The Use of Referential Context and Grammatical Gender in Parsing: A Reply to Brysbaert and Mitchell (2000)

Jos J.A. van Berkum,[1,2] **Peter Hagoort,**[1] **and Colin M. Brown**[1]

Based on the results of an event-related brain potentials (ERP) experiment (van Berkum, Brown, & Hagoort. 1999a, b), we have recently argued that discourse-level referential context can be taken into account extremely rapidly by the parser. Moreover, our ERP results indicated that local grammatical gender information, although available within a few hundred milliseconds from word onset, is not always used quickly enough to prevent the parser from considering a discourse-supported, but agreement-violating, syntactic analysis. In a comment on our work, Brysbaert and Mitchell (2000) have raised concerns about the methodology of our ERP experiment and have challenged our interpretation of the results. In this reply, we argue that these concerns are unwarranted and, that, in contrast to our own interpretation, the alternative explanations provided by Brysbaert and Mitchell do not account for the full pattern of ERP results.

KEY WORDS: referential context; grammatical gender; syntactic ambiguity; parsing; ERP.

INTRODUCTION

An important issue in psycholinguistics is how readers and listeners incrementally assign a syntactic analysis to the unfolding sentential input and what types of information can be brought to bear on syntactic parsing at which moments in time. Based on the results of an event-related brain potentials (ERP) experiment, we recently argued that referential context information can be taken into account by the parser at a very early moment in

time (van Berkum, Brown, & Hagoort, 1999a, b), in line with so-called "context-sensitive" accounts of parsing. From the same ERP results, we also inferred that grammatical gender information, although unequivocally used within a few hundred milliseconds after having become available to our subjects, is not always used quickly enough to prevent the parser from considering a discourse-supported syntactic analysis that actually violates a gender agreement constraint.

In a critical comment on our work, Brysbaert and Mitchell (2000) have recently challenged the validity of both interpretations of our data. With respect to the impact of referential context on parsing, Brysbaert and Mitchell argue that, in part because of certain features of our methodology, our results do not provide evidence against so-called "syntax-first" models of syntactic ambiguity resolution. With respect to the impact of grammatical gender, Brysbaert and Mitchell (henceforth, B&M) propose an alternative, semantics-based explanation for our data, and, furthermore, present the results of a questionnaire study that suggest that readers sometimes do not take gender into account at all.

We take up each of these issues below. To anticipate the main conclusions, we will argue that B&M's methodological concerns are unwarranted, that their alternative account of our gender-related findings fails to explain the complete pattern of results, and that their gender-related questionnaire findings are not necessarily at odds with our data.

CONTEXT-SENSITIVE INITIAL PARSING

In van Berkum *et al.* (1999a), we presented ERP evidence for very early contextual modulations of the way the parser deals with a local syntactic ambiguity. The evidence came from a study in which Dutch subjects read target sentences beginning like *David vertelde het meisje dat.* . . (*David told the girl*$_{NEU}$ *that.* . .). As in English, a local syntactic ambiguity arises at the word *dat,* which can introduce a complement clause (example 1), as well as a relative clause (example 2).

(1) David vertelde het meisje *dat*   **er**   visite kwam.
    [*David told*   *the girl*$_{NEU}$ *that*$_{COMPL}$ **there**   *would be some visitors.*]

(2) David vertelde het meisje *dat*   **had**   zitten bellen op te hangen.

    [*David told*   *the girl*$_{NEU}$ *that*$_{RELPR(NEU)}$ **had**   *been phoning to hang up.*]

Following Crain, Steedman, and Altmann (e.g., Crain & Steedman, 1985; Altmann, 1988), we placed these temporarily ambiguous target sentences in two types of discourse context. A so-called *two-referent* context introduced two plausible referents for the critical noun phrase (*het meisje* in examples 1 and 2), and, as such, favored a relative-clause continuation. The otherwise identical one-referent context introduced a single unique referent for the NP and, as such, favored a complement–clause continuation instead. According to context-sensitive theories of parsing, the parser can make immediate use of such discourse-level information when resolving a local syntactic ambiguity (e.g., Altmann, 1988; Crain & Steedman, 1985; Spivey-Knowlton & Tanenhaus, 1998; Vosse & Kempen, 2000). According to syntax-first theories of parsing, however, this information would initially be ignored (e.g., De Vincenzi & Job, 1995; Ferreira & Clifton, 1986; Frazier, 1987; Mitchell, Cuetos, Corley, & Brysbaert, 1995).

To assess the parser's commitments made at *dat,* we disambiguated the sentence at the very next word (shown in boldface in 1 and 2), while recording event-related brain potentials from our subjects. Based on earlier consistent observations that a word at which the parser must abandon its hitherto preferred syntactic analysis elicits a so-called P600/SPS effect in the ERPs (for reviews, see Brown, Hagoort, & Kutas, 2000; Hagoort, Brown, & Osterhout, 1999; Osterhout & Hagoort, 1999), we reasoned that to the extent that a two-referent context would lead the parser to pursue the relative-clause analysis at *dat,* subsequent disambiguation as a *complement* clause (at *er*) should elicit this P600/SPS "garden-path" effect. Also, to the extent that a *one-referent* context would bias the parser to pursue the complement-clause analysis instead, this should lead to a P600/SPS effect at *relative* clause disambiguation (at *had*) in this context. It was exactly this cross-over pattern of results that we observed in the ERP waveforms. We, therefore, took our findings to indicate that discourse-level referential context can very rapidly influence the analysis of a syntactic ambiguity.

The Wrong ERP Effect?

B&M challenge our interpretation of these findings on several grounds. One of their concerns is "to what extent the P600/SPS is a marker of the initial syntactic analysis as conceived by syntax-first models." With respect to this issue, B&M report that "although van Berkum *et al.* give the impression it is, a closer look at the ERP literature shows that this is not a generally accepted assumption." To our knowledge, no ERP specialist working in this field has, indeed, ever claimed that the P600/SPS is a marker of the initial syntactic analysis as conceived by syntax-first models. Certainly none

of us has ever claimed this. In fact, in the main article describing the work at hand, we explicitly state 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" (van Berkum *et al.,* 1999a, p. 152). In contrast to what B&M suggest, the latter (much weaker) assumption *is* generally accepted by the ERP community, even by researchers who take the P600 to more generally reflect syntactic difficulty (which subsumes dead ends; Kaan, Harris, Gibson, & Holcomb, 2000), and those who have challenged the uniqueness of the P600/SPS effect vis-à-vis the family of P300 effects (e.g., Coulson, King, & Kutas, 1998a, b; Gunter, Stowe, & Mulder, 1997; see Osterhout and Hagoort, 1999, for discussion). As for the P300 debate, note that even if the P600/SPS effect *would* be a member of the P300 family, this would not in any way affect our assumption that within the domain of language comprehension, critical words at which the parser runs into a syntactic dead end reliably elicit this particular ERP effect.

Not Enough Fillers?

An important concern raised by B&M is that the pattern of ERP results that we observed may be an artifact caused by a shortage of filler trials. B&M report that we had only 40 such trials in the experiment, but, in fact, we had no noncritical filler trials at all. That is, all 240 story trials in our experiment participated in the study's parsing design. At first sight, this would seem to fuel the concern even more ("no fillers at all!"). However, we believe that the absence of filler trials is by no means necessarily a flaw in one's design.

First of all, critical trials of a particular type can very well act as fillers for critical trials of some other type. Our study contained four syntactically different sentence types, crossed with two very different lexical means of disambiguation (see van Berkum *et al.,* 1999a for details). A structured posttest revealed that our subjects had completely missed these alternations, as well as the existence of local syntactic ambiguities. To further distract our subjects, we also deliberately included semantically anomalous words in the tails of 40 critical trials (which, as the posttest revealed, indeed attracted the subjects' attention in particular). In addition, and in part because there would be no other task than to attend to our stories, we took great pains to create short discourses that were inherently interesting and variable enough to engage our subjects in a natural language processing mode. Finally, we avoided that undesirable (e.g., predictive) strategies would "pay off" by fully crossing all our factors such that, for example, a two-referent context would be followed by as many complement clauses as relative clauses, or

by as many neuter-gender NPs as common-gender NPs (which determined the later point of syntactic disambiguation; see below). In all, we believe that we have taken reasonable and presumably sufficient precautions to avoid the use of strategies based on critical alternations in the materials.[4]

Word-by-Word Presentation Too Slow?

A related concern raised by B&M involves the fact that we presented our critical target sentences word by word, with a new word every 600 ms. The use of this relatively slow Serial Visual Presentation (SVP) procedure would, according to B&M, "seem likely to have introduced priming effects and to have allowed time for the use of tactics involving reanalysis of the material." As explained elsewhere (van Berkum *et al.,* 1999a), the procedure we used is a common and well-advised one for ERP experiments that try to cover new ground, because it allows for a relatively unambiguous time-locking of the effects to particular words. The a priori validity of this research strategy is backed up by a small, but growing, database of ERP studies in which a variety of sentence-processing effects obtained with relatively slow (600 ms/word) SVP were found to be highly comparable to those obtained with faster (250 ms/word) SVP rates, as well as with natural speech input (see, for example, van Berkum, Hagoort, Brown, & Zwitserlood, 1999 and van Berkum, Hagoort & Brown, 1999, for discourse-level N400 effects; Kutas, 1993 and Hagoort & Brown, 2000a, for sentence-level N400 effects; Hagoort & Brown, 2000b for sentence-level P600/SPS effects; and Kutas, 1997, for memory-related slow ERP shifts). Obviously, the existence of several consistent ERP findings across input rate and modality does not imply that every other ERP finding will also generalize across these input parameters. However, we do take this database to indicate that a new sentence-processing result obtained with a 600 ms/word SVP procedure should not be lightly discarded simply *because* of that paradigm.

Although we had good reasons to rely on the validity of our written-language findings, we nevertheless conducted an exact spoken-language

---

[4] Many of our subjects did actually notice that expressions sometimes failed to refer successfully. We believe this to be an unavoidable consequence of the fact that the explicit and successful identification of referents is central to language comprehension. People simply notice referential ambiguity right away. The consequence, however, is that although having just 5% of such referential ambiguities in our study would have avoided the "no fillers" criticism, it would not have hidden the phenomenon from view. All one can do is to disperse the subject's attention over a wider range of other salient phenomena and to make sure that the development of a strategy based on the referential phenomenon does not pay off.

variant of our experiment, with fully natural and fully connected speech versions of the same materials (see Brown, van Berkum, & Hagoort, 2000, for details). The resulting ERP data are somewhat more noisy because of the use of connected speech. Nevertheless, we obtained a pattern of ERP results that largely overlapped with the van Berkum *et al.* 1999a written-language findings (see van Berkum, Brown, & Hagoort, submitted, and Brown *et al.,* 2000 for this replication). In addition to providing converging evidence for our interpretation of the written-language data, these spoken-language results also underline our claim that there is no reason to simply put aside sentence-processing data just because they have been obtained with a 600 ms/word SVP procedure. In the absence of data to the contrary (which B&M do not present), B&M's suggestion about undesirable side effects due to the procedure can thus be put aside.

What Would Be the Alternative?

In all, we believe that the concerns that B&M raise over "the density of experimental trials (relative to fillers) and the slow presentation rate (600 ms per word)" are unwarranted. In this respect, we also note that there was no account of *how* these specific features of our experiment would have led to the intricate and highly selective pattern of referential context effects that we observed. That is, B&M do not in any way expand on their comment that these features "seem likely to have introduced priming effects and to have allowed time for the use of tactics involving reanalysis of the material," nor do they explain how "sentence to sentence priming" and "short-term tuning" could be responsible for our *discourse*-dependent effects. In the face of a highly specific pattern of ERP results that exactly matches a coherent theoretical a priori prediction, simply mentioning "tactics" and "priming" is, in our opinion, not a realistic alternative.

Probing Too Late?

The "reanalysis" remark does point to a perennial methodological problem for research on parsing, which is that any referential (or other nonsyntactic) context effect on how the parser resolves a syntactic ambiguity is necessarily observed at *some* temporal delay relative to the onset of the ambiguity. In the 1980s, such context effects were usually demonstrated by probing the parser's syntactic commitments several words downstream. This opened up a real possibility that the observed effects did not reflect a contextual modulation of the parser's *initial* resolution, but a modulation of later syntactic reanalysis.

After a careful analysis of this problem, Mitchell, Corley, and Garnham (1992) concluded that the best way to overcome it, is to probe for the parser's structural commitments immediately after the first word in the ambiguous region, to as such "minimize the chances that performance will be distorted by the influence of early reanalysis" (p. 85). We, therefore, designed our ERP study in just the way Mitchell *et al.* had suggested and disambiguated at the very next word after *dat* in two of our four critical sentence types (see examples 1 and 2). Needless to say, we were thus somewhat surprised by Brysbaert's and Mitchell's current concern that our effects "could well be due to reanalysis." Note that the early probes exemplified in (1) and (2) leave very little time for such reanalysis, as the latter would have to be finished before the parser attempts to assign the next word.[5]

(3) David vertelde   de vrouw   **dat**   er   visite kwam.
    [*David told   the woman*COM **that**COMPL *there would be some*
                                                            visitors.]

Furthermore, the ERP results we obtained with so-called immediate-complement sentences, exemplified in (3), clearly suggest that a context-induced reanalysis would have to be finished a lot earlier. In contrast to their early-complement counterparts in (1), immediate-complement sentences contained a *common* gender NP. As the form of the relative pronoun for such common gender nouns is *die,* not *dat,* this formally ruled out a relative pronoun reading of the subsequent word (*dat*). However, native speaker intuitions (see van Berkum *et al.,* 1999b for details) led us to consider the possibility that in on-line processing the lexically ambiguous word *dat* would 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 should then be able to modulate the parser's preferences, just as it would at *dat* in the two early-probe sentence types. The clear prediction that followed from this for the immediate-complement sentences, a P600/SPS effect in the ERP waveform to *dat* in a two-referent

---

[5] Although our early probes were presented 600 ms after the onset of the word *dat,* this does not necessarily imply that our subjects had much more time to ponder the ambiguity than, say, those participating in Mitchell *et al.*'s (1992) own reading time study. In the latter, the time a subject would spend on the critical word *that* was actually under their own control (and not separately registered). Also, the mean reading times observed on the subsequent /had been/ disambiguation ranged from 600 to 1063 ms. We do not wish to discredit the Mitchell *et al.* results on these grounds. However, we also do not see why in terms of the immediacy of probing, our study would be in an entirely different ballpark.

context, was confirmed (also see Brown *et al.,* 2000, for a spoken-language replication).[6]

We believe that, in terms of probing initial parsing commitments, this is as immediate as one can get. It certainly meets—and even surpasses—the recommendations of Mitchell *et al.* (1992). Note that the relevant P600/SPS effect emerged at about 450–500 ms after presentation of the critical word *dat.* A syntax-first account would, therefore, need to assume that word recognition, emergence of a structural complement/relative-clause ambiguity, *initial* syntax-first (i.e., context-free) resolution toward the complement clause alternative, context-based evaluation and rejection, subsequent diagnosis, *secondary* commitment to the relative-clause alternative, and detection of the resulting gender agreement violation all take place within the 450–500 ms allowed for between the onset of *dat* and the onset of the P600/SPS.

As long as the time needed to make and subsequently revise a parsing commitment is not specified in syntax-first theories, this could, in principle, be argued for. This raises the issue of whether one could ever probe early enough to falsify a syntax-first account (see also Mitchell, 1994).[7] More important, perhaps, is to ask whether a syntax-first account adjusted to accommodate our findings is still an interesting and plausible story. First of all, why would the parser ignore potentially useful context information in favor of a syntax-based heuristic if the system can bring context informa-

---

[6] Our exact prediction followed from a five-step processing account (see van Berkum *et al.,* 1999a, b) which, as B&M correctly point out, is in certain respects quite similar to an account previously argued for by Mitchell (e.g., 1989). However, B&M do mislocate the similarities, arguing that "the only difference between Mitchell (1989) and van Berkum (1999a, b) is that Mitchell placed steps 4 and 5 in the second, re-analysis stage, whereas van Berkum *et al.* considered them as part of the initial analysis." In fact, we were very explicit in *not* considering steps 4 and 5 as necessarily part of the initial analysis. After all, one of the two main gender-related implications of our work (see van Berkum *et al.,* 1999b for details) was that gender, although acting fast, apparently only came into play after a first commitment had already been made on the basis of a discourse bias. Furthermore, we explicitly argued that the P600/SPS effect generated in step 5 need not *itself* directly reflect initial parsing operations—all we took it to reflect was a syntactic dead end (which, as such, did betray a discourse modulation of the initial parse). In all, it seems that our steps 4 and 5 are actually in reasonable agreement with the Mitchell (1989) lexical-filtering account. The truly critical difference between the two accounts, of course, is in whether discourse-level factors are allowed to play a role in step 3.

[7] Frazier (1995) has suggested that syntax-first accounts can be falsified by demonstrating a context-induced garden-path effect in what would be the preferred structure under that account. With the complement-clause alternative being the preferred structure under currently formulated syntax-first theories, we take our P600/SPS effects in early- and immediate-complement sentences (examples 1 and 3, respectively) to demonstrate exactly this.

tion to bear on parsing so quickly? If discourse-level context information can be used to revise initial parsing commitments 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 architectural principle, initially make those commitments on the basis of structural information only (cf. Altmann, 1988, for a similar argument). Note that a syntax-first heuristic based on structure-associated information (e.g. frequency, simplicity) is not necessarily "more exact" than a context-sensitive heuristic based on discourse-level referential information—both can simply be wrong. It is therefore relevant to ask why, as syntax-first theorists would have to maintain, the parsing system would have evolved (biologically and/or developmentally) to systematically ignore part of the "circumstantial evidence" that is available to probabilistically improve its decisions.

One of the classical arguments for why the parser would initially ignore discourse-level (and other nonsyntactic) context was that it would just take too much time and computation to bring this principally unbounded knowledge source to bear (Fodor, 1983). However, whatever difficulty we may have in capturing these effects in our computational models without a combinatorial explosion or other such disaster, "on-line" measurements (e.g., this study; Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; van Berkum, Hagoort, & Brown, 1999; van Berkum, Hagoort, Brown, & Zwitserlood, 1999; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Federmeier & Kutas, 1999a, b; Sedivy, Tanenhaus, Chambers, & Carlson, 1999) simply tell us that referential and semantic information about the linguistic and non-linguistic context *can* be brought to bear on language processing extremely quickly.

RAPID (BUT NOT TOO RAPID) USE OF GENDER

The referentially induced P600/SPS effect on *dat* in sentences like (3) played an important role in our case for very early referential context effects on parsing. In addition, the effect clearly suggested two things about the use of grammatical gender in parsing. First, the presence of a P600/SPS effect induced by the relative-clause biasing two-referent context revealed that the parser was somehow able to, at least momentarily, pursue a relative-clause analysis, although this analysis violated the rules of grammatical gender. This indicated that grammatical gender did not immediately block an ill-formed analysis from consideration. Second, the fact that this P600/SPS effect was elicited by the word *dat* (and not by the subsequent word, e.g., *er* in example 3) revealed that grammatical gender, although not used to immediately block the ill-formed analysis, *was* used by the parser rapidly enough to dispose of

it within only a few hundred milliseconds, at least for the syntactic structures under consideration (see van Berkum *et al.,* 1999b, for details).

A Semantic Reinterpretation?

Brysbaert and Mitchell express considerable doubts about our gender-related interpretation of this P600/SPS effect. As part of their argument, they present an alternative explanation of our gender results, findings that according to B&M "may have little to do with syntactic parsing because they may be more in line with a semantic interpretation." Furthermore, B&M present results of a questionnaire that suggest that readers sometimes do not seem to use grammatical gender at all, and in those cases focus on semantic factors instead.

The latter finding is clearly of interest. In fact, we were aware of these questionnaire data when we designed our ERP experiment in early 1996. They confirmed earlier reports of Dutch linguists (e.g., Verhoeven, 1990) that semantic factors seemed to invade the Dutch grammatical gender system at specific points, particularly with respect to the use of the relative pronoun *die.* It was partly because of these questionnaire findings and linguistic observations that we designed our critical conditions around the more reliable neuter-gender relative pronoun *dat.* The immediate-complement result, a P600/SPS effect on *dat* in a two-referent context, was exactly as predicted under our context-sensitive parsing account and, as such, also confirmed our working assumption that with respect to the *neuter* relative pronoun the Dutch grammatical gender system was still sufficiently intact.

In an attempt to unify their questionnaire data with our own ERP data, B&M propose a rather different account of our findings, arguing that "the reason why van Berkum *et al.* thought they had found an effect of grammatical gender may very well have been the fact that a relative pronoun interpretation of their complementizer *dat* was semantically incongruent with the preceding noun phrase, which always referred to a human (or a human-like) creature." We are puzzled by this, for this account cannot explain why we only observed a P600/SPS effect on the word *dat* after a common-gender NP (de vrouw dat, the woman$_{COM}$ that), and *not* after a neuter-gender NP (het meisje dat, the girl$_{NEU}$ that), both of which involve human or humanlike nominal referents throughout the set of items. Furthermore, the B&M account also cannot explain why our critical P600/SPS effect is brought about *by a modulation of the number of referents* for the noun at hand. Along the lines of our own model, and momentarily abstracting away from the first problem, one could perhaps argue that the two-referent context first lured the parser into pursuing the incorrect relative-clause alternative and that B&M's "semantic incongruity" would subsequently lead to an

ERP effect in this condition. However, note that such an account would actually confirm the very essence of our interpretation.

Is Gender Often Completely Ignored?

The questionnaire data that B&M present also led them to conclude that "the grammatical gender information of a relative pronoun in Dutch is often completely ignored." These results were taken to be at odds with our proposal that grammatical gender, although not blocking ill-formed phrase structure alternatives from being considered, is very rapidly used to dispose of them (van Berkum *et al.,* 1999b). In supporting their argument, B&M additionally refer to earlier self-paced reading and eye-monitoring studies with NP1-(of)-NP2-RC materials (Brysbaert & Mitchell, 1996).

Although it is perhaps somewhat premature to conclude that in Dutch the gender of a relative pronoun is often completely ignored, we agree that these results are at odds with our own. However, in contrast to on-line measurements, off-line questionnaire data only reveal the products of people's final interpretation of a sentence. The fact that this interpretation is modulated by semantic factors, such as whether an NP refers to a human or non-human entity, thus does not necessarily reflect anything about the syntactic parser.[8] This said, the B&M questionnaire data do suggest that the final interpretation of sentences with an NP1-(of)-NP2-RC construction is not completely governed by the availability of disambiguating grammatical gender information.

Perhaps more convincing in terms of an on-line syntactic parsing account is that the Brysbaert and Mitchell (1996) reading-time experiments also suggest a failure to use this information. We agree with B&M that the discrepancy between the Brysbaert and Mitchell (1996) findings and our own results can in this particular case not be explained in terms of a difference in the sensitivity of the respective response measures (reading times vs. ERPs). One possibility, suggested in van Berkum *et al.* (1999b), is that the absence of an effect in the Brysbaert & Mitchell (1996) study may have been caused by relying on the relative pronoun *die* in half of the trials, a pronoun whose gender-marking appears to be particularly subject to erosion (Verhoeven, 1990). If this is the case, then a reanalysis of the Brysbaert and Mitchell

---

[8]B&M anticipated this response to their questionnaire data (which is, in effect, the strongest possible version of their own concern about probing the parser's initial commitments too late). To make their data relevant to on-line parsing, B&M therefore argued that "parsing studies typically show that in the absence of conflicting evidence, readers tend to stick to an initial or early analysis of the material rather than abandoning it in favor of an alternative." This line of argumentation is not convincing, because B&M actually argue elsewhere in their paper that their questionnaire data do indeed reveal the interaction of several conflicting sources of evidence.

(1996) data for their *dat*-items only ought to reveal a clearer impact of grammatical gender in parsing.

Another possibility, noted by B&M as well, is that the impact of grammatical gender on parsing in fact depends on the type of construction (and associated ambiguity) involved. In this respect, it is of interest to observe that other ERP experiments conducted with Dutch in our laboratory have revealed immediate ERP effects of violating agreement between a noun and a gender-marked determiner (Hagoort & Brown, 1999), as well as between a noun and a gender-marked adjective (van Berkum, Zwitserlood, Brown, & Hagoort, 2000). Also, to the extent that Dutch antecedent–pronoun agreement violations of the form "*Anna saw himself*" are taken to reflect the same agreement system (but see van Berkum, 1996, Ch. 2, and van Haeringen, 1954, for a different account), there is additional ERP evidence to suggest that in Dutch, gender is checked very rapidly for such constructions as well (van Berkum, Brown, & Hagoort, 1999c; see Osterhout & Mobley, 1995, for similar data in English). Based on the limited evidence obtained sofar, therefore, it appears that, in Dutch at least, the NP1-(of)-NP2-RC construction may be the odd one out.[9]

CONCLUSIONS

Based on the results of an event-related brain potentials experiment (van Berkum, Brown, & Hagoort, 1999a, b), we have recently argued that discourse-level referential context can be taken into account by the parser extremely rapidly. Moreover, our ERP results indicated that local grammatical gender information, although available to our subjects within a few hundred milliseconds from word onset, is not always used quickly enough to prevent the parser from considering a discourse-supported, but agreement-violating syntactic analysis. In a comment on our work, Brysbaert and Mitchell (2000) have challenged the validity of both claims. With respect to the impact of grammatical gender, B&M propose an alternative account based on semantics, and present the findings of a questionnaire study that suggest that readers sometimes fail to adequately take gender into account at all. We

---

[9]Strictly speaking, evidence for the rapid *detection* of gender violations does not necessarily suggest that grammatical gender is effectively *used* by the parser. That is, one may want to distinguish effects that "merely" reveal that violations of agreement are detected at some level of the system (e.g., van Berkum *et al.,* 2000; Hagoort & Brown, 1999) from effects that additionally reveal this detection to have an impact on subsequent parsing (van Berkum *et al.,* 1999a, b; Brown, van Berkum, & Hagoort, 2000) or not (Brysbaert & Mitchell, 1996). Note that it is difficult to see, however, why a system would check for adequate gender agreement *without* subsequently acting upon the detected violations.

have shown that the alternative semantic explanation that B&M propose is an implausible and insufficient account of our gender-related data. We also note that their gender-related questionnaire findings are not necessarily at odds with our own, and that in the face of other gender-related datasets, the use of gender in the NP1-(of)-NP2-RC constructions that Brysbaert and Mitchell focus on may be somewhat atypical.

With respect to our referential context effects on parsing, Brysbaert and Mitchell primarily challenged the methodological adequacy of our experiment, criticisms that we have argued to be unwarranted. Brysbaert and Mitchell also pointed out that regardless of how rapidly referential context can affect the provisional resolution of a syntactic ambiguity, the *initial* syntax-first phase of this process can still have occurred before. Given the temporal constraints imposed by our ERP results, we take this to primarily reveal a problem with the falsifiability of syntax-first theories, and not so much a problem with our design. Perhaps more important, the retreat into a shorter and shorter initial phase in which syntactic ambiguity resolution is syntax-first eliminates much of the theoretical appeal (and empirical relevance) of a syntax-first position.

In this respect, it seems more reasonable to take our findings as revealing referential context effects on the *initial* parsing of a syntactic ambiguity. More generally, however, and independent of whether we have really tapped initial parsing or just missed it, our written- and spoken-language ERP findings are a straightforward demonstration of very early effects of discourse-level referential context in syntactic processing. In addition, our ERP findings reveal that such discourse-level information can operate *before* the parser has had time to bring a hard grammatical gender constraint to bear on its syntactic analysis. The implications of this for models of language comprehension obviously extend beyond the "initial-parsing" debate.

REFERENCES

Altmann, G. T. M. (1988). Ambiguity, parsing strategies, and computational models. *Language and Cognitive Processes, 3,* 73–97.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition, 76,* B13–B26.

Brown, C. M., van Berkum, J. J. A., & Hagoort, P. (2000). Discourse before gender: An event-related brain potential study on the interplay of semantic and syntactic information during spoken-language understanding. *Journal of Psycholinguistic Research, 29,* 53–68.

Brown, C. M., Hagoort, P., & Kutas, M. (2000). Postlexical integration processes in language comprehension: Evidence from brain-imaging research. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 881–895). Cambridge, Massachusetts: MIT Press.

Brysbaert, M., & Mitchell, D.C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *The Quarterly Journal of Experimental Psychology, 49A,* 664–695.

Brysbaert, M., & Mitchell, D.C. (2000). The failure to use gender information in parsing: A comment on van Berkum, Brown, and Hagoort (1999). *Journal of Psycholinguistic Research, 29,* 453–466.

Coulson, S., King, J., & Kutas, M. (1998a). Expecting the unexpected: Event-related brain responses to morphosyntactic violations. *Language and Cognitive Processes, 13,* 21–58.

Coulson, S., King, J., & Kutas, M. (1998b). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes, 13,* 653–672.

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: Cambridge University 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.

Eberhard, K., Spivey-Knowlton, M., Sedivy, J., & Tanenhaus, M. (1995). Eye movements as a window into real-time spoken language processing in natural contexts. *Journal of Psycholinguistic Research, 24,* 409–436.

Federmeier, K.D., & Kutas, M. (1994a). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41,* 469–495.

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

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, Massachusetts: MIT Press.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 559–585). Hove and London, UK: Erlbaum.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research, 24,* 437–468.

Gunter, T.C., Stowe, L.A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology, 34,* 660–676.

Hagoort, P., & Brown, C.M. (1999). Gender electrified: ERP evidence for the syntactic nature of gender processing. *Journal of Psycholinguistic Research, 28,* 715–728.

Hagoort, P., & Brown, C.M. (2000a). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia, 38,* 1518–1530.

Hagoort, P., & Brown, C.M. (2000b). ERP effects of listening to speech compared to reading. The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38,* 1531–1549.

Hagoort, P., Brown, C.M., & Osterhout, L. (1999) The neurocognition of syntactic processing. In C.M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 273–316). Oxford: Oxford University Press.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15,* 159–201.

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.

Mitchell, D.C. (1989). Verb guidance and other lexical effects in parsing. *Language and Cognitive Processes, 4,* 123–154.

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.

Osterhout, L., & Hagoort, P. H. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King, and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes, 14,* 1–14.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34,* 739–773.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71,* 109–147.

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.

van Berkum, J. J. A. (1996). *The Psycholinguistics of grammatical gender: Studies in language comprehension and production.* Nijmegen, The Netherlands: Nijmegen University Press.

van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999a). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41,* 147–182.

van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999b). When does gender constrain parsing? Evidence from ERPs. *Journal of Psycholinguistic Research, 28,* 555–571.

van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999c). Tracking referential processes with event-related (brain potentials. *Proceeding of the 5th Conference on Architectures and Mechanisms of Language Processing* (AMLaP-99), Edinburgh, September 23–25, 1999.

van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (submitted). ERP effects of referential ambiguity in spoken-language comprehension. Manuscript submitted for publication.

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

van Berkum, J. J. A., Hagoort, P., Brown, C. M., & Zwitserlood, P. (1999). Relating spoken sentences to prior discourse: Evidence from the N400. *Proceedings of the 7th Annual Meeting of the Cognitive Neuroscience Society* (CNS-2000), San Francisco, April 9–11, 2000.

van Berkum, J. J. A., Zwitserlood, P., Brown, C. M., & Hagoort, P.(2000). (The computation of gender and number agreement in parsing: An ERP-based comparison. *Proceedings of the 6th Conference on Architectures and Mechanisms of Language Processing* (AMLaP-2000), Leiden, September 21–23, 2000.

van Haeringen, C. B. (1954). *Genus en geslacht: Het voornaamwoordelijk gebruik in de gesproken taal* (Gender and sex: Pronominal reference in spoken language). Amsterdam: Meulenhoff.

Verhoeven, P. R. F. (1990). Voornaamwoordelijke aanduiding in het hedendaagse Nederlands (Pronominal reference in modern Dutch). *De Nieuwe Taalgids, 83,* 494–513.

Vosse, Th., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition, 75,* 105–143.