(COM)</sub>	itself	had clung	not to fall.

*Note.* All disambiguating words are in bold. The two context stories have the critical referents denoted by a neuter gender noun (for early-disambiguation targets) with the alternative common gender noun (for immediate-disambiguation targets) in brackets. COMPL = complementizer; RELPR(NEU) = relative pronoun for neuter gender nouns; RELPR(COM) = relative pronoun for common gender nouns. Translations are approximate, to preserve overall fluency and meaning.

mean age 24, range 21–26 years), recruited from the Max Planck Institute subject pool. All had normal or corrected-to-normal vision and were right handed (10 subjects reported having left-handed relatives). None of the subjects had any neurological impairment, had experienced any neurological trauma, or used neuroleptics. Also, none of them had participated in any of the pretests (see below).

**Materials**

We constructed 60 early-complement, 60 early-relative, 60 immediate-complement, and 60 immediate-relative target sentences and created both a one-referent and a two-referent discourse context for each sentence. To avoid non-intended systematic biases in the preceding discourse, we made sure that each of the contexts could also be followed by any of the three other target sentence types (by actually con-

structing the latter too). Table 1 illustrates the resulting “full paradigm” for one item—a second example is given here in Table 2. Note, once again, that only *one* of the four target sentences created for each item was used in the actual experiment, in both a one- and a two-referent context. The full set of Dutch target sentences and context stories used in the experiment is available from the authors.

*Target sentences.* All target sentences followed the template <Subject-NP> <matrix-verb> <indirect-object-NP> *dat/die* <early-disambiguation> <remainder>. We used eight different singular past tense matrix verbs (*vertelde, waarschuwde, beloofde, verweet, voor-spelde, schreef, leerde, antwoordde*—roughly, *told, warned, promised, reproached, predicted, wrote, taught, answered*), each of which subcategorizes for an indirect-object NP followed by a direct-object complement clause headed by *dat*

(e.g., *X vertelde [het meisje]<sub>NP</sub> [dat Z]<sub>CP</sub>*). Each verb also had at least one relatively accessible alternative argument structure (e.g., with the direct object slot taken by an NP, a PP, or a complement clause not headed by *dat*), allowing us to wrap up relative-clause target sentences without using a *dat*-complement in the remainder and thus avoid a predominance of such constructions in the experiment.

In the 60 early-complement sentences the complement/relative-clause ambiguity was disambiguated in favor of a complement-clause reading by a word that immediately followed *dat*. In the 60 early-relative sentences it was resolved in favor of a (restrictive) relative clause, again by the next word after *dat*. Two different disambiguation schemes were used for these early-probe sentence types, illustrated in Tables 1 and 2 respectively. The expletive/auxiliary scheme used the expletive pronoun *er* (*there*) to resolve the ambiguous fragment as a complement clause and the auxiliary verb *had* (*had*, occasionally also the comparable Dutch auxiliaries *was*, *zou*, *lag*, *kon*, or *zat*) to resolve it as a relative clause. The nominative/reflexive pronoun scheme used the nominative third-person singular pronouns *ze* and *hij* (*she* and *he*) to resolve the ambiguous fragment as a complement clause and the reflexive pronoun *zich* (*herself* or *himself*) to resolve it as a relative clause. When using the latter disambiguation scheme, the referents of the subject and the direct object always differed in sex (e.g., *David told the girl that she . . .*), such that the nominative third-person pronoun could only be taken to refer to the direct object referent. Both schemes exploit particular constraints on Dutch phrase structure to achieve effective disambiguation (as confirmed by a sentence completion pretest; see below). In each set of 60 early-probe target sentences, half used the expletive/auxiliary scheme and half used the nominative/reflexive pronoun scheme.

For the third and fourth target sentence types, the neuter noun in indirect object position, e.g., *het meisje* (*the girl*<sub>NEU</sub>), was replaced by a semantically comparable noun of common gender, e.g., *de vrouw* (*the woman*<sub>COM</sub>). In the 60 immediate-complement sentences this com-

mon-gender noun was followed by *dat*, which in this construction disambiguated in favor of a complement-clause reading (because [<sub>NP</sub> *the woman*<sub>COM</sub> *that*<sub>RELPR(NEU)</sub> . . .] violates grammatical gender agreement). Finally, in the 60 immediate-relative sentences the common-gender noun was followed by the unambiguous relative pronoun *die* for such nouns.

To minimize eye movement artifacts in the EEG records and to provide good time-locking for later ERP analysis and interpretation, the target sentences were displayed with serial visual presentation, with every new word centered on the screen. Every critical word had at most 10 letters, all others had at most 12 letters, and the longest target sentence contained 12 words (average length = 10.5 words).

*Context stories.* Every story introduced the agent of the later target sentences (e.g., *David*) and two potential recipients, of some later message (e.g., two girls or a girl and a boy), after which the story provided discriminative qualifications for each (e.g., X had stayed in bed all morning, Y had been on the phone all the time) so that a restrictive relative clause could felicitously apply. We took great care to avoid foregrounding one candidate referent at the expense of the other one to such an extent that referential ambiguity might be eliminated. The two candidates were of roughly equal salience and qualified such that the later target sentence message (*David told X/Y that Z*) could equally plausibly be directed at either one. Also, in half of the stories in each condition the referent selected in the target sentence was mentioned first, and in half it was mentioned second. Furthermore, to not confound the effects of syntactic disambiguation with those of referential disambiguation, we carefully phrased our stories such that the early-relative disambiguation resolved the syntactic ambiguity without yet selecting either of the referents.

For each item, the one- and two-referent context stories only differed in the number of referents made available for the later critical NP (e.g., either a girl and a boy or two girls for *het meisje*; see Tables 1 and 2, upper half). To avoid a lexical priming confound, the noun that would become critical in the target sentence was

used twice in both the one- and the two-referent story.

To attenuate the visual strain associated with serial visual presentation and ERP recording conditions, context stories were presented auditorily. They were read by a female native speaker who used a normal speaking rate and intonation while minimizing prosodic differences between the two phrases that qualified the candidate referents as well as between the two consecutive stories of every item. All stories were digitally recorded in a single session, beginning with the one-referent story of half of the items (randomly selected) and with the two-referent story of the other half. A native listener monitored all story pairs for possibly disruptive prosodic differences, and problematic recordings were redone. The DAT-recordings were subsequently sampled at 16-kHz mono and stored on disk. An average recorded story lasted approximately 11 s.

*Pretests.* We evaluated the effectiveness of the expletive/auxiliary (*er/had*) and nominative/reflexive (*ze/zich*) disambiguation schemes in two written questionnaire pretests. In the first, 24 subjects were asked to complete 20 isolated target sentences that had been truncated right after *er*, *had*, *ze*, and *zich* with five items of each type. These sentences were randomly intermixed with 119 filler sentences truncated at other points in two different random orders. In the second pretest, 26 subjects were asked to complete 30 target sentences, again truncated right after *er*, *had*, *ze*, and *zich*, but now presented in a one- or two-referent context in two random orders.

When presented in isolation, targets truncated with *er* or *ze* both elicited 100% complement-clause continuations, and those truncated with *had* or *zich* elicited 97.5% and 99.2% relative-clause continuations respectively. When presented in a discourse context, targets truncated with *er* or *ze* elicited 92.3 and 99.3% complement-clause continuations respectively, and those truncated with *had* or *zich* elicited 98.0% and 97.3% relative-clause continuations respectively, all averaged across referential contexts. The latter factor did have its own small numerical effects, but never did an “unfavorable” con-

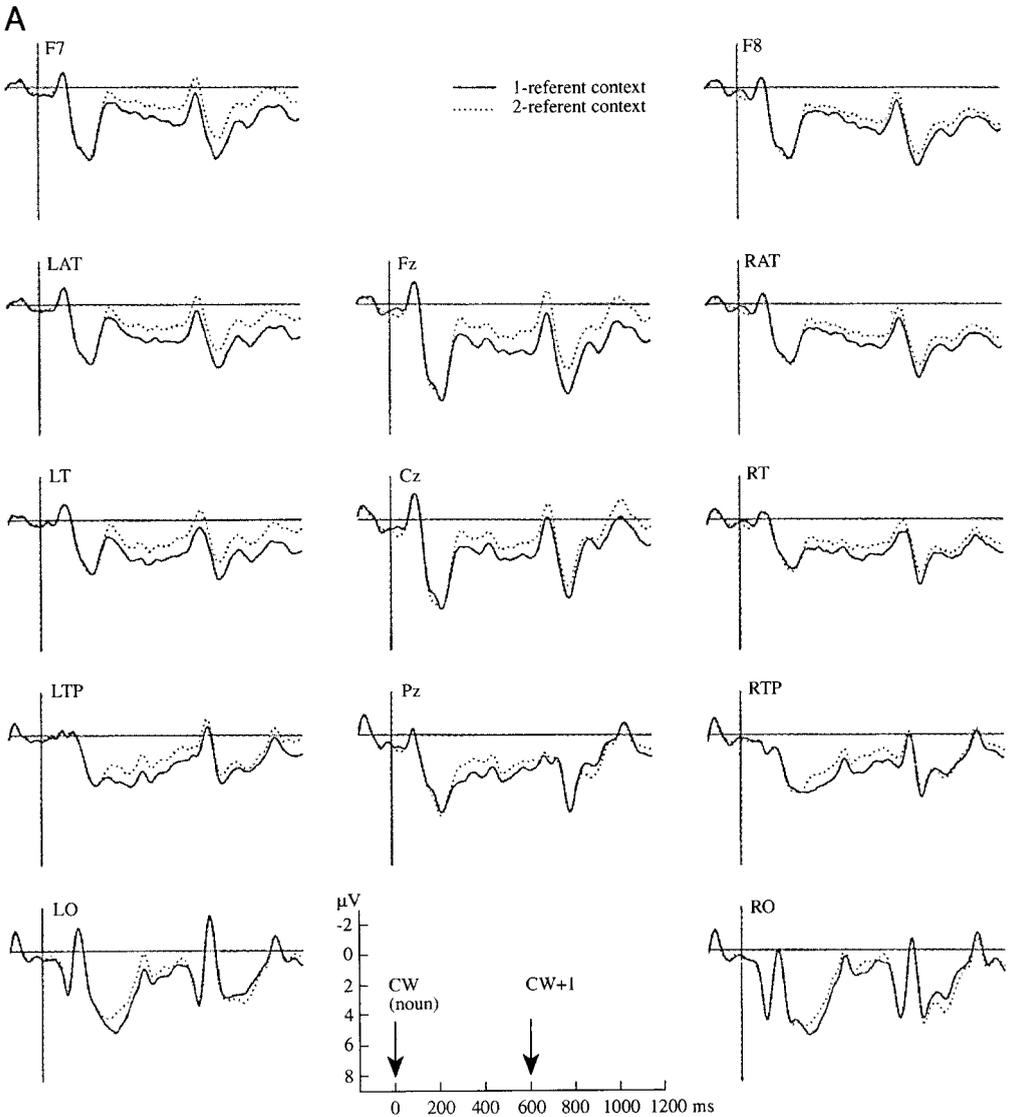
text cause a disambiguating construction to drop below 88.5% responses of the intended type. Thus, both disambiguation schemes operated as intended to a sufficient degree, even under opposing discourse biases.

*Randomizations.* We constructed two different trial lists, one for 12 subjects each. Every list contained the  $4 \times 60$  critical target sentences. For the first list, half of each target sentence subset was pseudorandomly paired with a one-referent context story, and the other half was paired with a two-referent story, completely crossing this factor with item disambiguation scheme and referent order. The 240 resulting trials were pseudorandomly mixed such that the same type of context, type of target sentence, disambiguation scheme, or referent order did not occur more than four times consecutively and such that trials of each type were matched on average list position. The second stimulus list was derived from the first by swapping one- and two-referent context stories only.

To address a research question orthogonal to current purposes (see van Berkum, Hagoort & Brown, in press), 40 of the 240 target sentences also contained a semantically odd word in the remainder of the sentence, i.e., after all immediate and early syntactic disambiguation. The oddity hinged on the semantic relation between a target sentence and its discourse context and was, for current purposes, merely expected to distract the subjects' attention from aspects of the design at hand. This factor was completely crossed with the referential manipulation.

### *Procedure*

Subjects were tested individually in a dimly lit sound-attenuating booth. They were seated in a comfortable reclining chair, instructed to move as little as possible, and told that they would be presented with a series of short episodes, with the first part of each played over headphones and the last sentence shown word-by-word on a computer screen. Subjects were asked to process each episode for comprehension. They were free to blink and move their eyes during the spoken part of each episode, but instructed to fixate on the screen and avoid all

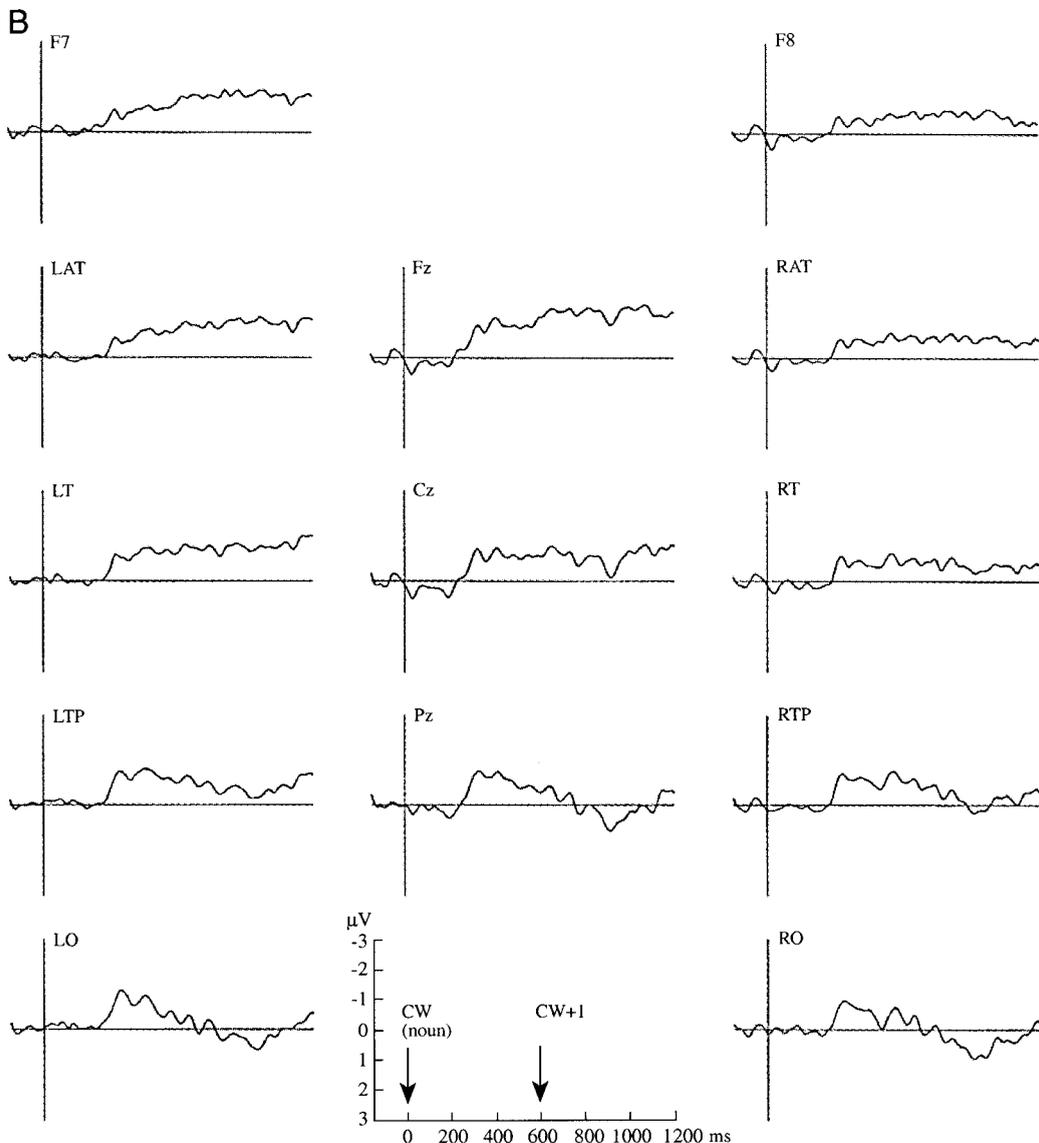


**FIG. 1A.** Referential ambiguity effect: Grand average ERPs elicited by singular nouns presented in a one-referent context (solid line) and a two-referent context (dotted line) across all sentence types. The onset of the critical noun is at 0 ms, the next word (CW + 1) follows at 600 ms. Each waveform averages over 24 subjects and approximately 2296 trials. Negativity is plotted upward in this and all following figures.

movement during the written part. No additional task demands were imposed.

Each trial began with a 300-ms auditory warning tone over headphones, followed by 700 ms of silence and then a spoken context story. At 1000 ms after offset of the spoken part, the visual presentation of the target sentence began, word-by-word in white lowercase Arial letters

(plus sentence-initial capitalization) against a dark background in the center of a VGA computer screen. Viewing distance was approximately 110 cm, and the largest word subtended a visual angle of about 3.1 degrees horizontally and 0.5 degrees vertically. Each word was presented for 300 ms, followed by a blank screen for another 300 ms, after which the next word



**FIG. 1B.** Referential ambiguity effect: two- minus one-referent difference waveforms.

appeared. The final word was presented together with a period sign, and 2500 ms after its offset the next trial began. To inform subjects when they were allowed to blink and move their eyes, an asterisk was displayed from 1600 ms after written target sentence offset to the offset of the next trial's spoken context story. After a short practice, the trials were presented in five blocks of 15 min, separated by rest periods.

#### *EEG Recording and Analysis*

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 mid-line 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

TABLE 3

Analyses of Variance (ANOVAs) of the Mean ERP Amplitude in the 300- to 600-ms Latency Range Following Onset of the Critical Noun

Source	<i>df</i>	<i>F</i>	<i>MS<sub>e</sub></i>	<i>p</i>
Omnibus ANOVA (13 electrodes)				
RC	1, 23	12.59	29.60	.002**
RC × Ge	1, 23	0.31	16.43	.586
RC × Cl	1, 23	0.02	13.78	.879
RC × El	2.8, 64.1	1.06	1.29	.369
RC × Ge × Cl	1, 23	0.37	13.22	.550
RC × Ge × El	3.5, 80.3	0.55	0.96	.679
RC × Cl × El	3.1, 70.8	2.41	0.93	.072
RC × Ge × Cl × El	3.1, 70.8	0.95	0.96	.422
Midline ANOVA (3 electrodes)				
RC	1, 23	7.12	17.16	.014*
RC × El	1.5, 33.9	0.79	0.98	.428
Lateral ANOVA (2 × 5 electrodes)				
RC	1, 23	14.80	17.20	.001**
RC × He	1, 23	2.56	1.82	.123
RC × El	1.4, 32.1	0.95	1.51	.368
RC × He × El	2.0, 45.6	0.43	0.23	.653

*Note.* RC = referential context; Ge = noun gender; Cl = (later) clause type; El = electrode; He = hemisphere. Table only displays omnibus ANOVA tests that involve RC, and midline/lateral ANOVA tests of additional interest. Fractionated degrees of freedom have been adjusted using the Box epsilon hat procedure.

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

and F8-T4 respectively), temporal (LT, RT, laterally to Cz, at 33% of the interaural distance), temporoparietal (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 sub-orbital 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 hi-cut of 30 Hz and a time constant of 8 s. Impedances were kept below 5 and 3 kOhm

for EOG and all other electrodes respectively. The EEG and EOG signals were digitized on-line with a sampling frequency of 200 Hz.

Prior to off-line averaging, all single-trial waveforms were screened for eye movements, electrode drifting, amplifier blocking, and EMG artifacts in a critical window that ranged from 150 ms before onset of the indirect object NP's determiner to 1200 ms after onset of the last critical word (CW). Trials containing such artifacts were rejected (18.3%). Next, average waveforms were computed for each subject, at each critical word in each referential condition, after normalizing the waveforms of the individual trials relative to a 150-ms prestimulus baseline interval preceding the critical word (i.e., after subtracting the mean amplitude in that interval from all sample points in the epoch at hand). Subsequent analyses of variance (ANOVAs) used mean amplitude values com-

puted for each subject in time windows that were suggested by earlier findings and by visual inspection of the grand average waveforms.<sup>2</sup> All parsing-related P600/SPS effects were tested in a latency window of 500–700 ms after onset of the critical word, which is where the P600/SPS is commonly observed. Univariate *F* tests with more than 1 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 rest of the design. 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

### *Referential Ambiguity Effect*

Figure 1A displays, for each electrode, the grand average waveforms elicited by a singular noun presented in a one-referent or a two-referent context, averaged across four sentence types, and using a pre-stimulus baseline of 150 ms preceding the critical noun. The corresponding difference waveforms, each computed by subtracting the grand average ERP in the one-referent condition from that in the two-referent condition, are displayed in Figure 1B. In these and all following figures, positive voltage is plotted downward, the critical word is presented at 0 ms, the next word follows at 600 ms, and the signals are normalized relative to the 150 ms interval preceding the onset of the critical word.

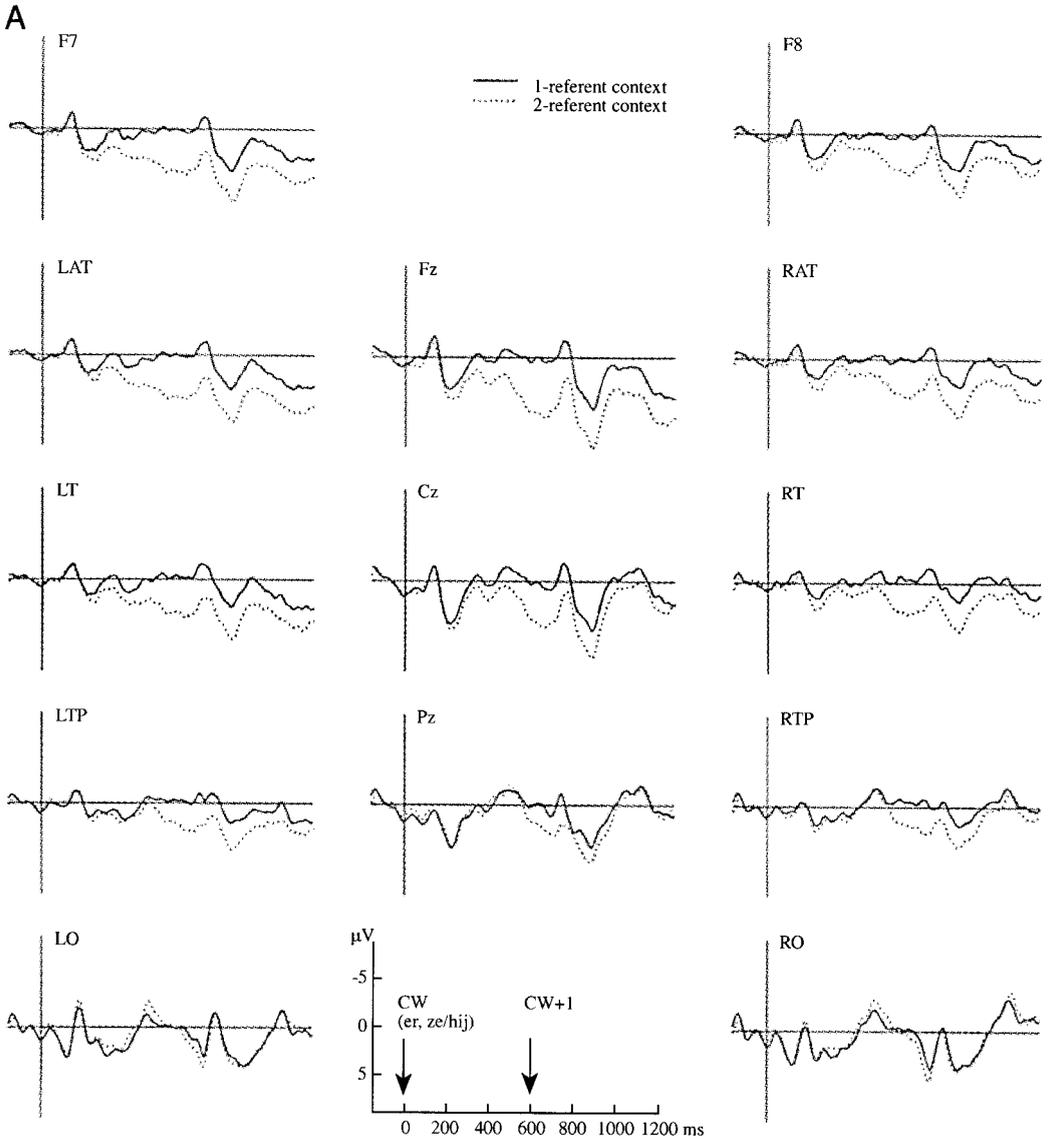
Each individual word can be seen to elicit an N1–P2 complex in the first 250 ms after its

onset. Also, a clear P1 component precedes the N1–P2 complex at occipital sites. These are typical ERP profiles for visually presented material. More interesting, the waveforms clearly show that the referential status of a singular noun has very rapid processing effects: when a singular definite NP such as *het meisje* is referentially ambiguous, the NP's head noun elicits a negative deflection in the average waveform that begins somewhere between 250 and 300 ms relative to a situation in which the same NP does have a unique referent in context. This effect shows up at all electrode sites, appears to be somewhat larger over the left than over the right hemisphere, lasts for several hundreds of milliseconds, and is particularly persistent at anterior and central locations, where it enters the time domain of the next word in the sentence.

As can be seen in Table 3, statistical analysis corroborates these observations. Using mean amplitude in the interval of 300–600 ms after onset of the critical noun, an omnibus ANOVA with referential context (one- or two-referent), noun gender (neuter or common), clause type (complement or relative-clause disambiguation in later sentence parts), and electrode (13 sites) yielded a significant effect of referential context. On average, the mean waveform amplitude in the critical 300–600 ms after noun onset was 0.8  $\mu$ V more negative for nouns presented in a two-referent context than for the same nouns presented in a one-referent context. In this latency window, referential context did not significantly interact with electrode site. As expected, it also did not interact with noun gender and/or clause type, which indicates that the size of the referential ambiguity effect did not differ statistically across the four target sentence types defined for the parsing issue. Similar results were obtained in the topographical analysis on midline and lateral electrode subsets (see Table 3).

Because the waveforms were normalized with a 150 ms baseline just before the noun, the effect at hand cannot be attributed to a differential influence of discourse context on the state of the sentence processor *before* it encountered the noun. In fact, no referential context effects

<sup>2</sup> Item analyses are usually not performed on ERP data in part because stable average ERPs per item and condition would require prohibitively large numbers of subjects. In the subject analyses, however, average ERPs per subject and condition always involve a large numbers of items, which reduces the probability that the results hinge on just a few odd items. For the current study, one should also note that we manipulated referential context within-items, so that differences between items, if any, were completely orthogonal to our critical comparison.

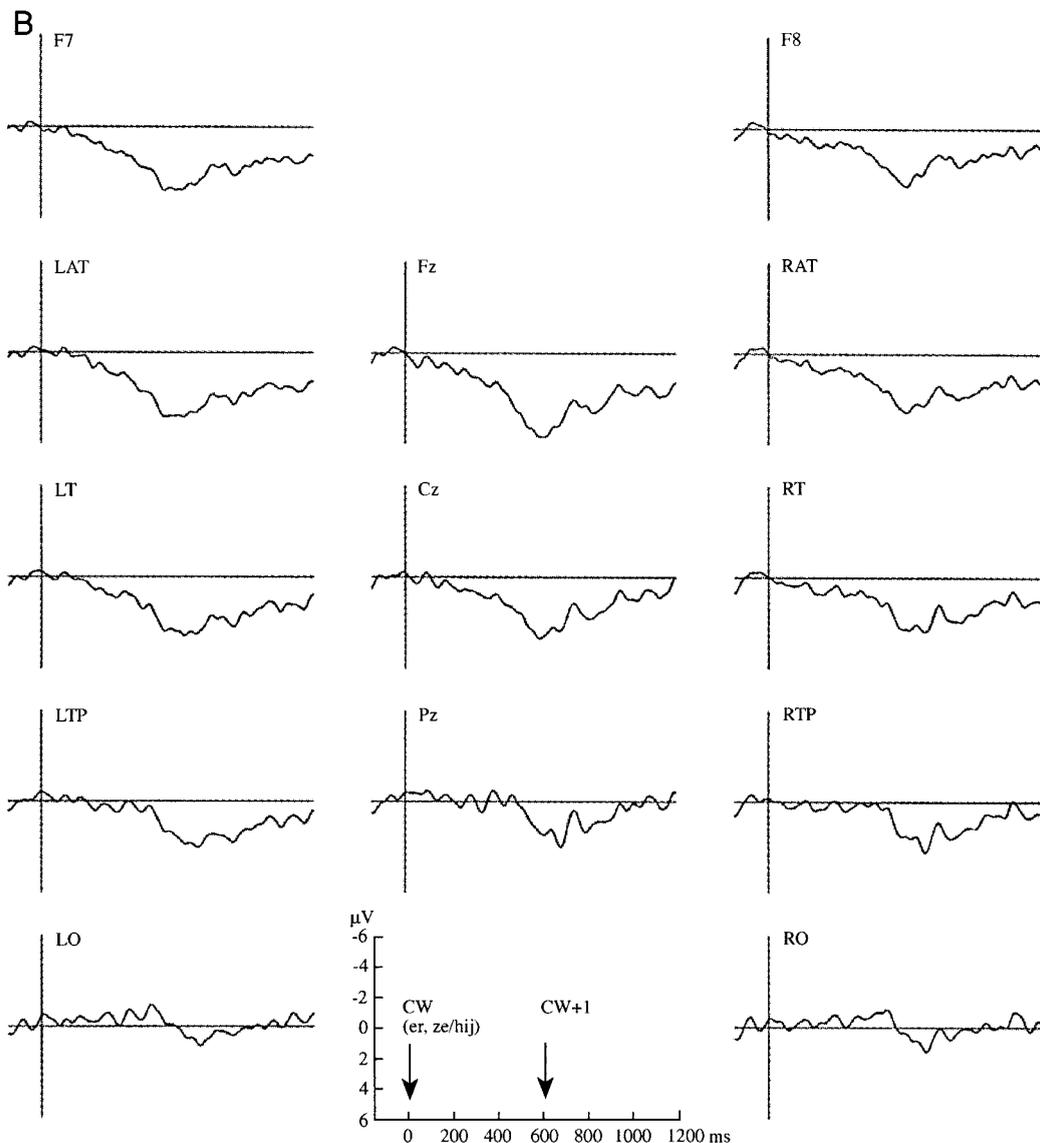


**FIG. 2A.** Early-complement sentences: Grand average ERPs elicited by early-complement disambiguation in a one-referent context (solid line) and a two-referent context (dotted line). The onset of the disambiguating word (*er* or *ze/hij*) is at 0 ms, the next word (CW + 1) follows at 600 ms. Each waveform averages over 24 subjects and approximately 574 trials.

could be observed at all in the recording epoch prior to the noun. This was confirmed by a control ANOVA on mean amplitude in the equivalent 300- to 600-ms interval after onset of the preceding article (context main effect:  $F(1,23) = 0.31$ ,  $MS_e = 15.70$ ,  $p = .585$ ).

Figures 1A and 1B suggest that referential context begins to exert a differential effect on

the average ERP somewhere between 250 and 300 ms after onset of the critical noun. To find out at what point in time the waveforms obtained in the one- and two-referent contexts begin to diverge significantly, we carried out two-tailed repeated-measures  $t$  tests on the mean amplitude in a 50-ms latency window that was for each test shifted 10 ms rightward over



**FIG. 2B.** Early-complement sentences: garden-pathing minus supportive context (two-minus one-referent) difference waveforms.

the signal (i.e., 200–250 ms, 210–260 ms, etc.). Taking the onset of the first of at least five consecutive latency ranges with a significant  $t$  value as the onset of the referential context effect, this analysis revealed that the waveforms begin to diverge significantly at about 280 ms after onset of the critical noun [ $t(23) = 2.42$ ,  $SD = 1.11$ ,  $p = .024$ ].

#### *Referential Context Effects in Parsing*

*Early-complement sentences.* For this sentence type, the complement/relative-clause ambiguity generated by *dat* was resolved as a complement clause by the next word (*er*, *ze*, or *hij*). Figure 2A shows, for each electrode, the grand average waveforms elicited by early-comple-

TABLE 4

Analyses of Variance (ANOVAs) of the Mean ERP Amplitude in the 500- to 700-ms Latency Range Following Onset of the Early-Complement Disambiguating Words *er* or *ze/hij*

Source	<i>df</i>	<i>F</i>	<i>MS<sub>e</sub></i>	<i>p</i>
Omnibus ANOVA (13 electrodes)				
RC	1, 23	45.50	25.24	.000***
RC × EI	3.1, 72.2	14.28	1.69	.000***
Midline ANOVA (3 electrodes)				
RC	1, 23	31.32	12.84	.000***
RC × EI	1.3, 29.8	13.07	2.43	.000***
Lateral ANOVA (2 × 5 electrodes)				
RC	1, 23	49.27	15.52	.000***
RC × He	1, 23	1.38	5.50	.252
RC × EI	1.4, 31.4	37.44	1.31	.000***
RC × He × EI	2.2, 51.2	2.12	0.39	.126

Note. RC = referential context; EI = electrode; He = hemisphere. Table only displays ANOVA tests that involve RC. Fractionated degrees of freedom have been adjusted using the Box epsilon hat procedure.

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

ment disambiguation in a one-referent or a two-referent context. For each of the difference waveforms in Fig. 2B, the ERP in the (supportive) one-referent condition was subtracted from that in the (potentially garden-pathing) two-referent condition, so that a predicted garden-path effect, if any, would show up in the correct polarity.

The grand averages in Figs. 2A and 2B reveal a P600/SPS in the waveform in a two-referent context, relative to a one-referent context, at all but the occipital locations. The P600/SPS can be seen most easily at parietal sites (LTP, Pz, RTP), where it has an onset of about 450–500 ms and a duration of several hundreds of milliseconds, both typical characteristics for this type of effect. But the difference waveforms in Fig. 2B show that it also emerges, at about the same time, at more anterior locations, albeit superimposed on a positive trend with a much earlier onset. In either case, the P600/SPS effect peaked at about 600–700 ms.

In Fig. 2B, the early positive trend at anterior electrode sites in the two-referent context can be seen to begin right at the onset of the current critical word, with effectively a 0-ms delay.

This means that it cannot plausibly have been triggered by the latter (which, after all, needs to be read and recognized first). Instead, the early trend appears to be the result of the earlier referential ambiguity effect interacting with inherent limitations of the normalization procedure.<sup>3</sup>

As can be seen in Table 4, an omnibus ANOVA of mean amplitude in the 500-

<sup>3</sup> The normalization procedure accurately adjusts for earlier effects as long as they are stable, i.e., do not substantially change in size when going from the prestimulus baseline interval to the subsequent epoch under investigation. The earlier referential ambiguity effect, however, had at the more anterior sites approximately reached its maximum in the current prestimulus baseline interval and began to decline immediately thereafter (in Fig. 2B, a hint of this can be seen in the baseline interval at anterior sites, e.g., Fz). Under such conditions, normalization actually leads to an increasing overcorrection, since it partials out, as a constant, the approximately *maximal* referential ambiguity effect size observed in the current normalization window from later parts of the waveforms where this effect is in fact declining. As referential ambiguity yielded a negative deflection in two-referent waveforms (see Fig. 1A), the net effect of the normalization at hand is one of subtracting an increasingly overestimated negativity from two-referent waveforms at the anterior sites.

700-ms interval after CW onset, by referential context (one- or two-referent) and electrode (13 sites), yielded a significant effect of referential context as well as a significant interaction of referential context and electrode site. On average, the mean waveform amplitude in the critical 500–700 ms after early-complement disambiguation was  $2.7 \mu\text{V}$  more positive for targets presented in a two-referent context than for the same targets presented in a one-referent context. As would be expected given Figs. 2A and 2B, both the midline and the lateral ANOVA showed that the referential context effect was significantly larger over more anterior regions of the scalp than over more posterior ones. The lateral analysis also revealed that referential context did not interact with hemisphere. In simple effects tests at individual electrode sites, significant context effects emerged at all but the occipital electrodes. Separate control analyses revealed that the specific disambiguation scheme did not matter.

The difference waveforms at the occipital sites suggest that, although the P600/SPS is still visible in the waveform morphology, it is preceded by a small negativity peaking at around 400 ms. In an ANOVA on mean amplitude in the 300- to 450-ms latency window, however, no significant differences were found.

At the preceding word *dat*, referential context did not have systematic effects [e.g., context main effect on mean amplitudes in the 500- to 700-ms latency window:  $F(1,23) = 0.91$ ,  $MS_e = 48.23$ ,  $p = .350$ ] over and above the sustained (and rather persistent) frontal negativity triggered by the noun before.

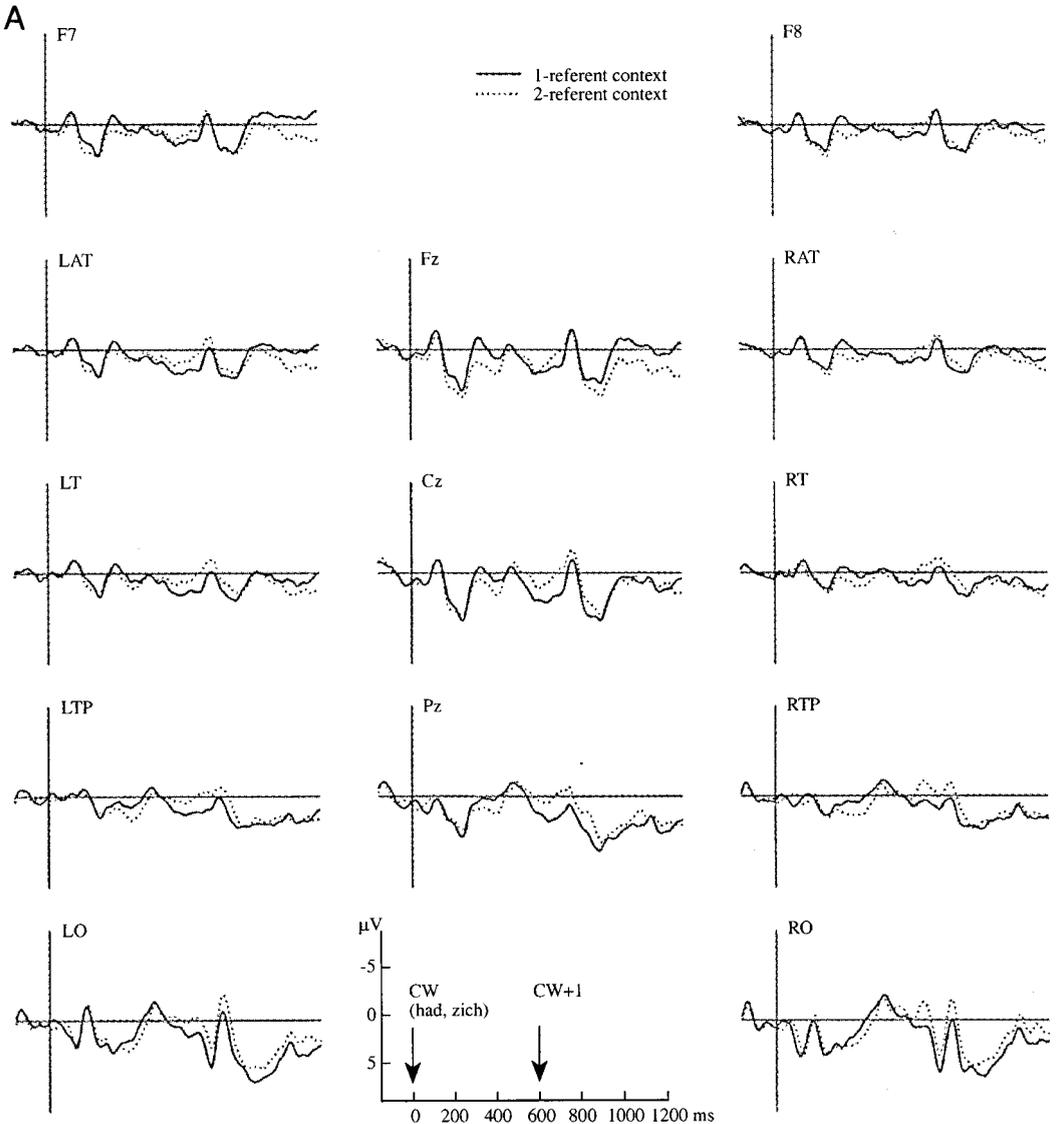
*Early-relative sentences.* In these sentences, the complement/relative-clause ambiguity generated by *dat* was resolved as a relative clause by the next word (usually *zich* or *had*). Figure 3A shows the average waveforms elicited by early-relative disambiguation in a one-referent or a two-referent context. For each of the corresponding difference waveforms in Fig. 3B, the ERP in the (now supportive) two-referent condition was subtracted from that in the (now potentially garden-pathing) one-referent condition, so that predicted garden-path effects would again show up in the correct polarity.

Early-relative disambiguation elicited a small but consistent P600/SPS in the waveform in a one-referent context, relative to a two-referent context, with an onset of about 450–550 ms, a peak amplitude around 650 ms, and a duration of several hundreds of milliseconds.

Table 5 displays the ANOVA results for mean amplitudes in the 500- to 700-ms interval after onset of the critical word. The omnibus ANOVA with all 13 electrodes revealed a significant effect of referential context, which did not vary significantly as a function of electrode site. On average, the mean waveform amplitude in the critical 500–700 ms after early-relative disambiguation was  $1.0 \mu\text{V}$  more positive for targets presented in a one-referent context than for the same targets presented in a two-referent context. The midline and the lateral ANOVA results confirmed that the referential context effect did not significantly depend on scalp region parameters. Separate analyses revealed that the specific disambiguation scheme did not matter. Also, the current effect of referential context is clearly in the opposite direction of that at early-complement disambiguation, as was confirmed by an additional ANOVA [referential context  $\times$  clause type:  $F(1,23) = 40.71$ ,  $MS_e = 26.04$ ,  $p < .001$ ].

As can be seen in Fig. 3A and, particularly, in Fig. 3B, there are also earlier negative deflections in the one-referent condition relative to the two-referent condition. At anterior sites, a negative shift develops right at the onset of the critical word and as such it resembles the early trend observed for early-complement sentences (compare, for instance, the waveforms at Fz in Figs. 2B and 3B, taking into account that the subtraction for Fig. 3B is the reverse of that used for Fig. 2B). This is again most likely a side-effect of normalization, as one would expect at these sites if our earlier account is correct.<sup>4</sup> Note, however, that the morphology of the

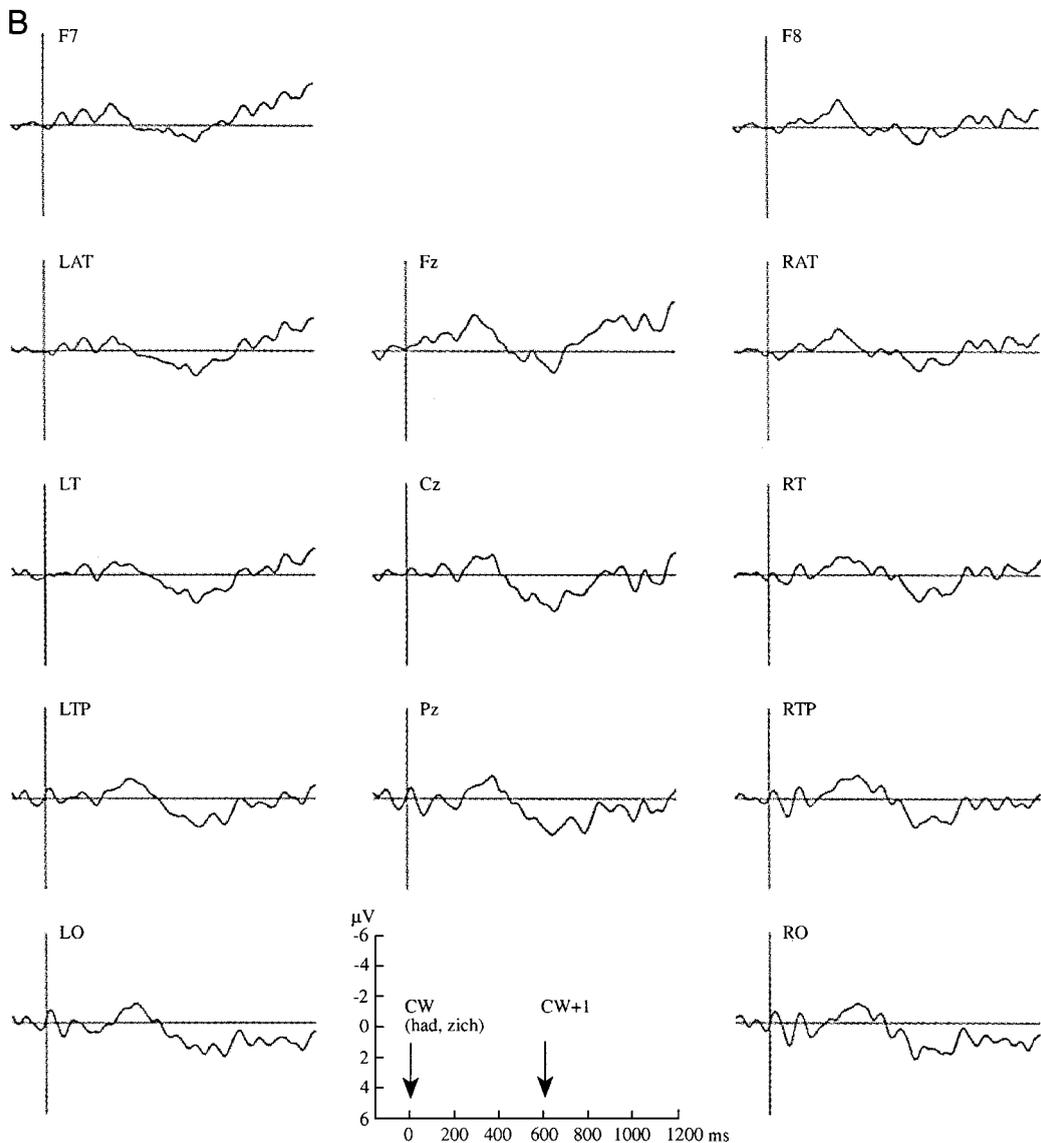
<sup>4</sup> See again footnote 3. Note, however, that whereas the early positive trend in two-referent contexts accumulated with the P600/SPS effect observed in two-referent contexts at early-complement disambiguation (see Fig. 2B), the corresponding early negative trend at hand here *works against* the P600/SPS effect observed in one-referent contexts (see Fig. 3B).



**FIG. 3A.** Early-relative sentences: Grand average ERPs elicited by early-relative disambiguation in a one-referent context (solid line) and a two-referent context (dotted line). The onset of the disambiguating word (*had* or *zich*) is at 0 ms, the next word (CW + 1) follows at 600 ms. Each waveform averages over 24 subjects and approximately 574 trials.

difference waveforms at the more posterior locations is qualitatively different; here, a distinct N400-like negative deflection shows up at about 200–250 ms after word onset. An ANOVA on mean amplitude in a 300- to 450-ms latency window revealed significant effects of referential context at LTP, RTP, and Fz, but no such effects at the other 10 locations.

As in the early-complement target sentences, referential context did not have systematic effects at the preceding word *dat* [e.g., context main effect on mean amplitudes in the 500- to 700-ms latency window:  $F(1,23) = 1.51$ ,  $MS_e = 22.01$ ,  $p = .232$ ] over and above the sustained frontal negativity triggered by the noun before.



**FIG. 3B.** Early-relative sentences: garden pathing minus supportive context (one-minus two-referent) difference waveforms.

*Immediate-complement sentences.* These were target sentences for which we assumed that the critical word *dat* both generated the complement/relative-clause ambiguity and, on the grounds of subsequently computed gender agreement, resolved it as a complement clause. Figure 4A displays the grand average waveforms elicited by *dat* in a one-referent or a two-referent context. For each of the difference

waveforms in Fig. 4B, the ERP in the (now again supportive) one-referent condition was subtracted from that in the (potentially garden-pathing) two-referent condition.

As can be seen in these figures, immediate-complement disambiguation elicited a P600/SPS in the waveform in a two-referent context, relative to a one-referent context, at all frontal and anterior temporal sites as well as at the

TABLE 5

Analyses of Variance (ANOVAs) of the Mean ERP Amplitude in the 500- to 700-ms Latency Range Following Onset of Early-Relative Disambiguating Words *zich* or *had*

Source	<i>df</i>	<i>F</i>	<i>MS<sub>e</sub></i>	<i>p</i>
Omnibus ANOVA (13 electrodes)				
RC	1, 23	5.32	27.76	.030*
RC × EI	3.1, 70.2	1.33	1.47	.271
Midline ANOVA (3 electrodes)				
RC	1, 23	4.46	15.30	.046*
RC × EI	1.4, 33.1	2.64	1.49	.101
Lateral ANOVA (2 × 5 electrodes)				
RC	1, 23	5.46	15.96	.029*
RC × He	1, 23	0.78	3.29	.387
RC × EI	1.4, 31.1	0.59	1.74	.496
RC × He × EI	2.0, 45.7	1.43	0.24	.250

*Note.* RC = referential context; EI = electrode; He = hemisphere. Table only displays ANOVA tests that involve RC. Fractionated degrees of freedom have been adjusted using the Box epsilon hat procedure.

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

right-temporal site. The P600/SPS effect has an onset of about 450–500 ms, a peak amplitude around 650 ms, and a duration of several hundreds of milliseconds.

Table 6 shows that the omnibus ANOVA on the mean amplitude in the interval of 500–700 ms after onset of the critical word did not yield a significant main effect of referential context. Instead, there was a significant interaction of referential context and electrode site, reflecting the specific anterior distribution of the effect. The midline and lateral ANOVAs both revealed that the referential context effect was indeed significantly larger over more anterior regions of the scalp than over more posterior ones. In addition, the lateral analysis confirmed a right–left asymmetry of the referential context effect. Simple effects tests at individual electrode sites yielded a significant context effect at F7, Fz, F8, RAT, and RT.

As before, we also tested for earlier N400-like negativities in a 300- to 450-ms latency window. The ANOVA on mean amplitude in this window revealed no significant effects of referential context at any of the 13 electrode locations.

Although immediate-complement sentences resolve the complement/relative-clause ambiguity at *dat*, we looked for additional effects at the subsequent words *er* and *ze* (or *hij*), which also rule out the relative-clause alternative, albeit redundantly. No relevant effects of referential context were observed in the waveforms, as was confirmed by mean amplitude ANOVAs in the 500- to 700-ms latency range following the onset of these words.

*Immediate-relative sentences.* Having the relative pronoun *die* instead of the ambiguous word *dat*, this final set of target sentences was not assumed to generate a complement/relative-clause ambiguity. Figure 5A shows the grand average waveforms elicited by *die* in a one-referent or a two-referent context. Figure 5B displays the corresponding difference waveforms, each computed by subtracting the ERP in the two-referent condition from that in the one-referent condition (as was done for early-relative sentences).

In the critical 500- to 700-ms latency range, immediate-relative *die* did not elicit any systematic difference in the waveforms obtained in one- and two-referent contexts respectively (see

Table 7, omnibus ANOVA). The midline and lateral ANOVAs suggest that the referential context factor to some extent interacted with topographical factors (particularly position along the midline), but simple effects tests at individual electrode sites yielded no significant context effect at any site.

The one-referent minus two-referent difference waveforms in Fig. 5B do reveal a clear early negativity at parietal and posterior locations, peaking at about 350 ms. The ANOVA on mean amplitude in a 300- to 450-ms latency window confirmed that there was a significant effect of referential context at LTP, Pz, RTP, LO, and RO. In view of the virtually immediate onset of this posterior effect, however, we hesitate to classify (all of) it as a regular N400 effect.

To be able to compare results across target sentence conditions, we also examined the ERP data at the subsequent words *had* and *zich*, which again enforce the relative-clause alternative, albeit redundantly. No consistent effects of referential context could be observed in the waveforms, as was confirmed by mean amplitude ANOVAs in the 500- to 700-ms latency range following the onset of *had* and *zich*.

*All critical sentence types.* Table 8 schematically displays the referential context effects observed at various critical and control word positions in the four target sentence types.

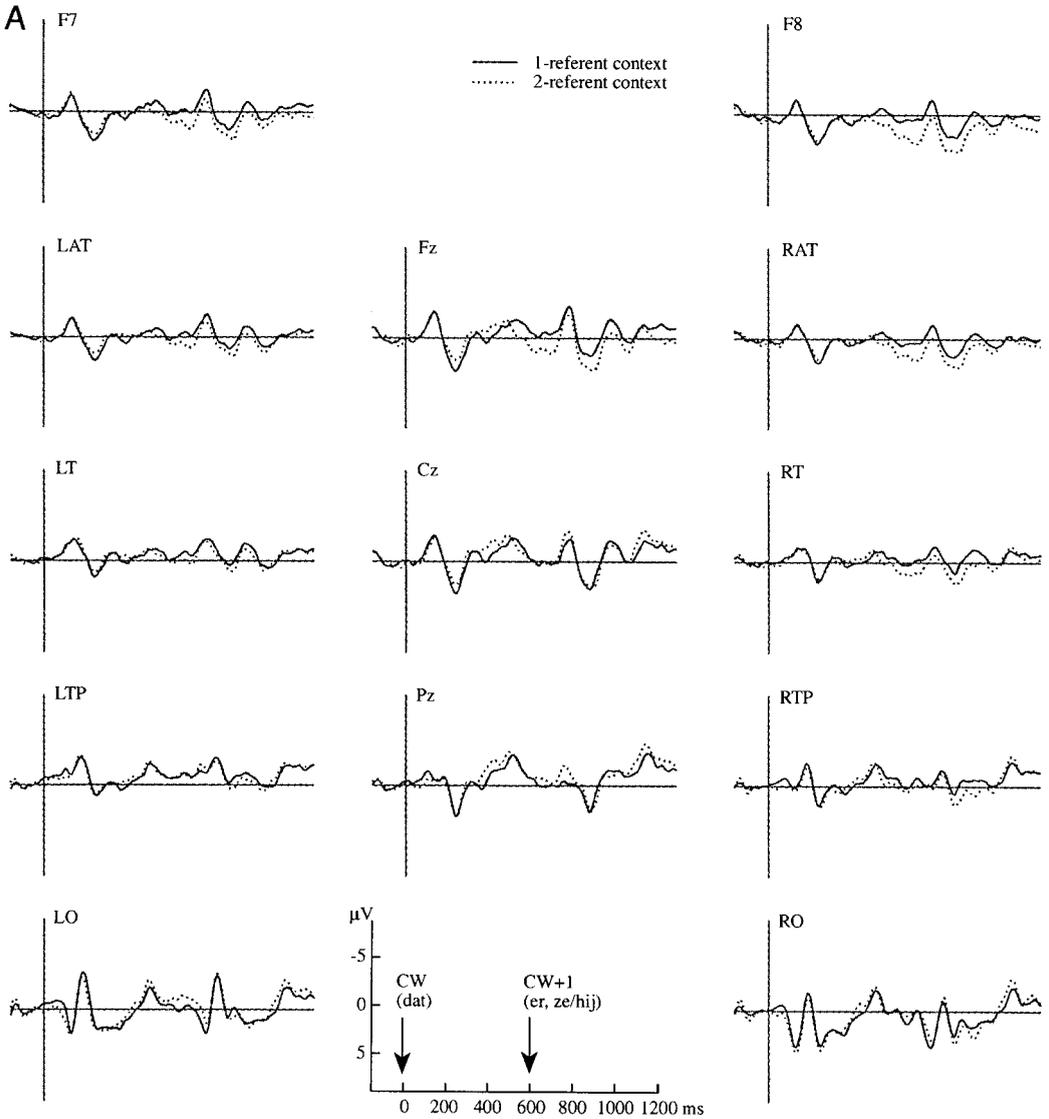
We observed reliable P600/SPS at early- and immediate-complement disambiguation in two-referent contexts and at early-relative disambiguation in one-referent contexts. The three effects had a very similar time course, in line with other P600/SPS effects reported in the literature: each emerged somewhere between 450 and 550 ms after onset of the critical word and lasted for several hundreds of milliseconds. Also note that all three effects peaked at about 600–700 ms. Nevertheless, the analyses did suggest differences in magnitude and scalp distribution. We therefore compared the three effects in several additional ANOVAs (in which the referential context factor was always recorded as garden-pathing versus supportive). Not surprisingly, a joint analysis combining all three showed that mean amplitude in the P600/

SPS window (500–700 ms after critical word onset) differed significantly across the early-complement, early-relative, and immediate-complement sentence types [referential context  $\times$  sentence type:  $F(2,0,44.9) = 6.54$ ,  $MS_e = 31.88$ ,  $p = .003$ ]. Pairwise analyses revealed significant differences between early-complement and early-relative sentences in the size of the mean referential context effect [ $F(1,23) = 8.76$ ,  $MS_e = 26.96$ ,  $p = .007$ ] and between early-complement and immediate-complement sentences [ $F(1,23) = 11.30$ ,  $MS_e = 33.06$ ,  $p = .003$ ], but not between early-relative and immediate-complement sentences [ $F(1,23) = 0.44$ ,  $MS_e = 35.60$ ,  $p = .514$ ].

Furthermore, the three effects differed in their scalp topography. After the appropriate data transformations for testing differences in scalp distribution (e.g., Rösler et al., 1993; see also McCarthy and Wood, 1985), a joint analysis with three sentence types revealed a referential context  $\times$  13-electrode  $\times$  sentence type interaction [ $F(7.4,171.1) = 4.10$ ,  $MS_e = 0.57$ ,  $p < .001$ ]. Pairwise analyses revealed significant differences in effect topography between early-complement and early-relative sentences [ $F(4.2,96.8) = 5.22$ ,  $MS_e = 0.61$ ,  $p = .001$ ], between early-complement and immediate-complement sentences [ $F(4.7,107.5) = 2.58$ ,  $MS_e = 0.51$ ,  $p = .034$ ], and between early-relative and immediate-complement sentences [ $F(3.9,90.3) = 4.24$ ,  $MS_e = 0.60$ ,  $p = .004$ ], each time reflecting distributional shifts along the anterior-posterior axis only.

## DISCUSSION

We recorded event-related brain potentials from subjects who were reading sentences in the context of short introductory stories. The stories varied in the number of suitable referents they introduced for a singular definite noun phrase that was embedded in the later target sentence (e.g., *David vertelde het meisje dat . . . , David told the girl that . . .*). A one-referent story presented just a single unique referent, whereas a two-referent story introduced two candidate referents for this noun phrase (cf. Crain & Steedman, 1985). The ERP results showed (a) that when processing the target sentence, readers very rapidly relate the



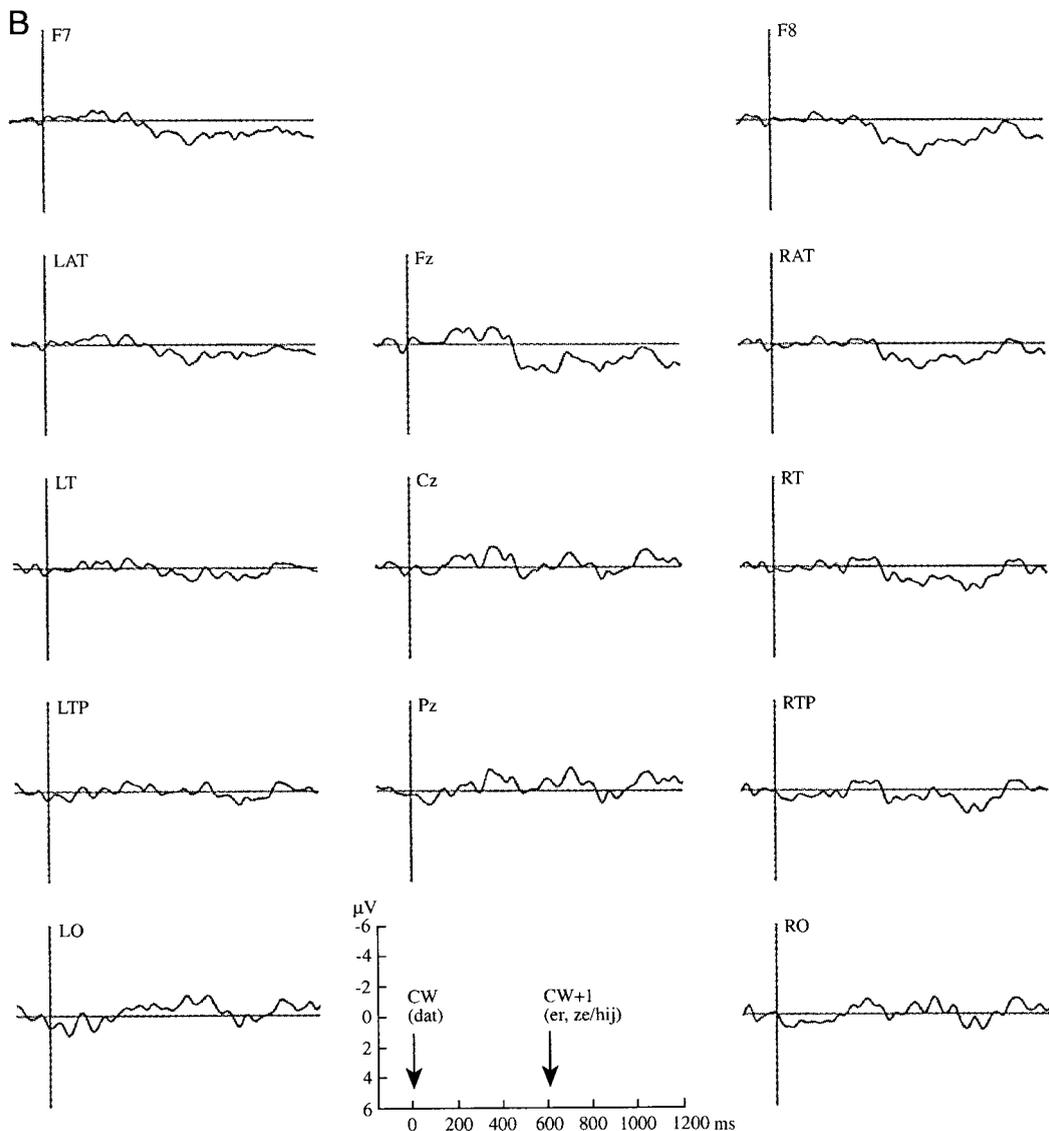
**FIG. 4A.** Immediate-complement sentences: Grand average ERPs elicited by immediate-complement disambiguation in a one-referent context (solid line) and a two-referent context (dotted line). The onset of the disambiguating word (*dat*) is at 0 ms, the next word (*er* or *ze/hij*) follows at 600 ms. Each waveform averages over 24 subjects and involves approximately 574 trials.

noun phrase to potential referents in their representation of the earlier discourse and (b) that they can immediately use the resulting information to parse a subsequent local structural ambiguity.

#### *Possible Methodological Concerns*

Before we discuss these findings, we address a number of possible concerns over aspects of

our methodology. First, the subjects in our experiment had no other task than to listen and read for comprehension. We did not use additional questions or secondary judgement tasks. Instead, we simply tried to make our materials as interesting and varied as possible within the constraints of the design and assumed that this by itself would engage the subject to a sufficient



**FIG. 4B.** Immediate-complement sentences: garden pathing minus supportive context (two- minus one-referent) difference waveforms.

degree. Although we see this as a valuable feature of our design, it may raise concerns as to whether our subjects indeed processed for comprehension.

To us, it seems that the pattern of results reported above—a negative shift associated with referentially ambiguous nouns and three context-induced P600/SPS effects—is most parsimoniously accounted for by assuming that subjects were processing the stories in a natural

way and were establishing reference as part of their natural language comprehension. Perhaps more compelling, though, is the fact that we obtained, in the same experiment and for the same subjects, a standard N400 effect to words that were designed to violate the semantics of the *global* discourse. These critical words occurred in the remainder of several of our target sentences, and each had been chosen such that, although the word was perfectly acceptable

TABLE 6

Analyses of Variance (ANOVAs) of the Mean ERP Amplitude in the 500- to 700-ms Latency Range Following Onset of the Immediate-Complement Disambiguating Word *dat*

Source	<i>df</i>	<i>F</i>	<i>MS<sub>e</sub></i>	<i>p</i>
Omnibus ANOVA (13 electrodes)				
RC	1, 23	1.50	28.68	.233
RC × EI	3.5, 80.2	6.68	1.07	.000***
Midline ANOVA (3 electrodes)				
RC	1, 23	0.39	14.10	.536
RC × EI	1.3, 30.5	8.03	1.55	.005**
Lateral ANOVA (2 × 5 electrodes)				
RC	1, 23	2.20	17.35	.151
RC × He	1, 23	5.80	1.35	.024*
RC × EI	1.4, 32.7	12.25	1.06	.000***
RC × He × EI	1.7, 38.6	0.16	0.34	.812

*Note.* RC = referential context; EI = electrode; He = hemisphere. Table only displays ANOVA tests that involve RC. Fractionated degrees of freedom have been adjusted using the Box epsilon hat procedure.

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

within this local sentence, it did not make sense given the earlier discourse context. For instance, for the first target sentence in Table 1, the word *visitors*, which makes good sense in the global discourse, would be replaced by something like *sharks*, which makes much less sense given what the discourse was about (see van Berkum, Hagoort & Brown, in press, for details). A separate control experiment confirmed that the resulting N400 effect hinged on the semantic relationship between the critical word and the earlier discourse context. Our subjects had thus indeed been processing the materials for comprehension. They were also clearly able to take written sentences as a continuation of spoken discourse.

Another possible concern relates to our use of serial visual presentation. Because the eye movements that occur during natural reading are detrimental to the recording of ERPs, written target sentences were displayed with a fixed 600-ms word-onset asynchrony. This is standard procedure for ERP research on sentence processing. But because readers normally fixate for about an average of 200–250 ms only, do so with much variability, and frequently skip

shorter words entirely (Rayner & Sereno, 1994), it is important to consider to what extent this mode of presentation may have had an impact on the current result.

We deliberately decided to begin our explorations of ERPs in discourse with written target sentences presented at a relatively slow presentation rate (compared to average natural reading rates) to minimize the probability that the ERP components elicited by two consecutive words would overlap. This is a common and desirable research strategy when using the ERP methodology in unexplored language territory, and it is of overriding importance when the hypotheses require that effects can unequivocally be ascribed to particular words. Given, for instance, that the P600/SPS usually begins at about 500 ms after onset of a critical word, the use of either a faster written word rate (e.g., a 250-ms SOA) or natural connected speech (two to three words per second on average; Levelt, 1989) would have made it difficult to unequivocally interpret our effects as P600/SPS effects elicited by a critical disambiguating word rather than, say, a very early positivity elicited by the *next* word.

As long as it is not *extremely* fast (e.g., Mason, 1983) or slow, though, serial visual presentation as such does not seem to be overly disruptive of the normal reading process (cf. Kutas & Van Petten, 1994). In our lab, for instance, we have observed that violations of syntactic constraints as well as of preferred structural assignments elicit highly comparable P600/SPS results regardless of whether sentences were presented with serial visual presentation at 600-ms SOA (Hagoort et al., 1993), at 250-ms SOA (Hagoort & Brown, 1998), or as natural connected speech (Hagoort & Brown, 1998). The semantic integration processes that are reflected by the N400 component also do not seem to depend much on whether one uses a serial visual presentation rate that approximates natural average reading times, a slower 600-ms SOA rate, or connected speech (Kutas, 1993; Hagoort & Brown, 1998). As another example, Kutas (1997) recently reported that embedded relative clause constructions presented either in natural connected speech or with serial visual presentation at a 500-ms SOA elicit remarkably similar ERP patterns in the slow-wave domain.

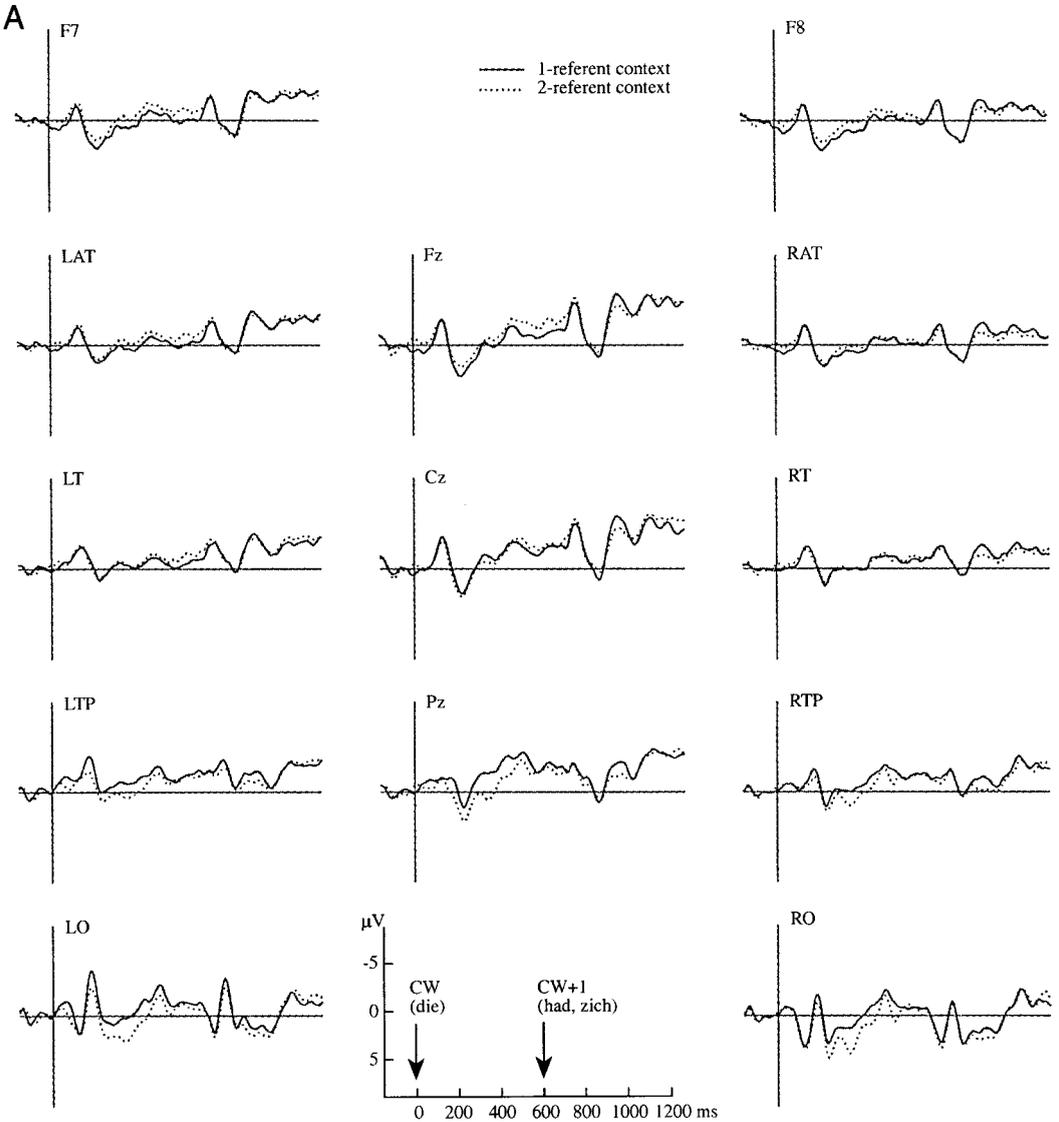
We take such findings to show that a new result solely obtained in a 600-ms SOA serial visual presentation paradigm should not be lightly discarded because of that paradigm. Of course, what the above-mentioned generalization studies clearly do *not* show is that such a new result will *necessarily* be identical to what would be obtained in natural connected speech. Modality-specific effects may appear, for a variety of interesting as well as less interesting reasons. Although we are confident that our current findings are not artifacts of the serial visual presentation procedure, an exact replication study with natural connected speech is currently running in our lab. Preliminary analyses suggest a highly comparable pattern of results (albeit more noisy, due to component overlap from critical words with very short SOAs).

#### *Establishing Reference in Discourse Context*

Our first objective was to study the nature of referential processes that relate sentences to earlier discourse. Do readers try to identify the

referent(s) of a noun phrase in an incremental fashion, i.e., as soon as they have a noun to work with, or do they delay such referential processing, e.g., because the next few words in the sentence might provide useful further information? The differential ERPs elicited by referentially ambiguous and referentially successful nouns clearly suggest that the search for referents was initiated at the noun and, moreover, that the processing system can very rapidly determine whether a noun is referentially ambiguous. In the constructions tested here, the system does not wait to see how the sentence develops (as suggested by Garrod & Sanford, 1994). Also, it does not wait until the word following the noun indicates that the NP can be closed (as was assumed by Perfetti & Britt, 1995). Rather, at least some basic referential processes are initiated, and deliver their result, within some 280 ms after noun onset (a large part of which will be needed to first recognize the noun and, presumably, parse it as modified by a definite article). In this respect, the mechanisms that establish reference are on a par with those that extract the syntactic and semantic structure of a sentence, mechanisms that also deliver vital information within only a few hundred milliseconds after the relevant lexical input. This finding is readily compatible with an incremental processing perspective in which every new word is immediately related to a representation of the preceding language input, at several levels. The result also accords well with other recent evidence that suggests that anaphoric expressions are associated with their linguistic antecedents without delay (e.g., Garnham et al., 1997).

The early onset of the differential ERP effect by itself reveals that at least some basic facts about reference become available very quickly. But referent identification is a complex process, much of which need not be visible in the ERP record. What exact aspect or consequence of this process is generating the current ERP effect? Because a referentially ambiguous noun may be less easy to integrate into a high-level semantic representation of the discourse, it might elicit a larger N400 than its unambiguous counterpart. But although the referential ambi-

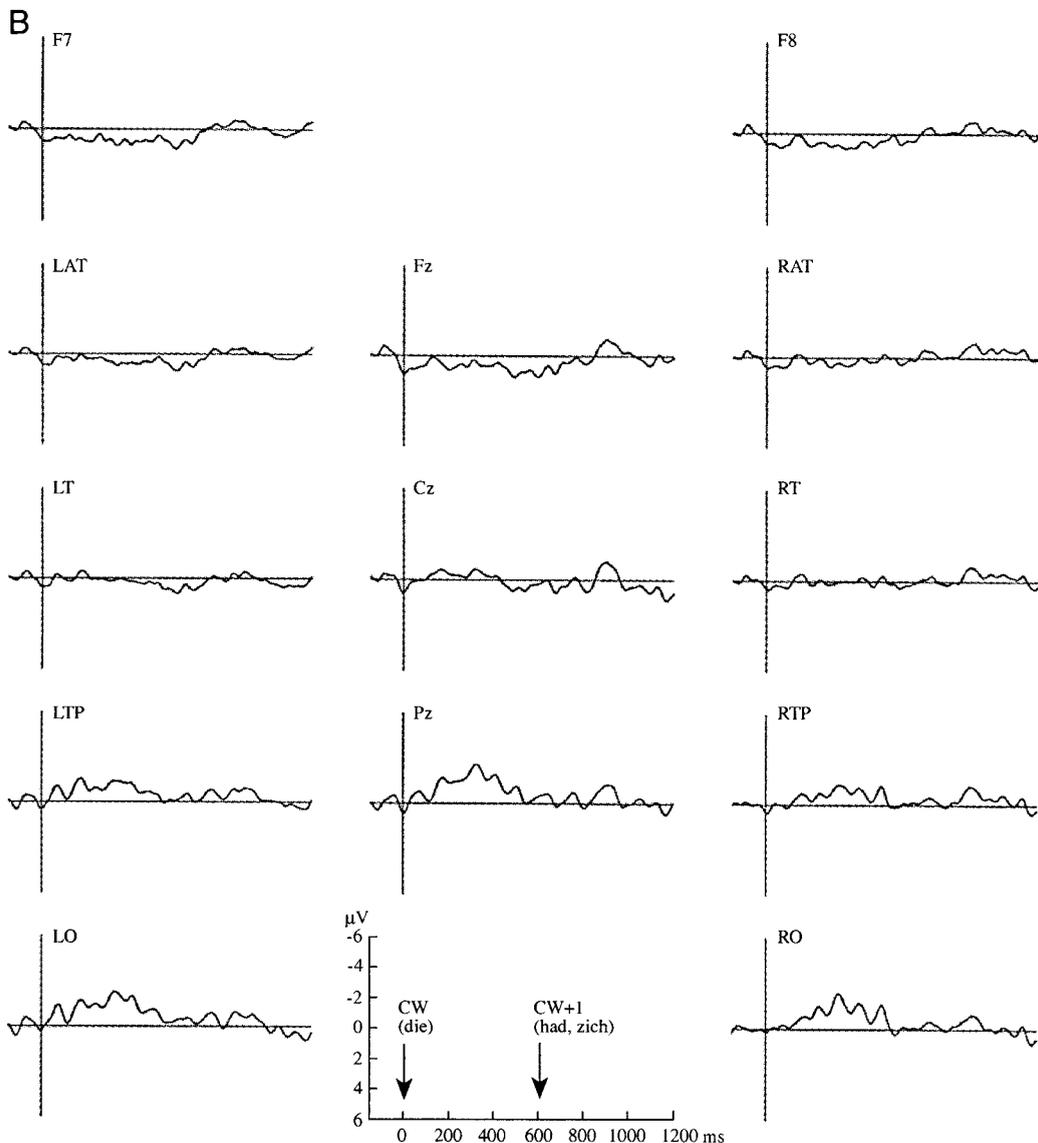


**FIG. 5A.** Immediate-relative control sentences: Grand average waveforms for the critical part of immediate-relative control sentences in a one-referent context (solid line) and a two-referent context (dotted line). The onset of the first control word (*die*) is at 0 ms, the next control word (*had* or *zich*) follows at 600 ms. Each waveform averages over 24 subjects and approximately 574 trials.

guity effect does develop around the relevant time, we cannot unambiguously classify it as an N400 effect. The latter is a modulation of the N400 component, and as such clearly peaked, fairly short-lived, larger posteriorly than anteriorly, and frequently also somewhat larger over the right hemisphere than over the left with written language stimuli (e.g., Kutas &

Van Petten, 1994). The current negativity, however, is much more persistent and less clearly peaked, particularly at central and anterior locations.

Another possibility is that our result reflects more extensive use of short- or long-term memory resources. Both have been observed to correlate with slow negative shifts in event-related brain



**FIG. 5B.** Immediate-relative control sentences: one- minus two-referent difference waveforms.

potentials (e.g., Friederici et al., 1996; King & Kutas, 1995; Kluender & Kutas, 1993; Kluender & Münte, 1998; Kutas & King, 1996; Rösler et al., 1993; Rösler, Pechmann, Streb, Röder, & Hennighausen, 1998). Referential ambiguity may entail that the system must maintain two candidate fillers (or pointers to them) for an unresolved single referential slot, thereby imposing an additional load on working memory (see Gibson, 1998, for a related account). The availability of

two candidate referents for a singular NP may also lead to additional search in memory of earlier discourse to see if other, more subtle, clues can help to select the most plausible referent (e.g., saliency, focus, recency of mention).

Two caveats are in order. First, our characterization of the current referential effect as a *negativity* associated with referential ambiguity does not follow from the data as such. Like most other effects obtained in psycholinguistic

TABLE 7

Analyses of Variance (ANOVAs) of the Mean ERP Amplitude in the 500- to 700-ms Latency Range Following Onset of the Immediate-Relative Control Word *die*

Source	<i>df</i>	<i>F</i>	<i>MS<sub>e</sub></i>	<i>p</i>
Omnibus ANOVA (13 electrodes)				
RC	1, 23	0.19	27.76	.664
RC × EI	3.7, 85.7	2.04	1.44	.100
Midline ANOVA (3 electrodes)				
RC	1, 23	0.27	17.23	.608
RC × EI	1.5, 33.8	3.62	1.64	.050
Lateral ANOVA (2 × 5 electrodes)				
RC	1, 23	0.14	15.15	.710
RC × He	1, 23	1.29	2.31	.268
RC × EI	1.5, 34.2	2.97	1.36	.079
RC × He × EI	1.8, 41.4	1.77	0.41	.186

*Note.* RC = referential context; EI = electrode; He = hemisphere. Table only displays ANOVA tests that involve RC. Fractionated degrees of freedom have been adjusted using the Box epsilon hat procedure.

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

research, the effect is a relative one, and it could in principle be characterized as a positivity associated with referential success instead. We

chose to influence referential processing by introducing referential ambiguity and therefore tend to view the effects of this as additional

TABLE 8

Schematic Overview of All Referential Context Effects

target sentence type and discourse context	<i>David vertelde . . .</i>			
Early-complement	. . . <i>het meisje</i>	<i>dat</i>	<i>er</i>	<i>visite kwam.</i>
1-Ref	—	—	—	
2-Ref	<b>NEG</b>	—	<b>P600/SPS</b>	
Early-relative	. . . <i>het meisje</i>	<i>dat</i>	<i>had</i>	<i>zitten bellen . . .</i>
1-Ref	—	—	<b>P600/SPS</b>	
2-Ref	<b>NEG</b>	—	—	
Immediate-complement	. . . <i>de vrouw</i>	<i>dat</i>	<i>er</i>	<i>visite kwam.</i>
1-Ref	—	—	—	
2-Ref	<b>NEG</b>	<b>P600/SPS</b>	—	
Immediate-relative	. . . <i>de vrouw</i>	<i>die</i>	<i>had</i>	<i>zitten bellen . . .</i>
1-Ref	—	—	—	
2-Ref	<b>NEG</b>	—	—	

*Note.* Example target sentence illustrates expletive/auxiliary disambiguation only. Words that resolve (or preclude) the complement/relative clause ambiguity are in bold. **NEG** = negative deflection associated with referential ambiguity of the noun. All effects are relative to the alternative discourse condition and tested after normalization in the 150-ms baseline preceding the current word.

processing consequences of this manipulation. The observed similarity to memory-related frontal negativities makes sense and as such supports our construal. Second, it is important to note that we only used *nominal* anaphors here. Pronouns may well be associated with partly different referential processes (see, e.g., Garrod, 1994, for interesting suggestions). Whether a referentially ambiguous pronoun also elicits a sustained frontal negativity is currently under investigation in our lab.

### *Parsing in Discourse Context*

It seems that people can very rapidly find out whether, in discourse, a noun is referentially ambiguous. But can such discourse-level knowledge also help them parse a subsequent local complement/relative-clause ambiguity? The full set of ERP results, schematically depicted in Table 8, is most compatible with an account of parsing in which discourse context can indeed influence the initial analysis of a structural ambiguity. Early-complement-clause disambiguation elicited a P600/SPS in two-referent contexts. Early-relative-clause disambiguation instead elicited a P600/SPS in one-referent contexts. Immediate-complement sentences again elicited a P600/SPS in two-referent contexts. Finally, no context-dependent P600/SPS effects were observed on any of the equivalent control positions in the structurally unambiguous immediate-relative sentences. We discuss each of these findings in turn.

Because the P600/SPS is known to reflect, within the domain of language comprehension, the processing consequences of a word that renders the preferred structural analysis untenable (e.g., Hagoort et al., 1993; Osterhout et al., 1994), we interpret the P600/SPS effect elicited by early-complement disambiguation in two-referent discourse contexts as evidence that such contexts had led the parser to prefer the relative-clause analysis, i.e., had lured it into a *syntactic* dead end. Such a context-induced garden-path phenomenon is most compatible with models of sentence processing that allow discourse factors to affect the initial parse (e.g., Crain & Steedman, 1985; Gibson, 1998; Lewis, 1993; MacDonald, Pearlmutter, & Seidenberg,

1994; Spivey-Knowlton & Tanenhaus, 1994; Tabor, Juliano, & Tanenhaus, 1997; Kempen, 1998; Kempen & Vosse, 1989). It is much less compatible with principled syntax-first accounts of parsing (e.g., Frazier, 1987; Mitchell et al., 1995), which predict that the parser initially always prefers the complement-clause analysis, regardless of the discourse context.

A discourse-induced P600/SPS effect also emerged when the structural ambiguity was resolved as a relative clause. But whereas early-complement disambiguation elicited a P600/SPS in two-referent contexts, the disambiguating word in early-relative sentences elicited a P600/SPS in one-referent discourse contexts. This crossover pattern of results shows that neither result is some unspecific side-effect of referential ambiguity (or uniqueness) as such. More important, it is again what one might expect under a context-sensitive parsing account. Taken together, the context effects obtained at early-complement and early-relative disambiguation can be interpreted as reflecting the same underlying phenomenon: when the parser encounters a local complement/relative-clause ambiguity, as in *David vertelde het meisje dat . . .* (*David told the girl that . . .*), it is more inclined to pursue a relative-clause analysis in a two-referent context, where the referentially ambiguous NP *het meisje* is in need of further modification, than in a one-referent context, where the NP is referentially complete.<sup>5</sup>

The debate between syntax-first and context-

<sup>5</sup> Due to differential contamination from the earlier referential ambiguity effect at anterior electrode sites, it is difficult to compare in our data the ERPs elicited by a context-supported (two-referent) relative clause disambiguation to those elicited by a context-supported (one-referent) complement clause disambiguation. However, it is of interest to observe that at the uncontaminated parietal and occipital sites, a supported relative clause does *not* elicit a P600/SPS compared to a supported complement clause, whereas when we presented the same target sentences in isolation in a control study (van Berkum, Brown, & Hagoort, 1998), relative-clause disambiguation *did* elicit such a P600/SPS effect at all sites. This suggests that, whatever factor was responsible for the complement clause preference in isolated sentences, it has no independent additional impact over and above that of appropriate discourse contexts.

sensitive accounts of parsing revolves around the parser's *initial* analysis of a local structural ambiguity and one should therefore probe the commitments of the parser at the earliest opportunity. Mitchell et al. (1992) took this to be the word that immediately followed the one giving rise to the structural ambiguity. The current early-complement and early-relative disambiguations qualify as such, with the probe word presented only 600 ms after the word that gave rise to the ambiguity (*dat*). One could argue, perhaps, that the preferences revealed by these early probe words, although very early ones, are not those that the parser *initially* entertained, but the result of a context-induced reanalysis of the initial structure-based commitment to the complement clause. Note, however, that the early probes leave very little time for such reanalysis, for it would have to be finished before the parser attempts to assign the next word.

The results obtained with the immediate-complement sentences in fact suggest that a context-induced reanalysis would have to be finished even earlier. In contrast to their early-complement counterparts, immediate-complement sentences contained a common gender NP (e.g., *de vrouw* in Table 8, third sentence), which formally ruled out a relative pronoun reading of the subsequent word (*dat*). We tentatively assumed that in on-line processing, however, the lexically ambiguous word *dat* might still give rise to a complement/relative-clause ambiguity, if only very briefly, before gender agreement information is used to eliminate the relative-clause alternative. In that short period of time, referential context could then modulate the parser's preferences, just as it would at *dat* in the two early-probe sentence types. Our findings were entirely consistent with this account (see van Berkum, Brown, & Hagoort, in press, for the specific implications with respect to grammatical gender).

In terms of probing initial parsing commitments, this is as immediate as one can get.<sup>6</sup>

<sup>6</sup> To avoid confusion, we should point out that the fact that P600/SPS effects emerge at about 500 ms after critical word onset is not relevant to the question of immediacy that we address. The issue is whether there really are no "immediate" context effects as defined for instance by Mitchell

Given that the P600/SPS effect emerged at about 450–500 ms after presentation of the critical word, a syntax-first account of this result can only be preserved by assuming that after recognizing the word in the first place, the system initially analyzes it as the head of a (simpler or more frequent) complement-clause analysis and subsequently overturns this structure-determined provisional commitment on the basis of referential discourse context, all within some 500 ms and all before the word's grammatical gender ultimately disconfirms the revised commitment. The findings of the current and other studies (e.g., St. George et al., 1994, 1997; van Berkum, Hagoort & Brown, in press) do suggest that the earlier discourse can in principle be brought to bear on sentence processing very rapidly. But if discourse-level information can be used to revise initial parsing commitments *that fast*, i.e., at the very same word where they are made in the first place, there does not seem to be all that much reason to, as a general principle, initially make those commitments on the basis of structural information only (cf. Altmann, 1988).

At this point, we should discuss an alternative reading of the data. Our current interpretation is that discourse context can affect the provisional analysis of a local structural ambiguity and can, as such, selectively cause the parser to "garden path" in a later part of the sentence. But one might argue that the results do not reflect context-induced garden paths in syntactic ambiguity resolution at all and are instead simply direct processing consequences of a syntactic construction that, once unambiguously clear, is infelicitous in that particular discourse context. To continue something like *David vertelde het meisje dat...* with a complement clause (which becomes clear at subsequent *er* or *ze*), for instance, is more felicitous in the one-referent contexts than in the two-referent contexts that we used. Alternatively, to continue it with a

et al. (1992), i.e., time locked to a particular lexical item (an early probe word rather than a late one). Given the procedures that we use, we can be certain that the obtained effects are indeed time locked to the processing of the relevant lexical items.

restrictive relative clause is more felicitous in two-referent contexts than in one-referent contexts. If these implications are sufficiently rapidly computed as soon as the correct syntactic structure has been identified, they may show up as people process the disambiguating words.

It is not inconceivable that, under particular circumstances, the processing consequences of an unambiguous infelicitous construction indeed show up in event-related brain potential recordings. But unless we make the strong assumption that such pragmatic factors can force the parser to *abandon the only structural analysis currently available*, i.e., that they are powerful enough to effectively render the sentence ungrammatical as it is being processed, there is no reason why the processing consequences of an infelicitous construction should yield an ERP effect that is generally accepted to selectively reflect a *syntactic* dead end. If anything, straightforward pragmatic infelicity would rather lead one to predict an N400 effect. In fact, we did actually obtain, in the same study and for the same subjects, an N400 effect that can at least in part be interpreted as an effect of pragmatic infelicity (e.g., an abrupt topic change; see earlier discussion, and see van Berkum, Hagoort & Brown, in press, for details). The fact that we now observe, at all three critical word positions, P600/SPS effects instead, strongly suggests that here it is indeed the *parser* running into syntactic dead ends, being led into them by certain referential contexts.

Furthermore, if the current P600/SPS elicited by early-relative disambiguation in one-referent contexts had merely resulted from the infelicitous use of a relative clause, it should in the same context also have been elicited by the equivalent word in *immediate*-relative sentences. But as schematized in Table 8 (fourth sentence), no context-induced P600/SPS was observed there, neither at the identical word *had* (or *zich*) nor at the word that first suggests that a restrictive relative clause might follow (*die*; also see Table 7, and Fig. 5b). This clearly suggests that the P600/SPS effect observed in the early-relative sentences hinges on the presence of a local structural ambiguity rather than on the pragmatic infelicity of using particular

constructions in particular discourse contexts. In all, it seems that an account based on the infelicity of particular constructions cannot easily explain the present pattern of P600/SPS results.

Before we can accept this pattern of results as evidence for context-sensitive parsing, we also need to address the fact that the three critical P600/SPS effects had statistically different scalp distributions. This is usually taken to reflect the involvement of (at least partly) different underlying neural generators and, hence, presumably, of (partly) different functional processes. Thus, it is important to consider to what extent the three results truly reflect the same phenomenon, i.e., the processing consequences of having to abandon a hitherto preferred structural analysis.

Although often reported to have a centroparietal maximum, the scalp distribution of the P600/SPS has been noted to vary before (e.g., Brown & Hagoort, in press; Hagoort & Brown, 1998; Osterhout & Holcomb, 1992). The reasons for this are as yet unknown, but it is conceivable that the P600/SPS is generated by a set of partly nonoverlapping syntax-sensitive neural generators, the exact composition of which may vary across conditions (Hagoort & Brown, 1998; Osterhout & Hagoort, 1999). Note, in this respect, that the three relevant conditions differ in the syntactic devices used to disambiguate the complement/relative-clause ambiguity as well as in the exact reanalysis that is subsequently required. Such differences may well be responsible for variability in the distribution of the P600/SPS. (In the current study, overlap with the earlier referential ambiguity effect also clearly contributes to—or perhaps better, obscures—variability in the distribution of the three P600/SPS effects.)

The existence of finer theoretical distinctions does not, however, weaken the interpretation of the P600/SPS as a class of ERP responses that within the domain of language processing all reflect the processing consequences of a word that renders the preferred *syntactic* analysis untenable, rather than, for example, the processing consequences of a word that complicates the semantic analysis of the language input. In view

of their similarity to P600/SPS effects reported in the literature (and the distributional variability observed there), we are confident to classify all three context effects at hand as belonging to the class of P600/SPS effects. Note also that the three P600/SPS effects in this study are not only specific in the sense that they occur exactly when (and only when) a context-sensitive parsing account led us to predict a syntactic dead end, but also in the sense that two other classes of “unexpected linguistic events”—referential ambiguity at a preceding noun and occasional later discourse-semantic violations—elicited qualitatively different ERP effects within the very same target sentences: a frontal sustained negativity at the noun and a discourse-dependent N400 effect further downstream (see van Berkum, Hagoort & Brown, in press, for the latter). Thus, within the domain of processing a single sentence, the P600/SPS effects selectively index *particular* unexpected events only (see Osterhout & Hagoort, 1999, for extensive discussion on the relation of the P600/SPS to the P300 family of ERP effects elicited by unexpected events in general).

Apart from the predicted P600/SPS effects, we also observed significant earlier negativities that peaked around 400 ms after early- and immediate-relative disambiguation in one-referent contexts as well as comparable but substantially smaller trends after early- and immediate-complement disambiguation in two-referent contexts. In view of the posterior focus of these N400-like effects, *they* perhaps do reflect the processing consequences of detecting an infelicitous construction, detecting it either directly, as in the immediate-relative sentences, or upon disambiguation, as in the other three sentence types. If so, then the larger effect size for relative-clause targets perhaps reflects an infelicity asymmetry: it may well be less felicitous to begin a restrictive relative clause when the referent is already uniquely identified (since this will almost inevitably lead to overspecificity) than to begin a complement clause when the referent is still ambiguous (since further referential cues can in principle still be given). Note, however, that the onset of the negativity observed with immediate-relative targets is unusu-

ally early for a visually evoked N400 effect. The exact interpretation of these early negativities must thus await further research.

In our view, the pattern of P600/SPS results presented here poses a clear challenge to syntax-first accounts of early parsing (e.g. Frazier, 1987; Mitchell et al., 1995). Note, in particular, that the garden paths induced by two-referent contexts in early- and immediate-complement sentences are elicited in what would under a syntax-first account be the structurally *preferred* alternative, which is where this phenomenon should never occur (Frazier, 1995). Furthermore, in most of the complement/relative-clause ambiguities used in our experiment (78%), the complement clause is an obligatory argument of the main clause verb. For this reason, our findings also appear to be at odds with two hybrid accounts of parsing according to which an obligatory argument, if available, should always be the initially preferred option (Construal Theory; Frazier & Clifton, 1996; Restricted Interactive Parsing Theory; Britt et al., 1993; Britt, 1994).

Our findings directly confirm the original referential predictions of Crain and Steedman (1985). We agree with Mitchell (1994), however, that referential factors alone cannot explain the full range of evidence on how the parser initially resolves structural ambiguities. In view of the evidence (see Tanenhaus & Trueswell, 1995, or Mitchell, 1994, for overviews) it is conceivable that the parser is in fact able to draw upon *any* type of information, as long as this reliably correlates with particular structural alternatives and becomes available rapidly enough to bear on initial processing. Although the notion of such liberal use of information sources is by no means at odds with symbolic and hybrid models of parsing (in particular, see Lewis, 1993; see also Kempen & Vosse, 1989; Kempen, 1998), it has recently become most strongly associated with the constraint-based or dynamic systems approach to sentence processing (e.g., MacDonald et al., 1994; Tabor et al., 1997; Tanenhaus & Trueswell, 1995; see Spivey-Knowlton & Tanenhaus, 1998, and Tanenhaus, Spivey-Knowlton, & Hanna, in

press, for models in which referential factors interact with other constraints).

We sympathize with the idea that once the structural alternatives have become available, no privileged single factor determines how the parser provisionally resolves the ambiguity at hand. However, before one can embrace a constraint-based account of our findings, one should note that it is as yet difficult to see why, in a constraint-based approach to language comprehension, some types of conflicts between multiple constraints generate an N400 effect in event-related brain potentials, while others elicit a P600/SPS (see Brown & Hagoort, in press, for further discussion). Also note that earlier failures to obtain rapid referential context effects in the English complement/relative-clause ambiguity have in this framework been explained, in part, as the result of an overly skewed availability of the two competing alternatives in English (Spivey-Knowlton & Tanenhaus, 1994). In written Dutch, the complementizer *dat* occurs about five times as often as the relative pronoun *dat* (uit den Boogaart, 1975), and conditional frequencies in the context of verbs like *vertelde* (*told*) will no doubt reveal an even larger asymmetry. Still, our results show that this does not prevent referential context effects from showing up in event-related brain potentials, suggesting that the above asymmetry account does not generally apply to referential effects in parsing sentences with a complement/relative-clause ambiguity.

In all, our evidence from event-related brain potentials suggests that when a sentence is processed in discourse, definite noun phrases are very rapidly associated with potential referents, and the outcome of this referent identification process can immediately help resolve a local structural ambiguity arising somewhat later in the same sentence. We do not deny that sentence processing is "syntax-first" in the sense that it is driven by word class and other central syntactic information associated with the incoming words. After all, it is this information that defines a definite noun phrase, or a particular structural ambiguity, in the first place. But as soon as a definite noun phrase emerges, people look for a referent in the discourse context.

And as soon as a structural ambiguity emerges, people evaluate the alternatives against the current discourse to help in resolving it. In these two ways, our findings testify to the highly incremental nature of sentence processing in which structural and referential aspects appear to go hand in hand.

## REFERENCES

- Ainsworth-Darnell, K., Shulman, H. G., & Boland, J. (1998). Dissociating brain responses to syntactic and semantic anomalies: Evidence from event-related potentials. *Journal of Memory and Language*, **38**, 112–130.
- Altmann, G. T. M. (1988). Ambiguity, parsing strategies, and computational models. *Language and Cognitive Processes*, **3**, 73–97.
- Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language*, **31**, 685–712.
- Altmann, G. T. M., Garnham, A., & Henstra, J.-A. (1994). Effects of syntax in human sentence parsing: Evidence against a structure-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 209–216.
- Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, **30**, 191–238.
- Altmann, G. T. M., Van Nice, K. Y., Garnham, A., & Henstra, J.-A. (1998). Late Closure in context. *Journal of Memory and Language*, **38**, 459–484.
- van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1998). *Rapid discourse context effects in sentence processing: ERP evidence*. Paper presented at the 11th Annual CUNY Conference on Human Sentence Processing, New Brunswick, March 19–21.
- van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (in press). When does gender constrain parsing? Evidence from ERPs. *Journal of Psycholinguistic Research*.
- van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (in press). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*.
- Boland, J. E. (1997). Resolving syntactic category ambiguities in discourse context: Probabilistic and discourse constraints. *Journal of Memory and Language*, **36**, 588–615.
- uit den Boogaart, P. C. (Ed.). (1975). *Woordfrequenties in geschreven en gesproken Nederlands*. Utrecht, The Netherlands: Oosthoek, Scheltema & Holkema.
- Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, **33**, 251–283.
- Britt, M. A., Gabrys, G., & Perfetti, C. A. (1993). A restricted interactive model of parsing. In *Proceedings of the 15th Annual Conference of the Cognitive Science*

- Society* (pp. 260–265). Univ. of Colorado: Boulder, CO.
- Britt, M. A., Perfetti, C. A., Garrod, S., & Rayner, K. (1992). Parsing in discourse: Context effects and their limits. *Journal of Memory and Language*, **31**, 293–314.
- Brown, C. M., & Hagoort, P. (in press). On the electrophysiology of language comprehension: Implications for the human language system. In M. Crocker, M. Pickering., & C. Clifton Jr. (Eds.), *Architectures and mechanisms for language processing*. Cambridge, UK: Cambridge Univ. Press.
- Brown, C. M., Hagoort, P., & Vonk, W. (1997). *Semantic effects on syntactic analyses: ERP evidence*. Poster presented at the Fourth Annual Meeting of the Cognitive Neuroscience Society, Boston, March 23–25, 1997.
- Brybaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *The Quarterly Journal of Experimental Psychology*, **49A**, 664–695.
- Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge Univ. Press.
- Clifton, C., Jr., & Ferreira, F. (1989). Ambiguity in context. *Language and Cognitive Processes*, **4**, 77–103.
- Crain, S., Ni, W., Schankweiler, D., Conway, L., & Braze, D. (1996). *Meaning, memory and modularity*. MIT Occasional papers in Linguistics 9 (Proceedings of the NELS 26 Sentence Processing Workshop).
- 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 Univ. Press.
- De Vincenzi, M., & Job, R. (1995). An investigation of late closure: The role of syntax, thematic structure, and pragmatics in initial and final interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 1303–1321.
- Dell, G. S., McKoon, G., & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior*, **22**, 121–132.
- Duffy, S. A., & Rayner, K. (1990). Eye movements and anaphor resolution: Effects of antecedent typicality and distance. *Language and Speech*, **33**, 103–119.
- Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, **25**, 348–368.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 559–585). Hove/London, UK: Erlbaum.
- Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research*, **24**, 437–468.
- Frazier, L., & Clifton, C., Jr. (1996). *Construal*. Cambridge, MA: MIT Press.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, **14**, 178–210.
- 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., Oakhill, J., & Cain, K. (1997). The interpretation of anaphoric noun phrases: Time course, and effects of overspecificity. *The Quarterly Journal of Experimental Psychology*, **50A**, 149–162.
- Garrod, S. C. (1994). Resolving pronouns and other anaphoric devices: The case for diversity in discourse processing. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 339–357). Hillsdale, NJ: Erlbaum.
- Garrod, S., Freudenthal, D., & Boyle, E. (1993). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory and Language*, **32**, 1–30.
- Garrod, S. C., O'Brien, E. J., Morris, R. K., & Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 250–257.
- Garrod, S. C., & Sanford, A. J. (1994). Resolving sentences in discourse context: How discourse representation affects language understanding. 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.
- Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**, 266–283.
- 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. (1997). *When syntax and semantics meet*. Poster presented at the Fourth Annual Meeting of the Cognitive Neuroscience Society, Boston, March 23–25, 1997.
- Hagoort, P., & Brown, C. M. (1998). *Semantic and syntactic ERP effects of listening to speech compared to reading*. Manuscript under review.
- Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (P600/SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, **8**, 439–483.
- Kempen, G. (1998). *A uniform cognitive architecture for human grammatical encoding and decoding*. Manuscript, Leiden University.

- Kempen, G., & Vosse, T. (1989). Incremental syntactic tree formation in human sentence processing: A cognitive architecture based on activation decay and simulated annealing. *Connection Science*, **1**, 273–290.
- 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.
- Kluender, R., & Münte, T. F. (1998). *ERPs to grammatical and ungrammatical subject/object asymmetries in German wh-questions*. Paper presented at the 11th Annual CUNY Conference on Human Sentence Processing, New Brunswick, NJ, March 19–21.
- Konieczny, L., Hemforth, B., & Voelker, N. (1994). The impact of context and semantic bias on constituent attachment in reading. In B. Hemforth, L. Konieczny, C. Scheepers, & G. Strube (Eds.), *First analysis, re-analysis, and repair* (Tech. Rep. No. STR-30/14-2). Freiburg: Albert-Ludwigs-Universität Freiburg.
- 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., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, **207**, 203–205.
- Kutas, M., & King, J. W. (1996). The potentials for basic sentence processing: Differentiating integrative processes. In T. Inui & J.L. McClelland (Eds.), *Attention and performance XVI* (pp. 501–546). Cambridge, MA: MIT Press.
- 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.
- Lewis, R. L. (1993). *An architecturally-based theory of human sentence comprehension* (Tech. Rep. No. CMU-CS-93-226). Pittsburgh: Carnegie Mellon University.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, **101**, 676–703.
- Masson, M. E. J. (1983). Conceptual processing of text during skimming and rapid sequential reading. *Memory & Cognition*, **11**, 262–274.
- McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, **62**, 203–208.
- Mitchell, D. C. (1994). Sentence parsing. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 375–409). New York: Academic Press.
- Mitchell, D. C., Corley, M. M. B., & Garnham, A. (1992). Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**, 69–88.
- 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.
- Murray, W. S., & Liversedge, S. P. (1994). Referential context effects in syntactic processing. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 359–388). Hillsdale, NJ: Erlbaum.
- 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.
- Ni, W., Crain, S., & Schankweiler, D. (1996). Sidestepping garden paths: Assessing the contribution of syntax, semantics and plausibility in resolving ambiguities. *Language and Cognitive Processes*, **11**, 283–334.
- Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King, and Kutas (1998) in the P600/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 brain potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes*, **8**, 413–437.
- Osterhout, L., & Holcomb, P. J. (1995). Event-related potentials and language comprehension. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind* (pp. 171–215). Oxford, UK: Oxford Univ. Press.
- Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 786–803.
- Osterhout, L., McKinnon, R. Bersick, M., & Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: Is the Syntactic Positive Shift a member of the P300 family? *Journal of Cognitive Neuroscience*, **8**, 507–526.
- Osterhout, L., McLaughlin, J., & Bersick, M. (1997). Event-related brain potentials and human language. *Trends in Cognitive Sciences*, **1**, 203–209.
- 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.
- Rayner, K., Garrod, S., & Perfetti, C. A. (1992). Discourse influences during parsing are delayed. *Cognition*, **45**, 109–139.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.
- Rayner, K., & Sereno, S. C. (1994). Eye movements in reading. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 57–81). New York: Academic Press.
- 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., Pechmann, T., Streb, J., Röder, B., & Henninghausen, 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.
- Spivey-Knowlton, M. J., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, **55**, 227–267.
- Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1994). Referential context and syntactic ambiguity resolution. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 415–439). Hillsdale, NJ: Erlbaum.
- Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24**, 1521–1543.
- Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology*, **47**, 276–309.
- St. George, M., Mannes, S., & Hoffman, J. E. (1994). Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience*, **6**, 70–83.
- St. George, M., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience*, **9**, 776–787.
- Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, **12**, 211–271.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1996). Using eye movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation. In T. Inui & J.L. McClelland (Eds.), *Attention and performance XVI* (pp. 457–478). Cambridge, MA: MIT Press.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., & Hanna, J. (in press). Modelling thematic and discourse context effects within a multiple constraints framework: Implications for the architecture of the language comprehension system. In M. Crocker, M. Pickering, & C. Clifton Jr. (Eds.), *Architectures and mechanisms for language processing*. Cambridge, UK: Cambridge Univ. Press.
- 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 Nice, K. Y., Garnham, A., & Altmann, G. T. M. (1997). *Referential context and the parsing of conjoined definite nouns*. Poster presented at the 10th Annual CUNY Conference on Human Sentence Processing, Santa Monica, CA, March 20–22, 1997.

(Received August 12, 1998)

(Revision received January 14, 1999)