

"Before" and "after": investigating the relationship between temporal connectives and chronological ordering using event-related potentials

Stephen Politzer-Ahles<sup>1,2\*</sup>, Ming Xiang<sup>3</sup>, Diogo Almeida<sup>4</sup>

<sup>1</sup>*Faculty of Linguistics, Philology & Phonetics, University of Oxford, United Kingdom*

<sup>2</sup>*NYUAD Institute, New York University Abu Dhabi, United Arab Emirates*

<sup>3</sup>*Linguistics Department, University of Chicago*

<sup>4</sup>*Psychology Program, New York University Abu Dhabi, United Arab Emirates*

*\*Address correspondence to*

Faculty of Linguistics, Philology & Phonetics

University of Oxford

Clarendon Press Centre

Walton Street

Oxford OX1 2HG

United Kingdom

E-mail: [stephen.politzer-ahles@ling-phil.ox.ac.uk](mailto:stephen.politzer-ahles@ling-phil.ox.ac.uk)

## **"Before" and "after": investigating the relationship between temporal connectives and chronological ordering using event-related potentials**

Sentence-initial temporal clauses headed by *before*, as in "Before the scientist submitted the paper, the journal changed its policy", have been shown to elicit sustained negative-going brain potentials compared to maximally similar clauses headed by *after*, as in "After the scientist submitted the paper, the journal changed its policy". Such effects reported thus far may be due to either one of two potential causes: *before* clauses may be more difficult than *after* clauses because they cause the two events in the sentence to be mentioned in an order opposite the order in which they actually occurred, or they may be more difficult because they are ambiguous with regard to whether the event described in the clause actually happened. The present study examined the effect of *before* and *after* clauses on sentence processing in both sentence-initial contexts, like those above, and in sentence-final contexts ("The journal changed its policy before/after the scientist submitted the paper"), where an order-of-mention account of the sustained negativity predicts a negativity for *after* relative to *before*. There was indeed such a reversal, with *before* eliciting more negative brain potentials than *after* in sentence-initial clauses but more positive in sentence-final clauses. The results suggest that the sustained negativity indexes processing costs related to comprehending events that were mentioned out of order.

Keywords: temporal connectives; electroencephalography; event-related potentials; sustained negativity; sentence comprehension

### **1 Introduction**

One of the hallmarks of human language is the ability to talk about events that are displaced in time and/or space from the speaker; this includes past events, events that have not happened yet, and possible events that did not actually happen (Hockett, 1960). Temporal connectives like *before* and *after* pose a special challenge to the language

comprehension system, as they express relationships between multiple events. Since events may have complicated internal structure—for instance, one event may begin after but end before another—which affects the way temporal expressions are used (Anscombe, 1964; Beaver & Condoravdi, 2003; Giannakidou, 1998, 1999; Merchant, 2015), the comprehension of temporal expressions, therefore, requires sophisticated temporal alignment between multiple events.

A well-known phenomenon in the comprehension of temporal connectives is that English sentences beginning with a temporal clause headed by *before* (1a) engender greater processing cost than those beginning with a temporal clause headed by *after* (1b).

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

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

In the seminal study on this phenomenon using event-related brain potentials (ERPs), which provide a measure of neural activity recorded at the scalp with precise temporal accuracy, Münte, Schilz, and Kutas (1998) showed that *before* sentences like (1a), relative to *after* sentences like (1b), elicited a negative-going ERP component over anterior portions of the scalp, which was sustained over the whole sentence. Anterior negativities are often argued to be elicited by stimuli or cognitive tasks which require greater working memory resources (Hagoort, Wassenaar, & Brown, 2003; King & Kutas, 1995; Vos, Gunter, Kolk, & Mulder, 2001; among others). The authors propose that the increased negativity elicited by *before* sentences is related to working memory demands and additional computation associated with having to construct a conceptual model in which the events occur in a different order in which they were presented in the sentence. In other words, (1a) describes a situation in which the first event that happened is the journal's changing its policy, and the second event is the scientist's

submitting her paper; in the sentence, however, these two events are mentioned in the opposite order (non-isomorphic order of mention), which leads to more difficult processing.

A variety of other research paradigms have shown similar costs for *before* sentences relative to *after* sentences. In behavioral experiments, sentences in which the order of mention of two events is different from the conceptual order in which they actually occurred are recalled less accurately (Clark & Clark, 1968), are read more slowly (Mandler, 1986), and are re-enacted less accurately by children in some experiments (Amidon & Carey, 1972; Natsopoulos & Abadzi, 1986; see, however, Clark, 1971, and Crain, 1982). Using ERPs, Nieuwland (2015) finds that downstream predictive processing was inhibited (as evidenced by a reduction in an N400 effect related to a truth-value manipulation) in *before* sentences, suggesting that *before* interfered with comprehension. With functional magnetic resonance imaging (fMRI), Ye and colleagues (Ye, Habets, Jansma, & Münte, 2011; Ye, Kutas, St. George, Sereno, Ling, & Münte, 2012) showed greater hemodynamic activation in the caudate nucleus and left middle frontal gyrus for *before* sentences compared to *after* sentences in healthy adults.

It is possible, however, that the processing cost for *before* clauses in this line of research is not due to non-isomorphic order of mention, but rather to different semantics and pragmatics of the words *before* and *after* themselves. There are several subtle asymmetries between the semantics of *before* and of *after* (see Beaver & Condoravdi, 2003; Heinämäki, 1972; Giannakidou, 1998, 1999), but the most important for present purposes is the licensing of *non-veridical* inferences: *after* presupposes that the temporal clause event happened, and *before* does not. That is to say, the *after* clause in (1b) necessarily means that the scientist did ultimately submit her paper (it presupposes

that the event described in the temporal clause is veridical). On the other hand, the *before* clause in (1a) is ambiguous: it might be the case that the scientist submitted her paper, but it might not, as in (2).

- (2) Before the scientist submitted the paper, she ripped it up and threw it away.

Thus, a *before* clause introduces temporary ambiguity as to whether or not the event described actually happened. This point was also noted by Xiang and colleagues (2014) and Baggio and colleagues (2015), who propose that the sustained ERP negativity observed by Münte and colleagues (1998) may be due not to the difficulty of realizing the conceptual order when it mismatches the order of mention, but rather may be due to ambiguity of the *before* clause and the concomitant working memory costs associated with holding multiple possible readings in working memory until it is possible to decide whether or not the event described in the *before* clause actually occurred. Consistent with this account, Xiang and colleagues (2014) replicated the sustained negative effect with sentences like (1a,b), but also showed that the effect disappeared when participants instead read sentences like (3a,b) in which real-world knowledge makes it clear that the event actually happened.

- (3) a. Before the Second World War broke out, John worked at a small factory.

- b. After the Second World War broke out, John worked at a small factory.

While this finding provides suggestive evidence that the sustained negativity may have been due to ambiguity, some details of the results are surprising. Notably, the sustained negativity in ambiguous sentences—and the corresponding *lack* of sustained negativity in unambiguous sentences—emerged right at the beginning of the sentences; there was no point early in the epoch where unambiguous *before* clauses elicited a

transient negativity. The point at which the unambiguous clauses would have been disambiguated to a veridical reading, however, was generally later in the clause, presumably around the temporal clause verb (for example, until the verb "broke out" was read, (3b) could have had a non-veridical continuation such as "Before the Second World War *caused the extinction of humankind, a peace treaty fortunately was signed*"). The fact that the unambiguous temporal clauses showed no sustained negativity at all, rather than an early emergence and later disappearance of a negativity, suggests that the lack of effect for these clauses may have been due to strategic factors as well as to unambiguity.

At present, therefore, it is difficult to adjudicate between the account of the sustained negativity based on non-isomorphic order of mention and that based on the ambiguity of the event described by *before*, as both accounts make the same predictions for sentences like (1a,b) without real-world disambiguating information. However, these accounts can be straightforwardly tested by examining sentences in which the temporal clause follows rather than precedes the main clause, such as (4a,b), which describe the same situations as (1a,b) but in the opposite order of mention:

(4) a. The journal changed its policy after the scientist submitted the paper.

b. The journal changed its policy before the scientist submitted the paper.

In this case it is the order of mention in the *after* sentence, not the *before* sentence, that is non-isomorphic with the event order. Thus, under the hypothesis that the sustained negativity is based on the incongruence between the conceptual order of the events and their order of mention, these sentences should show the opposite of the effects described above: over the temporal clause (*before/after the scientist submitted the paper*), an increased negative ERP should be observed for *after* clauses compared to *before* clauses.

On the other hand, the hypothesis that the sustained negativity is based on the ambiguity of the *before* clause does not predict such a reversal of the ERP effect. Rather, under such an account, one would make the following predictions. First, it is possible that *before* clauses would still elicit a greater negativity than the *after* clauses. This is because changing the order of mention, as in (4b), does not necessarily eliminate the veridicality ambiguity in the before-clause. For instance, (4b) is still ambiguous as to whether the scientist actually submitted his paper or not. It is also possible, however, that seeing the main clause first helps to reduce the ambiguity (if not completely eliminating it), since a comprehender would have more information to work with when incrementally making veridicality inferences about the before-clauses. In this case the *before* and *after* clauses should pattern similarly to each other. Crucially, in neither case would a larger negativity on the *after* clauses relative to the *before* clauses be predicted.

Thus far, only behavioral experiments have examined sentence-final temporal clauses like (4a,b). Most such studies have found a reversal (in terms of reading times [Mandler, 1986], act-out accuracy [Natsopoulos & Abadzi, 1986, child data], or recall accuracy [Clark & Clark, 1968]) as predicted by non-isomorphic order of mention account: better performance in sentence-final *before* clauses compared to sentence-final *after* clauses. On the other hand, adults in the study by Natsopoulos and Abadzi (1986) showed better performance on *after* than *before* across the board, regardless of the order of mention, consistent with *before*-ambiguity account. Thus, the extant behavioral literature is somewhat equivocal between the two accounts. The present study tests the order-of-mention and ambiguity hypotheses by examining ERPs elicited while participants read sentences with sentence-initial temporal clauses like (1a,b) and sentence-final temporal clauses (4a,b) for comprehension.

## 2 Methods

### 2.1 Participants

Twenty native speakers of English (14 women, mean age = 26, SD = 8.2, range 18-47) were included in the final analysis. All were right-handed as assessed by the Edinburgh Handedness Questionnaire (Oldfield, 1971). All participants provided their informed consent and were paid for their participation, and experimental procedures were approved by the Institutional Review Board of New York University Abu Dhabi. Detailed demographic information about the participants is available in Supplementary File 1. An additional nineteen participants took part in the study but were not included in the final data analysis: ten were removed because of excessive artifact in their data,<sup>1</sup> six for being early bilinguals,<sup>2</sup> one for being left-handed, and two to ensure that the

---

<sup>1</sup> While the proportion of participants removed from data analysis for artifact was high compared to many studies, this is not surprising given that we analyzed a large epoch (see section 2.4, Data acquisition and analysis) and had to exclude trials of data including artifacts anywhere within the relatively long epoch.

<sup>2</sup> Given the very heterogeneous language profile of our participant population in Abu Dhabi, at the outset of the study we recruited anyone who self-reported as a native English speaker because we were worried we would not find sufficient participants if we only used monolinguals. When it became clear later in the data collection process that there would be enough monolingual participants, we decided to exclude bilinguals from the analysis given that the different temporal clause structures in their language (for instance, many of these participants were speakers of languages with head-final temporal clauses, where the equivalent of *before* or *after* would come at the end of the clause) may influence their processing strategy. Nonetheless, exploratory analysis of the dataset with these participants included showed the same pattern of results as that reported below.



same number of participants completed each list of the design (see section 2.3, Procedure).<sup>3</sup>

## 2.2 Materials

The experimental stimuli comprised 154 two-clause sentences of the format shown in (1) and (4). The materials were adapted from Ye et al. (2012) and Xiang et al. (2014). Each item comprised two clauses which were not causally related and did not contain any pronoun-antecedent dependencies across clauses (of the original 160 items, six which were later noticed to include dependencies were excluded from data analysis). The four conditions were created by heading the temporal clause with either *before* or *after*, and by placing the temporal clause either before the main clause (and following it with a comma) or after the main clause (and preceding it with *and*). Thus the experiment followed a 2×2 design: CONNECTIVE (*before* vs. *after*) × STRUCTURE (sentence-initial temporal clause vs. sentence-final temporal clause). The full list of critical stimuli is available in Supplementary File 2. An additional 160 sentences from a separate experiment, including different kinds of *wh*-islands and resumptive pronouns, served as fillers.

## 2.3 Procedure

Participants were seated in an electrically-shielded and sound-attenuated booth, in front of a 59 cm, 1920×1080 pixel LCD monitor. They read the 320 stimulus sentences (in yellow 32-point Courier New font on a black background) word-by-word for

---

<sup>3</sup> Lists were balanced by removing the participants with the lowest number of trials left from the lists that had extra participants. Exploratory analysis of the dataset with these two participants included showed the same pattern of results as that reported below.

comprehension as the electroencephalogram (EEG) was recorded. The experiment was controlled using Presentation (Neurobehavioral Systems). Each trial began with a 64-point fixation cross presented at the center of the screen for 500-800 ms, after which the sentence was presented word by word (for the filler sentences, some short phrases were presented in single chunks). Each word remained on screen for 300 ms (except for the final word of the sentence-initial temporal clauses, which was presented for 500 ms together with a comma, and for the final word of each sentence, which was presented for 800 ms together with a period; these increased durations were used to accommodate for potential end-of-clause wrap-up processes) and followed by a blank screen for 200 ms.

One-third of the items were followed by a comprehension question, which probed various portions of the sentence. For each question, two possible answers were displayed on the screen (the sides were determined randomly at runtime), and participants indicated the correct answer with their right hand using a gamepad. Trials with no comprehension question were simply followed by the message "(press any button to continue)". In either case, the next trial began as soon as the participant pressed a button.

The 320 items were presented in a fully random order after a three-sentence practice. The experiment was divided into five blocks, with 64 sentences per block, and optional break times in between.

Overall, the experimental session (including the completion of consent and demographic forms, applying the EEG cap, the EEG experiment, a working memory test [see Supplementary File 5], and debriefing) took less than 1.5 hours per participant.

## 2.4 Data acquisition and analysis

EEG was continuously sampled (1000 Hz, 0.1-250 Hz analog filter) from 34 Ag/AgCl electrodes (actiCAP, Brain Products) in a 10/20 layout. FCz served as the online reference and AFz as the ground. Up to three bad channels per participant, if present, were interpolated offline, and the continuous data were then re-referenced to the average of both mastoids, high-pass filtered (Hamming windowed sinc FIR filter, 6600 samples filter order, in EEGLAB [Delorme & Makeig, 2004]), and segmented into epochs from -200 ms to +2500 ms relative to the onset of the temporal connective. This epoch window was chosen to encompass the shortest temporal clauses. Trials containing artifact were removed from subsequent analysis based on visual inspection. The cleaned data were then demeaned using the -200 to 0 ms pre-stimulus baseline and subjected to a 30 Hz low-pass filter (Hamming windowed sinc FIR filter, 440 samples filter order, in EEGLAB).

Statistical analysis was carried out using spatiotemporal clustering (Maris & Oostenveld, 2007), implemented in the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011).<sup>4</sup> Compared to traditional analysis of mean ERP amplitudes over pre-defined time windows and channel selections, this method is more neutral to researcher choices, and also addresses the multiple comparisons problem. Spatiotemporal clusters between -200 and +2500 ms with a significant CONNECTIVE×STRUCTURE interaction were identified, using a cluster  $\alpha$  level of 0.3 (based on our *a priori* expectation to observe effects that would be subtle in amplitude but long-lasting). Cluster-level p-values were estimated from 500 random permutations of the data. The CONNECTIVE×STRUCTURE interaction was in fact coded such that a negative test statistic

---

<sup>4</sup> For the sake of comparison with previous studies we also carried out a traditional analysis based on mean amplitudes. This analysis is reported in Supplementary File 4.

would represent a cluster where the simple effect of CONNECTIVE (*before* – *after*) was more negative in sentence-initial clauses than sentence-final clauses, and a positive test statistic would represent a cluster where the effect was more positive.<sup>5</sup>

An additional exploratory analysis was conducted in order to test whether item-wise variation in veridicality bias (i.e., the likelihood that a given sentence would be interpreted as implying that the event described in the *before* clause actually happened) influenced ERPs. This exploratory analysis is documented in more detail in Supplementary File 5.

### 3 Results

#### 3.1 Behavioral

The lowest accuracy score on the comprehension task for any participant was 83.3%, indicating that participants were attending to the stimuli. Mean accuracy was 93.2% for sentence-initial *after* items, 96.0% for sentence-initial *before*, 89.5% for sentence-final *after*, and 88.9% for sentence-final *before*. A generalized (logistic) linear mixed-effects model with fixed effects of CONNECTIVE, STRUCTURE, and their interaction, and crossed random intercepts for participants, items, and lists (Baayen, Davidson, & Bates, 2008) yielded a marginal CONNECTIVE×STRUCTURE interaction in model comparison ( $\chi^2(1) = 0.093$ ). The interaction indicated that accuracy was marginally higher for *before* than *after* sentences when the temporal clause was sentence-initial ( $b = 0.77$ ,  $z = 1.78$ ,  $p = .072$ ) but not when the temporal clause was sentence-final ( $b = -0.09$ ,  $z = -0.30$ ,  $p = .768$ ); or, alternatively, that accuracy was significantly higher for sentences with

---

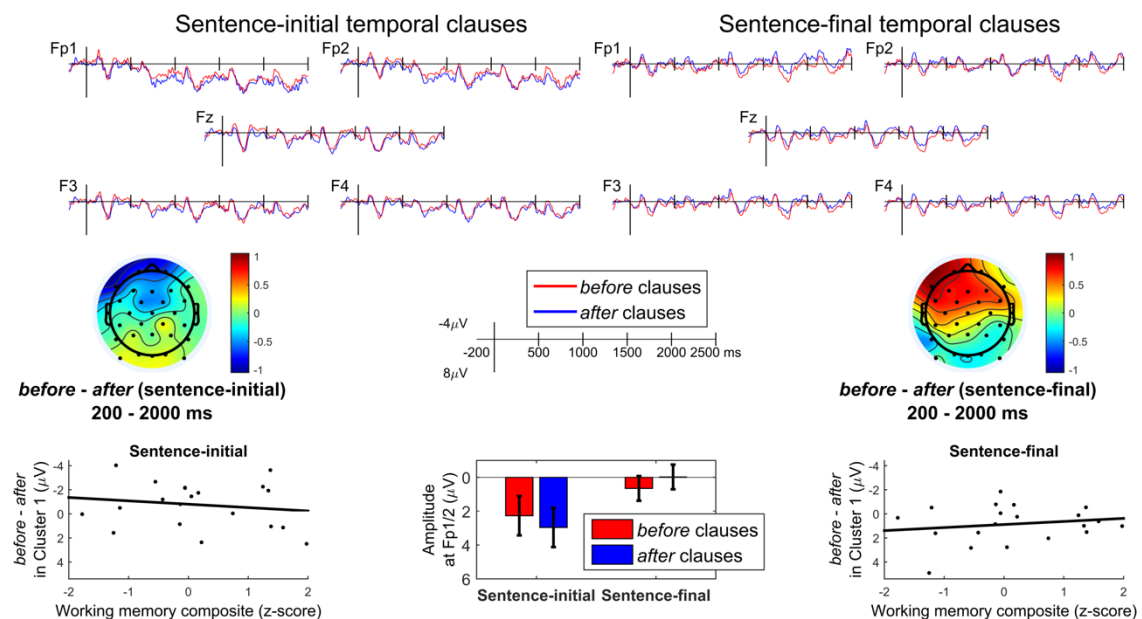
<sup>5</sup> See [http://www.fieldtriptoolbox.org/faq/how\\_can\\_i\\_test\\_an\\_interaction\\_effect\\_using\\_cluster-based\\_permutation\\_tests](http://www.fieldtriptoolbox.org/faq/how_can_i_test_an_interaction_effect_using_cluster-based_permutation_tests) regarding the coding of factorial interactions in FieldTrip. For a similar analysis see Almeida & Poeppel (2013).

sentence-initial than sentence-final temporal clauses when the connective was *before* ( $b = 1.49, z = 3.69, p < .001$ ) but only marginally so when the connective was *after* ( $b = 0.63, z = 1.90, p = .057$ ).

### 3.2 ERPs

After artifact exclusion, the minimum number of trials retained in any cell was 14 (see Supplementary File 1). A generalized linear mixed model showed that significantly more trials were retained in sentence-final temporal clause configurations than sentence-initial clause configurations ( $\chi^2(1) = 22.62, p < .001$ ), but there was no difference based on CONNECTIVE and no interaction ( $ps > .615$ ).

The ERPs for each condition at a selection of frontal electrodes, along with topographic maps for the mean amplitude across most of the epoch, are shown in Figure 1; the ERP averages are available in Supplementary File 3. The figure suggests that in sentence-initial position, clauses with *before* elicited a subtle but sustained anterior negativity relative to clauses with *after*, whereas in sentence-final position, it is clauses with *after* that elicit a negativity relative to clauses with *before*. Statistical analysis confirmed these observations.

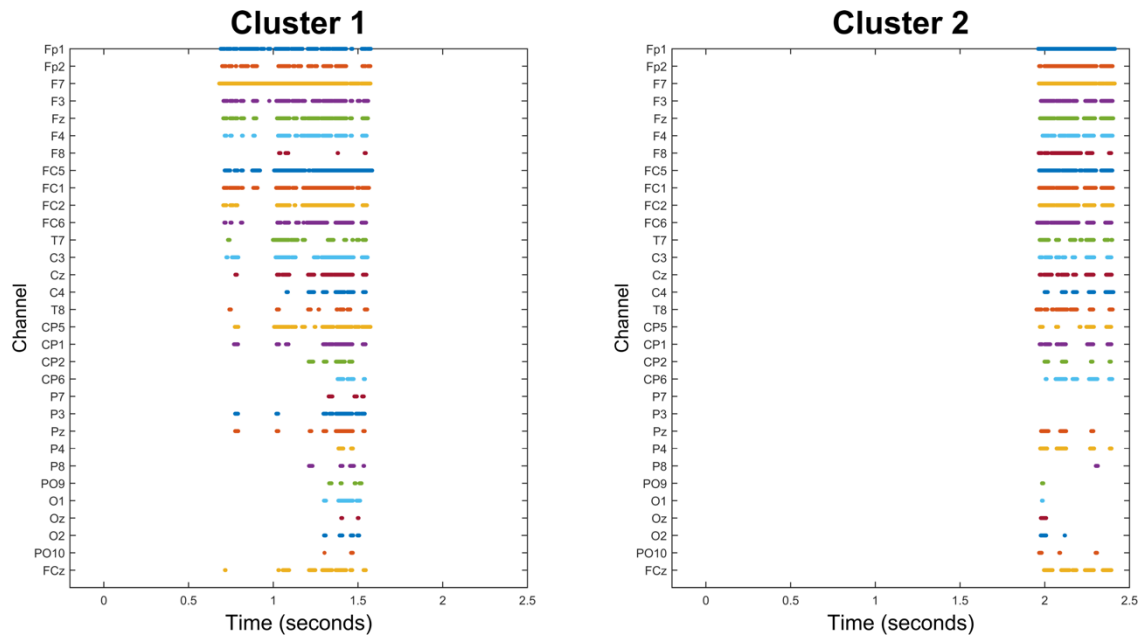


**Figure 1.** ERPs at frontal electrodes (top portion) for the sentence-initial temporal clauses (left) and sentence-final temporal clauses (right). Topographic maps of the *before* – *after* difference averaged over the 200-2000 ms time window are shown below, as well as barplots of this difference at the front-most electrode region. The bottom left and right portion of the figure shows the correlation between working memory scores and ERP effect sizes for both the sentence-initial temporal clauses (left) and sentence-final temporal clauses (right); see Supplementary File 5 for description and discussion of the working memory data.

The cluster analysis for the CONNECTIVE×STRUCTURE interaction yielded a marginal negativity ( $p = .084$ ) driven by a cluster with the spatiotemporal distribution illustrated in the raster plot on the left side of Figure 2; i.e., it extended from about 700 to about 1600 ms in the frontal channels, was more sustained in the left channels than the right channels, and emerged in centro-posterior channels only towards the end of this time window. Averaging together the amplitudes of all <channel,time> samples within this cluster and conducting pairwise  $t$ -tests on the averages revealed that the ERPs elicited by sentence-initial *before* clauses were more negative than those elicited by sentence-initial *after* clauses ( $t(19) = -2.22$ , 95% CI =  $-1.76 \dots -0.05$ ,  $p = .039$ ), whereas sentence-final *before* clauses were marginally more *positive* than sentence-final *after* clauses ( $t(19) = 1.96$ , 95% CI =  $-0.05 \dots 1.52$ ,  $p = .065$ ).<sup>6</sup>

---

<sup>6</sup> Note that this is not a non-independent analysis (Baker, Hutchinson, & Kanwisher, 2007), as the follow-up analysis tested simple effects, rather than the interaction test which was used as the basis for identifying the cluster. The purpose of the follow-up tests was not to reiterate the significance of the interaction, but to further clarify the nature of the interaction (e.g.,



**Figure 2.** Raster plots showing the spatiotemporal extents of the most significant interaction cluster ( $p = .084$ , left side) and second-most significant interaction cluster ( $p = .133$ , right side). Each row represents a channel, and each colored dot along that row represents a timepoint during which that channel was included in the cluster.

There was another negative trend ( $p = .133$ ) due to a cluster that emerged later, as illustrated in Figure 2. In this cluster, sentence-initial *before* clauses were marginally more negative than sentence-initial *after* clauses ( $t(19) = -1.77$ , 95% CI =  $-1.64 \dots 0.14$ ,  $p = .092$ ) and sentence-final *before* clauses significantly more positive than sentence-final *after* clauses ( $t(19) = 2.44$ , 95% CI =  $0.12 \dots 1.57$ ,  $p = .025$ ).

---

while in our dataset the interaction emerged because the *before-after* effect was negative in sentence-initial clauses and positive in sentence-final clauses, it could have been the case that both effects were negative and the sentence-initial effect was simply more negative; it also could have been the case that neither simple effect was significantly different than zero even though they were different from one another).

There were no other noteworthy trends in either direction ( $ps > .262$ ). There were also no significant main effects of CONNECTIVE ( $ps > .487$ ). There were significant main effects of STRUCTURE (initial  $>$  final  $p = .026$ , initial  $<$  final  $p = .006$ ), which not of interest because they involve direct comparison across clauses at different portions of the sentence.

#### 4 Discussion

The present study examined why sentence-initial *before* clauses are more difficult to process than sentence-initial *after* clauses (in, e.g., "Before/after the scientist submitted the article, the journal changed its criteria"), as indexed by an enhanced sustained negative ERP over frontal scalp locations. The traditional account for this effect is that the *before* clauses cause the events in the sentence to be mentioned in a different order than the order they actually occurred in (Münste et al., 1998). We compared this to an alternative hypothesis which attributes the difficulty observed in *before* clauses to interpretational ambiguity with respect to whether the event described by *before* actually happened (Baggio et al., 2015; Xiang et al., 2014). While these two accounts make the same predictions for temporal clauses in sentence-initial position, they make distinct predictions for clauses in sentence-final position (e.g., "The journal changed its criteria before/after the scientist submitted the article"): the order-of-mention account predicts the effect to reverse, with *after* clauses becoming more difficult than *before* clauses, whereas the account based on interpretational ambiguity does not. In the present experiment, the first to use the ERP method to investigate the processing of temporal connectives in both sentence-initial and sentence-final position, we indeed observed a reversal of the ERP effect: in sentence-initial position, *before* clauses elicited more negative ERPs than *after*, replicating previous findings, whereas in sentence-final position it was *after* that elicited more negative ERPs than *before*. This finding provides



support for the traditional order-of-mention account, and suggests that the comprehension of temporal expressions triggers increased processing cost when the mapping between order of linguistic mention and the actual order of events is non-isomorphic.

Although our results provide new evidence to support the order of mention account, they do not completely disconfirm the possibility that inferential ambiguity influences the online comprehension of temporal clauses. The current result is in fact compatible with an account that allows both the order of mention and inferential ambiguity to affect online comprehension. In particular, since inferential ambiguity may have been reduced in some sentence-final temporal clauses, the effect of order would have been more salient in the present study.

In summary, the present study showed direct ERP evidence that conflict between the order in which events are mentioned in a linguistic expression and the order in which the events actually occurred in the world contributes to the processing costs that are observed in the comprehension of temporal clauses. Open questions remain regarding the nature of the cognitive functions underlying this difficulty (e.g., whether temporal clauses that occur out of order trigger working memory operations or other kinds of operations) and the role played by the different semantics of *before* and *after* (especially the ambiguous veridicality of events in a *before* clause). Nonetheless, the current results demonstrate that language comprehenders construct conceptual models of events online as a sentence is unfolding, and that the order of mention of events in a linguistic expression can help or hinder the mapping between a linguistic model and a conceptual model of the world.

## Acknowledgements

This research was supported in part by grant G1001 from the NYUAD Institute, New York University Abu Dhabi.

## References

- Almeida, D., & Poeppel, D. (2013). Word-specific repetition effects revealed by MEG and the implications for lexical access. *Brain and Language*, 127, 497-509.
- Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Verbal Learning and Verbal Behavior*, 11, 417-423.
- Anscombe, G. (1964). Before and after. *The Philosophical Review*, 74, 3-24.
- Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- Baggio, G., van Lambalgen, M., & Hagoort, P. (2015). Logic as Marr's computational level: four case studies. *Topics in Cognitive Science*, 7, 287-298.
- Baker, C., Hutchinson, T., Kanwisher, N. (2007) Does the fusiform face area contain highly selective subregions for nonfaces? *Nature Neuroscience*, 10, 3-4.
- Beaver, D., & Condoravdi, C. (2003). A uniform analysis of *before* and *after*. *Semantics and Linguistic Theory*, 8, 37-54.
- Clark, E. (1971). On the acquisition of the meaning of *before* and *after*. *Journal of Verbal Learning and Verbal Behavior*, 10, 266-275.
- Clark, H., & Clark, E. (1968). Semantic distinctions and memory for complex sentences. *Quarterly Journal of Experimental Psychology*, 20, 129-138.
- Crain, S. (1982). Temporal terms: mastery by age five. *Papers and Reports on Child Language Development*, 21, 33-38.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, 134, 9-21.
- 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.

- Hagoort, P., Wassenaar, M., & Brown, C. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research*, 16, 38-50.
- Heinämäki, O. (1971). Before. *Proceedings from the 8th Regional Meeting of the Chicago Linguistic Society*, 139-151. Chicago: University of Chicago.
- Hockett, C. (1960). The origin of speech. *Scientific American*, 203, 88-111.
- Mandler, J. (1986). On the comprehension of temporal order. *Language and Cognitive Processes*, 1, 309-320.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, 177-190.
- Merchant, J. (2015). Individual anchors for tenses: How Keats learned to read before Shakespeare. *Linguistic Analysis*, 39, 3-4: 415-421
- Münté, T., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, 395, 71-73.
- Natsopoulos, D., & Abadzi, H. (1986). Understanding linguistic time sequence and simultaneity: a literature review and some new data. *Journal of Psycholinguistic Research*, 15, 243-273.
- Nieuwland, M. (2015). The truth before and after: brain potentials reveal automatic activation of event-knowledge during sentence comprehension. *Journal of Cognitive Neuroscience*, 27, 2215-2228.
- Oldfield, R. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9, 97-113.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 156869.
- Vos, S., Gunter, T., Kolk, H., & Mulder, G. (2001). Working memory constraints on syntactic processing: an electrophysiological investigation. *Psychophysiology*, 38, 41-63.
- Waters, G., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, 49A, 51-79.
- Xiang, M., Hanink, E., & Vegh, G. (2014). Before and after—processing presuppositions in discourse. *Poster presented at the Sixth Annual Meeting of the Society for the Neurobiology of Language*.
- Ye, Z., Habets, B., Jansma, B., & Münte, T. (2011). Neural basis of linearization in speech production. *Journal of Cognitive Neuroscience*, 23, 3694-3702.

Ye, Z., Kutas, M., St. George, M., Sereno, M., Ling, F., & Münte, T. (2012). Rearranging the world: neural network supporting the processing of temporal connectives. *NeuroImage*, 59, 3662-3667.

## Supplementary file information

**Supplementary file 1.** Demographic information for participants (.xls), numbers of trials left per condition for each participant in the ERP analysis, and working memory recall and accuracy scores.

**Supplementary file 2.** Critical stimuli (.xls). Highlighted rows indicate items that were later removed from the analysis. Veridicality bias ratings for each item are also given; for a description of how these were collected see Supplementary File 5.

**Supplementary file 3.** ERP averages (.mat). The file contains objects giving the channel location information (a structure), time in milliseconds corresponding to each sample (a vector of integers), the identifiers for participants (a cell array of strings), and two structures called 'subject\_averages' and 'item\_averages' each containing the ERPs averaged by subjects or by items, respectively. Each 'averages' structure has a separate field for each condition, and within that a field for each participant; the ERP for a given condition/participant combination is a Channels  $\times$  Samples matrix. Note that the item averages for the clause-initial *before* condition contain missing values for item 10, which contributed no data to this condition (for each participant who saw this item in this condition, there was artifact in the corresponding trial). These files are provided for researchers interested in verifying the claims of this paper through reanalysis of the data; researchers interested in using the data for other purposes (such as to test new research questions) should obtain prior agreement from the authors (per section 8.14 of the American Psychological Association's *Ethical Principles of Psychologists and Code of Conduct* <http://www.apa.org/ethics/code/index.aspx?item=11#814>).

**Supplementary file 4.** Summary of statistical results based on time-window mean amplitudes at predefined regions of interest, as opposed to spatiotemporal clustering (.pdf).

**Supplementary file 5.** Description of additional analyses involving between-participant variation in working memory and between-item variation in veridicality bias (.pdf).