Before and after, temporal connectives in discourse processing

Ming Xiang, Emily Hanink and Genna Vegh

Linguistics Department, University of Chicago

Corresponding author:

Ming Xiang

Language Processing Lab

Linguistics Department, University of Chicago, IL, Chicago, 60637

mxiang@uchicago.edu

phone: 773-702-8023

fax: 773-834-0924

Word count: 2991

Abstract

Temporal connectives *before* and *after* carry discourse presuppositions about the veridicality of the subordinate event. In an ERP study, we show that the comprehension system is highly sensitive to such presuppositions, leading to more processing cost on *before*-initial clauses than *after*-clauses, but only when the discourse context introduces more uncertainty about the veridicality of the event in the *before*-clause. Our results challenge previous account that attributes the processing asymmetry between *before* and *after* to simply the temporal ordering function of these connectives, and highlight a discourse processing mechanism that is highly incremental and sensitive to multiple sources of discourse information.

Introduction

Temporal connectives, such as *before* and *after*, serve to establish temporal ordering between events. Informally speaking, a temporal clause introduces the reference time for event *E1*, and the reference time for the main clause event *E2* is then anchored upon E1. Temporal clauses that begin with the connective *before* indicate that E1 takes place after E2, while those beginning with the connective *after* indicate the opposite. In their seminal paper, Münte et al. (1998) found that 500ms after the onset of a temporal connective, as in the examples below, *before*-initial clauses elicited a prolonged left frontal negativity that was larger than that elicited by *after*-initial clauses. Furthermore, the effect was highly correlated with comprehenders' working memory (WM) capacity – people with larger WM spans showed a larger negativity difference between *before* and *after* clauses.

(1) a. After the scientist submitted the paper, the journal changed its policy.

b. Before the scientist submitted the paper, the journal changed its policy.

Münte et al. (1998) argued that examples like (1b) present the temporally later event in the initial clause, and the temporally prior event in the following main clause, making the linear sequence of the two events reversed from how events are temporally ordered in the real world. The linear sequence of *after*-initial sentences (e.g. 1a), on the other hand, adheres to the real-world event sequence, where the temporally prior event is expressed first. More computations are therefore evoked while comprehenders integrate linguistic information and real-world knowledge to establish a discourse model for the *before*-initial sentences. A similar asymmetry between *before* and *after* clauses was also observed behaviorally in children (Natsopoulos and Abadzi, 1986; Troberg, 1982), and in patients with Parkinson's disease (Natsopoulos et al, 1991). In a recent fMRI study, Ye et al. (2012) identified the caudate nucleus and the left middle frontal gyrus as the neural network that supports the additional computations in *before*-initial sentences. In almost all of these studies, the temporal sequencing function of *before* and *after* is taken to be the primary source to explain the processing asymmetry.

Although the basic semantic function of *before* and *after* seems to be a mere matter of how these connectives temporally order two eventualities, there is actually important pragmatic information that they bring to the discourse. In particular, *before* and *after* clauses give rise to different veridicality inferences of the event expressed in the temporal clause (Heinämäki, 1972; Lascarides & Oberlander, 1993; Condoravdi, 2010; Krifka, 2010; Giannakidou, 1998, 1999). Consider the examples in (1) again. There is a very strong inference for both (1a) and (1b) that the scientist did submit his/her paper.

However, an important difference between *before* and *after* clauses is that although the events described by *after*-clauses must be veridical, *before*-clauses allow certain degree of variability—while (1a) necessarily describes a situation in which the scientist must have submitted the paper, (1b) is actually ambiguous. In the most salient reading of (1b), the scientist indeed submitted the paper after the journal changed its policy; but it is also possible that the scientist originally was planning to submit the paper, but he/she did not do so because the journal changed its policy.

As *before*-clauses allow both veridical and non-veridical events, their ultimate interpretation is largely determined by the overall discourse context. In (1b), the discourse is compatible with both the veridical and the non-veridical reading of the event in the *before*-clause, though the veridical reading is dominant. In the examples below, however, the discourse context is only compatible with a non-veridical reading of the *before*-clause.

(2) a. Before the scientist submitted the paper, the journal was closed down.

b. Before the bomb exploded, it was defused.

The more appropriate readings of these *before*-clauses are all non-veridical: the scientist did not submit the paper; the bomb of course did not explode. Otherwise, the resulting discourse would be incoherent.

We therefore hypothesize that the comprehension of the temporal connectives *before* and *after* triggers computations about discourse presuppositions, including whether a veridicality inference of the subordinate event is satisfied and whether the ultimate discourse is coherent. We further hypothesize that the additional processing cost on *before*-initial clauses is due to heightened uncertainty of the veridicality of the *before*-

event. Encountering a temporal connective *before/after* strongly signals to the comprehension system about the presence of a subordinate event. For *after*-initial sentences, the comprehension system can simply update the discourse context by adding a veridical subordinate event to the context. For *before*-initial sentences, until the comprehender finishes processing the main clause, the discourse model can not be updated with certainty—both the veridical and the non-veridical readings of the *before*-event are possible, and only the main clause can help determine whether the overall discourse is compatible with these possible readings. The prolonged frontal negativity observed in previous ERP studies, we hypothesize, indexes the working memory cost of holding multiple possible interpretations alive in the discourse model.

The pragmatic hypothesis then, in contrast to the temporal ordering hypothesis in Münte et al. (1998), makes the prediction that once the uncertainty about the veridicality of the *before*-event is removed, *before*-initial sentences should NOT be more costly than *after*-initial sentences, even though the temporal relations that *before* and *after* express remain the same. We test this hypothesis below by contrasting what we call "Ad-hoc" events and "Real-world" events.

Methods

Stimuli and Design

<u>Stimuli Creation</u>

All of our experimental items contained two clauses. The first clause is a temporal clause that starts with a temporal connective; and the second clause is a main clause. There were two sets of stimuli, each with one hundred and twenty items. The first set contains the "Ad-hoc" events in the temporal clauses and the second set contains the

“Real-world” events. “Ad-hoc” events are made up events, as the examples in (1). This set was largely adopted/translated from the German stimuli provided in Ye et al. (2012). We made slight modifications to make the English stimuli sound natural. The “Real-world” events were created by selecting historical or cultural events that were judged to be culturally salient enough that comprehenders would recognize them immediately as true events that have occurred in history, thus obviating the need to finish the whole sentence before finding out if the events in the temporal clause have taken place. The two critical conditions under each event type are the *before* and *after* conditions. An example is given below in Table 1. For each item, we also included a *when*-clause condition, but this was only a filler condition. We constructed the *when*-sentence in exactly the same way as the critical *before* and *after* conditions such that the experimental and filler sentences were matched as closely as possible. We did not have any a priori predictions on the *when*-sentences, and therefore did not analyze them in the results section below.

The Ad-hoc and Real-world stimuli sets were each divided into three lists according to a Latin Square design. We then combined these into three final experimental lists. Each list contained 80 *before*-sentences (40 ad-hoc and 40 real-world), 80 *after*-sentences (40 ad-hoc and 40 real-world) and 80 *when*-sentences as fillers (also 40 ad-hoc and 40 real-world). Each participant therefore saw 240 sentences in total.

Ad-hoc	**Clause 1**	**Clause 2**
Before (4.6)	1a. Before the psychologist submitted his article,	the journal changed its criteria.
After (4.8)	1b. After the psychologist submitted his article,	the journal changed its criteria.
When (4.6)	filler: When the psychologist submitted his article,	the journal changed its criteria.
Real-world		.
Before (4.3)	2a. Before Tiger Woods won the Masters,	Jake started playing golf.
After (4.8)	2b. After Tiger Woods won the Masters,	Jake started playing golf.
When (4.6)	filler: When Tiger Woods won the Masters,	Jake started playing golf.

Table 1: Example stimuli and coherence rating for each condition

We performed a norming task on the coherence of the three lists of sentences with twelve participants (four per list). Participants were asked to judge the coherence of each sentence on a 1-7 scale, 7 being the most coherent. The average rating for each condition is shown in Table 1. A two-way ANOVA revealed an effect of Connective ($F(2, 22)=9.5$, $p<.01$). But this effect was mainly driven by the fact that the *before*-sentences were rated slightly lower than the *after* and *when*-sentences under the Real-world events ($ps<.05$); no difference was found between connectives under the Ad-hoc events.

ERP Experiment

Participants and Stimulus Presentation

Thirty-seven native English speakers (18-28 years old) participated in the ERP study. Stimuli were presented on the monitor in black font centered on a white background. Before each trial, the text "Ready..." was displayed on the screen until the participant began the trial by pressing a button on the response pad. A cross “+” then appeared in the middle of the screen for 500ms before the sentence appeared. Most words were presented for 400ms each, but to accommodate for the clause/sentence final wrap up processing, the last word of the first clause was presented for 600ms, and the very last word for the whole sentence was presented for 900ms. In addition, since the second region of the first clause sometimes was a long phrase (e.g. due to the fact that some historical/cultural events or figures have long names), we always presented this region for 600ms. All words were followed by a 150ms blank white screen. The first clause of every item was always presented in five regions, so that the amount of time to present the first clause was constant (3150ms counting from the onset of the connective). Participants were instructed to read the sentences quietly to themselves for comprehension. For about 1/3 of the trials, after the final word of the second clause in each trial, a simple yes-no comprehension question appeared on the screen. As the goal of this experiment was to examine whether participants were sensitive to the veridicality of the event described in the first clause, we made sure the comprehension questions only probed the content of the second clause of each scenario. Before the experiment began, participants completed five practice trials. Overall, two hundred and forty trials were presented to each participant.

EEG Recording

The EEG response was recorded from 32 electrodes including the two mastoids electrodes (Brain Vision EasyCap). Two additional pairs of electrodes were placed on the face, one above and below the left eye to monitor blinks and vertical eye movements, and the other on each temple to monitor horizontal eye movements. The EEG signals were amplified through the Brain Vision QuickAmp amplifier, with a band pass of 0.5-70 Hz. It was continuously sampled at 1000Hz and the impedance was kept below 5kOhm.

RSPAN task administration

Before the ERP recording session, each participant performed a Reading Span (RSPAN) task (Kane et al., 2004). Participants read aloud sentences that were either plausible or implausible, and then had to give judgments about whether each sentence was "TRUE" or "FALSE". A letter then immediately appeared for the participants to hold in memory. The sentence-letter combinations are presented in sets. After each set, participants are cued to recall the letters from the set in the sequence they have appeared. The RSPAN scores are calculated by summing up the total number of correct letters recalled from the correct sequential position. (Conway et al., 2005).

Data analysis

For ERP data analysis, the raw data was filtered offline at 40Hz. We first performed an automatic artifact rejection procedure to remove artifact due to excessive ocular and muscular movements (maximal allowed voltage step 100µV; maximum allowed absolute difference 100µV in a 100ms window). We also manually inspected each individual subject to check for additional artifact. Six subjects were removed from data analysis due to a larger than 50% loss of their data. Three more subjects were removed because they have a low accuracy in comprehension questions (<80% correct). Twenty-eight subjects were included in the final data analysis. For data analysis, we first divided the scalp into two large regions and carried out two initial ANOVAs on each region separately: the midline region (four electrodes, Fz, Cz, Pz and Oz); and the whole scalp excluding the midline (24 electrodes). For the analysis excluding the midline four electrodes, we further divided the scalp into three smaller AP (anterior to posterior) areas: frontal (8 electrodes, Fp1, Fp2, F3, F4, F7, F8, Fc5, Fc6), central (8 electrodes, FC1, FC2, CP1, CP2, C3, C4, T7, T8) and posterior (8 electrodes, P3, P4, O1, O2, CP5, CP6, P7, P8). Each AP area is also divided into a left and a right part.

Following Münte et al. (1998), we did our analysis on a long and extended time window, starting from the onset of the connective (i.e. *before* or *after*). Averaged ERPs were calculated relative to a 200ms pre-stimulus baseline that is time-locked to the onset of the connective word. The total epoch shown in Figure 1 extends for 6200ms after the onset of the connective (excluding the 200ms baseline time window). The first 3150ms encompasses the temporal clause, i.e. clause 1 in Table 1; the remaining 3050ms is the main clause (i.e. clause 2 in Table 1)[1].

Statistical analysis was performed over the average amplitude within the 1000-6200ms time window, using the open source statistical package R (R core team, 2014). The three within-subject variables are: Connective (2 levels: *before* and *after*), Events (2 levels: Ad-hoc and Real-world) and Clause (clause 1 or clause 2). We first carried out a Connective x Events x Clause x electrode ANOVA for the midline four channels, and a Connective x Events x AP x Clause x Hemisphere omnibus ANOVA for the rest of the channels. Follow-up analyses were then carried out for subsets of the electrodes when the omnibus ANOVA revealed a reliable main effect or interaction. In all analyses, the Greenhouse and Geisser (1959) correction was applied to repeated measures with more than one degree of freedom.

Results

[1] During the stimulus presentation, the second clause was not completed yet for some trials at this point.

For the midline four electrodes, there is no main effect of Connective or Events ($Fs<1.5$, $p>.2$). But there is a significant Connective x Events interaction ($F(f,27)=7.6$, $p<.05$). We found no other three-way or four-way interactions that involve Connective, Events, Clause, or channels. For the rest of the scalp, neither Connective nor Events had a main effect ($Fs<2$, $ps>.1$), but there is an AP x Connective x Events interaction ($F(2,54)=4.1$, $p<.05$). No other interactions between Connective/Events and Clause or Hemisphere were significant. The three-way AP x Connective x Events interaction was driven by the fact that the Connective x Events interaction was significant in frontal ($F(1, 27)=8.22$, $p<.01$) and central ($F(1,27)=7.64$, $p<.05$) areas, but not in the posterior area ($F(1, 27)=2.32$, $p>.1$). Since there is no effect of Clause or Hemisphere, we next combined all the frontal and central channels (including the three central-frontal midline channels) to form a large 19-channel region, and ran a two-way ANOVA on the 1000-6200ms time window, with only Connective and Events as the two within-subject variables. The Connective x Events interaction is again significant ($F(1,27)=7.1$, $p<.05$; Figure 1).

We carried out planned comparisons over the mean amplitude within the 1000-6200ms time window for the Ad-hoc and Real-world events separately. As shown in Figure 1, for Ad-hoc events, *before*-clauses elicited larger negativity over central-frontal scalp area, throughout the long and extended time window (between 1000-6200ms). In contrast to this, there is no difference between the *before* and *after* clauses under the Real-world events. For Ad-hoc events, the difference between *before* and *after* clauses is also significantly correlated with participants' working memory capacity, as measured by the RSPAN test. As shown in Figure 1F, participants with higher working memory scores showed larger negativity difference (r=-0.44, p<.05; the correlation remained significant even after the highest negativity difference score was excluded). No correlation was found for the Real-world events (r=-0.27, p>.1). For visualization of the correlation, we also split the participants into a high WM and a low WM group, based on the median of their WM scores, and plotted the mean amplitude for each group separately in Figure 1G.

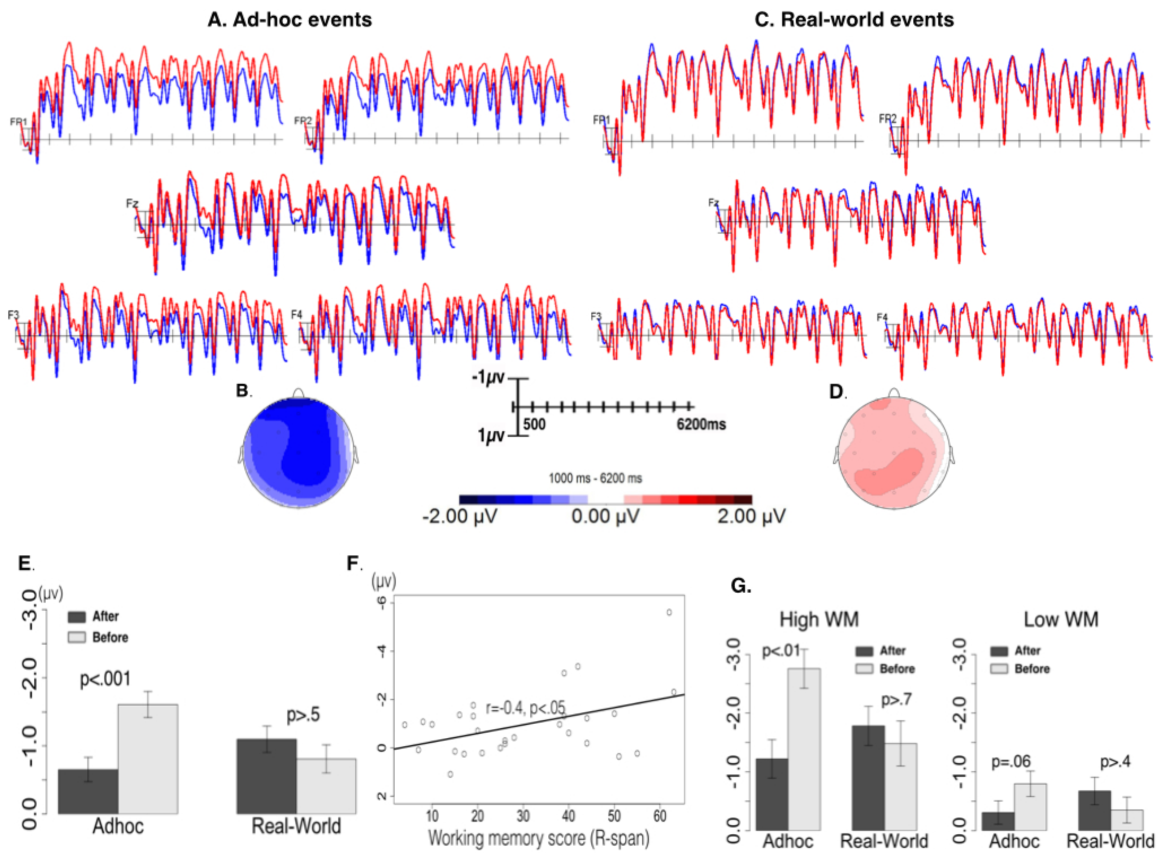

Figure 1: 1A&B: wave forms and voltage map for the Ad-hoc events; **1C&D**: wave forms and voltage map for the Real-world events; for both voltage maps response from the *after*-sentences were subtracted from the *before*-sentences; **1E**: average ERP amplitudes over 19 central-frontal electrodes, averaged over the 1000-6200ms time window; **1F**: Correlation between the negativity differences (Y-axis, *after*-clauses subtracted from the *before*-clauses) and the Rspan scores (X-axis). **1G**: Results plotted separately for the high WM and low WM groups (groups are divided along the median split of the Rspan scores).

Discussion and conclusion

In the current study, we showed that *before*-initial clauses evoked larger prolonged negativity than *after*-initial clauses, but only when the veridicality of the subordinate event was uncertain. Prolonged negativities have been previously associated with heightened working memory cost (Kluender and Kutas, 1993; Hagoort and Brown, 1994; Kaan and Swaab, 2003; King and Kutas, 1995), and our current results also revealed a significant correlation between each individual participant's working memory span and the size of the negativity effect. Taken together, we suggest that the processing cost observed for *before*-initial clauses reflect the cost of maintaining ambiguous discourse representations in working memory while the discourse model is being updated. We assume discourse context is incrementally updated, with each incoming sentence or part of a sentence contributing new information to the discourse model (Kamp, 1981). Each sentence is interpreted in the context of preceding discourse, and at the same time modifies and updates the discourse context. In the case of temporal clauses, when the event described by the temporal clause is added to the discourse context, the veridicality of the event could be determined by a number of factors, including the semantic and pragmatic information contributed by the temporal connective itself, speakers' world knowledge of the events (e.g. the *real-world* events are necessarily veridical given our world knowledge), and an overall evaluation of the discourse model when both the subordinate and the main events are integrated together. Our results showed that the comprehension system is extremely sensitive to all these different sources of information while performing a rapid incremental updating of the discourse model.

Acknowledgments

We thank Katina Vradelis, Stefan Bartell, and Wesley Jones for their help in stimuli creation and data collection; and Itamar Francez for discussion on the semantics of temporal connectives.

References:

Condoravdi, C. (2010) NPI Licensing in Temporal Clauses. *Natural Language and Linguistic Theory* 28, 877–910.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769–786. doi:10.3758/BF03196772

Giannakidou, Anastasia. 1998. *Polarity Sensitivity as (Non)veridical Dependency*. John Benjamins, Amsterdam-Philadelphia. 281pp.

Giannakidou, Anastasia. 1999. Affective dependencies. *Linguistics and Philosophy* 22: 367-421.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 95–112. doi: 10.1007/BF02289823

Hagoort, P., & Brown, C. M. (1994). Brain responses to lexical ambiguity resolution and parsing. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45-80). Hillsdale,NJ: Erlbaum.

Heinämäki, O. (1972) Before. *Proceedings from the 8th Regional Meeting of the Chicago Linguistic Society*, pp. 139–151. University of Chicago, Chicago.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analy- sis: An electrophysiological differentiation. Journal of Cognitive Neuroscience, 15, 98–110.

Kamp, H. 1981. A theory of truth and semantic representation. In: J.A.G. Groenendijk, T.M.V. Janssen, and M.B.J. Stokhof (eds.), *Formal Methods in the Study of Language. Mathematical Centre Tracts* 135, Pp. 277-322. Amsterdam.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working-memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189–217.

King, J. W., & Kutas, M. (1995). Who did what when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*, 376- 395.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERP's on the processing of unbounded dependencies. Journal of Cognitive Neuroscience, 5(2), 196–214.

Krifka, M. (2010) Before and after without coercion. Comment on the paper by Cleo Condoravdi. *Natural Language and Linguistic Theory*, 28, 911-92

Lascarides, A. and J. Oberlander, [1993] Temporal Connectives in a Discourse Context, in *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL93)*, pp260--268, Utrecht, The Netherlands.

Natsopoulos, D. and Abadzi, H. (1986). Understanding linguistic time sequence and simultaneity: a literature review and some new data. *Journal of Psycholinguistics Research*, 15, 243–273.

Natsopoulos, D., Mentenopoulos, G., Bostantzopoulou, S., Katsarou, Z., Grouios, G. and Logothetis, J. (1991). Understanding of relational time terms before and after in Parkinsonian patients. *Brain and Language*, 40, 444–458.

Münte T.F., Schiltz, K., and Kutas, M. (1998). When temporal terms belie conceptual order. *Nature* 395, 71–73.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Trosberg, A., 1982. Children's comprehension of 'before' and 'after' reinvestigated. *Journal of Child Language*, 9, 381–402.

Ye, Z., Kutas, M., St. George, M., Sereno, M.I., Ling, F. and Münte, T.F. (2012) Rearranging the world: Neural network supporting the processing of temporal connectives. *Neuroimage*, 59, 3662-3667. doi: 10.1016/j.neuroimage.2011.11.039.