

## Computing and recomputing discourse models: An ERP study

Giosuè Baggio<sup>a,\*</sup>, Michiel van Lambalgen<sup>b</sup>, Peter Hagoort<sup>c,d</sup>

<sup>a</sup> *F. C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, P.O. Box 9101, NL-6500 HB Nijmegen, The Netherlands*

<sup>b</sup> *Institute for Logic, Language and Computation University of Amsterdam, The Netherlands*

<sup>c</sup> *F. C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, The Netherlands*

<sup>d</sup> *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

Received 11 December 2007; revision received 11 February 2008

Available online 28 April 2008

---

### Abstract

While syntactic reanalysis has been extensively investigated in psycholinguistics, comparatively little is known about reanalysis in the semantic domain. We used event-related brain potentials (ERPs) to keep track of semantic processes involved in understanding short narratives such as ‘The girl was writing a letter when her friend spilled coffee on the paper’. We hypothesize that these sentences are interpreted in two steps: (1) when the progressive clause is processed, a discourse model is computed in which the goal state (a complete letter) is predicted to hold; (2) when the subordinate clause is processed, the initial representation is recomputed to the effect that, in the final discourse structure, the goal state is not satisfied. Critical sentences evoked larger sustained anterior negativities (SANs) compared to controls, starting around 400 ms following the onset of the sentence-final word, and lasting for about 400 ms. The amplitude of the SAN was correlated with the frequency with which participants, in an offline probe-selection task, responded that the goal state was not attained. Our results raise the possibility that the brain supports some form of non-monotonic recomputation to integrate information which invalidates previously held assumptions.

© 2008 Elsevier Inc. All rights reserved.

*Keywords:* Semantics; Discourse; Non-monotonicity; ERPs; Sustained anterior negativity

---

### Introduction

In the past three decades experimental research using event-related potentials (ERPs) has provided numerous insights into word, sentence and discourse comprehension. However, as has been noted, “a cognitive neuroscience approach to language has not as yet merged with linguistic and psycholinguistic research programmes”

(Brown & Hagoort, 1999). One linguistic research program that may contribute to understanding the basis of meaning in the human brain is semantic theory. Logicians and formal semanticists since the ‘dynamic turn’ (Peregrin, 2003) have shifted their attention from describing semantic competence to modeling cognitive update and information exchange. A case in point is a recent proposal by van Lambalgen and Hamm (2004) which regards comprehension as an incremental, yet non-monotonic process whereby temporary structures are set up in working memory and may be later revised on the basis of further discourse information. Although

---

\* Corresponding author. Fax: +31 024 3610989.

E-mail address: [g.baggio@fcdonders.ru.nl](mailto:g.baggio@fcdonders.ru.nl) (G. Baggio).

evidence for semantic reanalysis exists (Carreiras, Garnham, Oakhill, & Cain, 1996; Sturt, 2007), the issue has arguably received less attention than it deserves. The purpose of the present study is to contribute filling this gap. We used ERPs to test a processing hypothesis proposed by Baggio and van Lambalgen (2007) as an application of the formal, non-monotonic semantics of tense and aspect of van Lambalgen and Hamm (2004).

#### ERP research on semantic processing

Event-related potentials have proved useful to address a number of issues concerning the relative complexity and time course of semantic processes. Kutas and Hillyard (1980) conducted the first ERP experiment in which linguistic factors were successfully manipulated, in this case the semantic plausibility of a word given the preceding sentence context:

- (1) a. The officer shot the man with a gun  
b. The officer shot the man with a moon

Compared to ‘gun’, the anomalous noun ‘moon’ resulted in a larger negative shift starting around 250 ms after word onset, peaking at 400 ms, and lasting for approximately another 150 ms. This ERP component, called N400 after its polarity and peak latency, is known not to be affected by other unexpected events, such as variations in the physical properties of the stimuli. Larger N400s are also triggered by semantically plausible words which are nevertheless judged as less preferred in a given sentence context (Hagoort & Brown, 1994; Kutas & Hillyard, 1984), for example ‘pocket’ in (2b):

- (2) a. Jenny put the sweet in her mouth after the lesson  
b. Jenny put the sweet in her pocket after the lesson

The amplitude of the N400 is also modulated by lexical items which provide information conflicting with the discourse context (van Berkum, Hagoort, & Brown, 1999; van Berkum, Zwitterlood, Hagoort, & Brown, 2003) or world knowledge (Hagoort, Hald, Bastiaansen, & Petersson, 2004). In sum, although every content word evokes an N400, the amplitude of the negative shift appears to be affected by the degree of semantic fit of a lexical item with the preceding context and the knowledge base relevant for its integration.

Semantics-related negativities different from the N400 have also been found. Van Berkum and colleagues recorded ERPs while subjects read (van Berkum, Brown, & Hagoort, 1999) and listened to (van Berkum, Brown, Hagoort, & Zwitterlood, 2003) discourses in which a particular NP in a target sentence could denote either

a single referent introduced in the preceding discourse or two equally suitable referents. For instance, (3c), containing the NP ‘the girl’, could follow either the single-referent context (3a) or the double-referent context (3b):

- (3) a. David had told the boy and the girl to clean up their room before lunch time. But the boy had stayed in bed all morning, and the girl had been on the phone all the time  
b. David had told the two girls to clean up their room before lunch time. But one of the girls had stayed in bed all morning, and the other had been on the phone all the time  
c. David told the girl that...

Referentially ambiguous NPs, such as ‘the girl’ in (3c) following (3b), elicited a sustained anterior negativity (SAN), emerging 300–400 ms after noun onset and lasting for several hundreds of milliseconds. The SAN differed from typical instances of the N400 in duration (‘sustained’) and scalp distribution (‘anterior’). Following earlier claims (Mecklinger, Schriefers, Steinhauer, & Friederici, 1995; Müller, King, & Kutas, 1997), the time-course and topographical profile of the observed ERPs are taken to suggest that “at least some of the processing consequences of referential ambiguity may involve an increased demand on memory resources” (van Berkum et al., 2003).

Another study (Münte, Schiltz, & Kutas, 1998) reported sustained anterior negativities. ERPs were recorded while subjects read narratives differing in the initial temporal connective:

- (4) a. After the scientist submitted the paper, the journal changed its policy  
b. Before the scientist submitted the paper, the journal changed its policy

‘Before’ sentences elicited a larger sustained negativity, maximal over left anterior sites. At the left frontal electrode, ERP responses to ‘before’ and ‘after’ diverged approximately 300 ms after sentence onset. The effect lasted throughout the sentence and was larger during the second clause. The difference of anterior negativity between ‘before’ and ‘after’ items was positively correlated with participants’ working memory span. Münte et al. argue that the slow negative shift evoked by ‘before’ sentences reflects working memory operations involved in computing a model for (4b) in which the events are represented in their actual order of occurrence. That is, in contrast with (4a), (4b) requires *additional* memory resources as the two events are mentioned in reverse temporal order.

The connection between sustained anterior negativities and working memory is explicit in van Berkum et al. (1999), van Berkum et al. (2003) and Münte

et al. (1998). However, there is no full agreement on a functional account, based on linguistically-informed notions, of these findings. For instance, while van Berkum et al. suggest that the sustained anterior negativity reflects ‘referential processing’, Münte et al. seem to implicate that ‘additional discourse-level computations’ of the temporal and causal profiles of the events described by ‘before’ and ‘after’ sentences are responsible for the observed slow negative shifts. Matters appear to be further complicated by the finding that sustained anterior negativities are elicited by constructions in which complexity at the syntax-semantics interface is affected, as in long-distance *wh*-dependencies (Felsler, Clahsen, & Münte, 2003; Fiebach, Schlesewsky, & Friederici, 2002; King & Kutas, 1995; Müller et al., 1997; Phillips, Kazanina, & Abada, 2005).

Despite the differences between the conditions in which sustained anterior negativities have been observed, the proposed processes can be brought under a single umbrella term, which we shall refer to as ‘computing a discourse model’. Formal semantics, at least since Discourse Representation Theory (DRT) (Kamp, 1981), has assumed that interpreting definite and indefinite NPs, resolving anaphoric pronouns, determining the order of events, establishing *wh*-dependencies and other cross-clause and cross-sentence processes concur in the construction of a discourse model, that is, a cognitive representation making a given narrative true. More recent proposals, which build upon DRT and add some sophistication to it, regard discourse comprehension as a process in which lexical meanings, context and world knowledge interact to produce consistent discourse representations (Hamm, Kamp, & van Lambalgen, 2006; van Lambalgen & Hamm, 2004). Pragmatic constraints and causal/world knowledge are brought to the fore by these accounts. Furthermore, discourse models as envisaged by the theory (called ‘minimal models’, see “Minimal models, inference in the limit and recomputation” section) can be efficiently computed by artificial neural networks, which account for some capabilities and limitations of working memory (Stening & van Lambalgen, 2005). Therefore, in this framework it becomes possible to raise and address a number of issues concerning the complexity of computing discourse models in working memory. To see in some detail how this could be done, we must first introduce the linguistic phenomenon with which we shall be concerned: the ‘imperfective paradox’.

### *The imperfective paradox*

Verb phrases (VPs) can be semantically classified as states (‘know’, ‘love’, etc.), activities (‘run’, ‘write’, etc.), accomplishments (‘write a letter’, ‘bake a cake’, etc.), achievements (‘finish’, ‘reach’, etc.) and points (‘flash’, ‘hop’, etc.) (Steedman, 1997). Accomplishments

involve the activity from which they are derived. For instance, ‘write a letter’ is constituted by the activity ‘write’ and the direct object ‘a letter’, which need not refer to an existing entity, but carries information about the goal toward which the writing activity is directed. Here we use the term ‘activity’ to denote both the aspectual class of VPs such as ‘write’ in the above classification and the atelic process involved in all accomplishments. We use ‘progressive’ and ‘imperfective’ interchangeably to allow the reader to see the connection between the semantics of the progressive and the imperfective paradox, although this is not entirely correct (Comrie, 1976).

Let us consider accomplishments first:

- (5) The girl was writing a letter when her friend spilled coffee on the tablecloth.

From (5) the reader would typically conclude that, barring unforeseen circumstances, the girl will attain the desired goal and would thus assent to the statement ‘The girl has written a letter’ (see Sections “Entailment questionnaire” and “Behavioral data” for evidence supporting this claim). Such an inference is based on the assumption that spilling coffee on the tablecloth is usually *neutral* with respect to the writing activity, that is, it is not a typical immediate cause leading to the termination of the activity. It is possible to imagine situations in which writing was temporarily interrupted or even terminated by the accident. However, as the data reported in Sections “Entailment questionnaire” and “Behavioral data” will demonstrate, failing to explicitly mention an obstacle in the discourse is sufficient to lead the reader to assume that there was no such obstacle to attaining the intended goal.

We hypothesize that the inference to a goal state is *defeasible* or *non-monotonic*, that is, it can be suppressed if the discourse describes an event which terminates the relevant activity:

- (6) The girl was writing a letter when her friend spilled coffee on the paper.

Assuming that writing was intended to occur on the same paper sheets on which coffee was spilled, the accident is sufficient to terminate the activity and it is therefore a *disabling* condition for obtaining a complete letter. Accordingly, on the basis of (6) the reader would assent to ‘The girl has written no letter’.

Suppression can obtain only with accomplishments, not with activities (Rothstein, 2004). In accomplishments, the object NP ‘a letter’ expresses the existence of a natural culmination point or ‘canonical goal’ toward which the writing activity is directed, namely a complete letter. Activities, for instance ‘writing letters’, do not involve any such canonical goal. The use of the

bare plural ‘letters’ indicates that the number of letters is (for the speaker and the hearer) unspecified and that, therefore, the activity has no natural culmination point. Accordingly, a narrative containing the activity VP ‘writing letters’ will be interpreted as entailing that ‘The girl has written one or more letters’ regardless of the consequences of the second event on the writing activity:

- (7) The girl was writing letters when her friend spilled coffee on the tablecloth.
- (8) he girl was writing letters when her friend spilled coffee on the paper.

There appears to be something paradoxical about (6) in its relation to (5), which is not found in the pair (7) and (8). Whereas it belongs to the meaning of the accomplishment ‘writing a letter’ that the writing activity is directed toward the goal state of a complete letter, the actual occurrence of that consequent state can be denied without contradiction. How can an essential component of the meaning be denied without destroying meaning itself? This is the so-called ‘imperfective paradox’. We now turn to its processing consequences.

#### *Minimal models, inference in the limit and recomputation*

Language processing amounts to incrementally computing a discourse representation given lexical, syntactic and contextual constraints (Hagoort, 2006). To make computation tractable, discourse models must be ‘minimal’, that is, in a precise mathematical sense (van Lambalgen & Hamm, 2004), the simplest possible structures making the narrative true. Minimal models behave like ‘closed worlds’, in which only those propositions which are asserted in discourse, or which can be inferred from it or from background knowledge, are represented as true in the model. For the remaining cases, a distinction must be drawn. Propositions which are mentioned in discourse, but are not asserted and do not follow from what is said or from background knowledge (e.g. the antecedent of a conditional), are represented as false in the minimal model. In logical terms, these propositions still belong to the finite language upon which the construction of the minimal model is based. But as long as nothing forces their truth, they will be taken as false. Propositions which are not part of the finite language—because they do not occur in the discourse context or in background knowledge—are not included in the minimal model, that is, they are not represented as being either true or false.

One important upshot of the theory is that the occurrence of a goal state can be inferred from a minimal model of a discourse containing an accomplishment in the past progressive. As soon as the sentence ‘The girl was writing a letter’ is processed,

the system constructs a minimal model in which the goal state (a complete letter) is attained at some time later than the interval referred to by the progressive. Two remarks concerning this crucial point are in order. First, interpretation is based on the ‘closed world assumption’: if no disabling condition is described in discourse (so far), it will be (temporarily) assumed that there is no obstacle interfering with the writing activity. Second, the conclusion that eventually a letter is accomplished is an instance of *predictive inference* or, more precisely, *inference in the limit*: given that writing is asserted to hold some time in the past, that it can be assumed there are no obstacles for the writing activity, that some form of inertia holds (writing continues if it is not hindered by external forces) and that a letter is a finite object, it can be expected that the process will converge—‘in the limit’—to a complete letter. This holds for both neutral (5) and disabled (6) accomplishments. Now, when the initial model is extended with a ‘when’ clause describing an event which terminates the writing activity (i.e. a disabling condition), the goal state inference will be suppressed. The subordinate clause ‘when her friend spilled coffee on the paper’ will lead to the retrieval of causal knowledge from semantic memory to the effect that the coffee accident terminated the writing activity. Spilling occurred *during* the writing process, from which follows that the accident took place *before* a complete letter was obtained. The writing event can be imagined as an open interval, where the goal state (a complete letter) is no longer part of the structure. We shall use the term ‘recomputation’ to refer to the suppression of the goal state inference when the subordinate clause in (6) is processed. Because (5) describes a neutral scenario, the goal state derived while processing the progressivized VP is maintained in the final model. In conclusion, whereas (5) involves an *extension* of the initial discourse model, (6) might induce a *recomputation*. Since (7) and (8) do not involve a canonical goal, they will require an extension only.

#### *Predictions for ERPs*

The only difference between neutral and disabled *activities* (e.g. ‘writing letters’) is the noun in the subordinate clause, ‘tablecloth’ or ‘paper’. In both cases the initial model is simply extended, thus we expect to observe only local ERP differences related to the integration of the differing nouns. As ‘tablecloth’ is less semantically related to the other lexical items occurring in the sentence compared to ‘paper’, we expect a larger N400 for the former compared to the latter word.

Processing a ‘when’ clause following an *accomplishment* (e.g. ‘writing a letter’) involves integrating the differing nouns *and*, in the disabling case, recomputing

the initial discourse representation. Also in this case, the neutral noun ‘tablecloth’ is predicted to evoke a larger N400 compared to the disabling ‘paper’, reflecting a lower degree of semantic relatedness with the preceding context. In our ERP study Dutch materials were used, where the verb in subordinate clauses occupies the sentence-final position (see “Materials” section). The temporal and causal information provided by verbs in ‘when’ clauses is necessary to initiate the recomputation process (Baggio & van Lambalgen, 2007). Thus, the ERP effects of what we have analyzed as recomputation are expected to surface at the sentence-final verb ‘spilled’ (‘morste’ in our Dutch stimuli, see “Materials” section and Table 1).

One additional prediction is that the amplitude of the ERP effect evoked by disabled accomplishments is correlated with the relative frequency with which readers infer that the goal state was not attained. Recomputation is expected to consistently evoke a time-locked shift in the EEG in each trial in which a negative judgment concerning the attainment of the goal is made. Therefore, the higher the frequency

of such inferences—the larger the number of trials in which recomputation occurred—the larger the amplitude of the ERP component. The method and results of an ERP study in which these predictions were tested are described below.

## Methods

### Materials

The set of Dutch materials used in the experiment included 160 test and 160 filler items (see Supplementary Material). Each test item included two context sentences providing a neutral setting for the events, four target sentences (A)–(D) and two probe pairs (E) and (F) (Table 1). Target sentences were constructed by manipulating the aspectual class of the progressive VP (activity or accomplishment) and the effects of the event introduced by the ‘when’ clause (neutral or disabling) on the event described in the main progressive clause. All progressive VPs were instances of the Dutch periphrastic

Table 1  
Examples of stimulus sentences

<i>Context sentences</i>	
	De deur van de woonkamer was gesloten. Binnen speelde de radio klassieke muziek. <i>The door of the living-room was closed. Inside played the radio classical music.</i> <sup>a</sup> ‘The door of the living room was closed. Inside the radio played classical music’. <sup>b</sup>
<i>Target sentences</i>	
(A)	Het meisje was brieven aan het schrijven toen haar vriendin koffie op het tafelkleed morste. <i>The girl was letters on the to-write when her friend coffee on the tablecloth spilled.</i> <sup>a</sup> ‘The girl was writing letters when her friend spilled coffee on the tablecloth’. <sup>b</sup>
(B)	Het meisje was brieven aan het schrijven toen haar vriendin koffie op het papier morste. <i>The girl was letters on the to-write when her friend coffee on the paper spilled.</i> <sup>a</sup> ‘The girl was writing letters when her friend spilled coffee on the paper’. <sup>b</sup>
(C)	Het meisje was een brief aan het schrijven toen haar vriendin koffie op het tafelkleed morste. <i>The girl was a letter on the to-write when her friend coffee on the tablecloth spilled.</i> <sup>a</sup> ‘The girl was writing a letter when her friend spilled coffee on the tablecloth’. <sup>b</sup>
(D)	Het meisje was een brief aan het schrijven toen haar vriendin koffie op het papier morste. <i>The girl was a letter on the to-write when her friend coffee on the paper spilled.</i> <sup>a</sup> ‘The girl was writing a letter when her friend spilled coffee on the paper’. <sup>b</sup>
<i>Probe sentences</i>	
(E)	Het meisje heeft een of meer brieven geschreven. <i>The girl has one or more letters written.</i> <sup>a</sup> ‘The girl has written one or more letters’. <sup>b</sup> Het meisje heeft geen brief geschreven. <i>The girl has no letter written.</i> <sup>a</sup> ‘The girl has written no letter’. <sup>b</sup>
(F)	Het meisje heeft een brief geschreven. <i>The girl has a letter written.</i> <sup>a</sup> ‘The girl has written a letter’. <sup>b</sup> Het meisje heeft geen brief geschreven. <i>The girl has no letter written.</i> <sup>a</sup> ‘The girl has written no letter’. <sup>b</sup>

<sup>a</sup> Literal translation.

<sup>b</sup> Paraphrase.

‘was/waren NP aan het  $V_{\text{inf}}$ ’ construction. This solution is to be preferred to the use of the Dutch simple past which, in some cases, is aspectually ambiguous between perfective and imperfective readings. Accomplishments differed from activities in the object NP only: an indefinite (‘een brief’/‘a letter’) was used for accomplishments, a bare plural (‘brieven’/‘letters’) for activities. Disabling and neutral subordinate clauses differed only in the prepositional or object NP, for instance ‘papier’ and ‘tafelkleed’. Neutral and disabling events were distinguished based on the experimenters’ judgment (but see “Entailment questionnaire” section for some data supporting these choices). Probe pairs (E) were used with activities and (F) with accomplishments (Table 1).

Fillers were 160 sentences of varying length, structure and content. Analogously to test items, fillers were preceded by two neutral context sentences and followed by a probe pair. Target sentences described an event consistently, as in (9), or inconsistently, as in (10), with factual knowledge (see (Hagoort et al. (2004) for an experiment based on these stimuli):

- (9) Dutch trains are white and very crowded.  
 (10) Dutch trains are yellow and very crowded.

Probes were of the type ‘Trains in the Netherlands are white.’/‘Trains in the Netherlands are yellow.’ These fillers were chosen to add variety to the materials while preserving the task used for test items.

Four test versions were constructed, comprising randomized lists of test and filler items. The task was identical for critical and filler sentences. Participants had to select the correct probe based on the information provided by the context and target items. Mean length, raw and lemma frequency of the differing nouns in the NP of subordinate clauses were matched using the CELEX Dutch corpus (Baayen, Piepenbrock, & Gulkers, 1996). Mean length was 7.9 letters ( $SD = 2.46$ ) for neutral and 7.75 ( $SD = 2.79$ ) for disabled cases, and was kept below 12 letters in any case. Raw frequency per million words was 1113 ( $SD = 2462$ ) for neutral and 1096 ( $SD = 2792$ ) for disabled cases. Lemma frequency per million words was 1730 ( $SD = 3559$ ) for neutral and 1666 ( $SD = 3585$ ) for disabled cases. The length of sentence-final verbs was identical across conditions and was kept below 12 letters in any case. Cloze probabilities of sentence-final verbs were normed in a dedicated pre-test discussed below.

### Pre-tests

#### Cloze probability test

In order to determine the cloze probabilities of sentence-final verbs, context sentences followed by a target sentence with the final word blanked were presented to a group of 32 native speakers of Dutch (mean age 22.5, 27

female). Participants were requested to fill in the blank with the first word that came to their mind. Four versions (40 items per condition), randomized and balanced across conditions, were constructed. Mean cloze probabilities were not different between the conditions (all comparisons using  $t$ -tests,  $P > .05$ ) in each test version as well as in the entire set.

#### Entailment questionnaire

A paper-and-pencil judgment task was also administered. Thirty six Dutch native speakers (mean age 22.5, 28 female) were presented with the context followed by a target sentence and a probe pair. The task was to select the appropriate probe. Negative probes were more frequently chosen for disabled accomplishments than for the other conditions. Neutral activities (S1) (see Table 1) showed the lowest mean of negative responses ( $M = 2.72$ ,  $SD = 3.22$ ), followed by disabled activities (B) ( $M = 8.06$ ,  $SD = 7.05$ ), neutral accomplishments (C) ( $M = 1.03$ ,  $SD = 9.23$ ) and disabled accomplishments (D) ( $M = 25.14$ ,  $SD = 8.02$ ) (see Baggio & van Lambalgen, 2007 for details).

#### Participants

Thirty-one students participated in the ERP experiment. Of these, seven were left out of the final analysis due to a high number (>20%) of trials contaminated by artifacts. This left us with twenty four participants (mean age 22.5, 17 female), with no history of neurological, psychiatric or cognitive disorders. Subjects were selected from the database of the F.C. Donders Centre for Cognitive Neuroimaging at the Radboud University Nijmegen. Participants received €8 per hour or course credits. None of the subjects who took part to the pre-tests participated in the ERP experiment.

#### Procedure

After applying the electrodes (see “Recording” section), participants were conducted into the experimental room and were asked to sit in front of a video monitor. The stimuli were presented as follows: the two context sentences were displayed together on a single screen (white on black background) for a variable duration (6, 7 or 8 s), depending on the length of the sentences themselves; next the target sentence, one of (A)–(D), was presented on the screen word-by-word (600 ms SOA, 300 ms word duration; white on black background); the target sentence was preceded and followed by a fixation cross, presented for 1500 ms; finally, the probe pair, one of (E) and (F), was shown on the screen (red on black background) and remained visible until the participant gave a button-press; the probes were followed by a fixation cross which lasted for 1500 ms. The same presentation parameters were used for fillers. Participants were

instructed to read each sentence carefully, to blink only when the fixation cross was shown and to select the correct probe by pressing one of two buttons (left or right on the button box) as quickly and accurately as possible. The position on the screen (top or bottom) of the positive and negative probe corresponded to the left and right button respectively, and was counterbalanced across test versions. In this way, participants could not prepare their motor response before the probe pair was presented on the screen. The experiment took about 2 h and was divided into 24 blocks of 10 trials each.

### Recording

EEG and EOG signals were recorded using Ag/AgCl electrodes. The EOG was measured from four electrodes: one at the outer canthus of each eye, one below and one above the left eye (FE). The EEG was measured from 28 electrodes, arranged according to American Electrophysiological Society conventions: Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, and O2. Two additional electrodes were placed on the left and right mastoids, the former serving as the reference during the measurement. All EEG and EOG electrodes were referenced offline to a linked mastoid. EEG electrodes were attached to an elastic cap, whereas EOG and reference electrodes were applied using two-sided adhesive decals external to the cap. Electrode impedance was kept below 5 k $\Omega$  throughout the experiment. The EEG/EOG was amplified by a multichannel BrainAmp DC system, with a 500-Hz sampling rate, a low pass filter set at 125 Hz and a 10-s time constant.

### Data analysis

Data analysis was conducted using FieldTrip,<sup>1</sup> a MATLAB package for processing EEG signals. The following transforms were applied to each subject's dataset. Segments corresponding to the noun and the sentence-final verb were extracted from the EEG with an interval of 200 ms preceding and 800 ms following stimulus onset. Baseline correction used the 200-ms interval preceding the onset of nouns, and the 100-ms interval following the onset of sentence-final verbs. The latter choice was effected so as to prevent ERP differences in the 400- to 600-ms interval following the onset of the nouns from biasing the baseline correction for the ERPs evoked by sentence-final verbs in the same time interval. The use of such a baseline seems acceptable on grounds that the expected recomputation effect at the verb would not affect such largely exogenous components as the N1.

Artifact rejection was based on two FieldTrip functions: the first detects and rejects all trials that contain activity exceeding a threshold of  $\pm 100 \mu\text{V}$ ; the second identifies and discards trials contaminated with eye movements or blinks by means of thresholding the  $z$ -transformed value of the raw data in the EOG channels, preprocessed using a band-pass filter of 1–15 Hz. A 30-Hz low-pass digital filter was applied to the segmented, artifact-free data. ERPs were obtained for each subject by averaging over trials in each experimental condition. A 5-Hz low-pass filter was used to produce the waveforms shown in Figs. 2–5. Topographical plots and statistical analyses were however based on the 30-Hz low-pass filtered data.

For the analysis of behavioral responses we employed two repeated-measures ANOVA models with Subject as the random effect, Aspectual Class (Activity/Accomplishment) and Subordinate Clause Type (Neutral/Disabling) as fixed effects, and the mean value of either negative judgments (negative probes selected in the response task) or decision times in each condition as the dependent variables.

Statistical analyses of ERP data used a non-parametric randomization procedure (Maris, 2004; Maris & Oostenveld, 2007) which took as input mean amplitude ( $\mu\text{V}$ ) values in each condition in time bins of 100 ms, starting from the onset of the relevant word and ending 800 ms after, and produced as output a cluster of electrodes (min. 1 and max. 28) in which the difference between the conditions was significant in each time bin, the sum of  $t$ -statistics in that cluster and Monte Carlo estimates of  $P$ -values.

For the correlation analysis (see Section Predictions for ERPs), we calculated the difference between the ERPs evoked by sentence-final verbs in subordinate clauses—disabled (D) minus neutral (C)—following accomplishments at anterior sites (Fp1, Fp2, F7-, F3-, Fz-, F4-, F8-averaged) in the 500- to 700-ms interval after the onset of the sentence-final verb (see “Event-related brain potentials” section for motivation). Pearson's product-moment correlation was computed to determine whether the amplitude difference in ERPs varied with the number of negative responses, quantified again as the difference of negative judgments between disabled (D) and neutral (C) accomplishments. The correlation analysis was done on a per-subject basis (i.e. each pair of data points in the correlation corresponded to a single subject's data).

## Results

### Behavioral data

Neutral activities (S1) showed the lowest mean of negative responses ( $M = 4.08$ ,  $SD = 2.87$ ), followed by disabled activities (B) ( $M = 5.83$ ,  $SD = 4.51$ ), neutral

<sup>1</sup> For more information, see <http://www2.ru.nl/fcdonders/fieldtrip/>.

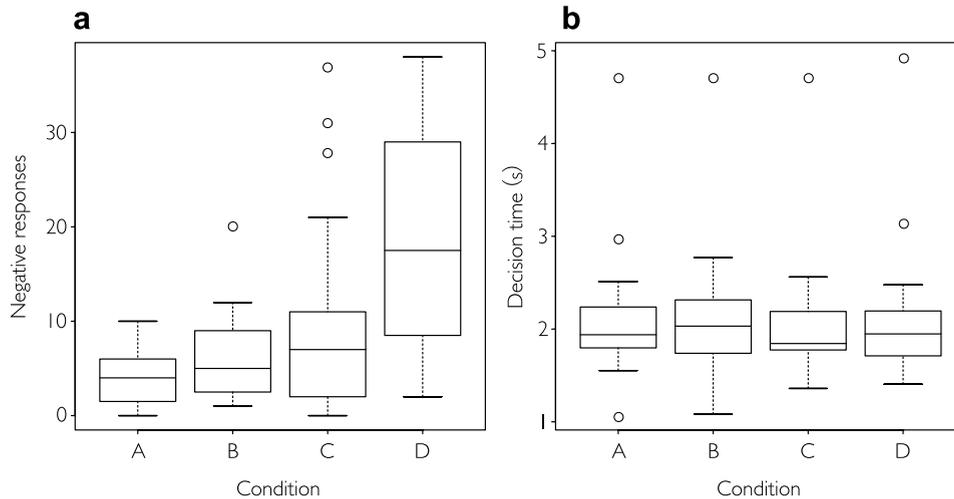


Fig. 1. Behavioral data. (a) Boxplot of the categorical responses. (b) Boxplot of the decision times. Conditions are represented on the abscissa (see Table 1 for the corresponding labels). Negative responses and decision times are plotted on the ordinate. The solid line within the boxes indicates the median, box height is equal to the interquartile range, whiskers represent adjacent values, and empty circles denote outliers. The maximum of potential negative responses is 40.

accomplishments (C) ( $M = 9.58$ ,  $SD = 9.96$ ) and disabled accomplishments (D) ( $M = 18.13$ ,  $SD = 11.16$ ). The distribution of the data in the different conditions appears rather similar, as indicated by box height and whisker length in Fig. 1a. However, disabled accomplishments have a more spread-out distribution, suggesting that inference patterns were less uniform across participants. ANOVAs revealed significant main effects of Aspectual Class and Subordinate Clause Type, and a significant interaction between the two factors (Table 2, Fig. 1a). The observed pattern of responses supports the linguistic views outlined above and replicates our previous findings (see “Entailment questionnaire” section). There is no difference in decision times (Table 2, Fig. 1b): (S1),  $M = 2111$  ms,  $SD = 677$  ms; (B),  $M = 2100$  ms,  $SD = 680$  ms; (C),  $M = 2070$  ms,  $SD = 646$  ms; (D),  $M = 2086$  ms,  $SD = 710$  ms.

#### Event-related brain potentials

##### Nouns

Fig. 2 displays the ERP topographies and waveforms elicited by *activities*. An N1 component peaking at

Table 2  
Summary of ANOVA statistics for behavioral data

	Categorical responses	Decision times
Aspectual class	$F(1, 23) = 21.65$ $P < .001$	$F(1, 23) < 1$
Subordinate clause type	$F(1, 23) = 23.60$ $P < .001$	$F(1, 23) < 1$
Aspectual class $\times$ Subordinate clause type	$F(1, 23) = 17.20$ $P < .001$	$F(1, 23) < 1$

approximately 100 ms is followed by a P2 component with a trough at about 200 ms. The amplitude of the N1 and F is not different between disabled and neutral clauses: no significant clusters were found between 0 and 300 ms (all contrasts,  $P > 0.1$ ). The N1–P2 complex is followed by an N400. The amplitude of the N400 is larger in neutral (‘tafelkleed’) than in disabling (‘papier’) clauses (Fig. 2b): significant clusters with a central distribution were found between 300 and 500 ms (Table 3, Fig. 2a). No difference between neutral and disabling clauses was found after 500 ms.

Fig. 3 displays the ERP topographies and waveforms elicited by *accomplishments*. Also in this case, an N1–P2 complex can be observed. There is no difference between neutral and disabling clauses, as no significant clusters between 0 and 300 ms were found (all contrasts,  $P > 0.1$ ). The amplitude of the N400 is again larger in neutral (‘tafelkleed’) than in disabling (‘papier’) clauses (Fig. 3b). The effect lasts longer than the N400 observed in activities: significant clusters with a central distribution were found between 300 and 600 ms (Table 3 and Fig. 3a). No difference between conditions was found after 500 ms.

There is no overall difference between the two aspectual classes. Cluster-based *t*-tests comparing the N400 effects in activities and accomplishments (corresponding to testing the main effect of Aspectual Class in a parametric model) produced no significant clusters between 300 and 600 ms from noun onset (all contrasts,  $P > 0.1$ ). No difference was found in any of the remaining time bins.

##### Sentence-final verbs

Fig. 4 displays the ERP topographies and waveforms elicited by *activities*. Contrary to what we had observed

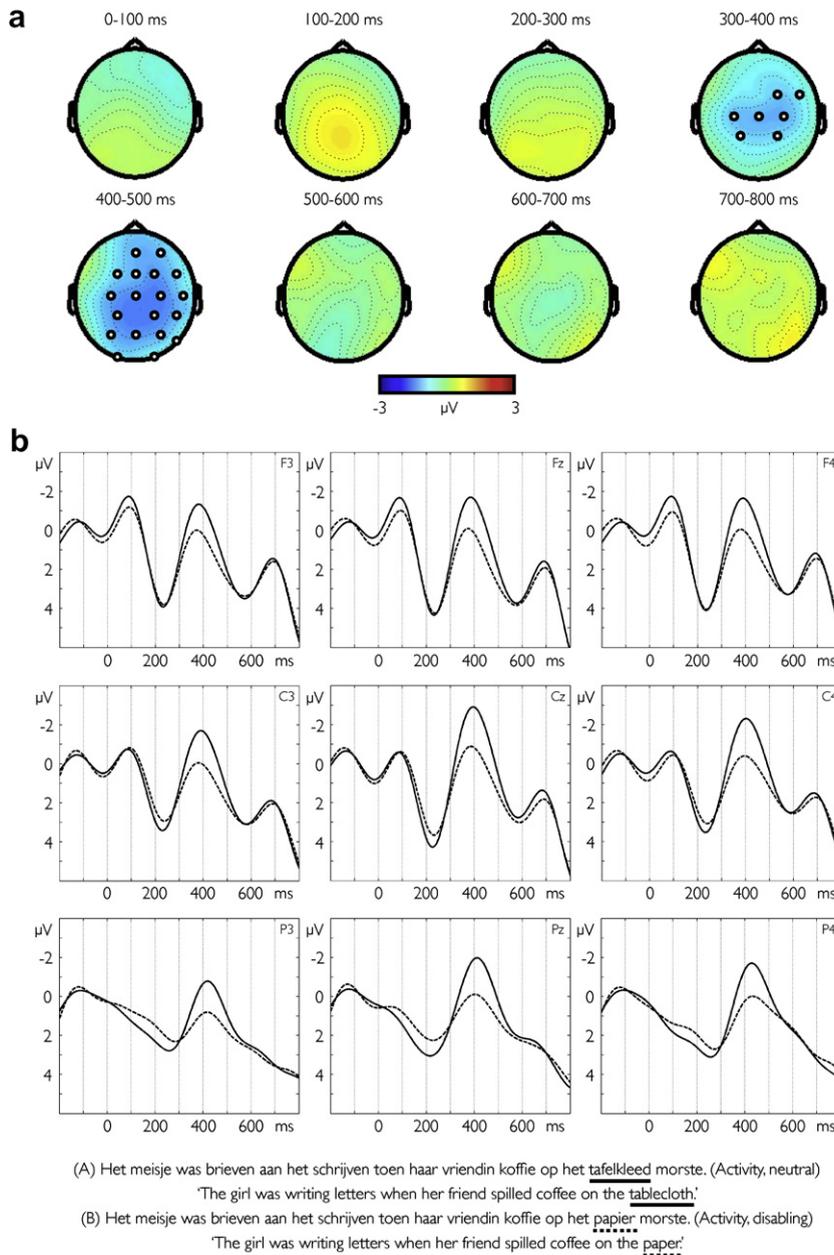


Fig. 2. Activities, noun. (a) Grand-average ( $N = 24$ ) topographies displaying the mean amplitude difference between the ERPs evoked by the noun in neutral compared to disabled activities. Circles represent electrodes in a significant cluster. (b) Grand-average ( $N = 24$ ) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the noun in neutral and disabled activities. Negative values are plotted upward.

at the noun, there is no difference between the N400 elicited by neutral and disabling clauses. Moreover, there is no difference between conditions in any of the remaining time bins (Table 3).

Fig. 5 displays the ERP topographies and waveforms elicited by *accomplishments*. No difference between disabling and neutral clauses was observed in either the

N1–P2 complex or the N400: no significant clusters between 0 and 400 ms were found (all contrasts,  $P > 0.1$ ). While disabled and neutral *activities* do not result in any robust differential effect in later time bins (400–800 ms; Table 3 and Fig. 4), disabling verbs following *accomplishments* evoked larger negative shifts compared to neutral verbs (Table 3, Fig. 5). The effect

Table 3  
Summary of cluster-based statistics for the ERP data

Time	Noun		Sentence-final verb	
	Activities	Accomplishments	Activities	Accomplishments
300–400 ms	$t(22) = -16.54$ $P = .026$	$t(22) = -60.11$ $P < .001$		
400–500 ms	$t(22) = -57.02$ $P < .001$	$t(22) = -78.69$ $P = .002$		$t(22) = -11.85$ $P = .034$
500–600 ms		$t(22) = -18.58$ $P = .022$		$t(22) = -71.09$ $P < .001$
600–700 ms				$t(22) = -39.16$ $P = .008$
700–800 ms				$t(22) = -16.92$ $P = .028$

Disabling and neutral clauses are compared at the noun and at the sentence-final verb, for activities and accomplishments, in bins of 100 ms starting from word onset. The first significant effects occurred at 300–400 ms. Empty cells denote the absence of significant clusters.

emerges at about 400 ms following the onset of sentence-final verb, lasts for approximately 400 ms, and is larger over the more anterior scalp sites, in particular of the left hemisphere. Based on its temporal profile and scalp distribution, we take this effect to be an instance of sustained anterior negativity (SAN). The magnitude of the SAN effect is correlated with the frequency of negative judgments in the response task ( $r = -0.415$ ,  $t(22) = -2.140$ ,  $P = .043$ ; Fig. 6): the higher the number of negative responses, the larger the amplitude of the SAN.

No difference between the two aspectual classes was found. Cluster-based  $t$ -statistics comparing mean ERP amplitudes in activities and accomplishments, again corresponding to testing the main effect of Aspectual Class in a parametric model, produced no significant clusters between 0 and 800 ms from noun onset (all contrasts,  $P > 0.1$ ).

## Discussion

The ERP results reported above can be summarized as follows: the N400 elicited by nouns is larger in neutral than in disabling clauses, following both activities and accomplishments. This can be explained by the lower degree of semantic association with the preceding words ('writing', 'letter' or 'letters') of the noun in neutral clauses ('tablecloth') compared to the noun in disabling clauses ('paper'). On the basis of our processing model, we predicted that disabled accomplishments would induce a different ERP response at the sentence-final verb compared to neutral accomplishments. This corresponds to the difference between the recomputation and the extension of the initial discourse model (see "Minimal models" section). The effect was expected to be (i) absent in activities and (ii) correlated with the frequency with which participants inferred that the goal state was not attained. These predictions were borne out. Disabled

activities did not modulate ERPs at the verb. Disabled accomplishments evoked sustained anterior negativities (SANs). Moreover, a correlation of the SAN amplitude with the frequency of negative judgments was observed. Taken together, our results would seem to offer some support for the recomputation hypothesis. Below we address a few alternative explanations of the data and some related outstanding issues.

### Alternative explanations and outstanding issues

#### Local integration

An alternative account of the data would relate the observed effect to difficulty in integrating the sentence-final verb into the 'local', clause-level context, rather than to suppressing a 'global', discourse-level inference. If this were correct, a modulation of the N400 should be expected, possibly correlated with differences in cloze probabilities. However, as reported above, cloze probabilities do not differ between conditions (see "Cloze probability test" section). Also, there was no difference in the N400s elicited by sentence-final verbs (see "Event-related brain potentials" section), which were also lexically identical across conditions. Following earlier work (Osterhout, 1997), we see the sustained anterior negativity as reflecting difficulty in constructing a discourse-level representation of disabled accomplishments. Supported by further experimental evidence, the recomputation hypothesis could provide a more explicit characterization of at least one instance of sentence-final 'wrap-up effects', in terms of restructuring the initial model.

#### Response frequency

Another alternative account would be based on the observation that sentences requiring a negative response (disabled accomplishments) are relatively less frequent than sentences requiring a positive one (activities and

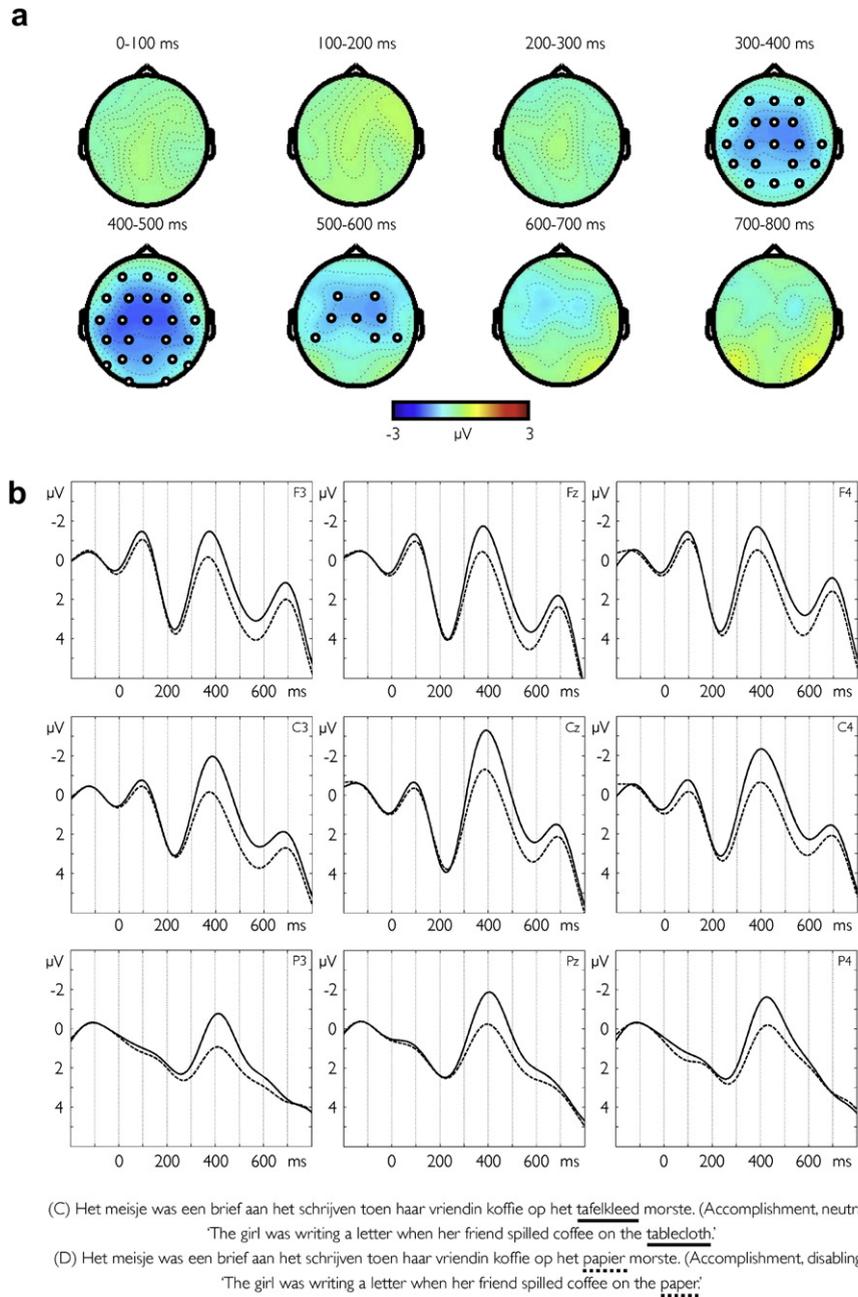


Fig. 3. Accomplishments, noun. (a) Grand-average ( $N = 24$ ) topographies displaying the mean amplitude difference between the ERPs evoked by the noun in neutral compared to disabled accomplishments. Circles represent electrodes in a significant cluster. (b) Grand-average ( $N = 24$ ) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the noun in neutral and disabled accomplishments. Negative values are plotted upward.

neutral accomplishments), the projected ratios being respectively 1/4 and 3/4 (see “Behavioral data” section for the actual behavioral data). On this view, a modulation of the P3 component (Donchin, 1981; Ruchkin, Johnson, Cacoune, Ritter, & Hammer, 1990) might be expected, inversely correlated with the frequency of nega-

tive judgments to disabled accomplishments: the less frequent the negative responses, the larger the amplitude of the P3. However, in our experiment no P3 response was observed and, moreover, the correlation was rather the inverse: the more frequent the negative responses, the larger the amplitude of the sustained anterior negativity.

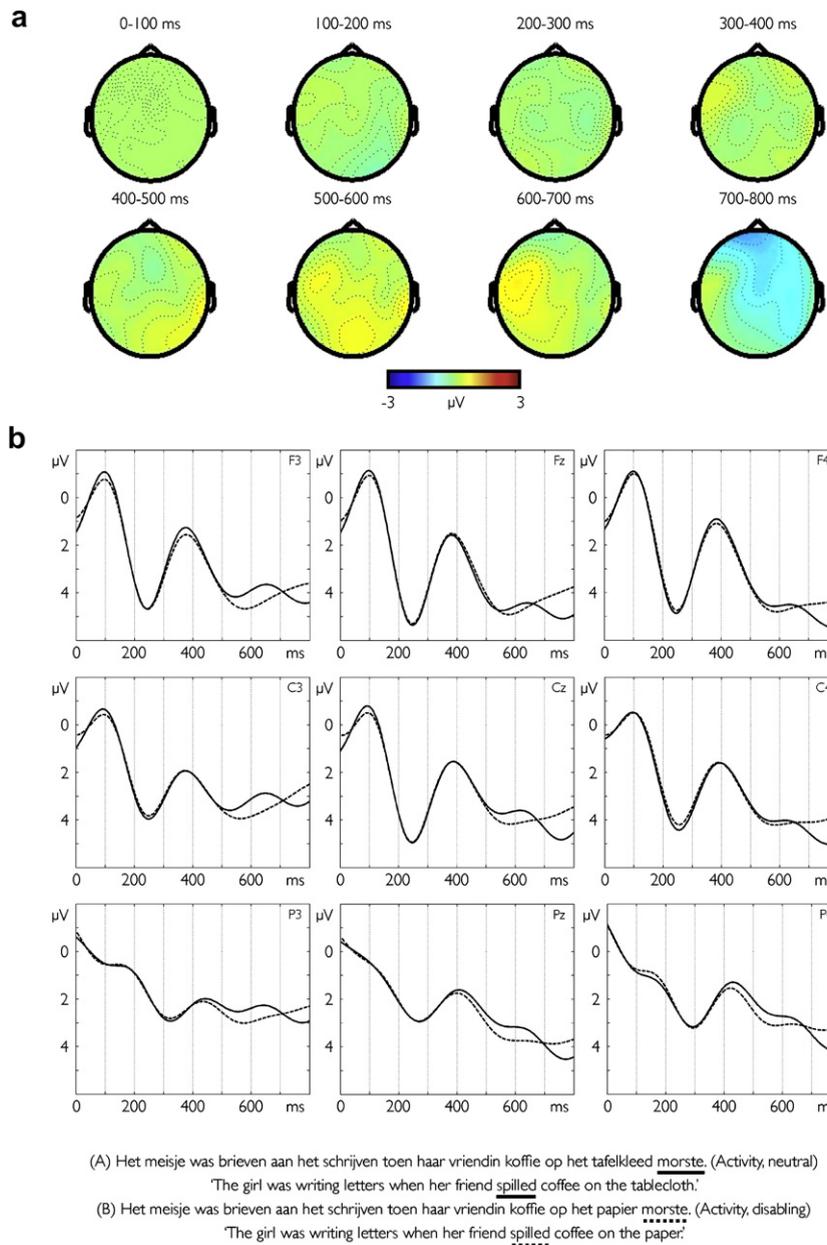


Fig. 4. Activities, sentence-final verb. (a) Grand-average ( $N = 24$ ) topographies displaying the mean amplitude difference between the ERPs evoked by the sentence-final verb in disabled compared to neutral activities. (b) Grand-average ( $N = 24$ ) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the verb in neutral and disabled activities. Negative values are plotted upward.

#### *Monotonicity and possible worlds semantics*

An important issue is whether the observed sustained anterior negativity can be explained by a monotonic account of the progressive. Similarly, it may be asked whether the data reported here constitute compelling evidence for non-monotonicity and against monotonicity. One such alternative explanation can be formulated in possible worlds semantics (Dowty, 1979; Kripke, 1963).

In possible worlds semantics, the progressive denotes a stage of a process which, if it does not continue in the actual world, has chances of continuing in some other possible world (de Swart, 1998). The latter may be called ‘inertia worlds’, courses of events in which the process is not disturbed by external forces and is therefore brought to a successful end. In his analysis of the progressive, Dowty (1979) claimed that the following are equivalent:

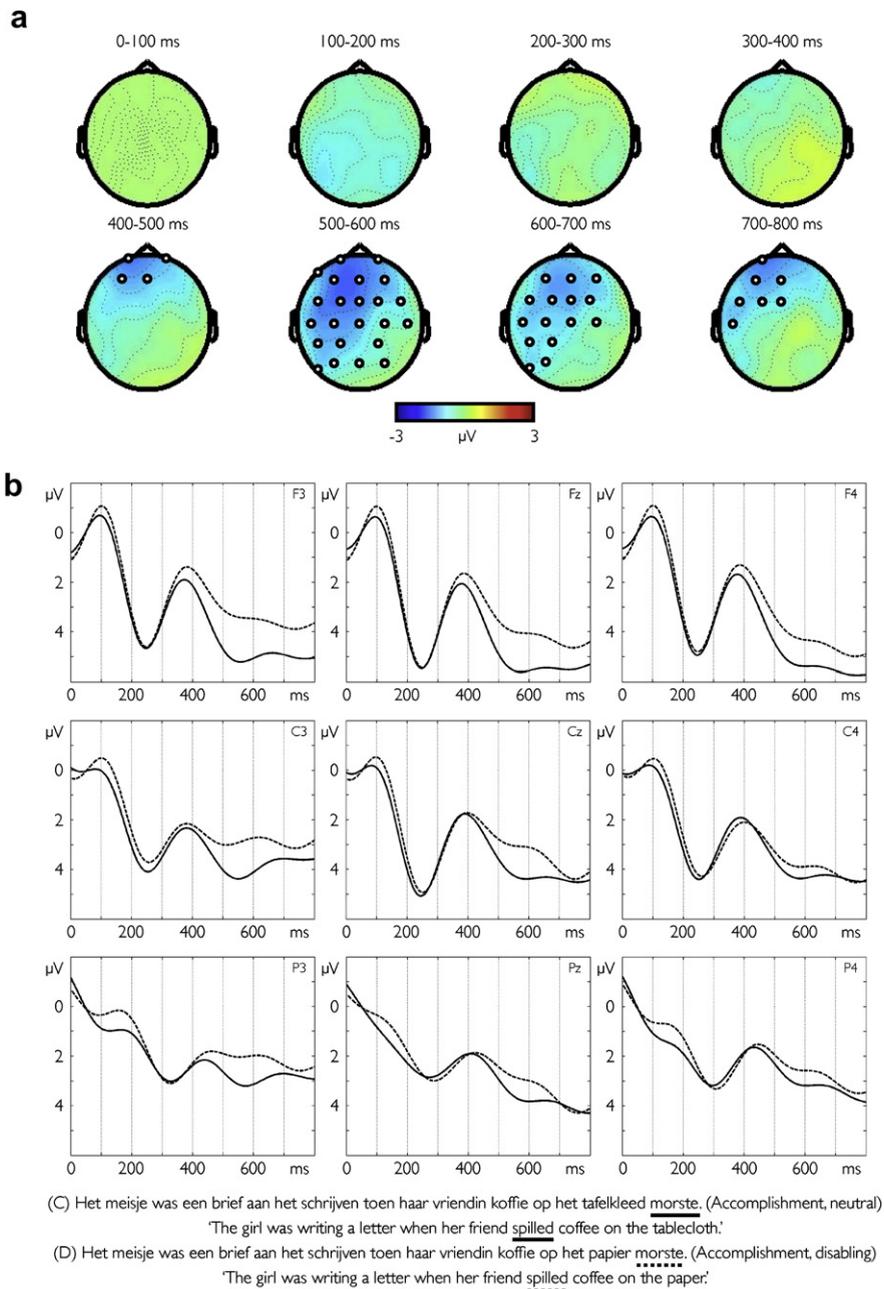


Fig. 5. Accomplishments, sentence-final verb. (a) Grand-average ( $N = 24$ ) topographies displaying the mean amplitude difference between the ERPs evoked by the sentence-final verb in disabled compared to neutral accomplishments. Circles represent electrodes in a significant cluster. (b) Grand-average ( $N = 24$ ) ERP waveforms from frontal, central and parietal electrode sites time locked to the onset (0 ms) of the verb in neutral and disabled accomplishments. Negative values are plotted upward.

1. 'The girl is writing a letter' is true in the actual world;
2. 'The girl will have written a letter' is true in all so-called 'inertia worlds', worlds which are identical with the present world until 'now', but then continue in a way most compatible with the history of the world until 'now'.

These insights can be rendered into processing terms. Processing neutral accomplishments involves moving from the actual world, a snapshot of which is provided by the progressive clause, to some inertia world, in which the goal state is eventually attained (the behavioral data reported in "Behavioral data" section show

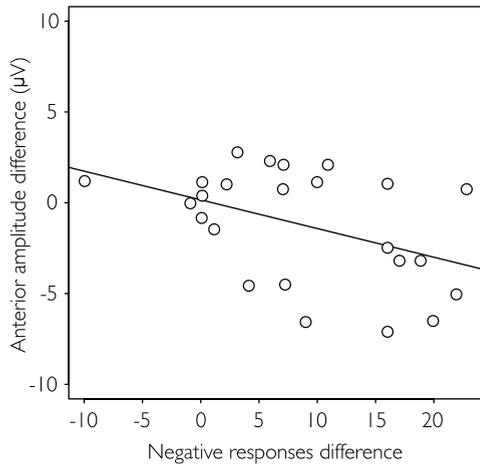


Fig. 6. Scatter plot displaying the correlation between the amplitude of the sustained anterior negativity elicited by disabled accomplishments and the frequency of negative responses ( $r = -0.415$ ,  $t(22) = -2.140$ ,  $P = .043$ ). The mean difference of negative responses between disabled and neutral accomplishments is plotted on the abscissa. The mean amplitude difference at fronto-polar and frontal electrodes between disabled and neutral accomplishments in the 500- to 700-ms interval following the onset of the sentence-final verb is plotted on the ordinate.

that such an inference is drawn). By contrast, processing disabled accomplishments amounts to proceeding from the actual world to a relatively unexpected ‘non-inertia world’, in which the process is disrupted by some event, such as spilling coffee on the paper. Accessing a world in which the goal state is not attained may be surprising.<sup>2</sup> The sustained anterior negativity may then be construed as an index of surprise or some other equivalent notion. This account is monotonic, as for both neutral and disabled accomplishments interpretation involves shifting from the actual world to another accessible world. A simple extension of the initial model is performed in both cases.

This analysis is seemingly in conflict with the non-monotonic one. Still, there is no real opposition between the two as regards these data. In both accounts, an initial commitment to the occurrence of the goal state is made. In the non-monotonic approach this takes the form of a defeasible inference leading to a minimal model, whereas in the monotonic theory it is rather a prior, positive expectation concerning the attainment of the consequent state, and thus a lower probability

<sup>2</sup> This particular combination of possible worlds semantics and surprise was proposed by an anonymous reviewer of this paper, hence our choice of discussing it here. The same applies (for different reviewers) to the issues discussed in “Interruption and termination of activities” and “Goal states and underspecification” sections.

assigned to its failure. This commitment is necessary, because accessing a possible world in which the goal state is not satisfied can be surprising only if there is such a prior expectation. Furthermore, in both accounts a subsequent revision of the initial commitment is made. In the non-monotonic analysis this is a recomputation of the initial minimal model, whereas in the monotonic one it is a recomputation of the initial (low) probability associated with the possible world in which the goal state is not attained.

So there is recomputation and non-monotonicity built into this ‘monotonic’ theory too. The possible worlds account is monotonic as far as *models* are concerned (models are always extended and never recomputed), but *expectations* change non-monotonically (probabilities are recomputed). While *prima facie* opposed, the two accounts are in this respect equivalent. Our reasons for preferring a non-monotonic account, one in which *models* are recomputed, derive from *a priori* considerations. First, there exist forms of non-monotonic inference formally strongly related to the non-monotonic reasoning in the progressive which cannot be captured by Bayesian updates (Stenning & van Lambalgen, 2008a). Second, minimal models and non-monotonic inference can be implemented in neural networks (see “Recomputation in working memory models” section). Last, our account of the progressive is embedded into a larger non-monotonic framework (Stenning & van Lambalgen, 2008b; van Lambalgen & Hamm, 2004), covering other phenomena in reasoning and language processing in children, adults, as well as patients with ADHD (Attention Deficit Hyperactivity Disorder) (van Lambalgen et al., 2008) and ASD (Autistic Spectrum Disorder) (van Lambalgen & Smid, 2004).

We must note however that our study was designed to test a particular non-monotonic theory of the progressive, and not to discriminate between monotonic and non-monotonic accounts of the same phenomenon. The latter task would require, for one, a well-specified *entirely* monotonic theory—that is, one which does not involve recomputation of models, probability values or other processing parameters—and, moreover, a set of predictions in which the two proposals would actually differ. This is admittedly hard, apart from being beyond the scope of the research reported above. Hence the need to emphasize the direction along the theory-observation path which is relevant here: although it can be argued that our non-monotonic theory leads to predictions that are consistent with the observed sustained anterior negativity, it is clearly not the case that the data support *only* this particular theory.

#### Interruption and termination of activities

It may be argued that, compared to disabled accomplishments, disabled activities are inherently simpler

because they involve at most an *interruption* of the activity, for example writing letters in (B), which may be continued on some other paper sheets. Accomplishments might leave a more definite ERP trace because they lead to the *termination* of the activity, for example writing a letter in (D), which cannot be continued being there only a single sheet. On this view, the sustained anterior negativity would not be related to model recomputation (as opposed to monotonic extension), but to the termination (as opposed to the interruption) of the activity. Such an explanation follows from the seemingly plausible notion that computing a model in which the effects of a given event are more ‘catastrophic’ should also be more difficult. Here semantic theory comes to our rescue and suggests that such notion is in fact misguided.

One issue that plays a role here is a type/token distinction concerning the noun ‘letter’. In the token interpretation of ‘letter’ as referring to some particular scribbles on a particular piece of paper, there is indeed a difference between interruption and termination. However, on a type interpretation of ‘letter’ as referring to particular content which can be inscribed on any piece of paper, the activity and the accomplishment case seem comparable, in that in both cases the girl has to reach for a new piece of paper. On the type reading, one wouldn’t even expect a difference in behavioral responses. Nonetheless, since a behavioral difference was observed, it seems the token reading is what subjects adopt. On this assumption, it can be shown that, contrary to the alternative proposal, there is *more* computation going on in the interruption case compared to the termination case—if goal states are not taken into account; if they are, the pattern is reversed as implied by the recomputation hypothesis. It seems harder to compute a model in which an activity is first interrupted, then re-initiated, compared to computing a model in which the activity is just terminated (van Lambalgen & Hamm, 2004). The alternative account would predict a larger sustained anterior negativity for activities compared to accomplishments, which does not fit the experimental results. Also in this case, however, we are ready to acknowledge that a different model, in which terminations are shown to be *more* costly than interruptions, and in which goal states are not invoked to account for such processing cost, may explain the observed sustained anterior negativity.

#### *Goal states and underspecification*

The processing model adopted here implies that, as soon as an accomplishment in the past progressive is encountered, the system constructs a semantic representation in which the goal state is satisfied. Processing the clause ‘The girl was writing a letter’ amounts to computing a minimal model in which the writing activity leads to a complete letter, which is therefore part of the resulting discourse structure. As we have hypothesized, such

computation is defeasible, that is, the model can be recomputed if further discourse information implies that the goal state is not satisfiable, as in (6). One may ask whether the claim that the goal is part of a minimal model of the progressive clause is at all tenable. A seemingly more plausible account would assume that an *underspecified model*, in which it is left undecided whether the goal state is attained or not, is computed while the progressive clause is processed, and a decision is made only at the subordinate clause.

The main problem with an underspecification-based account is that, while it is true that the information provided by the progressive clause is insufficient for determining whether the goal was attained (which would motivate the construction of an underspecified model at that stage), it is *not* the case that sufficient information is contributed by the subordinate clause. While disabling clauses provide evidence that the activity was terminated, and thus license the inference that the goal was not attained, no evidence concerning the satisfaction of the goal state is derivable from neutral clauses. This is a consequence of the well-known ‘frame problem’ (McCarthy & Hayes, 1969), which implies that it is impossible to enumerate all the effects and non-effects of an event. For example, that ‘spilling coffee on the tablecloth does not affect the writing activity’ (if that is the case) is not stored in declarative memory, but must be inferred. This is an instance of ‘closed world reasoning’, which was described above (see “Minimal models, inference in the limit and recomputation” section). In a ‘closed world’, it is assumed that no obstacle to attaining the goal state occurred. Therefore, a letter was completed. The behavioral data reported above show that subjects draw this inference or, equivalently, they are more likely to give positive responses to neutral accomplishments. Processing models based on underspecification—or on parallel processing, for that matter—would have to explain why that very same conclusion (‘the girl wrote a letter’) is not drawn when the system is faced with the relevant input (the VP in the progressive), and is instead delayed until the end of the sentence, where critical information is nonetheless still missing. The hypothesis that the goal state inference is drawn when the input is given seems to be more consistent with the available evidence on immediacy and incrementality in discourse processing (Hagoort & van Berkum, 2007).

This line of reasoning speaks also to the issue of the potential influence of the primary response task on online interpretive processes. It can be argued that the system may have carried out a number of inferences online in order to facilitate a response when the probes were presented, but would have processed the same sentences in an underspecified manner if no response task was administered. The brain would therefore compute representations which are merely ‘good enough’ for the task at hand, striking a balance between efficiency and

cost minimization (Douglas & Martin, 2007; Ferreira, Ferraro, & Bailey, 2002; Ferreira & Patson, 2007). We grant that this is a possibility, which cannot be excluded based on either our data or our processing model. It can however be suggested that, although comprehension probes do not occur in actual language use, it is possible to imagine ‘language games’ in which hearers are required to make interpretive commitments and form a belief concerning the potential outcomes of a process described using the progressive. Our experiment may be taken as a laboratory study of such real world situations, but is not intended as a realistic account of all situations in which progressive constructions are uttered and understood. Further work is needed to investigate the influence of the response task on online ERP measures.

#### *Recomputation in working memory networks*

Minimal models can be regarded as the stable states of associated neural networks. It has been shown that recurrent networks can compute or approximate (depending on the expressiveness of the logical formalism) the semantic operators based on which minimal models are constructed (Hitzler, Hölldobler, & Seda, 2004; Stenning & van Lambalgen, 2005, 2008a). In this framework, recomputation can be modeled as the readjustment of connection strengths driven by a simple form of back-propagation called ‘perceptron learning’ (Rosenblatt, 1962). Computing a minimal model of the progressive clause will correspond to the network settling into one such ‘attractor’ or stable state. Further computation on the initial model brings the network from its initial stable state to another stable state, corresponding to the new minimal model. Importantly, there is a large difference in the overall pattern of network activity in disabled compared to neutral accomplishments. If the initial model is monotonically extended, as in the neutral case (4), a number of units will be activated which were previously silent, while the activation state of the remaining units, including those representing the goal state (the complete letter), will remain unaltered. But if the initial minimal model is recomputed upon encountering the subordinate clause in (6), units which were silent will be activated *and* the activation patterns across some units which were previously active will be readjusted. For instance, the units representing the goal state (the complete letter) will no longer be active. In the neural network this is achieved by successive applications of perceptron learning.

Even though in both cases the network processes the subordinate clause by settling into a new attractor state, the transition in the disabling case requires an extensive adjustment of the connection weights of the units representing the goal state. Recomputation thus results in a more costly state transition. It remains an open question

whether biologically plausible networks can also approximate the semantic operators which give rise to minimal models. Firing rate models, for instance, have been used to implement operations in connectionist networks (e.g. multilayer perceptrons) of the kind required by the construction of minimal models (de Kamps & van der Velde, 2002). Interestingly, recurrent excitation in firing-rate models can account for several aspects of persistent activity in prefrontal cortex neurons during working memory tasks (Durstewitz, Seamans, & Sejnowski, 2000). Recurrent networks thus suggest a plausible mechanistic link between recomputation and sustained anterior negativities, and in general between working memory processes and sustained anterior negativities (Felser et al., 2003; Fiebach et al., 2002; King & Kutas, 1995; Müller et al., 1997; Münte et al., 1998; Phillips et al., 2005; van Berkum et al., 1999, 2003).

As we noted in Introduction, a cognitive neuroscience of language needs to bridge the gap between psycholinguistic and formal models of specific aspects of language on the one hand, and the neural architecture underlying neurophysiological measures on the other hand. For a number of reasons (Poehpel & Embick, 2005) this is a daunting task, which we do not claim to have adequately solved. However, tentatively the following can be said. There is no indication or proof that the sustained anterior negativity is a language-specific ERP effect. Most likely, it reflects the recruitment of neurophysiological activity that might be generated in prefrontal cortex, and is triggered by different cognitive operations which build upon working memory capacity. For this purpose, the prefrontal cortex is a plausible candidate from a neurobiological point of view. In the light of our model, the sustained anterior negativity is taken to index the recomputation following the blocking of the goal state in accomplishments, and the recruitment of working memory resources required for this recomputation. In other cases, the demand might be triggered by different cognitive operations, as in the work by Münte et al. (1998). In general, what we seem to obtain with ERPs, is a many-to-one mapping from cognitive models to neuronal implementation. This however in no way invalidates our interpretation, which is based on combined constraints from the cognitive and neuronal levels of analysis.

The research presented in this paper extends the range of phenomena to which ERPs can be applied, by testing a processing hypothesis suggested by a formal semantics of tense and aspect. Our results raise the possibility that the brain supports some form of non-monotonic recomputation to integrate information which invalidates previously held assumptions. It is a task for future research to provide more stringent tests of non-monotonic as opposed to monotonic models of semantic processing and cognitive update more generally.

## Acknowledgments

We thank Oliver Bott, Travis Choma, Bart Geurts, Fritz Hamm, Karl Magnus Petersson, Jos van Berkum, Roel Willems and three anonymous reviewers for many useful comments on earlier versions of the paper. We are grateful to the Netherlands Organization for Scientific Research (NWO) for support under Grant 051.04.040.

## References

- Baayen, R., Piepenbrock, R., & Gulikers, L., (1996). *CELEX2*. Philadelphia: Linguistic data consortium.
- Baggio, G., & van Lambalgen, M. (2007). The processing consequences of the imperfective paradox. *Journal of Semantics*, 24, 307–330.
- Brown, C., & Hagoort, P. (1999). The cognitive neuroscience of language: Challenges and future directions. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 3–12). New York: Oxford University Press.
- Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology*, 49A, 639–663.
- Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.
- de Kamps, M., & van der Velde, R. (2002). Implementation of multilayer perceptron networks by populations of spiking neurons using rate coding. *Neurocomputing*, 353–358.
- de Swart, H. (1998). Aspect shift and coercion. *Natural Language and Linguistic Theory*, 16, 347–385.
- Donchin, E. (1981). Surprise!...surprise? *Psychophysiology*, 18, 493–513.
- Douglas, R., & Martin, K. (2007). Mapping the matrix: The ways of the neocortex. *Neuron*, 56, 226–238.
- Dowty, D. (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.
- Durstewitz, D., Seamans, J., & Sejnowski, T. (2000). Neurocomputational models of working memory. *Nature Neuroscience*, 3, 1184–1191.
- Felser, C., Clahsen, H., & Münte, T. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. *Brain and Language*, 87, 345–354.
- Ferreira, F., Ferraro, V., & Bailey, K. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11–15.
- Ferreira, R., & Patson, N. (2007). The ‘good enough’ approach to language comprehension. *Language and Linguistics Compass*, 1, 71–83.
- Fiebach, C., Schlesewsky, M., & Friederici, A. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German wh-questions. *Journal of Memory and Language*, 47, 250–272.
- Hagoort, P. (2006). The memory, unification, and control (MUC) model of language. In A. Meyer, L. Wheeldon, & A. Krott (Eds.), *Automaticity and control in language processing* (pp. 243–270). Philadelphia: Psychology Press.
- Hagoort, P., & Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45–81). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438–441.
- Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society of London, B*, 362, 801–811.
- Hamm, F., Kamp, H., & van Lambalgen, M. (2006). There is no opposition between formal and cognitive semantics. *Theoretical Linguistics*, 22, 1–40.
- Hitzler, P., Hölldobler, S., & Seda, A. (2004). Logic programs and connectionist networks. *Journal of Applied Logic*, 3, 245–272.
- Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language*. Amsterdam: Mathematical Centre.
- King, J., & 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.
- Kripke, S. (1963). Semantical considerations in modal logic. *Acta Philosophica Fennica*, 16, 83–94.
- Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163.
- Maris, E. (2004). Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology*, 41, 142–151.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG and MEG data. *Journal of Neuroscience Methods*, 164, 177–190.
- McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, 4, 463–502.
- Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory and Cognition*, 23, 477–494.
- Müller, H., King, J., & Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research*, 5, 193–203.
- Münte, T., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, 395, 71–73.
- Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, 59, 494–522.
- Peregrin, J. (Ed.). (2003). *Meaning: The dynamic turn*. Oxford: Elsevier.
- Phillips, C., Kazanina, N., & Abada, S. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22, 407–428.
- Poeppel, D., & Embick, D. (2005). Defining the relation between linguistics and neuroscience. In A. Cutler (Ed.),

- Twenty-first century psycholinguistics: Four cornerstones* (pp. 103–118). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rosenblatt, R. (1962). *Principles of neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. New York: Spartan Books.
- Rothstein, S. (2004). *Structuring events*. Oxford: Blackwell.
- Ruchkin, D., Johnson, R. J., Cacoune, H., Ritter, W., & Hammer, M. (1990). Multiple sources of P3b associated with different types of information. *Psychophysiology*, *27*, 157–175.
- Steedman, M. (1997). Temporality. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (pp. 895–938). Amsterdam: Elsevier.
- Stenning, K., & van Lambalgen, M. (2005). Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science*, *29*, 919–960.
- Stenning, K., & van Lambalgen, M. (2008a). *Human reasoning and cognitive science*. Cambridge, MA: MIT Press.
- Stenning, K., & van Lambalgen, M. (2008b). The logical response to a noisy world. In M. Oaksford (Ed.), *The psychology of conditionals*. Oxford: Oxford University Press.
- Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, *105*, 477–488.
- van Berkum, J., Brown, C., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*, 147–182.
- van Berkum, J., Brown, C., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology*, *40*, 235–248.
- van Berkum, J., Hagoort, P., & Brown, C. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*, 657–671.
- van Berkum, J., Zwitserlood, P., Hagoort, P., & Brown, C. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, *17*, 701–718.
- van Lambalgen, M., & Hamm, R. (2004). *The proper treatment of events*. Oxford: Blackwell.
- van Lambalgen, M., & Smid, H. (2004). Reasoning patterns in autism: Rules and exceptions. In L. Perez Miranda & J. Larrazabal (Eds.), *Proceedings of the eighth international colloquium on cognitive science Donostia—San Sebastian*. Dordrecht: Kluwer.
- van Lambalgen, M., C. van Kruistum, & M. Parriger, (in press). Discourse processing in attention-deficit hyperactivity disorder (ADHD). *Journal of Logic, Language, and Information*.