Switching modalities in a sentence verification task: ERP evidence for embodied language processing

Lea A. Hald[1]*, Julie-Ann Marshall[1,2], Dirk P. Janssen[3] and Alan Garnham[4]

[1] Department of Applied Social Sciences, Canterbury Christ Church University, Canterbury, UK
[2] Health Care of Older People, William Harvey Hospital, Ashford, UK
[3] Academy for Digital Entertainment, NHTV Breda University of Applied Sciences, Breda, Netherlands
[4] School of Psychology, University of Sussex, Brighton, UK

In an event related potential (ERP) experiment using written language materials only, we investigated a potential modulation of the N400 by the modality switch effect. The modality switch effect occurs when a first sentence, describing a fact grounded in one modality, is followed by a second sentence describing a second fact grounded in a different modality. For example, "A cellar is *dark*" (visual), was preceded by either another visual property "Ham is pink" or by a tactile property "A mitten is soft." We also investigated whether the modality switch effect occurs for false sentences ("A cellar is *light*"). We found that, for true sentences, the ERP at the critical word "*dark*" elicited a significantly greater frontal, early N400-like effect (270–370 ms) when there was a modality mismatch than when there was a modality-match. This pattern was not found for the critical word "*light*" in false sentences. Results similar to the frontal negativity were obtained in a late time window (500–700 ms). The obtained ERP effect is similar to one previously obtained for pictures. We conclude that in this paradigm we obtained fast access to conceptual properties for modality-matched pairs, which leads to embodiment effects similar to those previously obtained with pictorial stimuli.

Keywords: ERP, N400, embodiment, language processing, veracity, modality, modality switch effect

INTRODUCTION

The idea that our conceptual system is grounded in modality-specific or embodied simulations has received support from many different areas of research including psychology, neuroscience, cognitive modeling, and philosophy (Lakoff and Johnson, 1999; Gibbs, 2005; Pecher and Zwaan, 2005, for reviews). The suggestion that modality-specific simulation also affects language processing has been put forward by a number of authors (Glenberg, 1997; Barsalou, 1999; Glenberg and Robertson, 1999, 2000; Zwaan, 2004; Zwaan and Madden, 2005). For example, Barsalou's (1999) theory of perceptual symbol systems suggests that modality-specific simulations arise from perceptual states and that these (simulated) states underlie the representation of concepts. Hence, all conceptual symbols are grounded in modality-specific states. Linguistic symbols develop alongside the perceptual symbols that they are linked to so that when we use or encounter words, we simulate the perceptual states that are linked to the linguistic information. Such a source of perceptual state simulations is called a simulator by Barsalou (1999).

Zwaan and Madden (2005) similarly assume language is grounded in perception and action via something akin to Barsalou's (1999) perceptual symbols. However, they focus specifically on how language guides the simulators. They assume that what we simulate is based on attentional frames (Langacker, 2001). In particular, within one attentional frame we construct a "construal": a simulation that includes time, spatial information, perspective, and a focal and background entity (for details see Zwaan and Madden, 2005). Furthermore, during construal, information from previous construals forms the context with which we integrate the information

from the current construal. This is how we understand connected discourse (Zwaan, 2004). For Glenberg (1997; Glenberg and Robertson, 1999, 2000) the key issue is that we use the perceptual symbols to derive affordances, in the sense of Gibson (1986), for the specific situation. Understanding a sentence is a result of meshing the affordances, which is guided by the syntax of the sentence.

Evidence for the modal grounding of conceptual and linguistic representations has been found using a variety of techniques and tasks. Only a few key findings relative to the current experiment will be reviewed here. Goldberg et al. (2006) measured fMRI BOLD responses while participants did a blocked property verification task. Participants had to press the button for each word that had the property "*green*" (visual), "*soft*" (tactile), "*loud*" (auditory), or "*sweet*" (gustatory). The results for visual and tactile decisions showed increased activation in visual and somatosensory cortex when compared to control, which supports the notion of modal grounding.

Using a behavioral measure and the same paradigm, Pecher et al. (2003) established that there is a cost to switching modalities. They presented participants with short sentences that consisted of a concept followed by a modal property (they used audition, vision, taste, smell, touch, and action). For example, after reading "*blender can be loud,*" participants were asked to decide whether "*loud*" is a typical property of "*blender.*" Crucially, half of the experimental trials were preceded by a trial of the same modality (matched modality, "*leaves can be rustling*" – "*blender can be loud*") while the other half were preceded by a trial of a different (mismatched) modality (e.g., "*cranberries can be tart*" – "*blender can be loud*"). Participants were able to verify the property of the concepts faster

and more accurately in matched modality trials than in mismatched modality trials. Similar modality switch effects have been found in other studies across both conceptual and perceptual processing tasks (e.g., Spence et al., 2001; Marques, 2006; Vermeulen et al., 2007; Van Dantzig et al., 2008).

If the mental simulations that are required for understanding involve the premotor areas, keeping these areas otherwise involved should interfere with language comprehension. This has indeed been demonstrated, for example by Zwaan and Taylor (2006), who found that reading about an action which involves clockwise turning (e.g., increasing the volume on a radio) interfered with the action of turning a knob counterclockwise. More abstractly, Glenberg and Kaschak (2002) showed that reading a sentence which involves transferring an object or information away from the participant ("You told Liz the story") interfered with that participant pressing a response button which was located toward their body as compared to a button which was located away from their body.

Evidence from tasks not involving large physical movements comes from sentence–picture verification tasks. For example, a sentence such as "*John pounded the nail into the floor*" was followed by a picture of a nail (Stanfield and Zwaan, 2001). Response times were faster when the picture matched the orientation implied in the sentence (vertical) compared to when there was a mismatch in orientation (see also Zwaan et al., 2002).

What is striking about the behavioral studies described here is the number of innovative tasks and procedures that were created in order to show that concepts and language are grounded in bodily states. While these and other sets of studies form a convincing body of literature, one might question how effects related to embodied cognition might be evident in other tasks that are more standardly used within language comprehension research. If simulating linguistic or conceptual material in terms of our bodily states is the norm, we should see evidence of it in any standard task if properly designed and analyzed. This is important as it could be argued that the use of tasks involving movement and pictures encouraged participants to use an imagery based strategy (e.g., Glucksberg et al., 1973), which would make the embodiment results specific to the tasks that are used in this field. Of course, exceptions already exist: Results from neuroimaging studies where participants read either single words or sentences referring to bodily actions support the embodied view by showing increased activation in the premotor and sometimes the primary motor areas of the cortex (for example, Hauk et al., 2004; Boulenger et al., 2009). Recent findings using the sentence–picture verification task also suggest that the results are not due to the use of imagery as a strategy (Pecher et al., 2009) and the study by Pecher et al. (2003) did use solely linguistic stimuli, albeit in a slightly unnatural task. If embodied simulation is a part of everyday language comprehension, we should be able to find evidence for it using the standard language comprehension techniques that do not involve pictures or movements. In the current study we will therefore use the well-studied paradigm of the sentence verification task (Meyer, 1970). Before discussing results related to this task, we will quickly outline our experiment to frame the discussion below.

The materials of the current study were adapted from the design used by Pecher et al. (2003) to a sentence verification task. We drew our materials from items that have previously been rated as having either a salient visual or a salient tactile property (Pecher et al., 2003; Van Dantzig et al., 2008; Lynott and Connell, 2009; Van Dantzig and Pecher, submitted; see Materials and Methods for details). From this set of concepts with salient modality features, we created true statements "*A cellar is dark*" (visual). As in Pecher et al. (2003), sentences were presented one by one and for the participants, appeared to be unrelated. However, the critical manipulation was that sentences that followed each other were either matched in the salient modality (e.g., visual–visual, "*Ham is pink*" – "*A cellar is dark*") or mismatched in the salient modality (e.g., tactile–visual, "*A mitten is soft*" – "*A cellar is dark*"). We crossed modality with veracity by making half of the experimental target sentences false, while maintaining the same modality information ("*A cellar is light*").

We will now first review the links between property verification and sentence verification and then discuss the previous findings on veracity. As mentioned above, Pecher et al. (2003) asked participants to perform a conceptual-property verification task for statements such as "*blender/can be/loud*" (slashes indicate line breaks on the computer monitor). The participants were asked to verify that the property (always shown on the third line) is "usually true" of the concept (always shown on the first line) and had to respond with a true or false response. In the current study, it was decided to change the task to sentence verification. Sentence verification is a similar but more general task and has been used extensively in early sentence processing literature (for a review see Carpenter and Just, 1975) and in event related potential (ERP) experiments (e.g., Fischler et al., 1983). In this task, sentences are presented and subjects respond with a true or false judgment at the end of the sentence. Some items are almost identical between tasks ("*A blender can be loud*"), others can only be used in the sentence verification ("*A baby drinks milk*"). In our version of sentence verification, the words are presented one by one in the middle of the screen, which leads to a relatively natural reading experience, while avoiding eye movements. Using the sentence verification task, the typical finding is that false sentences take longer to verify than true sentences (e.g., Fischler et al., 1983).

The majority of the response time literature on veracity investigates the time to decide whether a sentence is a true or false representation of a corresponding picture ("*The dots are red*" with a picture showing either red or blue dots). In this situation, true sentences have been consistently shown to be verified faster than false sentences (for example, Trabasso et al., 1971; Clark and Chase, 1972; Wason, 1980). The primary explanation for this is that readers match the color red to the color of the dots. When this is congruent, readers are facilitated; when the colors are incongruent there is a slow down (Carpenter and Just, 1975; see also Fischler et al., 1983).

In this paper, we will try to obtain further empirical evidence for an embodied approach and we will discuss how an embodied language comprehension system can explain the current and past findings through the process of simulation. In an embodied view, determining the veracity of a statement depends on the outcome of a simulation and the comprehension process should be modulated by direct or indirect effects of simulation.

The reaction time effects for modality switch are quite subtle so we decided to use a more sensitive technique for this study: To explore the processing dynamics of modality switching, veracity,

and their interaction, we will look at the presence and significance of modulations in the ERP. If embodied simulation is an automatic process that occurs when we understand language, evidence of modality switching, veracity effects, and their interaction should be evident in ERPs. Predictions relative to modality switching are discussed below followed by veracity predictions.

One possible prediction of the effect of modality switching would be a modulation of the N400 effect. Although often incorrectly thought of as an increased negativity that occurs only to semantic anomalies (e.g., "*He spread the warm bread with socks/butter*"; Kutas and Hillyard, 1980), a large body of research suggests that semantic anomalies are nether necessary or sufficient to elicit an N400 effect (see Kutas et al., 2006, for review). Instead results show that a (small) N400 occurs as a response to each meaningful word as part of normal processing (Van Petten, 1995). The amplitude of the N400 is sensitive to many different semantic and linguistic factors [for example, Cloze probability (Taylor, 1953), word frequency, word class, and discourse context]. Furthermore, relative to the veracity, a consistently larger amplitude N400 is seen for words that change the veracity of a single sentence (at the critical word, here shown in bold for "a ham is **blue**" versus "a ham is **pink**"; Fischler et al., 1983; Hagoort et al., 2004).

Given this range of meaning effects modulating the N400 and the behavioral findings that switching modalities leads to a processing cost (Pecher et al., 2003), a reasonable expectation is that the N400 effect could be modulated by modality switching. We think it is a priori unlikely that a modality switch would trigger a sizeable N400 by itself, as "*a cellar is dark*" is usually a true, semantically coherent statement, even after a tactile context like "*a mitten is soft.*" However, the N400 is sensitive to the *integration* of incoming semantic information into the ongoing representation: Assuming that the ongoing representation is indeed embodied, a switch in the modality may lead to an earlier effect than the N400 (because modality switching should occur before integration), and the modality switch may modulate (enhance or suppress) the N400 itself. The effect of modality switch on the N400 may not be linear, as is known to be the case for word frequency and context (Van Petten and Kutas, 1990). Specifically, one may predict that a match in modality may lead to easier simulation and therefore a reduction or absence of the N400 for integration. Alternatively, there could be an ERP effect that occurs earlier than the N400, which is specifically indicative of the simulation itself.

A second question addressed in our study is what happens when the target sentence is false or commonly false ("*a cellar is light*"). We know that the veracity of the sentence can modulate the N400 (Fischler et al., 1983, with sentence verification task; Hagoort et al., 2004, with no task given). Similar N400 modulation results were found using a task where participants were required to determine whether a probe word was related conceptually to the precious sentence (for example, "flute" following "Mozart was a musical child prodigy"; Nieuwland and Kuperberg, 2008). To better understand the effects of modality switching, we investigate whether there is an interaction between effects for the veracity of the sentence and effects of the modality switch. Barsalou (1999) suggests that when a false sentence is read the simulation fails, which means that the meaning of the sentence cannot be successfully mapped onto reality. After a simulation fails presumably a new simulation is carried out

that is grounded in the failed simulation (Barsalou, 1999, p. 601). However, as in the response time literature on veracity discussed above, Barsalou (1999) discusses false sentences in a context where one compares the sentence to a situation (or picture) immediately in front of people, not what would happen when something is false based on background knowledge. Nonetheless, if false sentences lead to a failure of simulation, this may lead to a different ERP modulation based on the point at which the simulation fails. Considering false sentences take longer to verify than true sentences, one might expect the ERP modulation relative to modality switching to occur later in the time course of processing.

MATERIALS AND METHODS
PARTICIPANTS
Sixteen native speakers of English recruited from Canterbury Christ Church University and the University of Kent participated in this experiment, 10 of whom were included in the final analysis (eight females; aged 18–22, mean = 19.7). They were paid a small fee for their participation. All participants had normal or corrected-to-normal vision, and normal hearing and all were right handed. None of the participants had any neurological impairment and none of them had participated in the pretests (see below). The six participants (37.5%) who were excluded from the final analysis were rejected for the following reasons: excessive artifacts (eye-movements, excessive noise from muscle tension, two participants, see EEG Recording and Analysis below for details), technical problems with recording (one participant), reaction time errors over 25% (two participants), and non-native English speaker (one participant). Ethical approval for the ERP study and the pretest was obtained from the Canterbury Christ Church University Faculty Research Ethics Committee, which follows the British Psychological Society guidelines for ethics on human subject testing. All participants signed a consent form prior to participating in the ERP experiment and the pretests.

STIMULUS MATERIAL AND DESIGN
The experimental materials comprised 160 pairs of sentences. Each pair consisted of a first sentence, which we will call the modality context sentence followed by a second sentence, the target sentence. The modality context sentences were always semantically correct, true statements which either described a salient tactile property (tactile context) of an object or a salient visual property (visual context) of an object. We selected a subset of the items that have been previously rated as having one modality that was clearly dominant in people's perception of that item (ratings from Pecher et al., 2003; Van Dantzig et al., 2008; Lynott and Connell, 2009; Van Dantzig and Pecher, submitted). The target sentence either matched the modality of the modality context sentence or mismatched. Additionally, the target sentence could either be true or false. False versions of the target sentences were created by using a word that was rated in a pretest to be the opposite of the salient feature of the object. For example, for "*A cellar is dark*" the word "*light*" was independently rated as the opposite of dark and it was used to create the false version. The false target sentences always contained a property in the same modality as the true target sentences. By using opposites we can keep the format of the true and false sentences

identical, but one may wonder what the effect of an opposite is. Furthermore, many of the studies looking at veracity have used opposites to create false sentences. For example, Nieuwland and Kuperberg (2008) looked at true and false sentences where all false sentences were created by using opposites (for example, "*With proper equipment, scuba-diving is very safe/dangerous…*") and found a typical N400 effect for false sentences compared to true (see also Hald et al., 2005 for similar use of opposites for creating false sentences.

The conditions modality-match and veracity of the target sentence were fully crossed, with 40 pairs in each of the four cells. Half of these 40 target sentences were visual, the other half tactile (see **Table 1** for example materials). Eighty false–false filler pairs were added to balance the number of true and false targets. The filler pairs also contained strongly modality related properties in half of the sentences, using tactile, visual, auditory, and gustatory modalities. The other half of the fillers were not based on modality-specific information but instead contained highly related words, while conveying false information (e.g., *A ball is refereed*; see Pecher et al., 2003, for similar use of semantically related filler items).

The critical words were matched across conditions on the following criteria: (i) word log (lemma) frequency (true–matched modality: 2.32; true–mismatched modality: 2.32; false–matched modality: 2.37; false–mismatched modality: 2.37; from Baayen et al., 1993); (ii) word length (true–matched modality: 4.8 letters; true–mismatched modality: 4.8 letters; false–matched modality: 4.5 letters; false–mismatched modality: 4.5 letters), (iii) word class (all adjectives). None of the critical words was over 12 letters in length.

The 240 pairs of sentences were presented in a pseudorandomized order specific to each participant (created using the program Mix, Van Casteren and Davis, 2006) using a fully within-participants design. The use of within-participants manipulation kept the design similar to that of Pecher et al. (2003), where matched versus mismatched modality was manipulated within-participants. Furthermore, in previous ERP sentence verification experiments a within-participants design was also utilized (Fischler et al., 1983; see also Hald et al., 2005, for a direct comparison of within and between-participants design using a sentence verification task).

Table 1 | Example materials for tactile and visual modality.

Veracity	Modality-match	Modality context	Target sentence
TACTILE TARGET SENTENCE EXAMPLE			
True	Mismatched	A leopard is spotted.	A peach is **soft**.
	Matched	An iron is hot.	A peach is **soft**.
False	Mismatched	A leopard is spotted.	A peach is **hard**.
	Matched	An iron is hot.	A peach is **hard**.
VISUAL TARGET SENTENCE EXAMPLE			
True	Mismatched	A mitten is soft.	A cellar is **dark**.
	Matched	Ham is pink.	A cellar is **dark**.
False	Mismatched	A mitten is soft.	A cellar is **light**.
	Matched	Ham is pink.	A cellar is **light**.

Critical words are shown here in bold for clarification.

PRETEST

In order to create false versions of the target sentences that maintained the same modality information, we decided to replace the adjectives with their opposites. For example, if the target sentence used the critical word "*dark,*" we had people rate on a 7-point scale (7 = strongly agree): "*The opposite of dark is light.*"

These opposites do not form anomalous sentences of the kind that were used for the first N400 experiments (Kutas and Hillyard, 1980; Kutas et al., 2006). However, previous research has shown that people can show N400 effects to sentences that are at odds with their basic world knowledge (Hagoort et al., 2004; Hald et al., 2007). By using opposites, we were able to construct an experiment without anomalous sentences and we were able to use properties from tactile and visual modalities for all experimental items.

We tested a total of 52 different candidate opposites to arrive at the final set of 40. In addition to looking at possible opposites of all critical words, we also included fillers of two different types in this pretest to make sure participants were using the full scale in their ratings. Twenty fillers had properties that are difficult to assign an opposite to, for example "*The opposite of checkered is striped.*" The other 20 fillers were based on words that are related but clearly not opposites "*The opposite of clean is polished.*"

For the visual modality, target words were often color terms (37.5% of the time). Although in one technical sense colors do have opposites (complementary colors), these opposites may not be conceived as such by ordinary language users in the same way as terms such as "*dark*" and "*light.*" For that reason we tested all color word opposites (such as "*Black is the opposite of white*") separately, in a list with fillers that were also all color words. We encouraged the participants to use the full scale by including fillers that were related but clearly not opposites ("*Magenta is the opposite of violet*") and fillers that were difficult to judge ("*Black is the opposite of fuchsia*"). For the non-color pretest 27 native English speakers (eight males; mean age = 31) and for the color terms 37 participants (11 males; mean age = 31) completed the ratings online using SurveyMonkey[1]. We selected the words that were rated most highly as opposite as the adjectives for the false condition. The mean rating for the non-color list was 5.75 (SD = 0.52) and for the color list 4.61 (SD = 0.74). Although the color words were rated lower (less opposite) than the non-color words, the key issue for the sentence verification task is that using these words makes the sentences false. Thus, we had a set of clearly false statements that retained the same modality as the true statements, and had very similar content.

PROCEDURE FOR THE ERP STUDY

Participants were asked to fill out a questionnaire about their language and basic health background. Additionally, participants filled out a handedness questionnaire (Oldfield, 1971) and signed a consent form. Participants were tested individually in quiet room, seated in a comfortable chair approximately 70 cm away from a computer monitor. Participants were asked to read the sentences for comprehension and decide whether each sentence was true or false. They were also asked to try not to move or blink during the presentation of the sentences on the computer screen. No other tasks were imposed.

---

[1]www.surveymonkey.com

The experimental stimuli were presented using E-Prime 2.0 (Schneider et al., 2002). The experimental session began with a practice block of 10 sentences, which were similar in nature to the experimental items. At the end of the practice block the participant had a chance to ask any questions they had about the task. The remaining sentences were split into six blocks lasting approximately 12 min each. A short break followed each block. Each block began with two filler items, which were similar in nature to the experimental items. These filler items were included to minimize loss of data due to artifacts after beginning a new block.

Each trial began with a fixation ("+++") displayed for 1 s in the middle of the computer screen. The participants were told they could blink their eyes during the fixation display, but to be prepared for the next sentence. After a variable time delay (randomly varying across trials from 300 to 450 ms), the sentence was presented word by word in white lowercase letters (Courier New, 18-point font) against a black background. The first word and any proper noun were capitalized and the final word of each sentence was followed by a period. Words were presented for 200 ms with a stimulus-onset asynchrony of 500 ms. Following the final word, the screen remained blank for 1 s, after which three question marks appeared, along with the text "1:true" and "5:false." Participants needed to press either "1" or "5" on the number keypad of a keyboard to indicate whether the sentence was true or false (half of the time, the numbers were reversed). If they responded incorrectly, "Wrong Answer" was displayed and if they took more than 3000 ms, "Too slow" was shown. Exactly the same presentation was used for context and target sentences, so that participants were not aware that sentences were presented in pairs.

Following the experiment, the participants were debriefed and a short questionnaire was given to determine if they were at all aware of the purpose of the experiment.

EEG RECORDING AND ANALYSIS

The EEG was recorded from a 64-channel WaveGuard Cap using small sintered Ag/AgCl electrodes connected to an ANT amplifier (ANT, Enschede, Netherlands). An average reference was used. Electrodes were placed according to the 10–20 standard system of the American Electroencephalographic Society over midline sites at Fpz, Fz, FCz, Cz, CPz, Pz, POz, and Oz locations, along with lateral pairs of electrodes over standard sites on frontal (Fp1, Fp2, AF3, AF4, AF7, AF8, F1, F2, F3, F4, F5, F6, F7, and F8), fronto-central (FC1, FC2, FC3, FC4, FC5, and FC6), central (C1, C2, C3, C4, C5, and C6), temporal (FT7, FT8, T7, T8, TP7, and TP8), centro-parietal (CP1, CP2, CP3, CP4, CP5, and CP6), parietal (P1, P2, P3, P4, P5, P6, and P7), and occipital (PO3, PO4, PO5, PO6, PO7, PO8, O1, and O2) positions. Vertical eye movements were monitored via a supra- to sub-orbital bipolar montage. A right to left canthal bipolar montage was used to monitor for horizontal eye movements. The signals were digitized online with a sampling frequency of 512 Hz, with a 0.01–100 Hz band-pass filter. Electrode impedance was maintained below 10 kΩ, mostly under 5 kΩ. The software package ASA was used to analyze the waveforms[2]. The EEG data were screened for eye movements, electrode drifting, amplifier blocking, and EMG artifacts in a critical window ranging from 100 ms before to 800 ms after the onset of the critical word. Trials containing such artifacts were excluded from

---

[2] www.ant-neuro.com

further analysis, with 88.25% of the epochs being included. Below, we will carry out region-specific analyses of the predicted effects, comparing anterior regions (frontal and fronto-central electrodes, also including midline electrodes Fpz, Fz and temporal electrodes FT7, FT8) versus posterior regions (centro-parietal, parietal, and occipital electrodes, also including midline electrodes CPz, Pz, and Oz). Electrodes TP7, TP8, and POz were not included in the region analyses to balance the number of electrodes in each region.

RESULTS

An overview of nine representative electrodes (out of 64 total electrodes) is shown in **Figures 1 and 2**. It is apparent that, in the true sentences (**Figure 1**), the Modality-Match conditions (abbreviated to ModMatch, levels match and mismatch) clearly differ from each other whereas they are visually almost identical for the false sentences (**Figure 2**). For the true sentences, there are clear difference between the magnitude and direction of the effects across the scalp that leads us to include an additional factor Region (levels Anterior, Posterior) in the analyses.

Based on established effects that have been found in the literature and visual inspection of the peaks of the ERP waveforms, we divided the analysis into four time windows: First, a very early window (160–215 ms) to capture the N1–P2 complex. Second, an early window (270–370 ms), which is positioned just before the classic N400 window. Third, a standard N400 window (350–550 ms). Fourth, a late window (500–700 ms) which should capture any late positive shift effects.

A three-way analysis of Modality-Match, Veracity, and Region (anterior, posterior) was carried out for all time windows. This analysis was followed by additional analyses split by Veracity, exploring the existence of a ModMatch effect and/or a Region effect for the subsets of true and false sentences.
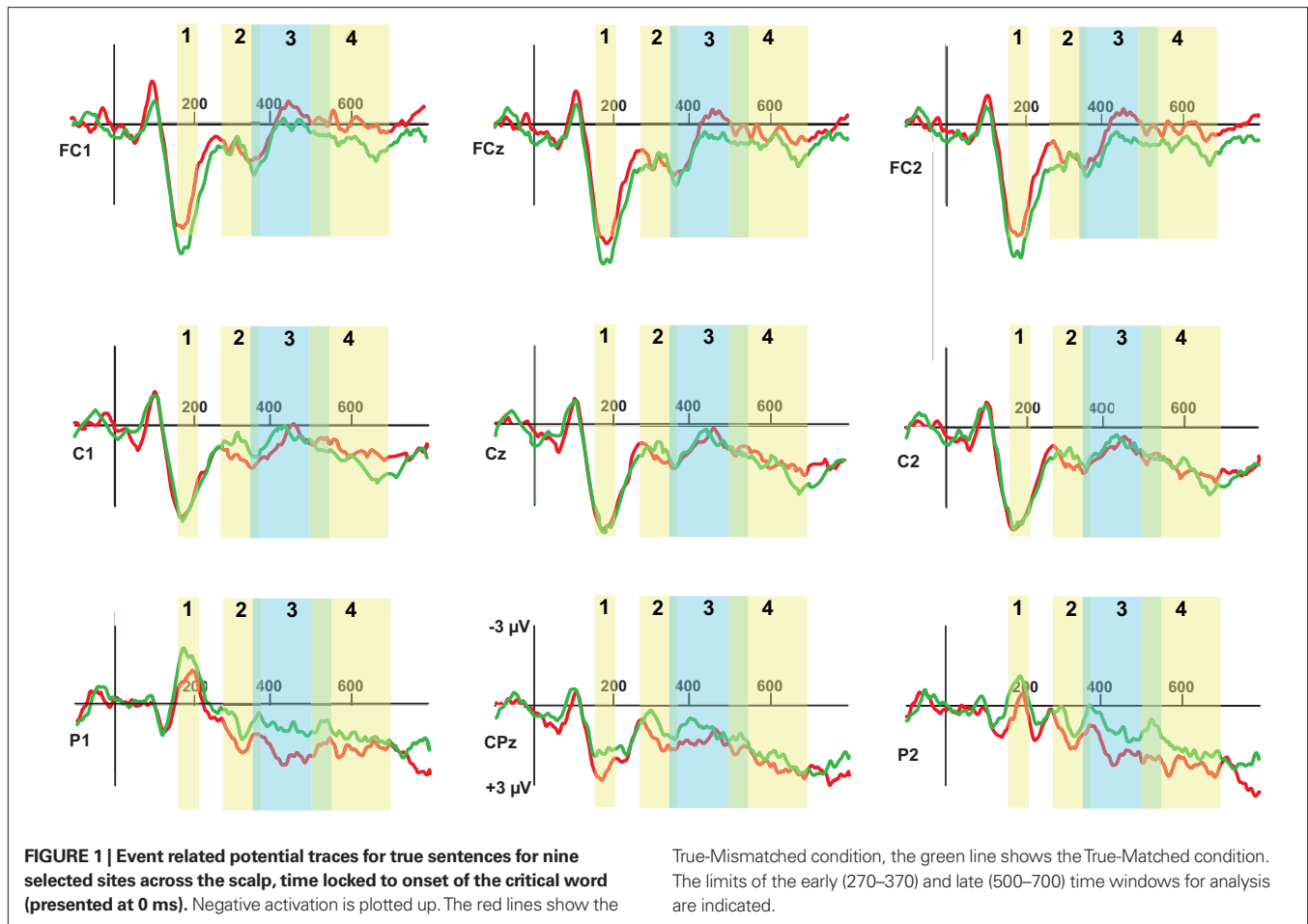
FIRST TIME WINDOW: N1–P2 COMPLEX, 160–215 MS

An N1–P2 complex is seen, which is typical for visual word presentation at this rate. We explored whether there was a difference between conditions in this very early time window. In the $2 \times 2 \times 2$ analysis, a significant effect of ModMatch was found and a significant interaction between Veracity, ModMatch, and Region (*F*-values and significance levels are reported in **Table 2** for easy reference; full details are in **Table A1** in Appendix). We explored this interaction by computing simple effects analysis for both levels of the Veracity condition: In the first follow-up analysis (for true sentences only), the factor ModMatch was again significant in this very early window (True-Match mean = 0.145 μV, True-Mismatch mean = 0.063 μV, difference = 0.082, see **Table 2** for significance levels). No significant effects were found in the second analysis (for false sentences).

SECOND TIME WINDOW: EARLY N400-LIKE EFFECTS, 270–370 MS

This time window was chosen after visual inspection of the ERP waveforms to capture the majority of differences that occur over the scalp, in all conditions. Given the theoretical and observed difference between our true and false sentences (see **Figure 3**), separate windows for true and false sentences could have been justified but we felt this would unnecessarily complicate the analysis (we carried out *post hoc* analyses on a number of other time windows but these analyses did not result in a different pattern of significance).

FIGURE 1 | Event related potential traces for true sentences for nine selected sites across the scalp, time locked to onset of the critical word (presented at 0 ms). Negative activation is plotted up. The red lines show the True-Mismatched condition, the green line shows the True-Matched condition. The limits of the early (270–370) and late (500–700) time windows for analysis are indicated.

In the overall $2 \times 2 \times 2$ analysis, we found that there was a significant difference between anterior and posterior electrodes and a significant three-way interaction between Veracity, ModMatch, and Region. We explored this interaction by computing simple effects analysis for both levels of the Veracity condition. In the first follow-up analysis (for true sentences only), the main effect of ModMatch was not significant, but the main effect of Region and the interaction ModMatch × Region were significant (see **Table 2**).
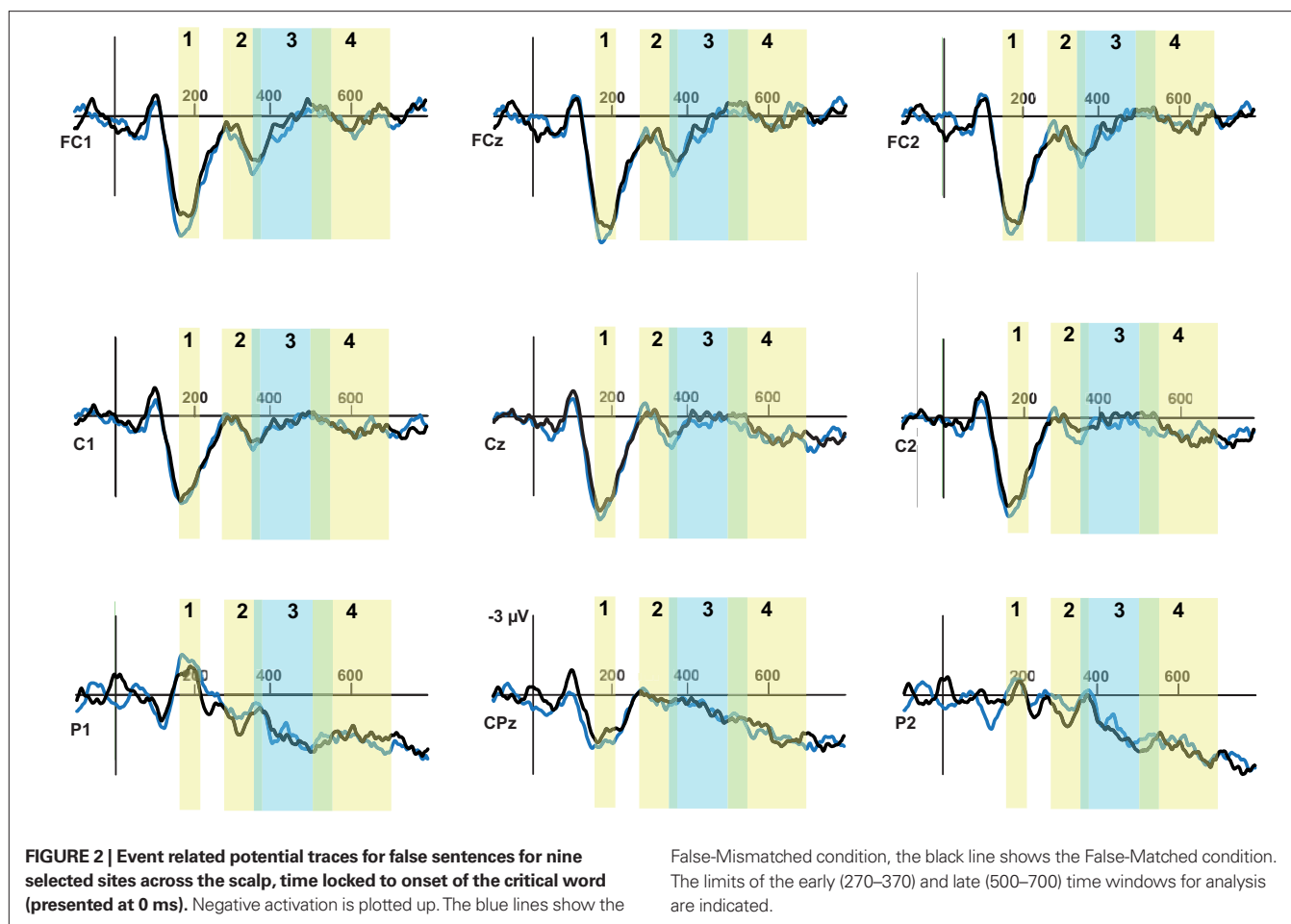
We further explored this two-way interaction for true sentences in a second follow-up and found that for true sentences, a significant ModMatch effect was found both on anterior electrodes [Anterior-True-Match mean = 2.35 μV, Anterior-True-Mismatch mean = 1.52 μV, difference = 0.83, $F(1,9) = 19.615$, MSE = 4.396, $p = 0.002$] and posterior electrodes [Posterior-True-Match mean = −2.00 μV, Posterior-True-Mismatch mean = −1.28 μV, difference = −0.72, $F(1,9) = 19.221$, MSE = 3.498, $p = 0.002$; see **Table 3**; and **Table A2** in Appendix]. Because the ModMatch effect for anterior electrodes has a different polarity than the effect for posterior electrodes, the effects cancel out in the first follow-up analysis, but they are significant in the second follow-up. In the third follow-up analysis (for false sentences only), the factors Region and ModMatch were included but only Region was significant; this effect is not of substantive interest.

THIRD TIME WINDOW: N400 EFFECTS, 350–550 MS

In this classic N400 window, we are probing the tail end of the modality effects reported above for the 270–370 ms time window. A first question is whether the modality-match effects persist; a second question is whether a classic N400 for false versus true sentences will be obtained.

In the initial $2 \times 2 \times 2$ analysis, we obtained a significant three-way interaction between Veracity, ModMatch, and Region. The main effect of Veracity was borderline significant ($F = 4.939$, $p = 0.053$); no other effects reached significance. We explored the three-way interaction by computing simple effects analysis for both levels of the Veracity condition. In the first follow-up analysis (for true sentences only), we found no significant effects; in the second follow-up analysis (for false sentences only), we also found no significant effects.
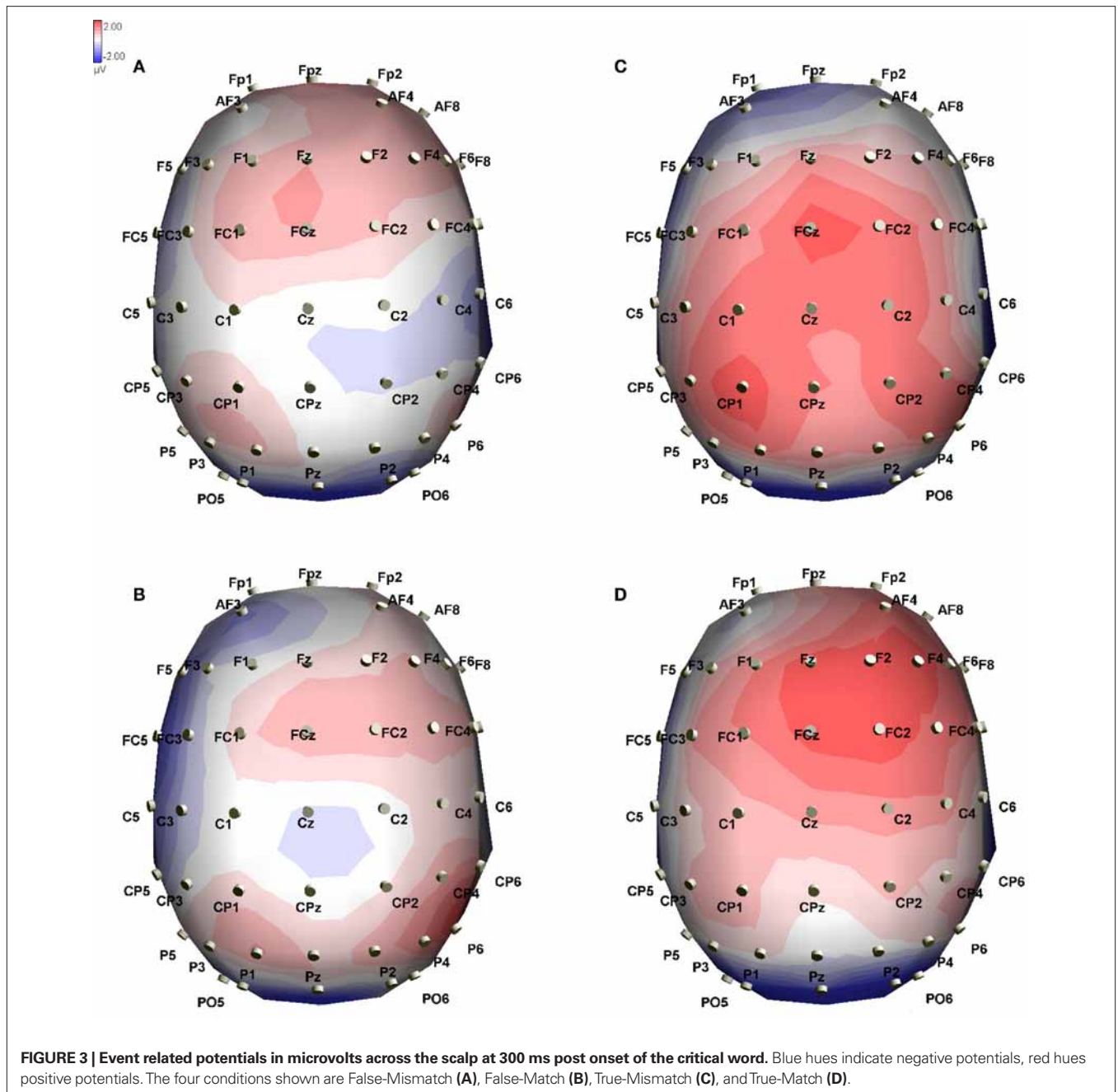
So although a significant three-way interaction (Veracity, ModMatch, and Region) was found, no significant effects of ModMatch and Region are found when the data are split by Veracity. However, one can also split the data by Modality-Match and look for effects of Veracity and Region. This analysis corresponds to looking for a veracity N400; the results are reported in **Table 3** and **Figure 4**. For modality-matched sentences, no effect of Veracity or Region were found (all $p > 0.25$). However, for modality-mismatched sentences, a significant effect of Veracity and a significant interaction of Veracity and Region were found [Anterior-True-Mismatch

FIGURE 2 | Event related potential traces for false sentences for nine selected sites across the scalp, time locked to onset of the critical word (presented at 0 ms). Negative activation is plotted up. The blue lines show the False-Mismatched condition, the black line shows the False-Matched condition. The limits of the early (270–370) and late (500–700) time windows for analysis are indicated.

Table 2 | Results of the Veracity × ModMatch × Region analysis for each window, with follow-up analyses for the results of ModMatch × Region within levels of Veracity (Vera).

Analysis	Effect	Window 1 (160–215 ms)	Window 2 (270–370 ms)	Window 3 (350–550 ms)	Window 4 (500–700 ms)
		F-value	*F*-value	*F*-value	*F*-value
2 × 2 × 2	Vera	0.126	2.133	4.939	4.066
	ModMatch	14.009**	2.059	0.003	0.057
	Region	0.010	18.376**	0.070	4.784
	Vera × ModMatch	3.300	1.517	1.892	1.986
	Vera × Region	0.091	1.040	1.419	0.305
	ModMatch × Region	0.440	0.487	0.424	1.798
	Vera × ModMatch × Region	8.550*	13.715**	5.399* (c)	3.496
For true	ModMatch	9.932*	2.488	0.755	2.017
	Region	0.042	22.042**	0.454	2.627
	ModMatch × Region	3.921	19.965** (a)	3.805	7.271* (b)
For false	ModMatch	0.301	3.353	0.698	0.230
	Region	0.000	15.176**	0.027	6.875*
	ModMatch × Region	0.421	1.714	2.094	0.491

*For each analysis, the significance of the effects of interest are reported by the F-values, with asterisks indicating levels of significance. Full details (MSE, df, p-values) are supplied in the Appendix. Notes: (a–c) refer to simple effects follow-up analyses, see text and **Table 3**.*
*p < 0.05, **p < 0.01, ***p < 0.001.

FIGURE 3 | Event related potentials in microvolts across the scalp at 300 ms post onset of the critical word. Blue hues indicate negative potentials, red hues positive potentials. The four conditions shown are False-Mismatch **(A)**, False-Match **(B)**, True-Mismatch **(C)**, and True-Match **(D)**.

mean $= -0.48$ μV, Anterior-False-Mismatch mean $= 0.38$ μV, Posterior-True-Mismatch mean $= 0.51$ μV, Posterior-False-Mismatch mean $= -0.17$, $F(1,9) = 8.519$, MSE $= 0.201$, $p = 0.017$; $F(1,9) = 6.358$, MSE $= 23.706$, $p = 0.033$]. This points at the presence of a classic Veracity N400 effect which is stronger in the anterior region and which is present in the classic N400 window.

FOURTH TIME WINDOW: LATE EFFECTS (500–700 MS)

This late time window was chosen to analyze the late negativity that is apparent for true sentences on the anterior electrodes (see **Figure 1**). In line with the analyses above, the $2 \times 2 \times 2$ was followed by two separate statistical analyses for true and false sentences.

significant effects of interest were found in the $2 \times 2 \times 2$ analysis, although three effects (Veracity; Region; three-way interaction) were marginally significant, $p < 0.10$. For true sentences, a significant interaction between ModMatch and Region was again found, follow-up analyses showed a ModMatch effect in both anterior and posterior regions [Anterior-True-Match mean $= -0.28$ μV, Anterior-True-Mismatch mean $= -1.14$ μV, difference $= 0.86$, Anterior $F(1,9) = 7.628$, MSE $= 12.131$, $p = 0.022$; Posterior-True-Match mean $= 0.23$ μV, Posterior-True-Mismatch mean $= 0.99$ μV, difference $= -0.77$, Posterior $F(1,9) = 6.803$, MSE $= 11.290$, $p = 0.028$]. The effect of ModMatch was numerically of the opposite sign in the regions so that the ModMatch main effect was non-significant.

Table 3 | Results of the simple effects follow-up analyses.

Analysis	Subset	Effect	*F*-value
(a) Simple effect of ModMatch within True sentences for Anterior/Posterior regions in window 2	For Anterior	ModMatch	19.615**
	For Posterior	ModMatch	19.221**
(b) Simple effects of ModMatch within True sentences for Anterior/Posterior regions in window 4	For Anterior	ModMatch	7.628*
	For Posterior	ModMatch	6.803*
(c) Effects of Veracity and Region for Matched/Mismatched sentences in window 3	For Mismatched	Veracity	8.519*
		Region	0.379
		Veracity × Region	6.358*
	For Matched	Veracity	0.561
		Region	0.000
		Veracity × Region	1.303

For each analysis, the significance of the effects of interest are reported by the F-values, with asterisks indicating levels of significance. Full details (MSE, df, p-values) are supplied in the Appendix.
*p < 0.05, **p < 0.01, ***p < 0.001.

This corroborates the findings in the 270–370 ms time window. Also similar to those earlier findings, no substantive significant effects were obtained for false sentences.

REACTION TIME DATA

Participants made a true/false judgment after each sentence was presented. Although there are enough participants for EEG analysis, the analysis of reaction times may lack the power to detect all differences. Note that Pecher et al. (2003) included 32 participants per between subject experimental condition, three times the number of participants in this study. The average reaction time and standard deviations are given in **Table 4**. One should keep in mind that, to keep the task as natural as possible, participants were not required to give a speeded response and this generally leads to large standard deviations. Additionally, to avoid movement artifacts we used a delay response (see Materials and Methods), which may also contribute to more variation. The means for the four conditions are very close to each other and do not differ significantly: In a ModMatch × Veracity ANOVA, we found no significant effects [Veracity $F_{(1,9)} < 1$; ModMatch $F_{(1,9)} < 1$; Veracity × ModMatch $F_{(1,9)} < 1$]. For this analysis, we included all correct responses to target sentences and removed responses faster than 200 ms and slower than 2500 ms. Similarly, accuracy for the conditions was very high and not significantly different: True-Match 94.25% accurate; True-Mismatch 95.25% accurate; False-Match 90% accurate; and False-Mismatch 90% accurate.
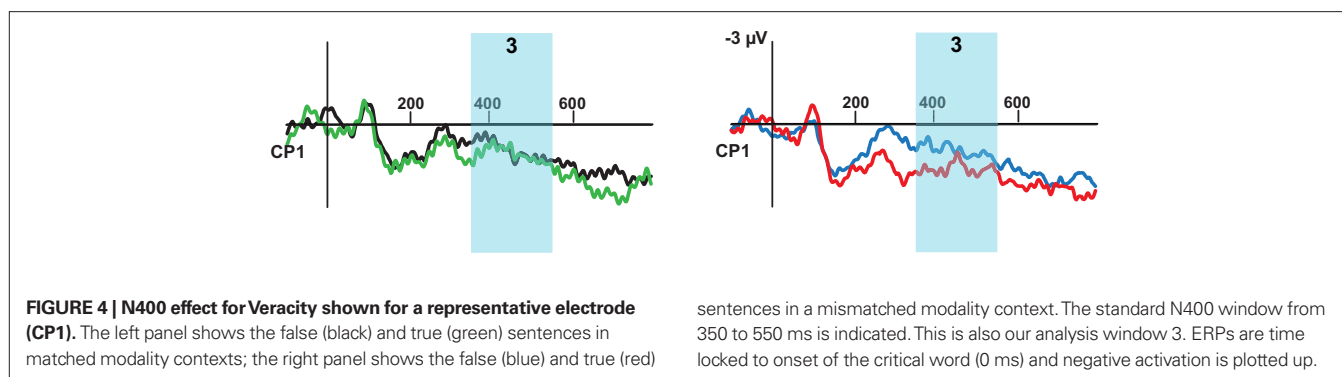
DISCUSSION

We conducted an ERP study where participants were exposed to written sentence pairs that either matched or mismatched in modality. We looked for an effect of modality-match in true and in false sentences. Previous research suggests that true and false sentences

are processed differently (Fischler et al., 1983). Our results indicate very different effects of modality for true and false sentences, see for example the scalp distributions in **Figure 3**. For true target sentences, we found a large early frontal N400-like effect for true, modality-mismatched pairs ("*A mitten is soft*" – "*A cellar is dark*") compared to true, modality-matched pairs ("*Ham is pink*" – "*A cellar is dark*") in our time window 1 (160–215 ms) and 2 (270–370 ms). In time window 1 (160–215) the anterior negativity effect did not significantly interact with region. However, in time window 2 (270–370 ms) this effect interacted with region such that true, mismatched sentences elicited a larger anterior negativity than true, matched sentences. True, mismatched sentences also elicited a larger positivity on posterior sites compared to true, matched sentences. The effects of modality on the true statements were replicated in a late time window (500–700 ms). As with the early time windows, more negativity is seen for the true, mismatched condition as compared to the true, matched condition across the frontal electrodes. Across the posterior electrodes, more positivity is seen for the true, mismatched condition compared to the true, matched condition. For false target sentences ("*A cellar is light*"), no significant effects of modality were seen at any time window. This is unlikely to be due to a lack of sensitivity, as the pattern for false sentences was numerically reversed compared to the true sentences (false, matched pairs eliciting a non-significant but larger anterior negativity in the waveforms than false, mismatched pairs).

We obtained one additional finding: False, mismatched sentences elicited a classical Veracity N400 in the 350–500 ms window when compared to true, mismatched sentences. This negativity interacted with region, such that it was strongest centro-posteriorly. No effect of veracity was found for the modality-matched sentences.

We will first discuss the relatively early time course of the effect and its distribution. The effect of modality-match on the frontal ERP sites begins in the first time window, as early as 160 ms, and is clearest in the second time window, around 300 ms. The presence of a modality-match effect in our earliest window (165–215 ms) indicates that modality switching is a precursor to and likely to be necessary for meaning integration. The modality-match effect develops further and becomes easily discernible across the scalp in the second time window (270–370 ms). This is the main effect of interest as the polarity of the effect reverses across the scalp, with mismatched pairs eliciting a larger anterior negativity and a larger posterior positivity. Both effects are much earlier than a standard N400 effect which typically begins around 250 ms and peaks around 400 ms (Kutas et al., 2006). In addition, the N400 typically has the strongest negativity on occipital and posterior sites. The distribution of the negativity in window 1 is mostly anterior. The distribution of the negativity in window 2 is also mostly anterior, but in window 2 we see an additional posterior positive distribution, which does not resemble the standard N400 at all.

We are not the first to find an anterior N400-like effect in an embodied context. For example, Van Elk et al. (2010) found an anterior N400 for the preparation of meaningful actions compared to meaningless actions, in a task that required participants to grasp objects. Interestingly, their N400 was largest for the preparation of meaningful actions. Holcomb et al. (1999) found that concrete words elicit a stronger anterior N400 than abstract words, an effect

FIGURE 4 | N400 effect for Veracity shown for a representative electrode (CP1). The left panel shows the false (black) and true (green) sentences in matched modality contexts; the right panel shows the false (blue) and true (red) sentences in a mismatched modality context. The standard N400 window from 350 to 550 ms is indicated. This is also our analysis window 3. ERPs are time locked to onset of the critical word (0 ms) and negative activation is plotted up.

Table 4 | Average reaction time (milliseconds) and standard deviation for the true/false judgments on target sentences.

Modality-match	Veracity	
	True	**False**
Matched	941 (445)	935 (471)
Mismatched	931 (444)	944 (507)

which they coined the concreteness-N400. However, we did not use abstract words in the current study so this particular N400 variant cannot explain our results.

The effect in windows 1 and 2 and 4 are quite similar to the ERP modulation that has been found for pictures and combined sentence–picture stimuli (Barrett and Rugg, 1990; Ganis et al., 1996). In the Ganis et al. (1996) study, the relevant experimental stimuli were sentence fragments that were followed by a picture. The picture was either semantically congruent or incongruent with the sentence semantics up to that point. It was found that, on the frontal electrodes only, incongruent pictures elicited a large negative deflection between 150 and 275 ms compared to the congruent pictures. Barrett and Rugg (1990) found a similar effect, which they called the N300. This effect is similar in time course and distribution to the window 2 effect we report. Ganis et al. (1996) also found that there was a larger anterior N400-like effect for pictures than for control words, and that this effect was reversed on the posterior sites. We found that our window 2 early anterior negativity also reverses on the posterior sites. Lastly, Ganis et al. (1996) report a late congruency effect from 575 to 800 ms whereby the incongruent pictures elicit a negativity at anterior sites and a positivity at posterior sites, which is similar to the findings in our fourth window (500–700 ms).

Ganis et al. (1996) suggest that their findings are specific to pictorial stimuli (see also Barrett and Rugg, 1990). However, we found a very similar effect using only language stimuli. We argue that our specific design, in which all the experimental stimuli refer to a highly salient modal (physical) aspect of an object, induces effects that are comparable in distribution and time course to those that have been obtained with pictures.

This explanation is somewhat consistent with the explanation of the so-called N700 effect proposed by West and Holcomb (2000). The N700 is very similar in time course and scalp distribution to the late effect that was obtained here in window 4. However, the N700 is sensitive to abstractness and shows a stronger anterior negativity and a stronger posterior positivity for concrete words than for abstract words. We do not want to argue that our mismatched stimuli were somehow more concrete but the interesting parallel is that the N700 is stronger in a mental imagery task than in two other tasks (lexical decision, letter spotting). This lead West and Holcomb (2000) to propose that the N700 reflects some image-based type of processing for purely linguistic stimuli.

The similarity between our results using sentences and those from previous work with pictures can best be explained in an embodied view of conceptual representation that uses simulation to arrive at semantic interpretation (Barsalou, 1999). It has been shown that reading action verbs can activate motor cortex (Hauk et al., 2004; Boulenger et al., 2009), presumably because participants were simulating the action. If our participants generated a mental simulation of the properties of the object (“*A cellar is dark*”), this could have produced activation that is very similar to actually seeing the object. Hence, we found effects that are very similar to those that so far have been exclusively found with picture presentation.

An embodied view of concepts would predict that there are no fundamental differences between representations derived from words and those derived from pictures, because each type of stimulus connects to underlying concepts that are grounded in modality-specific representations (in contrast to a dual coding view, such as Paivio's, 1986). Normally, access to concepts happens earlier for pictures (a long-held assumption, e.g., Caramazza et al., 1990; Schriefers et al., 1990) and effects of embodiment and modality are therefore commonly obtained with pictorial stimuli (Stanfield and Zwaan, 2001). In the paradigm used here, modality is primed and access to concepts and modality information is very fast (see also below), which leads to ERP effects that are comparable to those obtained with pictures.

The above is indirect evidence for an embodied view. We also found direct evidence for such a view in the clear ERP differences between true sentences with matched and mismatched modalities. This is also consistent with results from Collins et al. (2011): Using a concept property verification task, modality switching lead to increased amplitude N400 for visual property verifications and a larger late positive complex for auditory verifications. These embodiment effects would not be predicted by models that assume that an abstract propositional representation is necessary

for language comprehension in general and for the sentence veri-fication task specifically (e.g., the Constituent Comparison model; Carpenter and Just, 1975).

The proposed similarity with pictorial stimuli makes it likely (but not necessary) that the modality mismatch effects are stronger for the visual than for the tactile dimension. The idea that different modalities may lead to different modality switch effects in the ERP is supported by Collins et al. (2011), where their results indicate dif-ferent ERP effects for visual and auditory verifications. Qualitative inspection of the frontal waveforms broadly supports this view, but unlike the Collins et al. (2011) study, the current design does not have the statistical power to investigate this matter quantitatively as there are only 20 items per cell.

MODALITY-MATCH FINDINGS ON TRUE SENTENCES
We offer the following, tentative, explanation for the findings on true sentences. Although the full range of mechanisms underlying the generation of an N400 is still not fully understood, integra-tion processes is one possibility (Brown et al., 2000). Increasing the difficulty of integration will produce a greater (more negative) modulation of the N400. Additionally or alternatively the ampli-tude may serve as an indicator of the ease or difficulty of retrieving stored conceptual knowledge related to a word. The modulation may be dependent on the stored conceptual representation as well as the preceding contextual information (Kutas et al., 2006). One way to integrate a word with the current discourse is to have a set of possible continuations at hand, which requires some type of prediction. In highly constraining contexts, strong predictive N400 effects have indeed been demonstrated (Van Berkum et al., 2005; see also DeLong et al., 2005). The experiment by Van Berkum et al. (2005) was conducted in Dutch, where adjectives must linguistically agree with nouns. The results showed an N400 effect to adjectives that did not agree with a strongly predicted noun.

In the current experiment, all experimental sentences speak about the visual or tactile modality and a half of the experimen-tal sentences are in the same modality as the preceding sentence. Hence, when a visual context is followed by the target sentence "the cellar is…," participants are likely to have "dark" as the highly activated top candidate in the set of possible continuations. This prediction is derived from being in a visual context and simulat-ing the visual experience of "cellar." When, in the true, matched condition, the word "dark" is read it is immediately integrated in the simulation.

At the onset of the critical word in a true, mismatched sentence, the most highly activated candidate is "moist," because the par-ticipant's simulation of the concept cellar is in the tactile modal-ity. When the critical word "dark," comes in, the modality of the simulation has to be changed which leads to a modality switch effect and the observed anterior negativity and posterior positivity in windows 1 and 2 (160–215, 270–370 ms). This switch takes time, as was evidenced by the behavioral results of Pecher et al. (2003).

MODALITY-MATCH FINDINGS ON FALSE SENTENCES
As is clear from the scalp distribution shown in **Figure 3**, a very different pattern of activation was obtained for false sentences than for true sentences. In the false conditions, the target sentence is "the cellar is light," preceded by either a tactile or a visual context

sentence. We obtained no differences in windows 1 or 2 between the modality-matched and mismatched conditions because "light" is never an expected word: Similar to the explanation above, a visual context would raise an expectation for "dark" and the tactile context for "moist." Therefore "light" is equally unexpected for both modali-ties and no difference between modality-match conditions is found.

How does this activation explanation fit with embodied theories of language? Barsalou (1999) discusses falsity only with regards to comparing a sentence to a given situation. For that case, Barsalou (1999) essentially suggests that a simulation of the sentence is made and compared to the scene at hand. If there is a mismatch, then the simulation fails. In our experiment, participants presumably compare the information from the simulation of the false sen-tence to background knowledge, as there is no scene to compare to. Following Barsalou's (1999) line of reasoning, we would conclude that simulation of the sentence fails.

However this is an incomplete explanation of falsity since it seems that making the simulation of the false sentence should still show a benefit of modality-match. We can explain our results more completely if it is assumed that simulation is based on our prior recent experiences (for example, Glenberg et al., 2009) and that it never fails, but simply takes longer to complete. When trying to simulate "a cellar is light" out of context, we are unable to immedi-ately activate the relevant perceptual/action/emotion information because we have limited experience with this. This is not to say we cannot simulate things we have no experience with, but this account would predict that such simulations take longer out of context. The modality switch effect is a small and subtle effect that cannot be observed in this case. In our experiment, that would mean that a false sentence cannot benefit significantly from the preceding modality-match at time windows 1 and 2, but we believe partici-pants still arrive at a simulation of "the cellar is light." After all, this is what is required to understand larger discourse.

The inclusion of the false condition and the findings we obtained for it rule out a semantic relatedness explanation for our true sen-tence pairs. Under a semantic relatedness explanation, the results we find for true sentences are not due to embodiment but to simple semantic field priming. If, for example, a visual context used a color term and the following target sentence also used a color term, facilitation could be expected. There are independent arguments against this explanation: Pecher et al. (2003) provide empirical evidence against it and semantic priming does not usually last long enough to produce such an effect (see for example McQueen and Cutler, 1998). However, we can also rule out this explanation from our data: The same semantic priming should have occurred in our false sentences as word priming is not sensitive to veracity, but we found no effect for false sentences.

VERACITY FINDINGS: MODALITY-MISMATCHED SENTENCES
Overall, no effect of veracity was found. However, when splitting the data by modality-match versus modality-mismatch, an effect of veracity (greater amplitude N400 for false sentences) was seen in the modality-mismatched conditions. As already suggested, at the onset of the critical word in a false, mismatched sentence, the participant has simulated the concept cellar in the tactile modality and the most highly activated candidate is "moist." When the critical word "light" comes in, the modality of the simulation changes and this

causes a delay as outlined above. By the time of the N400 window (350–500 ms), the modality of the simulation may have switched to visual, but the simulation of "*light*" is minimal (assuming that simulation is based on our prior recent experiences) therefore a standard veracity N400 is observed. We tentatively conclude that a delayed minimal simulation leads to the difficulty in integrating "*light*" in the N400 time window.

VERACITY FINDINGS: MODALITY-MATCHED SENTENCES

The situation in the false, matched sentences is slightly more complex. At the onset of the critical word, the participant has simulated the concept Cellar in the visual modality and the most highly activated candidate is "*dark*." When the critical word "*light*" comes in, the modality of the simulation does not need to be changed and a wider simulation can be done, which will arrive at "*light*" as a possible property of cellars: Hence, no Veracity N400 is observed. In other words, although simulation is delayed due to falseness, some benefit occurs from the modality-match that occurs too late to show an effect in time windows 1 and 2, but by the N400 time window, the simulation is rich enough to provide support to the processing of the critical word "*light*," making it less difficult to integrate. This means the modality context modulates the N400 observed for veracity. We have previously provided evidence showing that the Veracity N400 can be modulated. In Hald et al. (2007), a three sentence context introducing new (supposed) facts about the world significantly reduced the N400 effect to objectively false sentences ("*Venice has many roundabouts*").

CONCLUSION

Our results fit well with the ideas of Zwaan and Madden (2005) and Glenberg and Robertson (1999) in that both sets of authors assume that, during comprehension, we build upon simulations constructed from the previous part of the discourse to integrate the ongoing information with the current simulation (Zwaan

calls this process construal). This idea applies most naturally to the comprehension of coherent discourse, but it should also apply to pairs of sentences such as our stimuli. It appears that the construction of a simulation in one modality for the context sentence can aid the simulation of the target sentence if it is in the same modality. A modality switch cost is incurred if the target sentence is of another modality, which leads to larger early anterior ERP effects.

Because the modality of previous sentences helps guide prediction, "*the cellar is…*" proceeded by a tactile context leads to a weaker activation of "*dark*" than when the preceding context is visual. Guided by the tactile context, the system is looking for a tactile property of "*cellar*" and this will lead to a modality switch negativity in our analysis windows 1 and 2 (160–370 ms) for true sentences. We think that the mismatch effect is not observed for false sentences because the comprehension system is engaged in efforts to integrate the false information (see above). Our finding suggests that the simulation process, which is central to embodied language processing, can be predictive (in line with Barsalou, 2009) and that that process will make stronger predictions when there is no modality switch.

REFERENCES

Baayen, R. H., Piepenbrock, R., and Van Rijn, H. (1993). *The CELEX Lexical Database (CD-ROM) [Computer Software]*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Barrett, S. E., and Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain Cogn.* 14, 201–212.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–609.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 364, 1281–1289.

Boulenger, V., Hauk, O., and Pulvermüller, F. (2009). Grasping ideas with the motor system: semantic somatopy in idiom comprehension. *Cereb. Cortex* 19, 1905–1914.

Brown, C. M., Hagoort, P., and Kutas, M. (2000). "Postlexical integration processes during language comprehension: evidence from brain-imaging

research," in *The New Cognitive Neurosciences*, 2nd Edn., ed. M. S. Gazzaniga (Cambridge, MA: MIT Press), 881–895.

Caramazza, A., Hillis, A. E., Rapp, B. C., and Romani, C. (1990). The multiple semantics hypothesis: multiple confusions? *Cogn. Neuropsychol.* 7, 161–189.

Carpenter, P. A., and Just, M. A. (1975). Sentence comprehension: a psycholinguistic processing model of verification. *Psychol. Rev.* 82, 45–73.

Clark, H. H., and Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cogn. Psychol.* 3, 472–517.

Collins, J., Pecher, D., Zeelenberg, R., and Coulson, S. (2011). Modality switching in a property verification task: An ERP study of what happens when candles flicker after high heels click. *Front. Psychol.* 2, 1–10.

DeLong, K., Urbach, T., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension

inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121.

Fischler, I., Bloom, P., Childers, D. G., Roucos, S. E., and Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology* 20, 400–409.

Ganis, G., Kutas, M., and Sereno, M. I. (1996). The search for "common sense": an electrophysiological study of the comprehension of words and pictures in reading. *J. Cogn. Neurosci.* 8, 89–106.

Gibbs, R. W. (2005). *Embodiment and Cognitive Science*. Cambridge: Cambridge University Press.

Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. London: Laurence Erlbaum.

Glenberg, A. M. (1997). What memory is for. *Behav. Brain Sci.* 20, 1–55.

Glenberg, A. M., Becker, R., Klötzer, S., Kolanko, L, Müller, S., and Rinck, M. (2009). Episodic affordances contribute to language comprehension. *Lang. Cogn.* 1, 113–135.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565.

Glenberg, A. M., and Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Process.* 28, 1–26.

Glenberg, A. M., and Robertson, D. A. (2000). Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning. *J. Mem. Lang.* 43, 379–401.

Glucksberg, S., Trabasso, T., and Wald, J. (1973). Linguistic structures and mental operations. *Cogn. Psychol.* 5, 338–370.

Goldberg, R., Perfetti, C., and Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *J. Neurosci.* 26, 4917–4921.

Hagoort, P., Hald, L., Bastiaansen, M., and Peterson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science* 304, 438–441.

Hald, L., Kutas, M., Urbach, T. P., and Parhizkari, B. (2005). "The N400 is not a brainwave. Negation and N400 effects for true and false sentences," in *12th Annual Meeting of the Cognitive Neuroscience Society*, New York, April 2005.

Hald, L., Steenbeek-Planting, E., and Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Res.* 1146, 210–218.

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307.

Holcomb, P. J., Kounios, J., Anderson, J. E., and West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 721–742.

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205.

Kutas, M., Van Petten, C., and Kluender, R. (2006). "Psycholinguistics electrified II: 1994–2005," in *Handbook of Psycholinguistics*, 2nd Edn, eds M. Traxler and M. A. Gernsbacher (New York: Elsevier), 659–724.

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Langacker, R. W. (2001). Discourse in cognitive grammar. *Cogn. Linguist.* 12, 143–188.

Lynott, D., and Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behav. Res. Methods* 41, 558–564.

Marques, J. M. (2006). Specialization and semantic organization: evidence from multiple-semantics linked to sensory modalities. *Mem. Cogn.* 34, 60–67.

McQueen, J. M., and Cutler, A. (1998). "Morphology in word recognition," in *Handbook of Morphology*, eds A. Spencer and A. M. Zwicky (Oxford: Blackwell), 406–427.

Meyer, D. E. (1970). On the representation and retrieval of stored semantic information. *Cogn. Psychol.* 1, 242–300.

Nieuwland, M. S., and Kuperberg, G. R. (2008). When the truth isn't too hard to handle: an event-related potential study on the pragmatics of negation. *Psychol. Sci.* 19, 1213–1218.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford: Oxford University Press.

Pecher, D., Van Dantzig, S., Zwaan, R. A., and Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *Q. J. Exp. Psychol.* 62, 1108–1114.

Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychol. Sci.* 14, 119–124.

Pecher, D., and Zwaan, R. A. (2005). *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*. Cambridge: Cambridge University Press.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.

Schriefers, H., Meyer, A. S., and Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: picture–word interference studies. *J. Mem. Lang.* 29, 86–102.

Spence, C., Nicholls, M. E. R., and Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Percept. Psychophys.* 63, 330–336.

Stanfield, R. A., and Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychol. Sci.* 12, 153–156.

Taylor, W. L. (1953). "Cloze procedure": a new tool for measuring readability. *Journal. Q.* 30, 415–433.

Trabasso, T., Rollins, H., and Shaughnessy, E. (1971). Storage and verification stages in processing concepts. *Cogn. Psychol.* 2, 239–289.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., and Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 443–467.

Van Casteren, M., and Davis, M. H. (2006). Mix, a program for pseudor-andomization. *Behav. Res. Methods* 38, 584–589.

Van Dantzig, S., Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2008). Perceptual processing affects conceptual processing. *Cogn. Sci.* 32, 579–590.

Van Elk, M., Van Schie, H. T., and Bekkering, H. (2010). The N400-concreteness effect reflects the retrieval of semantic information during the preparation of meaningful actions. *Biol. Psychol.* 85, 134–142.

Van Petten, C. (1995). Words and sentences: event-related brain potential measures. *Psychophysiology* 32, 511–525.

Van Petten, C., and Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Mem. Cogn.* 18, 380–393.

Vermeulen, N., Niedenthal, P. M., and Luminet, O. (2007). Switching between sensory and affective systems incurs processing costs. *Cogn. Sci.* 31, 183–192.

Wason, P. C. (1980). "The verification task and beyond," in *The Social Foundations of Language and Thought*, ed. D. R. Olson (New York: Norton), 28–46.

West, W. C., and Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. *J. Cogn. Neurosci.* 12, 1024–1037.

Zwaan, R. A. (2004). "The immersed experiencer: toward an embodied theory of language comprehension," in *The Psychology of Learning and Motivation*, ed. B. H. Ross (New York: Academic Press), 35–62.

Zwaan, R. A., and Madden, C. J. (2005). "Embodied sentence comprehension," in *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, eds D. Pecher and R. A. Zwaan (Cambridge: Cambridge University Press), 224–245.

Zwaan, R. A., Stanfield, R. A., and Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychol. Sci.* 13, 168–171.

Zwaan, R. A., and Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *J. Exp. Psychol. Gen.* 135, 1–11.

APPENDIX: FULL STATISTICAL REPORTS ON THE ANOVAS

Table A1 | Full results of the Veracity × ModMatch × Region analysis for each window, with follow-up analyses for the results of ModMatch × Region within levels of Veracity.

Analysis	Effect	Window 1 (160–215 ms)	Window 2 (270–370 ms)	Window 3 (350–550 ms)	Window 4 (500–700 ms)
2 × 2 × 2	Vera	$F(1,9) = 0.126$, MSE = 0.544, $p = 0.731$	$F(1,9) = 2.133$, MSE = 0.431, $p = 0.178$	$F(1,9) = 4.939$, MSE = 0.301, $p = 0.053$	$F(1,9) = 4.066$, MSE = 0.592, $p = 0.075$
	ModMatch	$F(1,9) = 14.009$, MSE = 0.075, $p = 0.005$	$F(1,9) = 2.059$, MSE = 0.180, $p = 0.185$	$F(1,9) = 0.003$, MSE = 0.303, $p = 0.955$	$F(1,9) = 0.057$, MSE = 0.403, $p = 0.817$
	Region	$F(1,9) = 0.010$, MSE = 218.497, $p = 0.923$	$F(1,9) = 18.376$, MSE = 369.775, $p = 0.002$	$F(1,9) = 0.070$, MSE = 124.872, $p = 0.797$	$F(1,9) = 4.784$, MSE = 228.636, $p = 0.057$
	Vera × ModMatch	$F(1,9) = 3.300$, MSE = 0.204, $p = 0.103$	$F(1,9) = 1.517$, MSE = 0.150, $p = 0.249$	$F(1,9) = 1.892$, MSE = 0.210, $p = 0.202$	$F(1,9) = 1.986$, MSE = 0.250, $p = 0.192$
	Vera × Region	$F(1,9) = 0.091$, MSE = 22.114, $p = 0.770$	$F(1,9) = 1.040$, MSE = 15.214, $p = 0.335$	$F(1,9) = 1.419$, MSE = 16.905, $p = 0.264$	$F(1,9) = .305$, MSE = 35.701, $p = 0.594$
	ModMatch × Region	$F(1,9) = 0.440$, MSE = 31.207, $p = 0.524$	$F(1,9) = 0.487$, MSE = 27.203, $p = 0.503$	$F(1,9) = 0.424$, MSE = 24.636, $p = 0.531$	$F(1,9) = 1.798$, MSE = 20.344, $p = 0.213$
	Vera × ModMatch × Region	$F(1,9) = 8.550$, MSE = 7.446, $p = 0.017$	$F(1,9) = 13.715$, MSE = 13.975, $p = 0.005$	$F(1,9) = 5.399$, MSE = 28.781, $p = 0.045$ (c)	$F(1,9) = 3.496$, MSE = 43.514, $p = 0.094$
For True	ModMatch	$F(1,9) = 9.932$, MSE = 0.172, $p = 0.012$	$F(1,9) = 2.488$, MSE = 0.237, $p = 0.149$	$F(1,9) = 0.755$, MSE = 0.290, $p = 0.408$	$F(1,9) = 2.017$, MSE = 0.182, $p = 0.189$
	Region	$F(1,9) = 0.042$, MSE = 97.613, $p = 0.841$	$F(1,9) = 22.042$, MSE = 139.620, $p = 0.001$	$F(1,9) = 0.454$, MSE = 67.963, $p = 0.517$	$F(1,9) = 2.627$, MSE = 168.700, $p = 0.140$
	ModMatch × Region	$F(1,9) = 3.921$, MSE = 17.413, $p = 0.079$	$F(1,9) = 19.965$, MSE = 7.657, $p = 0.002$ (a)	$F(1,9) = 3.805$, MSE = 32.378, $p = 0.083$	$F(1,9) = 7.271$, MSE = 23.239, $p = 0.025$ (b)
For False	ModMatch	$F(1,9) = 0.301$, MSE = 0.079, $p = 0.596$	$F(1,9) = 3.353$, MSE = 0.056, $p = 0.100$	$F(1,9) = 0.698$, MSE = 0.178, $p = 0.425$	$F(1,9) = 0.230$, MSE = 0.529, $p = 0.643$
	Region	$F(1,9) = 0.000$, MSE = 146.185, $p = 0.999$	$F(1,9) = 15.176$, MSE = 248.720, $p = 0.004$	$F(1,9) = 0.027$, MSE = 75.609, $p = 0.872$	$F(1,9) = 6.875$, MSE = 95.893, $p = 0.028$
	ModMatch × Region	$F(1,9) = 0.421$, MSE = 21.812, $p = 0.533$	$F(1,9) = 1.714$, MSE = 33.282, $p = 0.223$	$F(1,9) = 2.094$, MSE = 20.794, $p = 0.182$	$F(1,9) = 0.491$, MSE = 40.976, $p = 0.501$

Table A2 | Full results of the simple effects follow-up analyses.

Analysis	Effect	Statistics
(a) Simple effect of ModMatch within True sentences for Anterior/Posterior regions in window 2		
Anterior	ModMatch	$F(1,9) = 19.615$, MSE = 4.396, $p = 0.002$
Posterior	ModMatch	$F(1,9) = 19.221$, MSE = 3.498, $p = 0.002$
(b) Simple effects of ModMatch within True sentences for Anterior/Posterior regions in window 4		
Anterior	ModMatch	$F(1,9) = 7.628$, MSE = 12.131, $p = 0.022$
Posterior	ModMatch	$F(1,9) = 6.803$, MSE = 11.290, $p = 0.028$
(c) Effects of Veracity and Region for Matched/Mismatched sentences in window 3		
Mismatched	Veracity	$F(1,9) = 8.519$, MSE = 0.201, $p = 0.017$
	Region	$F(1,9) = 0.379$, MSE = 50.505, $p = 0.553$
	Vera × Region	$F(1,9) = 6.358$, MSE = 23.706, $p = 0.033$
Matched	Veracity	$F(1,9) = 0.561$, MSE = 0.310, $p = 0.473$
	Region	$F(1,9) = 0.000$, MSE = 99.002, $p = 0.985$
	Vera × Region	$F(1,9) = 1.303$, MSE = 21.979, $p = 0.283$