This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier’s archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright
From time to time: Processing time reference violations in Dutch

Olga Dragoy\textsuperscript{a,b,c}, Laurie A. Stowe\textsuperscript{a,d}, Laura S. Bos\textsuperscript{a,d}, Roelien Bastiaanse\textsuperscript{a,e,*}

\textsuperscript{a} Center for Language and Cognition Groningen (CLCG), University of Groningen, PO Box 716, 9700 AS Groningen, The Netherlands
\textsuperscript{b} Moscow Research Institute of Psychiatry, Potoshnyaya st. 3, 107076 Moscow, Russia
\textsuperscript{c} Research University “Higher School of Economics”, Myasnitskaya st. 20, 101000 Moscow, Russia
\textsuperscript{d} Neuroimaging Center (NIC), University of Groningen, PO Box 136, 9700 AD Groningen, The Netherlands
\textsuperscript{e} University Medical Center Groningen (UMCG), PO Box 30001, 9700 RB Groningen, The Netherlands

A R T I C L E   I N F O

Article history:
Received 3 October 2010
revision received 7 September 2011
Available online 5 October 2011

Keywords:
Language processing
Time reference
Tense
Discourse-linking
P600
Sentence-final negativity

A B S T R A C T

Time reference in Indo-European languages is marked on the verb. With tensed verb forms, the speaker can refer to the past (wrote, has written), present (writes, is writing) or future (will write). Reference to the past through verb morphology has been shown to be particularly vulnerable in agrammatic aphasia (Bastiaanse, 2008; Yarbay Duman & Bastiaanse, 2009) and both agrammatic and non-brain-damaged individuals have longer RTs with verb forms referring to the past than with verb forms referring to the present (Faroqi-Shah & Dickey, 2009). It has been argued that these results are due to the discourse-linking nature of past tense (Bastiaanse et al., 2011). This article reports ERP and behavioral (reaction time and acceptability rating) data on the processing of time reference violations in which verb forms do not match a time frame previously set by an adverb (present adverb–past tense verb; past adverb–present tense verb). The results show that violation by a present tense verb yields a P600 time-locked to the verb. No such response is found for violation by a past tense verb. These ERP results are similar to ERP findings on locally bound and discourse-linked pronominal processing and when related to behavioral findings on identical violations, support the claim that in present tense processing co-reference is established with the speech time (local binding), while past tense processing involves co-reference with some other event time (discourse-linking).

© 2011 Elsevier Inc. All rights reserved.

Introduction

The way in which a speaker can refer to a past, present or future time frame differs between languages. In Indo-European languages the verb may be inflected for tense and aspect, with or without auxiliaries. In English, for example, reference to the past can be made with past tense, imperfect aspect (‘he walked’); present tense, perfect aspect (‘he has walked’); past tense, perfect aspect (‘he had walked’); and with continuous forms (‘he has been walking’; ‘he was walking’). In other languages, like Chinese, verbs are not inflected for time reference, but free standing grammatical morphemes are used to refer to past, present and future, as shown in (1).

\begin{equation}
\text{zhe ge ren du le yi fong sin}
\end{equation}

\text{The man read the letter.}

Recent aphasiological data show that across language types comprehension and production of time reference through both verb inflection (with or without auxiliaries) and free grammatical morphemes is impaired in agrammatic individuals. The accuracy of agrammatic performance is reduced whenever it relies on time reference processing (Faroqi-Shah & Thompson, 2006). Although...
In linguistic theory a distinction is made between binding relations and discourse-linked coreference. The most widely discussed application of this distinction is to pronouns. Binding relations are established within a sentence, such as the relation between an antecedent and a reflexive noun. Binding relations are established within a sentence, widely discussed application of this distinction is to pronouns. A referential approach similar to that for anaphora has been applied to tenses. Both ternary (Reichenbach, 1947) and binary (Enç, 1987) accounts use the notion of reference to define tense systems. In Reichenbach’s terms, in present tense speech time, reference time and event time co-occur, and in past tense reference time and event time precede speech time. According to Enç (1987), tenses are represented as referential indices between event time and time anchor (which is equivalent to speech time): for present tense the two times are co-indexed, for past tense they have different indices. The referential nature of tenses was also stressed by Partee (1973). She argued that tenses behave similarly to pronouns: they can either be obligatorily bound to sentence-internal antecedents or co-referential in a wider discourse, an antecedent being a temporal anchor. The idea of tense as discourse-linked has been recently adopted by Avrutin (2000) who claimed that tense processing does not consist solely of morphosyntactic computations, but rather requires access to the discourse representation, that is, tense in general is discourse-linked. However, treating all tenses as anaphoric discourse-linked is both too wide and too restrictive to cover the data. It is too wide, because it does not make the distinction between past and present which was argued for above. On the other hand, the approach in which tense is only discourse-linked when it is morphologically marked is too restrictive, because reference to the past is more difficult than reference to present through not only tense, but also through periphrastic verb forms (‘has walked’ and through aspectual adverbs, as in Chinese. This has been demonstrated for periphrastic forms with longer RTs in non-brain-damaged individuals (Faroqi-Shah & Dickey, 2009) and with more comprehension and production problems in agrammatic aphasic individuals for all forms of time reference (Bastiaanse, 2008; Bastiaanse et al., 2009; Faroqi-Shah & Dickey, 2009; Yarbay Duman & Bastiaanse, 2009). Thus, we would like to rephrase Avrutin’s (2000) claim: reference to the past through grammatical morphemes requires discourse-linking, while reference to the present is locally bound to speech time.

Augmenting Avrutin’s (2000) account with the idea of a discourse-related distinction between present and past, we can model the difference between the processing of these time frames as establishing co-reference at the morphosyntactic level on the one hand (local binding of present tense) and co-reference at the discourse level on the other (discourse-linking of past tense). Successful processing of a present event expressed through any morphosyntactic

Time reference and discourse-linking

In linguistic theory a distinction is made between binding relations and discourse-linked coreference. The most widely discussed application of this distinction is to pronouns. Binding relations are established within a sentence, such as the relation between an antecedent and a reflexive in ‘the boy, is washing himself’. Having the same index indicates co-referentiality of ‘the boy’ and ‘himself’: they must refer to the same person. Some types of pronouns (reflexives) must be locally bound, others are free to be discourse-linked. In a sentence such as ‘the boy, is washing him’, ‘him’ refers to someone else than ‘the boy’ and a relation has to be established between ‘him’ and a person who is not mentioned in this sentence. In other words, ‘him’ is not locally bound. In order to interpret this pronoun, a co-reference relation needs to be made with an antecedent which may be outside the sentence, although it does not have to be. This relation is called discourse-linking.

A referential approach similar to that for anaphora has been applied to tenses. Both ternary (Reichenbach, 1947) and binary (Enç, 1987) accounts use the notion of reference to define tense systems. In Reichenbach’s terms, in present tense speech time, reference time and event time co-occur, and in past tense reference time and event time precede speech time. According to Enç (1987), tenses are represented as referential indices between event time and time anchor (which is equivalent to speech time): for present tense the two times are co-indexed, for past tense they have different indices. The referential nature of tenses was also stressed by Partee (1973). She argued that tenses behave similarly to pronouns: they can either be obligatorily bound to sentence-internal antecedents or co-referential in a wider discourse, an antecedent being a temporal anchor. The idea of tense as discourse-linked has been recently adopted by Avrutin (2000) who claimed that tense processing does not consist solely of morphosyntactic computations, but rather requires access to the discourse representation, that is, tense in general is discourse-linked. However, treating all tenses as anaphoric discourse-linked is both too wide and too restrictive to cover the data. It is too wide, because it does not make the distinction between past and present which was argued for above. On the other hand, the approach in which tense is only discourse-linked when it is morphologically marked is too restrictive, because reference to the past is more difficult than reference to present through not only tense, but also through periphrastic verb forms (‘has walked’) and through aspectual adverbs, as in Chinese. This has been demonstrated for periphrastic forms with longer RTs in non-brain-damaged individuals (Faroqi-Shah & Dickey, 2009) and with more comprehension and production problems in agrammatic aphasic individuals for all forms of time reference (Bastiaanse, 2008; Bastiaanse et al., 2009; Faroqi-Shah & Dickey, 2009; Yarbay Duman & Bastiaanse, 2009). Thus, we would like to rephrase Avrutin’s (2000) claim: reference to the past through grammatical morphemes requires discourse-linking, while reference to the present is locally bound to speech time.

Augmenting Avrutin’s (2000) account with the idea of a discourse-related distinction between present and past, we can model the difference between the processing of these time frames as establishing co-reference at the morphosyntactic level on the one hand (local binding of present tense) and co-reference at the discourse level on the other (discourse-linking of past tense). Successful processing of a present event expressed through any morphosyntactic
means requires the event to be bound to a locally sited antecedent, e.g. the time of speech, which also must be consistent with an explicit temporal adverb. In contrast, processing of a past event requires linking it to an antecedent via a different process which takes place at the discourse level entirely and does not directly involve morphosyntactic agreement. According to this account, sentences with reference to the past and to the present impose qualitatively different requirements on morphosyntactic and discourse-related knowledge integration.

Binding and discourse-linking in ERPs

It is of interest to see whether such a theoretically grounded and behaviorally supported distinction between local binding and discourse-linking in general and between present and past time reference in particular finds support at the neurophysiological level. For the study of quick and often overlapping processes like those involved in brain activities that underlie different aspects of language processing, using event-related brain potentials (ERPs) provide the method of choice, with an excellent temporal resolution (tens of milliseconds). ERPs allow to measure the time course of evoked brain activity. The obvious prediction from the distinction between local binding vs. discourse-linking is that their processing evokes different brain responses, which should be reflected in different ERP patterns.

The exact characteristics of the ERP responses related to the processing of discourse-linking and of local binding respectively are not known for time reference, since this question has never been addressed directly. However, the claim that present tense processing involves binding via morphosyntax implies that a tense violation in this case will elicit a morphosyntactic processing pattern. The classic response to morphosyntactic processing difficulties is a P600 often accompanied by a left anterior negativity (LAN). Further, if the processing of past tense is handled via discourse-linking and not as morphosyntactic agreement, those ERP responses are not expected.

Furthermore, we do have access to additional information about how the distinction between locally bound and discourse-linked elements is instantiated neurally. There is some ERP evidence supporting the distinction. Most of it comes from the literature on pronominal processing (Burkhardt, 2005, 2007, 2008; Hammer, Jansma, Lamers, & Münte, 2008; Harris, Wexler, & Holcomb, 2000; Osterhout, Bersick, & McLaughlin, 1997; Osterhout & Mobley, 1995; Streb, Rössler, & Hennighausen, 1999). Additionally the study of Baggio (2008) provides evidence that the ERP effects elicited by present time reference processing are similar to those for locally bound pronominal processing.

Looking at the pronominal literature, violations of agreement for locally bound pronouns typically elicit the P600, a positivity generally apparent by 600 ms after the critical word onset which is elicited by a variety of (morpho)syntactic violations (Coulson, King, & Kutas, 1998; Osterhout & Holcomb, 1992). Osterhout and Mobley (1995) found that reflexive–antecedent number and gender agreement violations (‘The hungry guests helped themselves/himself to the food’ and ‘The successful woman congratulated herself/himself on the promotion’) evoked a P600 effect in the time window 500–800 ms after the presentation of the reflexive. Critically, in sentences with a pronoun–antecedent mismatch (‘The aunt heard that she/he had won the lottery’) a local (P600) effect was only found for a subset of participants who judged those sentences to be unacceptable, but not for the ‘acceptable’ group who interpreted the pronoun as being co-referential with a non-mentioned antecedent in the discourse context. This group showed instead a non-significant trend toward a local negativity and a significant negativity at the end of the sentence. Similarly, Harris et al. (2000) showed that a mismatch of a locally bound reflexive in an argument position (‘John’s brothers like himself’) elicited a P600, while a mismatch of a reflexive in a non-argument position (logophor: ‘John’s brothers like Bill and himself’), which can be interpreted as a co-referent instead of as a bound pronoun, did not cause any local ERP effect.

The findings of Osterhout and Mobley (1995) and Harris et al. (2000) are consistent with the claim that a mismatch in local binding necessarily violates morphosyntactic rules and thus results in a P600 response. The P600 serves here as a marker of local binding disruption. In contrast, a discourse-linked mismatch need not be perceived as a morphosyntactic violation and does not necessarily cause a local effect, since the item can be interpreted at a higher level of discourse.

A number of other studies have addressed discourse-referential processing difficulties, as opposed to violations. These typically show a negativity with a latency of about 300 ms, which may either be sustained or short-lived, but in any case differs from the P600 seen with morphosyntactic processing difficulties (Kaan, Harris, Gibson, & Holcomb, 2000). These negativities also differ from the N400 effect which is a known marker for lexical-semantic integration difficulties (Kutas & Federmeier, 2000; Kutas & Hillyard, 1980), as this ‘referential negativity’ is predominantly more anterior in scalp distribution. A review of the literature shows that negativities are apparent for discourse-linked elements, but not to locally bound ones: (1) for logophors which require establishment of a discourse-based dependency as compared to reflexives (Burkhardt, 2005); and (2) for pronouns in contrast to non-referential proper names, but not for reflexives vs. proper names (Burkhardt, 2007; Streb et al., 1999). In yet other studies, a reference related negativity showed up following a pronoun incongruent in gender with its inanimate antecedent (Hammer et al., 2008). Lastly a series of studies has shown a sustained frontal negativity for processing of pronouns with more than one referent (Nieuwland & van Berkum, 2008; Van Berkum, Brown, & Hagoort, 1999; Van Berkum, Brown, Hagoort, & Zwitserlood, 2003). This result strongly suggests that the negativity is due to the attempt to establish discourse-linking to a referent.

In general discourse-linking effects are seen at the point of violation or ambiguity. However, in some cases the
effect was only significant after a delay, time-locked to the final word of the sentence (Osterhout & Mobley, 1995). The sentence-final negativity was more posterior or more broadly distributed across the scalp than the anterior negativity of the immediate responses for referential processing. With some variation in latency and scalp distribution, such a negativity was also apparent for a locally bound reflexive mismatch (Osterhout & Mobley, 1995; Osterhout et al., 1997) and for a range of syntactically, morphosyntactically or semantically anomalous sentences (Friederici, Pfeifer & Hahne, 1993; Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992, 1993; Osterhout & Nicole, 1999). These data show such a sentence-final negativity can be elicited by various difficulties in processing, not just discourse-linking, possibly with some variability in latency and scalp distribution depending on the exact nature of the difficulty.

Despite a difference in scalp distribution, this negativity seems to have much in common with a sustained frontal negativity found for referential ambiguity processing mentioned above, which the authors related to additional use of memory resources while maintaining two antecedents for one unresolved referential slot. Similarly, disruption of co-reference assignment due to a missing antecedent might impose an extra memory load for a search for a successful referential link and/or repair of semantic interpretation of the sentence. The onset of the negativity might vary depending on the specific linguistic materials used in the study: in some cases this search could start immediately after the critical word is presented, in others only at the end of the sentence when it becomes clearer that the processing difficulty has not been solved within the sentence as in the Osterhout and Mobley's (1995) study.

In contrast to research on pronominals, a very limited amount of work has been done targeting present and past time reference assignment. The majority of previous ERP studies focusing on verb violations addressed morphological processing rather than time reference per se. The effects measured, and it is consistent with the claim that processing present and past time reference might parallel the distinction between locally bound and discourse-linked pronominals. Baggio (2008) examined a conflict between a lexical adverb referring to the past and a verb with present tense in Dutch sentences like ‘Afgelopen lente +wint/won Julian een literatuur prijs in Frankrijk’ (‘Last spring Julian wins/won a prize in literature in France’). He found that such a violation elicited two ERP responses time-locked to the verb. Both are typically evoked by a morphosyntactic mismatch: a LAN reflecting automatic word form anomaly detection, and a P600 which has frequently been linked to sentence reanalysis and repair processes (Friederici, Hahne, & Mecklinger, 1996; Hagoort, Brown, & Osterhout, 2001).

Additionally, a sentence-final negativity was found in the violation condition contrasted to the correct condition. The results are strikingly similar to the findings on locally bound reflexives (e.g. in particular Osterhout and Mobley (1995)). In Baggio’s experiment, a verb referring to the present just like a mismatched reflexive cannot establish local co-reference with the past adverb, because the latter is not related to the speech time. That failure results in an ERP response of a morphosyntactic nature. The difference in LAN may possibly be due to the particular kind of materials used to create a morphosyntactic violation. At the end of the sentence it is evident that co-reference is not repaired even with a further context, which causes an extra search for a successful referential link in the whole discourse and subsequent extra memory load resulting in a sentence-final negativity.

Present experiments

The ERP evidence from studies on pronominals suggests how we can best investigate the hypothesis that the discourse-linking vs. local binding distinction is relevant for time reference processing as well as pronominal processing. The study of Baggio (2008) proved that processing present time reference marked on verb in a past time reference context is accompanied by the same ERP effects as processing locally bound pronominals. However, the experimental paradigm used was not complete from a time reference point of view. To support the claim that past tense is discourse-linked while present tense is locally bound, it is necessary to also test the processing of the past tense in an incongruous context and furthermore to show that the processing of this violation is qualitatively different from the processing of the present tense. Therefore we carried out a new study with non-brain-damaged speakers of Dutch using violation in both directions, containing a lexical adverb referring to the past followed by a present tense verb (2) and a lexical adverb referring to the present followed by a past tense verb (3).

1 In Fonteneau, Frauenfelder, and Rizzi (1998) another genuine time reference violation seems to be studied (reference to the future provided by a temporal adverb was violated with the past verb form), but since the publication contains very little information about materials, methods or results, and future may be regarded as mood rather than tense (Aronson, 1977; Partee, 1973), we will focus only on the present/past time reference distinction in this paper.
Embedded relative clauses were used because Dutch is a subject–object–verb language and we preferred to use clauses in base word order to avoid interference with grammatical complexity. Additionally it made the presence of the preceding time reference adverb natural. Since we wanted to measure ERP effects at the end of the sentence too, center-embedded relatives were the most suitable.

The main goal of the study was to discover if there is a difference between present and past time reference processing and provide evidence whether it is related to the distinction between locally bound and discourse-linked co-reference. To do this, we collected three kinds of data: ERPs, RTs and acceptability ratings.

In the ERP experiment, which constitutes the essence of the study, neurophysiological correlates of present and past time reference were investigated. Based on the previous ERP literature on locally bound and discourse-linked pronominals and Baggio’s results (2008) on time reference violation using a present tense verb, different ERP patterns for processing sentences such as (2) and (3) were expected. We hypothesized that a locally bound present verb form in (2) would be perceived as a violation of morphosyntactic agreement and thus result in P600 or LAN + P600 response, typically evoked by such mismatches. In contrast, a discourse-linked past verb form in (3) would not be perceived as a morphosyntactic violation and would not cause a P600 effect and possibly no local effect at all. Both kinds of violations should result in further attempts to establish a successful co-reference at the discourse level presumably reflected in a sentence-final negativity.

The RT experiment provided a deeper insight into online processing of time reference violations. On the basis of the results of Faroqi-Shah and Dickey (2009) we predicted that detecting a violation with verb forms referring to the past would yield longer RTs and more errors than with verb forms referring to the present. A possible reason for such a pattern of performance would be that the violation of local binding of the verb in (2) facilitates recognition of the ill-formedness of the sentence as a whole, while leaning on a wider discourse, needed to evaluate the violation in (3), is a more effortful and time-consuming operation and also offers more options for less common interpretations (leading to the “incorrect” judgment that the sentence is acceptable).

Acceptability data were added to investigate how the amount of context available to the comprehender influences the difference between local and discourse-based co-reference. If present tense verbs are locally bound, the relative clause from (2) should be rated as unacceptable, because a local morphosyntactic relation is violated here. No additional context is necessary. If past tense verbs are uniformly discourse-linked, and readers are content to wait for more context information before concluding that they are unacceptable, that would result in greater acceptability of the relative clause seen in (3), since raters could wait for a broader context in which the co-reference would be successfully established. Alternatively, relative clauses like those in (2) and (3) might be judged as equally acceptable, which would suggest that comprehenders prefer to link to a locally available time reference co-referent when a wider context is not readily available, as they do for pronominals presented in sentences with no available co-referent (Osterhout & Mobley, 1995).

**ERP experiment**

**Method**

**Participants**

Forty Dutch native speakers participated in the study, 8 of which were excluded from analysis due to strong artifacts in their EEG signal. The remaining 32 subjects (all currently receiving a university education; 12 male, 20 female; mean age 22, 7 years, age range 18–31) were distributed over four lists (3 male, 5 female on each list). All of them were right handed, had normal or corrected to normal vision, no diagnosed neurological impairment or psychiatric disorder, and reported no usage of alcohol, recreational drugs or medications that could influence their performance in the experiment. They signed an informed consent according to the Declaration of Helsinki under a procedure approved by the Medical Ethics committee. They received €20 for participating in the experiment.

**Materials**

Each subject read 80 experimental, 80 control and 140 filler sentences. Experimental sentences, as illustrated in Table 1, were presented in four conditions: correct sentences with an adverb preceding a verb, both referring to the present (PrPr); violation sentences with an adverb referring to the past preceding a verb referring to the present (PsPr); correct sentences with an adverb preceding a verb, both referring to the past (PsPs); violation sentences with an adverb referring to the present preceding a verb referring to the past (PrPs).

Conditions were matched on plausibility using ratings collected for the correct versions of the embedded clauses of the experimental sentences (conditions PrPr and PsPs) on a scale from 1 (highly implausible) to 5 (highly plausible). Ratings were obtained from 110 undergraduate students of the University of Groningen who did not take
part in the subsequent ERP experiment. These ratings were initially collected for a study (Bos, 2010) in which the meaning of the continuation of the matrix clause which follows directly after the critical verb in the embedded clause was not important for the grammaticality judgment task, therefore this part of the sentence was not included in the plausibility ratings. For the plausibility ratings the embedded clause was introduced by the neutral phrase ‘Hij ziet’ (He sees). Sentences in conditions PrPr and PsPs were rated as equally plausible (mean rate 4.00 (SD = 0.37) and 4.01 (SD = 0.32) respectively).

Experimental conditions were matched on syntactic structure (embedded relative sentences with the critical clause containing an adverb and a following target verb). The experimental conditions were also matched for noun phrase animacy (the subject was always animate, the direct object was inanimate if the semantics permitted), concreteness (no abstract nouns before the critical verb), and length (3–10 letters). Lists were matched for target verb frequency (again taken from CELLEX and target verb imageability (imageability ratings from Jonkers and Bastiaanse (2007) were used). Of the experimental sentences, 58 contained a transitive target verb and 22 an intransitive target verb. In sentences with an intransitive verb, an adverbal phrase was added to match in length with sentences containing a transitive verb. Of the 80 verbs used, 14 had an irregular past tense2.

The control sentences, shown in Table 2, were included to check whether sentence-final negativity like the one Baggio (2008) found, could be equated to the processing of a semantic violation (N400 sentences). Further it served to see whether morphosyntactic violations (P600 sentences) also evoke this negativity, or whether it is restricted more particularly to reference violations. Both types of control sentences had a correct version (N400c and P600c) and a version containing a violation (N400v and P600v). Anomaly of N400-sentences was manipulated by replacing the target verb with a semantically unrelated verb. Anomaly of P600-sentences was created by a morphosyntactic violation (instead of a past participle an infinitive was used and vice versa in the two experimental conditions).

### Table 1
Examples of four experimental conditions.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Time reference in adverb</th>
<th>Time reference in verb</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>PrPr</td>
<td>Present</td>
<td>Present</td>
<td>De kelner die nu de peper maait krijgt geen fooi</td>
</tr>
<tr>
<td>PsPr</td>
<td>Past</td>
<td>Present</td>
<td>De kelner die zonet de peper maait krijgt geen fooi</td>
</tr>
<tr>
<td>PsPs</td>
<td>Past</td>
<td>Past</td>
<td>De kelner die zonet de peper maalde krijgt geen fooi</td>
</tr>
<tr>
<td>PrPs</td>
<td>Present</td>
<td>Past</td>
<td>De kelner die nu de peper maalde krijgt geen fooi</td>
</tr>
</tbody>
</table>

### Table 2
Examples of control sentences in two conditions per type (c = control; v = violation).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>N400c</td>
<td>De grote bus werd door de chauffeurs bestuurd zodat wij konden slapen</td>
</tr>
<tr>
<td>N400v</td>
<td>De grote bus werd door de chauffeurs gezaaid zodat wij konden slapen</td>
</tr>
<tr>
<td>P600c</td>
<td>We hebben de hele middag op hem gewacht terwijl hij aan het drinken was</td>
</tr>
<tr>
<td>P600v</td>
<td>We hebben de hele middag op hem wachten terwijl hij aan het drinken was</td>
</tr>
</tbody>
</table>

---

2 The high number of factors the sentences were controlled for laid constraints on the availability of suitable verbs, so that there were also irregular verbs included. Since Newman et al. (2007) and Steinhauer and Ullman (2002) found that ERP effects for regular and irregular verbs can differ, a separate analysis with the 66 regular verbs was carried out and reported in footnote 7.
versa). The control sentences were taken from previous experiments on Dutch (P600-sentences from Sabourin (2003); N400 sentences from Hoeks, Stowe, and Doedens (2004)). In the original studies the critical verb was in sentence-final position. For the present study, a complementizer introducing an embedded clause was added after the verb.

To ensure an optimal match between conditions, the experimental and control sentences were created in 80 quadruplets of time reference sentences (320 in total), and 80 pairs of control sentences: 40 pairs to evoke an N400 and 40 to evoke a P600 (160 sentences in total). To avoid repetition effects, sentences were assigned to four lists. Each list contained only one version of the experimental and control sentences. An equal number of sentences for each experimental and control conditions appeared in each of the four lists. These materials were interleaved with 80 sentences from a different experiment, half of which included violations as well. Together with 60 filler sentences which were all correct and structurally different from experimental sentences, each list contained 300 sentences less than a third of which contained violations. The order was pseudo-randomized so that the conditions were evenly spread across the list to avoid effects of learning or attention loss. Lists were then divided into six blocks of 50 sentences. Blocks were matched on the number of experimental and control conditions, target verb frequency, transitivity and imageability.

Procedure

Programming and presentation of the experimental stimuli was done using E-prime (Psychology Software Tools Inc., 2001). Participants were tested in a dimly lit sound-proof cabin sitting at approximately 80 cm distance from a computer screen. Sentences were presented visually, word by word, in the middle of the computer screen, in black on a white background, with 12 pt font size, which allowed comfortable reading. At the beginning of each sentence, an asterisk marking a fixation point appeared for 500 ms. Each word was presented for 240 ms, followed by a blank screen of the same duration. At the end of each sentence a screen with a row of asterisks appeared for 1750 ms signaling to participants the opportunity to blink; the blink interval was introduced to diminish the number of blinks during sentence presentation. Participants were instructed to read each sentence for comprehension and to answer questions on the previous sentence which followed 25% of the sentences randomly to check attention and comprehension. The questions were simple in order to avoid distraction because of answering difficulties. The question disappeared as soon as the response was given or after 10 s of presentation. Participants were asked to respond by pressing buttons ‘1’ or ‘2’ on the keyboard which indicated ‘Yes’ or ‘No’ respectively. The six blocks lasted 7–10 min each, depending on the speed with which the participant responded to questions. At the end of each block participants were given a short break. The total testing time was approximately 1 h.

EEG recordings

The EEG activity was collected from 64 electrode sites, using the extended International 10–20 system via tin electrodes mounted in an elastic cap (Electro-Cap International Inc.), and two mastoid electrodes. Data were digitized using the Ref8a-64 amplifier (TMS International) to 22-bit accuracy and sampled at a rate of 500 Hz,4 employing an online average reference. EEGs were recorded from bipolar tin electrodes situated on the outer canthi of the left and the right eye (horizontal EOG) and above and below the left eye (vertical EOG). The ground electrode was placed on the sternum. Electrode impedance was kept below 10 kΩ.

Data analysis

Offline data processing in BrainVision Analyzer (Brain Products, Munich, Germany) included re-referencing to the average of the mastoid electrodes, lowpass and high-pass filtering with cut-off frequencies of 0.01 Hz and 40 Hz respectively. ERPs were calculated for a 1920 ms interval (480 ms before and 1440 ms after the target verb onset) encompassing a target word, one preceding and two following words.5 Blinks and horizontal movements were corrected using the algorithm of Gratton, Coles, and Donchin (1983). Trials containing artifacts were detected automatically and excluded from the analysis. Overall, 2% of the trials taken into analysis were rejected. Averaged waveforms were computed per person for each condition for each electrode aligned to a 200 ms prestimulus baseline. A total of 43 electrodes chosen for the analysis6 were grouped in 15 regions of interest (ROIs): frontal left (AF3, F3, F7), frontal right (AF4, F4, F8), fronto-central left (FC1, FC3, FC5), fronto-central right (FC2, FC4, FC6), central left (C1, C3, C5), central right (C2, C4, C6), centro-parietal left (CP3, CP5), centro-parietal right (CP4, CP6), parietal left (P1, P3, P5), parietal right (P2, P4, P6), parietal-occipital left (PO3, PO7, O1), parietal-occipital right (PO4, PO8, O2), anterior midline (Fpz, AFz, Fz), central midline (FCz, Cz, CPz), posterior midline (Pz, P0z, Oz).

Mean amplitudes were analyzed in three time windows: 300–500 ms, 500–700 ms and 700–1000 ms using a repeated-measures ANOVA. The same time windows were used for the control and experimental sentences to

---

1 We also suggested that participants could blink during the question answering period if necessary.

4 Because of built-in anti-aliasing filters in the amplifier, effectively frequencies above 140 Hz were not sampled. Additionally, for reasons of computational speed, after filtering the data were resampled to 250 Hz before proceeding with segmentation and artifact rejection.

5 The third word following the target was a final word of the sentence in 36% of the time reference sentences and was not included in the analysis because the time reference violation effect could be confounded with the end-of-sentence effect.

6 The inclusion criteria were: no less that 17 trials per condition (out of initial 20) are left after artifact rejection; if one or more electrode of the 43 electrodes included in the analysis was excluded, every ROI contains at least two such electrodes and homologous electrodes are included in the corresponding left and right ROIs. Note that since different electrodes met the inclusion criteria in different participants, the set of electrodes within each ROI was not identical.
compare latency and duration of time reference violation effects with standard effects of semantic and morphosyntactic violations in the same group of participants. Separate analyses were carried out for the N400, P600 and time reference sentences. Lateral and midline ROIs were analyzed separately. For contrasts, the verb was held constant. In all analyses, a factor of violation (correct vs. violated) was included. For the analysis of time reference sentences, a factor of target verb time reference (present vs. past) was also employed. An additional factor of posteriority investigated scalp distribution with three levels (anterior vs. central vs. posterior) for the midline electrodes and six levels (frontal vs. fronto-central vs. central vs. centro-parietal vs. parietal vs. parietal–occipital) for the lateral analysis. In the analysis of the lateral electrodes a second scalp distribution factor of laterality was added with two levels (left vs. right). Within-ROI electrode position was not included as a factor as interactions between electrode and the other scalp distribution factors was not of interest: each ROI consisted of the mean of the electrodes within it. Only the main effect of violation, time and their interactions with the effects of posteriority and lateralization will be reported, since the scalp distribution effects per se do not provide information of interest in this study. The degrees of freedom for the F tests were adjusted using Greenhouse–Geisser correction to correct for violations of the sphericity assumption where appropriate. We report the corrected p-values and original degrees of freedom. The significance criterion was p < .05.

Results

Before presenting results on time reference processing, we report the behavioral data and the ERP data on the control N400- and P600-sentences.

Behavioral data

On average, participants answered 96% of the randomly presented questions correctly (SD = 0.025). These accuracy data suggests that participants read the sentences carefully. No participant was therefore excluded from the analysis on the basis of behavioral results.

Control sentences

Control N400 and P600 conditions were tested in order to compare with the similar effects expected in the experimental sentences. The results from control sentences confirmed the findings of previous studies using similar stimuli (Hoeks et al., 2004; Sabourin, 2003): sentences with violations elicited robust N400 and P600 responses in the corresponding conditions.

N400-sentences. The main effect of semantic anomaly (violation) was significant in the 300–500 ms and 500–700 ms time windows over midline ROIs (F(1,31) = 31.54, p < .001, F(1,31) = 15.02, p < .001 respectively), as well as over lateral ROIs (F(1,31) = 36.91, p < .001, F(1,31) = 10.28, p < .01). The violation factor interacted with the posteriority factor in the 300–500 ms time window over midline ROIs (F(2,62) = 4.90, p < .02) and over lateral ROIs (F(5,155) = 3.57, p < .05).

P600-sentences. The main effect of morphosyntactic violation was significant in the 500–700 ms and 700–1000 ms time windows over midline ROIs (F(1,31) = 32.22, p < .001, F(1,31) = 27.03, p < .001 respectively), as well as over lateral ROIs (F(1,31) = 35.69, p < .001, F(1,31) = 31.27, p < .001). In the same time windows the violation factor interacted with the posteriority factor over midline ROIs (F(2,62) = 5.83, p < .005, F(2,62) = 13.96, p < .001) and over lateral ROIs (F(5,155) = 7.67, p < .01, F(5,155) = 15.93, p < .001). The interaction between the factors of violation and laterality was also significant in the 300–500 ms and the 700–1000 ms time windows (F(1,31) = 7.68, p < .01, F(1,31) = 3.73, p < .01).

Fig. 1 displays grand average ERP waveforms elicted by a critical word in sentences with morphosyntactic violations and their correct counterparts. As compared with the correct critical words, violations elicited a positive deflection starting at about 400 ms post stimulus and continuing on in all further time windows, peaking at about 600–700 ms after the stimulus onset. The effect of violation is larger at the central and posterior sites with no lateral preference, explaining the interaction with posteriority. The time course and the distribution of the N400-effect seen in the control sentences thus closely resemble the standard N400 response characteristics (Kutas & Hillyard, 1980).

Time reference sentences

In this section, we report the results for experimental sentences with target verb referring to the present and target verb referring to the past, time locked to the critical verb, to the sentence-final word and to the main clause verb.  

An analysis on regular verbs only did not yield different results from the analysis including all verbs reported in the main text. Since there were only 14 irregular critical verbs, no separate analysis for irregular verbs was carried out. The results can be considered typical for regular verbs, minimally.
Overall analysis. The main effect of violation was statistically significant in the 500–700 ms and 700–1000 ms time windows over midline ROIs ($F(1,31) = 6.55$, $p < .02$ and $F(1,31) = 4.84$, $p < .05$ respectively), as well as over lateral ROIs ($F(1,31) = 6.17$, $p < .02$, $F(1,31) = 5.87$, $p < .02$). Critically, the factor of violation interacted with the factor of the target verb form in all three time windows (300–500 ms, 500–700 ms, 700–1000 ms) over midline ROIs ($F(1,31) = 6.34$, $p < .02$, $F(1,31) = 7.74$, $p < .01$, $F(1,31) = 12.33$, $p < .001$ respectively) and over lateral ROIs ($F(1,31) = 7.35$, $p < .01$, $F(1,31) = 6.36$, $p < .02$, $F(1,31) = 13.52$, $p < .001$). Scalp distribution factors will be discussed later in this section.

The results suggest that from 500 ms after the onset of the critical verb sentences with a time reference violation are processed differently in comparison to correct sentences. Moreover, as early as 300 ms post-stimulus continuing through all time windows analyzed, sentences containing a violation with the verb referring to the present are processed differently in comparison to sentences containing a violation with the verb referring to the past. In order to find out how the violation effect is reflected in sentences with different target verb forms, separate analyses for contexts with the verb referring to the present and verb referring to the past were performed.

Target verbs referring to the present. In this analysis, correct sentences with both the adverb and the verb referring to the present (condition PrPr) were contrasted with sentences in which past time reference context provided by the adverb was violated with a verb referring to the present (condition PsPr). This contrast was chosen instead of comparison between sentences with correct and violated present context (condition PrPr vs. condition PrPs) to avoid an additional confound related to different target verb forms and the possibility that they could be processed differently in principle, for whatever reason (e.g. word form frequency, length, etc.).

The main effect of violation was significant in all time windows (300–500 ms, 500–700 ms, 700–1000 ms) over midline ROIs ($F(1,31) = 7.97$, $p < .01$, $F(1,31) = 16.46$, $p < .001$, $F(1,31) = 17.19$, $p < .001$ respectively) and over lateral ROIs ($F(1,31) = 8.19$, $p < .01$, $F(1,31) = 12.25$, $p < .001$, $F(1,31) = 17.83$, $p < .001$). The effect of violation also interacted with the effect of posteriority in the 500–700 and 700–1000 ms time windows over midline ROIs ($F(2,62) = 4.83$, $p < .02$, $F(2,62) = 9.62$, $p < .001$). Overall analysis. The main effect of violation was statistically significant in the 500–700 ms and 700–1000 ms time windows over midline ROIs ($F(2,62) = 4.83$, $p < .02$, $F(2,62) = 9.62$, $p < .001$). Over lateral ROIs a tendency toward this same interaction was observed in the 500–700 ms time window ($F(5,155) = 2.87$, $p < .08$), and in the 700–1000 ms time window it was statistically significant ($F(5,155) = 7.29$, $p < .01$).

The effects are illustrated by grand average ERP waveforms for correct and violated sentences with the target verb form referring to the present (Figs. 3 and 5). In comparison with correct sentences with both adverb and verb referring to the present, violation of the past time context by the present tense of the verb elicits a positive wave, starting as early as 300 ms, continuing through all analyzed time windows, and peaking at about 600–700 ms after the stimulus onset. The effect is larger at the central and posterior electrode sites with no difference between hemispheres, explaining the interaction with posteriority.

Target verbs referring to the past. In contrast to the violation by the present verb form, violation by the past verb form showed no significant effect in any time window. Fig. 4 illustrates that violation of the present time context by a verb referring to the past shows no change in brain responses in comparison to sentences with both adverb and verb referring to the present. With the goal of investigating a potential delayed effect, a post hoc analysis was performed.

As can be seen in Fig. 5, the acceptable past form is somewhat more positive than the acceptable present form. We chose to set up our contrasts with the direct comparison of the same form to make sure that any difference was not due to variation in factors such as length and frequency between the word forms employed, and these remain candidate explanations for the difference seen here. What is most important is that the figure clearly shows that the lack of an effect of the violation associated with the past tense verb cannot be regarded as a ceiling effect: it is clearly possible for additional positivity to be elicited by these sentences.

From a discourse-linking perspective, the main verb (directly after the critical verb) is where the information most immediately relevant for discourse-linking comes in. To explore whether a delayed effect of time reference violation for the past tense verb became apparent at this point, an analysis of this time period was also carried out. Time windows equivalent to the later time windows in the critical verb analysis were chosen and the same factors were included. The analysis revealed no significant effect for the matrix verb in the PrPs condition, from which it can be concluded that the new information relevant for the discourse level interpretation provided by the main clause verb does not immediately lead to a process which reflects the anomaly of the entire sentence.

End of the sentence. To investigate whether there was a sentence-final negativity segments time-locked to the final
word of experimental sentences (typically three or four words after the target verb) were identified and averaged. ERPs were calculated for a 1480 ms interval (480 ms before and 1000 ms after the target verb onset) encompassing a target word, one preceding word and 520 ms after the end of the sentence. The stages of data processing and

Fig. 3. P600 response time-locked to the target verb for time reference violations with present tense verbs.

Fig. 4. No response time-locked to the target verb for time reference violations with past tense verbs.
analysis were identical to the sentence-medial procedure, except time window choice. Since a blinking interval started 480 ms after onset of the final word, the statistical comparisons were only carried out in the 300–500 ms time window.

Overall analysis of the four time reference conditions, that is, sentences containing verbs referring to the present and sentences containing verbs referring to the past, showed that the main effect of violation was significant in the 300–500 ms time window over midline ROIs ($F(1,31) = 6.50$, $p < .02$) and over lateral ROIs ($F(1,31) = 5.16$, $p < .05$). A three-way interaction between the factors of violation, posteriority and laterality was also found in the lateral analysis ($F(5,155) = 3.38$, $p < .05$). There was no interaction between the factors of violation and the factors of the target verb form, which suggests that, in comparison to correct sentences, violations by both target forms elicited similar brain responses.

Fig. 6 illustrates the effect of time reference violations time-locked to the final word of the sentence for both contrasts, since no statistical difference was found between them. Violations elicit a negative-going deflection that starts from 300 ms, peaks at 400 ms after the stimulus onset and appears to sustain. The effect is larger at the central electrode sites, and more lateralized to the right hemisphere, mainly at the frontal and central electrodes.

To check whether this sentence-final negativity was specific to (time) reference violations, we performed the same analysis in semantically and morphosyntactically anomalous control sentences. These anomalies did not elicit any indication of a sentence-final negativity in the N400- and P600-control sentences.

**ERP summary**

With the aim of finding out if violation of reference to the present differs from violation of reference to the past in terms of brain responses an ERP study was performed. The results confirmed that processing present time reference and processing past time reference are associated with non-identical ERP effects. Violating preceding past time reference context by the present verb form, as in ‘The boy who just before a letter writes’, elicited a P600-like response time-locked to the target verb. No ERP effects time-locked to the target verb were found in sentences in which present time reference was violated with the past verb form, as in ‘The boy who now a letter wrote’. This remained true when time windows were examined in which the tense of the main clause verb would be taken into account. However, a sentence-final negativity was found in both experimental conditions, but was absent in the control sentences with semantic and morphosyntactic anomalies.

**RT study**

The lack of violation effect in sentences with a past tense verb might have different causes, one of them being that the violation was simply not detected. This is unlikely, since there was a sentence-final negativity, just like the one found in the sentences violated by a present tense verb. This sentence-final effect was not found in the control sentences, hence, it is not just a general wrap-up effect. The sentence-final negativity implies that the past tense violation was detected, though late, suggesting that processing a violation by past tense takes longer than processing a violation by present tense (as reported by Faroqi-Shah and Dickey (2009)). Such a behavioral delay
could cause a delay in a P600 response qualitatively similar to that found in the present tense violation sentences, although this possibility seems unlikely given the lack of a P600 effect in the very late time windows tested in the main verb analysis (see Time Reference sentences). To evaluate this possibility for the materials used in this experiment, which differ substantially from those used in the Faroqi-Shah and Dickey’s study, a reaction time experiment was performed. The results should also be informative about whether comprehenders are capable of recognizing the violation before the end of the sentence.

Method

Participants

Seventeen Dutch native speakers (none of whom participated in the previous ERP study) took part in this experiment; one was excluded from analysis due to an excessive number of errors in detecting violations. The remaining 16 subjects (4 male, 12 female; mean age 22 years, age range 18–25 years) met the same inclusion criteria as in the ERP experiment and were evenly distributed across four lists (1 male, 3 female on each list). They signed an informed consent according to the declaration of Helsinki and received €5 for participating in the experiment.

Materials

Each subject read 32 experimental and 64 filler sentences. A subset of the experimental sentences was randomly selected from the set of 80 experimental sentences used in the ERP study with the same distribution of sentence variants and conditions over four lists (See Table 1). Filler sentences (32 acceptable, 32 containing a violation) differed from the experimental sentences both in syntactic structure and in types of violations. Either lexical-semantic anomalies or erroneous interjections of one or two nouns, articles, prepositions or conjunctions were used in unacceptable filler sentences. Together with fillers each list contained 96 sentences.

Procedure

Participants were tested in a dimly lit sound-attenuated room sitting at approximately 80 cm distance from a computer screen. Programming and presentation was done using E-prime (Psychology Software Tools Inc., 2001). Presentation of the sentences was similar to the ERP procedure: they were presented visually, word by word, in the middle of the computer screen, in black on a white background, with 12 pt font size. At the beginning of each sentence an asterisk marking a fixation point appeared for 500 ms. Each word was presented for 240 ms, followed by a blank screen of the same duration. At the end of each sentence a blank screen appeared for 3000 ms, giving the participant time for any delayed responses. Participants were instructed to read each sentence carefully and to respond as soon as they detected an error by pressing the spacebar on the keyboard. Sentences were presented in three blocks of 6 min each. At the end of each block participants were given a short break. The total testing time was approximately 20 min.

Results

The response times and accuracy of participants’ responses to time reference violation sentences were analyzed. Since the point at which it becomes clear that sentences are not violated is extremely different these RTs were not included in the analysis. The mean values of both variables in the two conditions are presented in Fig. 7.

Participants judged correct experimental sentences (conditions PrPr and PsPs) as acceptable in 88% and 89% of trials (i.e., participants correctly did not respond that the sentence was incorrect). This demonstrates that acceptable sentences with present and past time reference used in the study are rated on-line as equally good, which corresponds to the off-line results of our plausibility ratings (see Materials, ERP experiment).

For experimental sentences containing a violation (conditions PsPr and PrPs), a clear difference was found between the conditions. First, RTs were analyzed. RTs exceeding two standard deviations from the mean RT for a participant in a condition were defined as outliers and excluded from further analysis. The results showed that violations of a present tense context by a past tense verb (condition PrPs) were detected later than violations of a past tense context by a present tense verb (condition PsPr). Mean RTs in the two conditions were 1760 ms (SD = 550 ms) and 1366 ms (SD = 545 ms) respectively, which is statistically significant ($F(1,15) = 7.71, p < .01; F2(1,30) = 10.27, p < .01$) as revealed by a repeated measured ANOVA analysis.

Additionally, an error analysis was performed. In sentences with an adverb referring to the past and a verb referring to the present (condition PsPr), time reference violations were detected in 90% (SD = 30%) of trials, comparable to the accuracy on correct sentences. However, in sentences with the adverb referring to the present and the verb referring to the past (condition PrPs), time
reference violations were detected in only 74% (SD = 44%) of the trials. A repeated measures ANOVA analysis revealed a significant effect of the target verb form, with participants being more accurate in detecting violations by verb forms referring to the present than by verb forms referring to the past ($F(1,15) = 7.98, p < .01, F(1,31) = 10.92, p < .01$).

**RT experiment summary**

Consistent with the ERP data, the violation of present time frame by a past tense verb was detected later than the violation of past time frame by a present tense verb. This replicates the results of Faroqi-Shah and Dickey (2009) with the current materials. Critically, although participants were slower and less accurate in detecting violations of a present time context by verbs referring to the past, judgments to both temporal mismatches were detected amply above chance, which proves that they are true violations. The difference in accuracy and RT suggests, however, that the way in which the judgment of violation is achieved is qualitatively different.

Since the ERPs only showed a difference between violated and correct past tense sentences at the final word, we were interested to see whether subjects could detect the error before the end of the sentence. The average response time was about 300 ms after the presentation of the last word for the sentences which contained three words and occurred before the presentation of the final word in the sentences which contained four words, suggesting that the decision was not contingent on presentation of the final word.

**Acceptability rating**

The ERP and RT results both suggest that present time reference violations by a verb referring to the past are not noticed as quickly as past time violations by a present tense verb. Even a word later with additional tense information does not elicit a response to the violation. Nevertheless the violation is clearly noticed by the end of the sentence. The question remains why there were longer RTs for violations by past as opposed to present tense verbs. Is it because processing past tense is more effortful? Alternatively is it because the judgment depends on information relevant to the interpretation of the past tense which comes later in the sentence, in the context of the matrix clause? Because the preliminary plausibility ratings using only the correct versions of experimental sentences (as described in the Materials section of the ERP experiment) do not supply sufficient information to answer this question, an additional acceptability rating was carried out. Ratings for the effects of the information available in the subject noun phrase was obtained and compared to the judgments based on the whole sentence.

**Method**

Binary acceptability ratings of time reference and P600 and N400 control sentences were obtained from 40 first-year students of the University of Groningen (mean age 20, range 18–31) who did not take part in the ERP and RT experiments. The same distribution of items into four lists as in the ERP experiment was used. For the first half of the items, participants got to see the item only up until the point of the critical verb and were instructed to rate whether the fragment could yield to a good sentence or not (sentence fragment task). Note that this task emphasizes that further information may be available later. For the second half of the sentences, participants had to judge whether the sentence as a whole was good or bad (full sentence task).

**Results**

For an overview of results on the acceptability rating, see Table 3. There is a strong main effect of violation in the time reference sentences; the violation sentences are much less often considered accepted than the correct sentences ($F(1,78) = 1361.67, p < .001; F(2,178) = 9672.33, p < .001$). There is also a main effect of tense ($F(1,78) = 6.73, p < .011; F(2,178) = 7.36, p < .01$). In the item analysis, an interaction between violation and judgment position (sentence fragment or full sentence) reaches significance ($F(2,178) = 10.87, p < .001$), but this is not the case for the subject analysis ($F(1,78) = 1.53, p < .22$). Violation interacts with tense ($F(1,78) = 12.00, p < .001; F(2,178) = 9.53, p < .01$). For both tenses, the adverb–verb mismatch causes a major violation, but for PrPs (14% found without violation, SD = 23%) this is somewhat milder than for PsPr (8% found without violation, SD = 19%).

A post-hoc analysis on the sentences without time reference mismatches did not show significant main effects for tense or position. A post-hoc analysis on the sentences with time reference mismatches revealed a main effect of tense ($F(1,78) = 10.84, p < .01; F(2,178) = 9.91, p < .01$). In the subject analysis, there was no main effect of position ($F(1,78) = 2.31, p > .1$) but in the item analysis it was significant ($F(2,178) = 17.53, p < .001$). Sentences in the PrPs condition are less often than sentences in the PsPr condition, but still in the vast majority, rated as containing a violation. This difference between the two violation conditions becomes smaller by the end of the sentence.

**Acceptability rating summary**

Acceptability rating suggests that violations of a present time context by verbs referring to the past appear milder than violations of a past time context by verbs referring to the present, but that they are nevertheless generally still noted even when only local context is available for co-reference assignment. At the broader discourse level (i.e., after the end of the experimental sentences has supplied a context for interpretation) the violation by a discourse-linked item is more readily identified than at the immediate clause level, as the acceptability rating indicates, but the fact remains that already at the clause level, at the end of the subject NP, time reference violations are noted for both present and past verbs.

In light of the rating data, the RT experiment findings that participants were less successful and slower in detecting
time reference violations by a past verb is taken as evidence that processing reference to the past requires extra attention or more complex processing. In the discussion below it will be argued that this is in line with the hypothesis laid out in Introduction that past tense requires discourse-linking and is therefore more demanding.

Discussion

The goal of the current experiments was to determine to what extent neurophysiological and behavioral evidence are consistent with the discourse-linking theory as applied to present and past time reference. First, we will discuss the ERP findings in relation to the previous ERP studies on time reference and pronominals (Baggio, 2008; Harris et al., 2000; Molinaro, Vespignani, & Job, 2008; Nieuwland & van Berkum, 2008; Osterhout & Mobley, 1995; Van Berkum et al., 1999; van Berkum et al., 2003). Then the results on the ERPs and behavioral experiments (RT and acceptability rating) will be discussed in relationship to each other. Finally, the overall results will be discussed in the framework of discourse-linking sketched in Introduction.

Brain responses evoked by time reference violations

The results of the ERP study support the idea that processing of present and past time reference relies on qualitatively different neural processes, as suggested by the theory that past but not present tense is discourse-linked. The participants’ brains responded to the violation of past time context by the present tense verb as soon as it was presented, as demonstrated by the P600 response time-locked to the target verb. In contrast, present time reference violation by a past tense verb was not accompanied by an immediate response whatsoever, even when the search for an effect was extended well into the main clause verb phrase. However, both sentence types elicited indistinguishable negativities later, at the end of the sentence, time-locked to the final word.

The findings are in general consistent with the previous ERP results on time reference processing. Baggio (2008) reported a LAN and a P600 time-locked to the target verb and a negativity time-locked to the final word of the sentence for past time contexts violated by the present verb form. Similar to those results, we found the P600 time-locked to the target verb and a sentence-final negativity, although there was no evidence of a LAN. This difference may be due to a difference in materials. In Baggio’s study, sentences with adjacent temporal adverbs and finite verbs were used (‘Afgelopen lente won Julian een literatuur prijs in Frankrijk’ – literally, ‘Last summer won Julian a literature prize in France’), while in the present study a direct object or an adjunct separated the temporal adverb and the verb (‘De kelner die zonet de peper maalt maalde krijgt geen fooi’ – literally, ‘The waiter who just before the pepper grinds/grounds gets no tip’). The adjacent position of the verb in Baggio’s study may have caused an earlier ERP response (such as the LAN) due to the increased activation of the temporal context provided by the adverb.

Critical similarities are also apparent between time reference and pronominal processing, as suggested by the hypothesis that a similar distinction between discourse-linking and local binding applies in both cases. Osterhout and Mobley (1995) and Harris et al. (2000) found a P600 for mismatches of locally bound pronouns. For pronouns which are discourse-linked, on the other hand, negativities have typically been found. In the present study, a similar dissociation was found for violations by present and past tense verbs, although in this case the discourse-linked elements did not elicit a negativity time-locked to it. Several studies have also failed to find an immediate effect for discourse-linked pronominal co-referents (Harris et al., 2000; Osterhout & Mobley, 1995). In addition to the dissociation at the target verb, we did find negativity, which may be comparable to that reported in a number of referentiality-focused ERP studies either immediately at the site of the violation or on sentence-final words (immediate: Burkhardt, 2005, 2007, 2008; Nieuwland & van Berkum, 2008; Van Berkum et al., 1999, 2003; sentence-final: Molinaro et al., 2008; Osterhout & Mobley, 1995). In full agreement with Osterhout and Mobley (1995), it showed up in both types of referential contrast, and for violations of what we have hypothesized to be discourse-linked elements it was the only significant effect found.

If present tense verbs which violate a past context cause difficulties in morphosyntactic processing (as opposed to discourse-linking), the response should be similar to that in response to other morphosyntactic processing problems. Therefore we compared the positivity to that elicited by our P600 control sentences with morphosyntactic anomalies that had previously been shown to elicit the standard P600 effects (Sabourin, 2003). The P600 effect was very obvious and highly significant in the control sentences. The comparison between this standard P600 response and the response elicited by time reference violations showed that both effects peak in between 600 and 700 ms and have a similar posterior scalp distribution, although it may be slightly more lateralized in the P600 control sentences. There appeared to be a slight difference in latency of onset, with the control violations only becoming more positive around 400 ms after presentation, while the tense violations appeared to already diverge by 300 ms post-stimulus.

Turning to the sentence-final negativity, as noted above, negativities are typical responses to referential violations in general, sometimes sentence-medially and sometimes sentence-finally. In the wide range of negativities with the same latency of about 200–300 ms reported in the literature, the effect for time reference violations found here is most similar to the negativities found by Baggio (2008) and Osterhout and Mobley (1995). These responses were sustained and posteriorly distributed across the scalp like the current results.

Although it is tempting to link the sentence-final effect to this class of referential negativities, there are several arguments against this step. First, sentence-final negativities

9 Note, however, that due to anatomical variability between individuals, in order to test whether scalp distributions differ, a within subject comparison is necessary.
The literature suggested that the decision is indeed more difficult, although the violation is noticed (Faroqi-Shah & Dickey, 2009). However, this assumption was based on an experiment using English materials and it seemed important to confirm this pattern for the sentences used in the current ERP experiment. The RT results showed that time reference violations by past tense verbs were indeed detected later than time reference violated by present tense verbs and were characterized by more errors, although not enough to suggest that subjects do not consider these sentences to contain a violation. These results are in line with the previous grammaticality judgment findings of Faroqi-Shah and Dickey who found that judging violations of a non-past temporal context by the past verb form takes more time than the other way around. Also, both findings are consistent with production data reported by Jonkers et al. (2007), who showed that it takes longer to complete a sentence with a past than with a present verb form. Critically, our RT data are in agreement with the latencies of the ERP effects: violations by present tense verbs are detected more quickly behaviorally and result in an immediate ERP response time-locked to the critical verb; violations by past tense verbs take longer to detect and only evoke a sentence-final ERP response. This is in contrast with results reported by Hagoort (2003b). In this study, ERP latencies could not be predicted by behavioral measures. Thus, both ERP and RT results are consistent with an account in which past and present are processed in a qualitatively different way, with violations by past tense verbs being related to the kind of processing that requires more effort and is therefore subject to RT delays and more errors.

An important issue, however, is whether comprehenders would tend, purely on basis of the subordinate clause itself, to decide that there is in fact a tense violation. As the acceptability rating showed, relative clauses with an adverb referring to the present and a verb referring to the past are indeed considered to be less unacceptable than sentences with an adverb referring to the past and a verb referring to the present, but generally both violations are noted without access to further information. More than 4 out of 5 of the relative clauses with past tense verb violations were judged to be unacceptable without any access to the main clause continuation. By the end of the sentence the difference between the violation conditions appears to decrease, although there is no significant interaction with position. This suggests that both kinds of time reference violation are detected at the local level, but participants find the combination of present adverb and past tense verb slightly easier to cohere into a meaningful sentence and do not always judge it to be a violation when a wider discourse context will become available. This willingness to wait is compatible with the discourse-linking view of past tense processing.

The ERP results do not entirely support the view that past tense is judged as incorrect only in terms of the main clause either. Primarily, this assumption predicts that at some point in the main clause we should see a delayed effect of the mismatch. If this occurs at the sentence-final word, for example, then it does not seem likely that the ERP effect of violation would be so similar to that found

have also been reported for a wide variety of violations in addition to clearly referential violations. For example, Molinaro et al. (2008) reported sentence-final negativities in response to gender mismatch between a subject noun phrase and predicative adjective which occurred earlier in the sentence. Although both have the same scalp distribution, the effect reported in Molinaro et al. is relatively short-lived (200–400 ms). Similarly, Hagoort (2003b) showed that after both gender agreement and semantic violations, early in the sentence, the final word showed an increased parietal negativity, which he interpreted as an N400 effect. Osterhout and Nicole (1999) similarly found sentence-final negativities following anomalous lexical items and morphosyntactic anomalies of verb number agreement which did not differ in any way. Although the current study did not show a significant sentence-final effect of semantic anomalies or non-tense-related morphosyntactic anomalies, these results from the literature suggest that the sentence-final responses are not necessarily specific to referential processing.

Effects of referential ambiguity processing also typically elicit a sustained negativity, which in this sense is similar to the sentence-final negativity found here. However, this effect also differs from the current results in two very important aspects. First and most important, it follows the ambiguous referential expression immediately rather than being delayed to the end of the sentence and secondly it is frontally distributed (Nieuwland & van Berkum, 2008; Van Berkum et al., 1999; van Berkum et al., 2003). Thus, although conceptually similar to the referential negativities found in previous studies, the negativity elicited by time reference violations appears to differ from them in critical ERP characteristics.

We also tested whether the sentence-final negativity was similar to the standard effect of semantic integration difficulties (the N400). This interpretation of the sentence-final negativity has been offered several times in the literature (e.g. Hagoort, 2003b; Osterhout, 1997). We tested this by examining the difference between the time reference negativity and the N400 control sentences with lexical-semantic anomalies. These appeared to differ, the former is a sustained negativity with a parietal maximum which is larger over the right hemisphere, while the latter is short-lived and has a centro-parietal scalp distribution, suggesting that these effects rely on non-identical neural resources related to dealing with integration difficulties with time-reference on the one hand and lexical semantics on the other.

The relation between the ERP and behavioral data

The behavioral experiments were carried out to further investigate why there was no immediate effect of violation by a past tense verb at the point of the violation in the ERP experiment. There are several possibilities. Most simply, it could be that the combination of present adverb and past verb tense is not considered to be a violation in and of itself. Alternatively, comprehenders may require additional time to process the violation, leading to a late effect, due to the nature of the process.
for the present tense violations sentences, in which the mismatch has been detected at an earlier point in time. It has been shown that a violation eliciting a positivity sentence-medially which occurs at the end of a sentence still elicits a positivity, this is not replaced by a negativity (Hagoort, 2003b; Osterhout, 1997). Also, the RT data suggests that people are able to make a decision about acceptability of the tense violation before the end of the sentence, as the mean decision time occurs before reading the sentence-final word for most sentences (those with more than three words in the main clause verb phrase) or extremely shortly (about 300 ms) after the presentation of the final word (for those with only three words). This would seem unlikely given that the decisions to the quite apparent locally bound mismatches takes well over a second to make. If the judgment were to be elicited on the basis of the tense information of the main clause verb, which is the information most likely to contribute to being sure that there is a tense violation, we would expect a response time-locked to this position, but this does not appear to be the case.

To sum up, we are left with a picture in which it appears that comprehenders are able to notice a violation due to the combination of a present tense context with a past tense verb, but respond to it in a qualitatively different way than they do to a present tense verb which violates a past tense context, the exact pattern predicted by the hypothesis that past tense is associated with discourse-linking while the present tense is locally bound.

**Bound and discourse-linked temporal processing**

The starting point for the current experiment was the dissociation between reference to the present and reference to the past as expressed through verb inflection and verb morphology. This dissociation has been reported for processing both in healthy and aphasic populations (Bastiaanse, 2008; Bastiaanse et al., 2009, 2011; Faroqi-Shah & Thompson, 2006; Jonkers et al., 2007). Based on the insight that time reference makes use of co-reference in a way very similar to pronouns (Partee, 1973; Webber, 1988), a theory was proposed by Zagona (2003) in which present tense processing requires establishing bound coreference with the speech time (local binding), while past tense processing involves co-reference with some other previous event time (discourse-linking). This theory provides a potential explanation for the dissociation found in aphasia (Bastiaanse et al., 2011). The current data provides support for this hypothesis: violations of the preceding temporal context by present and past tense verbs differ in their acceptability, their processing is characterized by different RT and error patterns, and they evoke different ERP responses. All the differences mentioned point to a substantive difference in the processing of the two types of violation, which is already explained by the distinction between local and discourse-related reference.

Each experiment provides some element necessary for understanding the processing of time reference as expressed through grammatical morphology from behavioral to neurophysiological levels. The acceptability rating provided evidence that the combination of a present adverb and a past tense verb is slightly more acceptable than the combination of a past adverb and a present tense verb. It suggests that a present tense verb is required to be locally bound to the speech time (realized in case of our experimental materials as the adverb 'now') and that this has immediate effects, while this is not true for the past tense, which rather can potentially co-refer to another event in a wider discourse. This results in RT and ERP differences in processing violations by present and past tense verbs. Since the past adverb–present verb combination is judged unequivocally unacceptable, this mismatch is immediately detected causing faster RTs and the ERP response time-locked to the critical verb. In contrast, a present adverb–past verb combination is judged to be not necessarily unacceptable, but with a greater chance to be redeemed by a successful referential link to some other event time expressed in further discourse. That moves the moment at which comprehenders are sure of the mismatch to the nearer to the end of the sentence and causes longer RTs and the appearance of the sentence-final ERP effect which is comparable to that found for the locally bound present tense violation.

In theoretical linguistic terms temporal processing can be modeled as straddling at the syntax–discourse interface. We suggest that an adequate model of processing has to be capable of modeling the distinction proposed here. For example, the Unification Model (Vosse & Kempen, 2000) has been employed by a number of researchers, including Hagoort (2003a) and Baggio (2008) to account for a range of sentence parsing effects. To illustrate possible mechanisms of temporal mismatch processing, local binding can clearly be encoded using the operation of unification (binding), which is the essence of the model: it links lexical frames and checks agreement features. One of the reasons why binding may fail is that an agreement check finds a mismatch in grammatical features of the elements to be bound. The failed check normally results in subsequent attempts to recover binding reflected in the P600 effect.

**Table 3**

Overview of percentage of items per condition rated as correct.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Overall</th>
<th>Sentence fragment</th>
<th>Full sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (%) SD (%)</td>
<td>Mean (%) SD (%)</td>
<td>Mean (%) SD (%)</td>
</tr>
<tr>
<td>PrPr</td>
<td>97 6</td>
<td>98 4</td>
<td>95 5</td>
</tr>
<tr>
<td>PsPr</td>
<td>8 19</td>
<td>10 10</td>
<td>6 8</td>
</tr>
<tr>
<td>Difference</td>
<td>89</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>PsPs</td>
<td>96 7</td>
<td>97 6</td>
<td>96 7</td>
</tr>
<tr>
<td>PrPs</td>
<td>14 23</td>
<td>19 16</td>
<td>10 10</td>
</tr>
<tr>
<td>Difference</td>
<td>82</td>
<td>87</td>
<td></td>
</tr>
</tbody>
</table>
(Hagoort, 2003a). A temporal mismatch between the adverb and the verb can be considered a kind of agreement violation that disrupts the local binding. In case of a past adverb and present verb this disruption is not recoverable, which causes the P600, just as for other kinds of local (morpho)syntactic mismatches. In contrast, the combination of a present adverb and past verb can be recovered if a wider discourse is forthcoming, since the past verb may be linked to a different node there. Thus, no P600 effect is observed for this kind of temporal mismatch. In both conditions, be it a disruption of local binding or a postponed linking, the parser is still open for a successful link until the end of the sentence when the attempt to adjust unification ultimately fails causing the sentence-final negativity.

**Issues for further research**

Needless to say, there are many issues left unresolved or not addressed at all which are raised by the current study. The issue of sentence-final negativity remains underspecified within the Unification Model. However, Baggio (2008) argues that within his experiment this effect reflects the readjustment of verb constraints, so that the whole set of constraints becomes satisfiable. In our experiment, such a readjustment may force a historic present (for the past adverb–present verb combination) or narrative past reading (for the present adverb–past verb combination). This claim is in line with the view that sentence-final negativity might reflect the extra memory load needed to support the search for a successful referential link and/or repair of the semantic interpretation of the whole discourse model constructed on the basis of the sentence. This view is also compatible with a single perspective for time reference, pronominal reference (Osterhout & Mobley, 1995) and referential ambiguity processing (Nieuwland & van Berkum, 2008). However, it remains an issue whether this negativity reflects such a specific integration, or rather a more general effect elicited by any incoherence in the sentence as a whole regardless of whether it implicates the discourse model or not. Such a conclusion would be interesting: although in the current study no sentence-final effect was found in control sentences with semantic and morphosyntactic anomalies, there are examples of sentence-final negativities elicited by both lexical semantic and morphosyntactic anomalies (e.g. Hagoort, 2003b). There is some variation in the scalp distributions and latencies of the sentence-final negativities discussed above, and thus more in depth research using within subject comparisons is indicated before the exact status of these negativities is clear.

We have suggested that the dissociation found in aphasia (e.g. Bastiaanse et al., 2009) can be explained by the difference between local binding and discourse-linking. A related issue for further research is whether the dissociation is due to a problem with discourse-linking or with the working memory resources necessary to carry out discourse-linking. The proposed involvement of extra memory load in establishing a discourse link while processing past time reference may well be the reason for a dissociation between present and past found in individuals with aphasia. It has been shown that linguistic and working memory deficits positively correlate in aphasia (Caspari, Parkinson, LaPointe, & Katz, 1998; Sung et al., 2009; Wright, Downey, Gravier, Love, & Shapiro, 2007). Those working memory limitations could make discourse-linked relations more vulnerable than locally bound relations. Also, effects similar to those apparent in the spontaneous speech or non-speeded behavioral experiments of individuals with aphasia become visible in healthy individuals under time pressure. In line with the working memory load interpretation of the sentence-final negativity mentioned above, in our RT experiment more errors were observed for sentences with verbs referring to the past: the speeded task imposed extra processing constraints, and healthy participants showed more difficulties in sentences for which they could otherwise successfully have determined whether a successful discourse-linking was possible. The role of working memory provides an attractive explanation of aphasic and normal error patterns, but its role in establishing discourse-linking requires additional clarification.

For the past discourse-linking theory discussed here, a necessary line for further research is the systematic comparison of ERP effects related to processing of locally bound and discourse-linked elements. This systematic examination has several facets. For one, we have appealed to a comparison between pronominal and tense processing to support the existence of a dissociation similar to that found in pronominal processing for tense processing. If we are really seeing the same effects in both modalities of reference, they should show comparable ERP responses. However, a real comparison of ERP effects is best done with the same group of subjects, since the scalp distribution of various effects can differ quite noticeably between different groups of participants.

Furthermore, we have barely scratched the surface of tense processing. The status of future tense is an interesting issue. Faroqi-Shah and Dickey’s (2008) results suggest that future tense patterns with present tense: that is, it is locally bound and relatively easy. Data from aphasia collected in a language with a more extensive morphologically marked tense system, on the other hand (Yarbay Duman & Bastiaanse, 2009), suggest that like the past tense, future is more difficult than present. To add to the possibilities, according to Zagona’s (2003) theoretical account, future is neither bound nor discourse-linked. Temporal relations conveyed by future time reference therefore urgently need further investigation.

Finally, the initial impulse for this line of study into time reference came from morphological processing studies. However, time can also be indicated by lexical adverbs and event sequences, which present another interesting direction for future research.

**Acknowledgments**

We thank the anonymous reviewers for their valuable comments on an earlier draft of this manuscript and most particularly the non-anonymous reviewer Giosuè Baggio for extensive discussion above and beyond the call of reviewership. We thank Peter Albronda for technical assistance with data collection and invaluable lectures on ERP.
hardware principles. We are also grateful to Zoltan Banreti for discussion on the topic of tense and anaphora. This work was supported by a grant from the Faculty of Arts of the University Groningen and by the Russian Foundation for the Humanities (Project 08-04-00165a).

References


