

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a postprint version which may differ from the publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/116714>

Please be advised that this information was generated on 2020-11-28 and may be subject to change.

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, 51(3), 437-447.

The contents of predictions in sentence comprehension: Activation of the shape of objects
before they are referred to

Joost Rommers^{1,2}, Antje S. Meyer^{1,3}, Peter Praamstra^{4,5}, Falk Huettig^{1,5}

¹ Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

² International Max Planck Research School for Language Sciences, Nijmegen, The
Netherlands

³ Radboud University Nijmegen, The Netherlands

⁴ Department of Neurology, University Medical Centre St Radboud, Nijmegen, The
Netherlands

⁵ Donders Institute for Brain, Cognition, and Behaviour, Nijmegen, The Netherlands

Correspondence should be addressed to Joost Rommers, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 XD Nijmegen, The Netherlands. Phone: +31-24-3521332. Fax: +31-24-3521213. Email: joost.rommers@mpi.nl.

Keywords: predictive sentence processing; visual representations; eye-tracking; event-related potentials (ERPs); N400.

Abstract

When comprehending concrete words, listeners and readers can activate specific visual information such as the shape of the words' referents. In two experiments we examined whether such information can be activated in an anticipatory fashion. In Experiment 1, listeners' eye movements were tracked while they were listening to sentences that were predictive of a specific critical word (e.g., “moon” in “In 1969 Neil Armstrong was the first man to set foot on the moon”). 500 ms before the acoustic onset of the critical word, participants were shown four-object displays featuring three unrelated distractor objects and a critical object, which was either the target object (e.g., moon), an object with a similar shape (e.g., tomato), or an unrelated control object (e.g., rice). In a time window before shape information from the spoken target word could be retrieved, participants already tended to fixate both the target and the shape competitors more often than they fixated the control objects, indicating that they had anticipatorily activated the shape of the upcoming word's referent. This was confirmed in Experiment 2, which was an ERP experiment without picture displays. Participants listened to the same lead-in sentences as in Experiment 1. The sentence-final words corresponded to the predictable target, the shape competitor, or the unrelated control object (yielding, for instance, “In 1969 Neil Armstrong was the first man to set foot on the moon/tomato/rice”). N400 amplitude in response to the final words was significantly attenuated in the shape-related compared to the unrelated condition. Taken together, these results suggest that listeners can activate perceptual attributes of objects before they are referred to in an utterance.

The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to

1. Introduction

In sentence comprehension, readers and listeners often anticipate upcoming information (Altmann & Kamide, 1999; DeLong, Urbach, & Kutas, 2005; Federmeier & Kutas, 1999; van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004). For instance, when a word is highly predictable, differential event-related brain potential (ERP) effects can be observed when a preceding adjective matches or mismatches with the expected word in grammatical gender (van Berkum et al., 2005; Wicha et al., 2004) or in phonological form (DeLong et al., 2005). Anticipation plays an important role in current views of sentence comprehension (Altmann & Mirkovic, 2009; Federmeier, 2007; Gibson, 1998; Kamide, 2008; Levy, 2008; Pickering & Garrod, 2007). Current theories focus on the underlying cognitive processes, with a prominent view being that predictions in comprehension are generated by the language production system (Pickering & Garrod, 2007).

An issue that has not received much attention concerns the kinds of information that listeners or readers pre-activate when they anticipate upcoming words. For understanding the mechanisms underlying prediction it is important to determine what listeners and readers do, or do not, predict. So far, studies have focused on functional semantic features or categories. Evidence for the involvement of this kind of information in predictions comes from anticipatory eye movements observed in the visual world paradigm, where participants listen to sentences and look at displays with multiple objects (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; for

review see Huettig, Rommers, & Meyer, 2011). For instance, in a seminal study by Altmann and Kamide (1999), participants heard sentences such as "The boy will eat the cake" or "The boy will move the cake" while viewing a display featuring several objects of which a cake was the only edible object. The authors found that, before the cake had been referred to, participants were more likely to initiate eye movements towards the cake when the verb was "eat" than when it was "move". This demonstrates that they predicted which object would be referred to on the basis of functional attributes (e.g., edibility) implied by the verb.

Other evidence for the involvement of functional semantic information in predictions comes from ERP studies. The component of interest is the N400, a centroparietally distributed component of negative polarity that peaks around 400 ms after onset of a content word and is considered a sensitive index of semantic processing (Kutas & Federmeier, 2000, 2011; Kutas & Hillyard, 1980). Federmeier and Kutas (1999) presented participants with contexts such as "They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of ...", which were followed by a predictable word (e.g., "palms"), or an unexpected word from the same semantic category (e.g., "pines"), or an unexpected word from a different category (e.g., "tulips"). N400 amplitude to the unexpected words was reduced when the word was related to the expected word (e.g., "pines") compared to when it was unrelated (e.g., "tulips"), even though in both of these conditions the predictable word was never presented. This suggests that semantic category information had been activated based on the context.

In the present study we were specifically interested in another kind of information that can be activated when words are processed, namely perceptual information - that is, physical attributes of objects. We focused on shape information. There is evidence from several sources that information about object shape can be activated *after* the relevant word has been processed. For instance, in sentence-picture verification experiments, Zwaan, Stanfield and Yaxley (2002) observed congruency effects between a shape implied in a sentence and a shape shown in a picture. Furthermore, several studies have observed perceptual priming in lexical decision, where a response to a target word (e.g., "coin") is faster when a preceding prime word has a referent with a shape similar to the target word's referent (e.g., "pizza") than when the prime word has a referent with a different shape (e.g., "table"; Moss, McCormick, & Tyler, 1997; Schreuder, Flores d'Arcais, & Glazenborg, 1984; but see Pecher, Zeelenberg, & Raaijmakers, 1998). Kellenbach, Wijers, and Mulder (2000) showed that at the neurophysiological level this shape priming was reflected in the N400, with N400 amplitude being smaller when prime and target had similarly shaped referents than when these were different. Finally, several visual world experiments showed that upon hearing a word such as "snake", listeners were more likely to move their eyes to objects with a shape similar to the referent (e.g., a cable) than to objects with a different shape (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007).

In sum, there is strong evidence that shape information can be activated when words are processed. However, it is unknown whether this type of information already becomes activated when words are anticipated. For instance, do listeners expecting to hear about a moon activate visual representations of the object before they actually hear

"moon"? As mentioned above, there is evidence for a high degree of specificity of predicted lexical information, since specific words can be anticipated (DeLong et al., 2005; van Berkum et al., 2005; Wicha et al., 2004). Listeners thus seem to prepare in advance for the input they might receive. Whether this degree of specificity extends to the level of perceptual attributes is unknown. One might, for instance, speculate that in many contexts shape information is not crucial for understanding the meaning of an utterance and would therefore be activated late or not at all. Alternatively, it could be the case that anticipation of a specific word entails some activation of all associated information, regardless of its relevance for understanding the current utterance.

A second issue we explored in the present study concerns the influence of the participants' task on the activation of shape information. Although previous studies have shown that shape representations can be activated in language comprehension, it is not known to what extent the reported shape effects depend on the participants' tasks. For lexical decision (Schreuder et al., 1984), it is conceivable that the activation of shape information stems from task properties, as being able to activate a mental image for a stimulus is a cue that distinguishes words from non-words. For sentence-picture verification (Zwaan et al., 2002), one could argue that the instruction to compare sentences to pictures encourages readers or listeners to create mental images corresponding to the sentence content. In the present study, participants were simply asked to listen attentively to spoken sentences. No additional task was given. Two online techniques were used to measure pre-activation of visual representations of upcoming words' referents: Eye-tracking during a visual world task, and recording of event-related

potentials during a passive listening task. We asked whether participants, in anticipation of specific critical words, would activate the shape of the referents.

In Experiment 1, we used the target-absent version of the visual world paradigm in which a fully matching target word referent is excluded from the visual display on some trials (cf. Huettig & Altmann, 2005; Huettig et al., 2011, for detailed discussion). Participants listened to sentences that were predictive of a specific word (e.g., "moon" in "In 1969 Neil Armstrong was the first man to set foot on the moon"). Five-hundred ms before the acoustic onset of the target word (i.e. "moon"), participants were shown displays of four objects. These were three unrelated distractor objects and one critical object. Depending on the condition, the critical object was either the target object (e.g., a moon), or a shape competitor of the target object (e.g. a tomato), or a control object with a different shape than the target (e.g. rice). The participants' eye movements were recorded. We expected to observe anticipatory eye movements to the target object (e.g., the moon) before it was referred to. More importantly, if specific object shape representations form part of the contents of predictions for upcoming words, the expectation of the target concept (e.g., the moon) should also lead to anticipatory eye movements to a shape competitor (e.g., to a tomato).

We indeed observed such anticipatory eye movements. However, the visual world experiment involved the presentation of pictorial stimuli, which may encourage the activation of information concerning the physical properties of the referent objects shape (for discussion, see Huettig et al., 2011; Mitchell, 2004; for experimental evidence, see Huettig & McQueen, 2011). To determine whether shape information is also preactivated in the absence of pictorial information, a second experiment was conducted.

This was an ERP experiment similar to Federmeier and Kutas's (1999) study. The participants listened to the same lead-in sentences as in Experiment 1. The final word of the sentence was either a highly predictable word (e.g., "moon"), or a word referring to an object with a similar shape (e.g., "tomato"), or an unrelated word (e.g., "rice"). We examined whether the N400 component in response to these words would reflect the semantic anomaly and, more importantly, would be sensitive to the referents' shape as well.

2. Experiment 1

2.1 Participants

Forty-five participants (34 women, mean age 21 years, range 18-29 years) from the Radboud University Nijmegen and the HAN University of Applied Sciences gave informed consent and were paid to take part in the experiment. All were native speakers of Dutch who had normal hearing, normal or corrected-to-normal vision and no history of language disorders.

2.2 Materials and Design

The experiment consisted of 96 experimental trials and 32 filler trials. On each trial the participants heard a sentence and saw a visual display featuring four objects. On experimental trials one of the four objects was, depending on the condition, the highly predictable target (e.g., a moon), or the shape competitor (e.g., a tomato) or an unrelated control object (e.g., rice). The remaining three objects were unrelated distractor objects. On filler trials four objects were presented of which one was referred to in the spoken sentences. Thus, half of the sentences referred to an object in the accompanying visual display.

The sentences were statements with an average length of 14 words. They were selected from an initial set of 408 sentences. The materials had been pretested on cloze probability to verify that the final words of the experimental sentences were highly predictable. Fifteen participants (13 female, mean age 20 years, range 18-23) were asked to complete each lead-in sentence, that is, the fragment up to the final word, with the first word that came to mind, without trying to be original. The cloze probability of the target words was the proportion of participants who chose to complete the sentence fragment with the word in question (Taylor, 1953). The average cloze probability for the selected sentences was 0.72 ($SD = 0.30$), indicating that the critical words were predictable. The cloze probability for the names of the shape competitors and control objects was zero; that is, none of the participants provided them as continuations of the lead-in sentences. The filler sentences were chosen from the same set and were therefore similarly predictable.

The selected sentences were spoken at a relaxed pace by a female native speaker of Dutch and recorded in a sound-attenuating booth in a single recording session (mono, 44 kHz sampling rate, 16 bit sampling resolution), along with additional sentences used in Experiment 2. The onset times of the critical words were marked using Praat (Boersma & Weenink, 2009).

To create the visual displays for the experimental trials, 96 sets of six objects were composed consisting of three critical objects (either the Target, Shape competitor or Control picture, depending on the sentence it was paired with) and three distractor pictures which were unrelated to the other pictures in shape, semantics, or phonology (example see Figure 1). In designing the experiment, we aimed to make sure that

differences in the participants' eye movements could be unambiguously attributed to the relationship of the critical objects to the sentence contexts, rather than properties of the sentence-final words or the corresponding objects themselves. Therefore, each critical object appeared in all three experimental conditions; that is, as a correct target on one list, as a Shape competitor on another list, and as a Control object on another list (see design below). In addition to the experimental sets, 32 four-object sets were composed to be used on filler trials.

The pictures were selected from Rossion and Pourtois (2004) and other sources or created by an artist. Colored pictures were used to facilitate the recognition of the objects (Rossion & Pourtois, 2004) and to ensure that participants would not confuse similarly-shaped pictures with the actual referents of the spoken words. When only a black-and-white picture was available in the picture data base, the artist colored the picture.

The pictures were pretested on three dimensions: familiarity, visual complexity, and name agreement. This was done to make sure that the critical pictures and the distractors did not differ in any obvious way and to establish that the objects were named as we anticipated. Fifteen participants (12 female, average age 20 years, range 18-26 years) from the same participant pool as used in the main experiment were paid for participation. Each picture was presented in the center of the screen until a response was recorded. First, the participants rated each picture for visual complexity on a scale from 1 ("very simple") to 7 ("very complex") by pressing the corresponding key on a keyboard. In the instructions, visual complexity was defined as the amount of details and the complexity of the lines in the pictures. The participants were asked to judge the complexity of the pictures rather than the complexity of the objects that the pictures

represented. They were encouraged to use the full 7-point scale. Next, the participants viewed the pictures again and indicated their familiarity with each object on a scale from 1 (“very unfamiliar”) to 7 (“very familiar”). Here, they were asked to indicate how often they came in contact with the object or thought about it. Finally, the participants typed in the name of each object as they would call it. The average visual complexity of target pictures (3.09, $SD = .80$) and distractor pictures (3.12, $SD = .76$) was very similar, $t(382) = .378, p = .706$. The same held for the average familiarity of the target pictures (4.70, $SD = .88$) and distractor pictures (4.67, $SD = .85$), $t(382) = .302, p = .763$. Average name agreement was high (distractor objects: 81%, $SD = 20\%$; target objects: 76%, $SD = 28\%$; $t(382) = 1.904, p = .058$).

All pictures were resized to fit into a square of 210 by 210 pixels, corresponding to 7 degrees of visual angle for the participant. They appeared in the four corners of the 1024×768 pixels screen at a distance of 256 or 768 pixels from the left and 192 or 576 pixels from the top.

The experimental and filler items were distributed across three presentation lists. In each list each experimental sentence was played once, accompanied by one of the three corresponding displays. In each list, each third of the 96 experimental sentences was combined with displays showing the target, the shape competitor or the unrelated competitor. Across the three lists, each sentence appeared once in each condition (Target, Shape competitor, Control). The 32 filler items appeared on all lists. Each list was used for 15 participants, who each received different randomizations of trials and picture positions on the screen. The same condition or the same target picture position never occurred on more than three consecutive trials.

*** Insert Figure 1 about here ***

2.3 Procedure

Participants performed a look-and-listen task as in previous experiments investigating anticipatory eye movements (e.g., Altmann & Kamide, 1999). They were asked to listen to the sentences carefully and told that they were free to look at whatever they wanted to, but that they should not take their eyes off the screen.

Participants were tested individually in a dimly illuminated room. They were seated in front of a screen with their chin on a chin rest. The movements of each participant's right eye were recorded with an Eyelink 1000 Tower Mount eye tracker sampling at 1000 Hz (except for one participant whose right eye could not be calibrated; the left eye was used instead). The spoken sentences were played through headphones.

Each trial started with a central fixation circle that remained on the screen until the participant fixated it. Then a blank screen appeared. A spoken sentence started playing, and 500 ms before the critical word was spoken four pictures appeared on the screen. Thus, in contrast to typical visual world experiments there was no long preview of the visual display before the critical target word was heard. This was done to avoid visual priming of the target concept (e.g. priming of the shape of the concept moon by the visual presence of a tomato in the display; see Huettig et al., 2011, for further discussion). The pictures remained on the screen until 2000 ms after sentence offset. Then a blank screen appeared for 500 ms. After every trial, a central fixation circle appeared, allowing for drift correction.

Data were coded in terms of fixations, saccades, and blinks using the algorithm provided in the Eyelink software. A fixation was coded as a fixation on an object if it occurred within a square interest area of 300 by 300 pixels around the centre of that object.

2.4 Results

Figure 2 shows a time-course graph of the proportions of fixations to the critical pictures (Target, Shape competitor, and Control) and the average proportions of fixations to the three distractor pictures across items and participants. Recall that each display included only one of the three critical objects. The three dotted lines show the average proportions of looks to distractors in displays featuring targets, shape competitors, and control objects, respectively. The proportions indicate the number of trials in which participants were fixating each location on the screen. Time zero represents the acoustic onset of the critical spoken word.

The graph shows that, unsurprisingly, participants were far more likely to fixate upon the targets than on the competitors or distractors. More importantly, they were also more likely to fixate upon the Shape competitors than upon the unrelated Control objects. Finally, from about 500 ms onward the Control objects were fixated somewhat more frequently than the corresponding distractors, perhaps because due to counterbalancing they were referred to in another sentence in the experiment. We had no hypotheses about fixations at this point in time.

Since it takes on average about 200 ms to program and initiate a saccadic eye movement (Saslow, 1967), we can safely assume that any fixation bias to either Target object or Shape competitor until 200 ms *after* the acoustic onset of the target word is of

anticipatory nature. Thus, for the statistical analyses we chose a 450-ms time window starting at 250 ms before word onset (the point in time when fixations were first directed to any of the objects, see Figure 2) until 200 ms after word onset (the approximate point in time when fixations could first be based on information retrieved from the spoken target word).

*** Insert Figure 2 about here ***

Fixation proportions within this time window were transformed to log odds, the appropriate scale for assessing effects on a categorical dependent variable, using the empirical logit function (Barr, 2008). The average log odds of looks to the unrelated distractors was subtracted from the log odds of looks to the Target/Shape/Control object to create the dependent variable, which indicates the strength of any bias toward each experimental picture over the unrelated distractor pictures. This dependent variable was analyzed using linear mixed-effects regression models, which allow for simultaneous inclusion of participants and items as random factors (Baayen, Davidson, & Bates, 2008). *P*-values were calculated assuming that the *t*-values were drawn from a normal distribution, which is justified for datasets the size of the current experiment (Barr, 2008). The full model included a fixed effect of Condition (Target, Shape competitor, Control) and the maximal possible random effects structure consisting of random intercepts and slopes for Condition by participant, sentence, and picture. The Control condition was mapped onto the intercept as a baseline to compare the Shape competitor and Target conditions to. This model was compared to the same model without the fixed effect of

Condition using a likelihood ratio test. Including Condition improved model fit, $\chi^2(2) = 61.741, p < .001$. There was no significant bias for the Control pictures (e.g., rice) over the unrelated distractors, $\beta = .04925, SE = .14250, t = .346, p = .730$. There was a higher bias for Target pictures (e.g., moon) than for Control pictures, $\beta = 2.02858, SE = .21299, t = 9.524, p < .001$. Importantly, participants also showed a higher bias for Shape competitors (e.g., tomato) than Control pictures, $\beta = 0.41310, SE = .19203, t = 2.151, p = .031$.¹

2.5 Discussion

Using a variant of the visual world paradigm, we observed that participants anticipated which object a speaker would refer to next. When listening to sentences such as "In 1969 Neil Armstrong was the first man to set foot on the moon", participants looked significantly more to the picture of a moon than to unrelated distractor objects before the word "moon" was heard. This is in line with earlier studies (e.g., Altmann & Kamide, 1999) demonstrating that listeners anticipate upcoming words. More importantly, we also found that participants looked significantly more at objects similar in shape to the target object (e.g. a tomato) than to unrelated objects (e.g., rice) before the word "moon" was heard. To our knowledge, the present research is the first to report such an effect of pre-activation of the visual form of the referents of upcoming words.

In this experiment, the visual displays only appeared 500 ms before the onset of the critical word. This preview period for the pictures is shorter than the preview commonly used in the visual world paradigm, which often spans at least the duration of

¹ A repeated-measures ANOVA on by-participant means yielded the same results, including a bias toward Target objects over unrelated distractors that was higher than that for Control objects, $F(1,44) = 153.177, p < .001$, and a bias toward Shape competitors over unrelated distractors that was higher than that for Control objects, $F(1,44) = 6.676, p = .013$.

the sentence. Nevertheless, a preview time of 200 ms has been used before, and a bias toward Shape competitors was observed (Huettig & McQueen, 2007). We chose a short preview period to minimize the impact of the visual stimuli on the processing of the spoken sentences, in particular to ensure that any priming of visual representations would be very minimal. However, to assess whether anticipatory activation of shape information would also occur in the total absence of any visual stimuli, a second experiment was conducted.

3. Experiment 2

In Experiment 2, participants listened to the same sentences as in Experiment 1 (Correct condition) and to sentences where the final highly predictable word was replaced by the name of the shape competitor (Shape condition) or the unrelated control object (Unrelated condition; see Table 1 for an example). The participants were asked to listen to the sentences for comprehension while their electroencephalogram (EEG) was recorded. The EEG component of interest was the N400 in response to the final word of the sentences. Since the sentences in the Shape and Unrelated condition both ended in a contextually anomalous word, we expected a more pronounced N400 in both of these conditions than in the Correct condition (e.g., Federmeier & Kutas, 1999; Kutas & Hillyard, 1980). Importantly, the N400 has also been shown to be sensitive to visual attributes of objects being referred to (Kellenbach et al., 2000). If shape representations are involved in predictions of upcoming meaning in the absence of pictorial information, the N400 amplitude should therefore be attenuated for the Shape condition relative to the Unrelated condition.

3.1 Participants

Twenty-four students (22 women, mean age 20 years, range 18-23 years) from the Radboud University Nijmegen and the HAN University of Applied Sciences gave informed consent and were paid to take part in the experiment. All were right-handed, native speakers of Dutch who had normal hearing, normal or corrected-to-normal vision and no history of neurological or language disorders.

3.2 Materials and Design

There were three conditions: Correct, Shape, and Unrelated (see Table 1). In the Correct condition, the same 96 sentences were presented as in Experiment 1. For the Shape and Unrelated condition, the final word of the sentences was replaced by the name of the Shape competitor or Control object, respectively. Thus, the Unrelated condition in Experiment 2 roughly corresponds to the Control condition in Experiment 1. In total, the materials consisted of 288 Dutch sentences, formed by recombining critical words and lead-in sentences (for an example, see Appendix A). Each critical word occurred once in each condition, thus controlling for lexical characteristics. For each sentence context the Shape and Unrelated word came from the same semantic category (e.g., *tomato* and *rice* are both food), thus controlling for previously demonstrated effects of semantic category on N400 amplitude (Federmeier & Kutas, 1999).

The mean spoken sentence durations in the three conditions were similar, Correct: 5334 ms ($SD = 1291$), Shape: 5325 ms ($SD = 1212$), Unrelated: 5413 ms ($SD = 1232$), $F(2,190) = 1.86, p > .1$. Written versions of the sentences and words were pretested on two dimensions, the shape similarity of the referents of the words and the plausibility of the sentences. Different participants took part in each pretest, and none of them participated in the main experiment.

The goal of the shape similarity ratings was to verify that the referents of the words in the Shape condition were more highly related in shape to the referents of the Correct word than the referents of the words in the Unrelated condition. The 96 shape-related pairs (e.g. tomato and moon) and the 96 unrelated pairs (e.g. rice and moon) were presented along with 80 other pairs of words. Twenty-four participants (13 women, mean age 25 years, range 21-32 years) rated the shape similarity of the referents of the word pairs on a scale of 1 ("completely different shapes") to 7 ("exactly the same shape"). Each person received a different random order of items, and the left-to-right order of the members of the pairs was counterbalanced across participants. Participants were instructed to use the entire scale from 1 to 7 and to ignore any similarities in meaning between the words. The ratings were averaged across participants for each word pair (pooling over the two left-to-right orders). The ratings confirmed that for the word pairs selected for the ERP experiment, the average ratings were higher in the Shape condition (mean = 4.2, $SD = 0.9$) than in the Unrelated condition (mean = 1.9, $SD = 0.7$), $t(94) = 14.1$, $p < .001$.

In addition, we wanted to ensure that the words from the Shape and Unrelated conditions would be equally difficult to integrate with the preceding context. In order to test this, plausibility ratings were obtained for the sentences. Forty-two participants (31 women, mean age 21 years, range 18-34) were randomly assigned to one of three lists. Every list consisted of 184 sentences in which no context or critical word was repeated and each condition was represented by 40 or 48 items depending on the list (this was at a stage before selecting the final set of 32 items per list). Forty-eight plausible sentences were used as fillers on every list, and each participant received the sentences in a

different random order. They were asked to indicate for every sentence, on a scale from 1 ("very implausible") to 7 ("very plausible"), how plausible the described situation would be in everyday life. The ratings were averaged across participants for each of the 96 selected critical words in each condition. These ratings were submitted to a repeated-measures analysis of variance (ANOVA) with Condition (Correct, Shape, Unrelated) as the within-word factor. This by-items analysis yielded an effect of Condition, $F(2, 190) = 3946.25, p < .001$. Planned comparisons indicated that, as expected, the sentences in the Correct condition ($M = 6.4, SD = 0.5$) were more plausible than those in the Shape condition ($M = 1.5, SD = 0.4$), $F(1, 95) = 5697.32, p < .001$, and also more plausible than those in the Unrelated condition ($M = 1.5, SD = 0.4$), $F(1, 95) = 5454.39, p < .001$. Importantly, the difference between the Shape and the Unrelated conditions was not significant, $F(1, 95) = 0.03, p = .870$. Also, on an item by item basis, shape ratings were not correlated with plausibility ratings, $r = .02, p = .749$. Thus, there was no evidence that one of the two types of semantic violation (Shape or Unrelated) was easier to integrate in the sentence context than the other.

The sentences were distributed across three lists. On every list each critical word appeared once, and each condition was represented by 32 items. Sixty-four filler sentences were added to every list, of which 48 were correct and 16 contained a semantic violation. In total, 50% of the sentences on every list contained a violation. Each list was used for eight participants, who received different randomizations of the items.

3.3 Procedure

Participants were tested individually in a dimly illuminated room. They were seated in front of a screen and two loudspeakers and were asked to relax, and to move

and blink as little as possible while carefully listening to the sentences. The experiment started with two practice blocks of eight trials each to familiarize the participants with the way the trials were structured. After the practice blocks the sentences were presented in twenty short blocks of approximately two minutes. Each trial started with a blank screen with a duration of 50 ms, followed by a short warning tone. 750 ms later, a fixation cross (+) appeared on the center of the screen, which participants were asked to fixate on. After 750 ms, a spoken sentence was played. The fixation cross remained on the screen until 1500 ms after sentence offset. Participants had been asked not to blink or move while the fixation cross was on the screen. Then three asterisks (* * *) were presented for 3000 ms. During this period participants were free to blink and move their eyes. After every block participants could take a break for as long as needed before resuming the experiment by pressing a button on a button box.

3.4 EEG recording and analysis

EEG was recorded from 128 active Ag/AgCl electrodes mounted in a cap according to the 10-5 system (Oostenveld & Praamstra, 2001). Recordings were performed relative to common mode sense (CMS) and driven right leg (DRL) electrodes placed just anterior to the Fz electrode. Horizontal eye movements were monitored using electrooculography (EOG) electrodes positioned laterally to the left and right eyes. Two electrodes were placed at the mastoids. The signals were amplified by Biosemi ActiveTwo amplifiers with a lowpass filter at 128 Hz and sampled with a frequency of 512 Hz.

The data were referenced to the average of the left and right mastoids. Bipolar horizontal EOG was computed as the difference between the left and right EOG

electrodes. A bandpass filter of 0.05-30 Hz was applied and the continuous EEG was segmented into epochs from 300 ms before until 1000 ms after critical word onset. These epochs were baseline-corrected through subtraction of the mean signal in the 150 ms before critical word onset. Trials containing blinks were removed using participant-specific thresholds for the three frontal electrodes (Fp1, Fpz, Fp2). In a second step, other artifacts were removed using a threshold of + and -100 μ V. This was done on a single-channel basis, as in high-density electrode setups the probability increases that some channels are contaminated by artifacts, which can easily lead to an unacceptably low number of trials when removing entire trials (cf. Junghöfer, Elbert, Tucker, & Rockstroh, 2000). Average ERPs were then computed for each participant and condition separately.

Mean voltage measures were taken in two predefined time windows: 300-500 ms (Federmeier & Kutas, 1999; van den Brink et al., 2006) and 500-700 ms after word onset (McCallum, Farmer, & Pocock, 1984). Besides having been used in previous studies, the 500-700 ms time window was based on the knowledge that the N400 can be longer in duration with auditory presentation compared to visual presentation (for discussion, see Kutas & van Petten, 1994). Furthermore, in comparison to most auditory ERP studies, some of our critical words were relatively long in terms of number of syllables (range 1-5, mean 2.1, *SD* 1.1) and duration in ms (range 289-1058, mean 603, *SD* 174). This might delay the point at which the words became recognized, and our effect of interest depended on recognition of the words.

The average ERPs were submitted to nonparametric cluster-based permutation tests (Maris & Oostenveld, 2007) using the Matlab toolbox Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011). The permutation test determined which channels showed a

significant effect by means of a clustering algorithm based on the physiologically plausible assumption that ERP effects are clustered together in space over neighboring electrodes. The test compares two conditions at a time and works as follows (for details see Maris & Oostenveld, 2007). First, a dependent-samples t -test compares the conditions at every data point (in this case, every electrode), and data points that do not exceed a significance level of .05 are zeroed. Adjacent non-zero data points (at neighbouring electrodes) are combined into clusters for each of which the cluster-level t -value is the sum of all t -values within the cluster. Then a null-distribution is created by randomly assigning subject averages to one of the two conditions a user-specified number of times, and computing the cluster-level statistics for each randomization. Finally, the observed cluster-level test statistics are compared against the null-distribution. When the observed statistic falls in one of the 2.5th percentiles of the null-distribution, the effect is considered significant. Note that this test can only compare two conditions at a time, so we tested each of the three possible contrasts between the three conditions.

In our analysis the distance at which channels were considered neighbors was set such that each channel had an average of 5.1 neighbors. The number of simulations was set to 500. The critical alpha level was set to .05 one-tailed, as the direction of the N400 effect was known and we expected an attenuated N400 effect when the semantically anomalous word had a referent with a similar shape (Shape condition) compared to the Unrelated condition. Reported t -values refer to the cluster-level statistics (“sum- t ”).

*** Insert Figure 3 about here ***

3.5 Results

As can be seen in Figure 3, after word onset the ERPs to the Correct condition remained essentially flat. In contrast, relative to the Correct condition, both types of semantic violation (the Shape and the Unrelated conditions) elicited a slow potential change of negative polarity between ~150 to 800 ms. These negative waves had an early onset, which could be a reflection of mismatches between expected and encountered segmental information similar to the N200 reported by van den Brink, Brown, and Hagoort (2001). These early negativities were not of interest for the present study, and further discussion will focus on comparisons between conditions in the later N400 time windows.

The results of the statistical analyses are shown in Figure 4. The Unrelated condition elicited a significantly more negative voltage than the Correct condition in a large cluster of electrodes covering the fronto-central scalp. This was the case in the first time window, between 300 and 500 ms after word onset, $\text{sum-}t = -418.14$, $p < .001$, as well as the second window, between 500 and 700 ms after word onset, $\text{sum-}t = -388.18$, $p < .001$. The ERPs for the Shape condition were also significantly more negative-going than those for the Correct condition in both the 300-500 ms and the 500-700 ms time windows, $\text{sum-}t = -318.88$, $p < .001$, and $\text{sum-}t = -220.32$, $p < .01$, respectively. All of these clusters were widely distributed across the scalp, consistent with an N400. However, there was a frontal emphasis which is slightly atypical, though the N400 tends to be more frontally distributed during listening than during reading (McCallum, Farmer, & Pocock, 1984) as well as for concrete words compared to abstract words (Kounios & Holcomb, 1994). The more frontal distribution might also be due to offsets of the ERPs

of the preceding word or to overlap with the contingent negative variation. Critically, the slow negative wave signaling a semantic mismatch was of lower amplitude in the Shape condition than in the Unrelated condition. The attenuation for the Shape condition had a discrete onset, as reflected in the waveforms of Figures 3 and 4 and in the relevant statistics. Thus, the difference between the Shape condition and the Unrelated condition was not significant in the early time window from 300-500 ms (no clusters were detected), but was significant in the later window from 500-700 ms, $\text{sum-}t = -46.70$, $p = .036$.² The cluster for the difference between the Shape and Unrelated condition was smaller than those for the other comparisons.³

*** Insert Figure 4 about here ***

3.6 Discussion

Semantic violations elicited N400 effects relative to correct words. Importantly, the N400 amplitude was significantly attenuated for the Shape condition relative to the Unrelated condition between 500 and 700 ms after word onset. The fact that the N400 amplitude in response to semantic violations was modulated by shape similarity between referents of presented words (e.g., "tomato" vs. "rice") and words that were predictable but not presented (e.g., "moon") is consistent with the hypothesis that predicted information can be very specific and perceptual in nature, involving the expected

² A repeated-measures ANOVA on ERPs averaged over this cluster of electrodes yielded the same pattern of results, including a difference between the Unrelated and Shape conditions in the 500-700 ms time window, $F(1,23) = 5.53$, $p = .028$.

³ Following suggestions made by reviewers of the paper, we further examined whether a more graded effect of shape-relatedness was detectable by correlating the amplitude of the N400 with the ratings of shape relatedness obtained for the individual items. These analyses, which did not yield any significant results, are reported in Appendix B.

referent's shape. The results fit with previous studies indicating high specificity of anticipation at the lexical level (DeLong et al., 2005; van Berkum et al., 2005; Wicha et al., 2004) and extend these findings to the semantic/conceptual level.

Before turning to the general implications of our study we will discuss four alternative explanations of the ERP results. First, one might argue that the shape effect (i.e., the difference between the Shape and the Unrelated condition) occurred relatively late (between 500 and 700 ms after critical word onset) and therefore does not reflect the 'typical' N400 component. The effect observed in the present study indeed occurred somewhat later than previously reported semantic category effects in the auditory modality (Federmeier, McLennan, De Ochoa, & Kutas, 2002), though time windows after 500 ms are very common in N400 studies using auditory stimuli (e.g., Holcomb & Andersson, 1993; Holcomb & Neville, 1991; McCallum, Farmer, & Pocock, 1984; Perrin & García-Larrea, 2003). In our study, the shape effect may have arisen relatively late because many of the critical words were polysyllabic, or because shape information might not receive the same degree of priority as semantic category information in facilitating the processing of unexpected words. Determining the origin of the latency difference between semantic and shape information is difficult however, given that the semantic and shape effects were seen in separate studies, differing in materials, languages, and participants.

Related to this, the difference between the expected and unexpected words was observed well before the shape effect. We interpret the early difference as an acoustic rather than a semantic effect, reflecting the differences between the expected and actual speech signal, similar to the N200 reported by van den Brink, Brown, and Hagoort

(2001). The ERPs elicited by the Shape and Unrelated conditions overlap initially because in both cases the expected and presented word mismatch acoustically. The later timing of the shape effect follows from the fact that, in contrast to the earlier difference, the shape effect requires access to the meaning of the presented word (e.g., "tomato"). We regard the clear electrophysiological difference between the Shape and Unrelated conditions as the main finding; whether or not this difference should be called an N400 effect is, in our view, of secondary importance.

Second, as we measured the consequences of expectations in terms of N400 amplitude in response to semantic violations, we cannot rule out that the effect reported here might reflect shape activation arising from the process of dealing with disconfirmed predictions. However, two observations argue against this interpretation. First, our plausibility ratings did not differ between the Shape and Unrelated conditions; both conditions were outright semantic violations that should lead to similar reanalysis processes. Second, reanalysis processes, particularly those involved in revising a strong prediction when encountering unexpected input, are usually associated with late positivities instead of the negativity observed here (cf. Federmeier, 2007).

A third alternative might be that our results are due to integration rather than semantic expectancy. We consider this to be very unlikely given that the plausibility ratings for our materials showed that the shape-related violations (e.g., "tomato" instead of "moon") were not more plausible substitutes for the correct words than the unrelated violations (e.g., "rice").

Fourth, it might be the case that the shape-related words were also associatively related to the targets. We tested this possibility, as far as possible, post-hoc using Dutch

association norms (de Groot & de Bil, 1987; Lauteslager, Schaap, & Schievels, 1986; van Loon-Vervoorn & van Bakkum, 1991) which had been collected from 100 participants with each participant stating the first association that came to mind, such that association strength ranges from 0 to 100. These revealed no associative relationships between the critical words in either direction for any of the 50 word pairs for which norms were available. Thus, we found no support for this explanation. A related possibility is that the shape effect could have been driven by associations between the shape-related words and words in the context sentence. We examined how strongly the content words in the context sentences (context words hereafter) were associated to each of the three critical words. Thus, for the sentence "In 1969 Neil Armstrong was the first man to set foot on the moon/tomato/rice", we examined whether any of the content words that were not names (e.g., "first", "man", "set", "foot") were associated to the critical words (e.g., "moon", "tomato", or "rice"). Of the 545 context words, 327 appeared in the norms. Of these, 60 words were associated to Correct critical words, with the average association strength being 8.583 (average across all 327 words: 1.490), one word was associated to a Shape-related critical word with an association strength of 11 (average across all words: .039), and three words were related to Unrelated critical words, with an average strength of 7 (average across all words: .047). These analyses show that there was some degree of associative relatedness between the words in the context and the Correct critical words, but not between the words in the other two conditions.

In sum, we found no evidence to support alternative explanations of our result. Instead, the ERP difference between the Shape and Unrelated conditions most likely arose because the referents of the critical words in the Shape condition were similar in

physical shape to the referents of the predicted words, which was not the case for the Unrelated condition.

4. General Discussion

The present study investigated the type of information activated during predictive sentence processing. More specifically, we asked whether listeners activate information about the shape of objects before the objects are referred to. A secondary question was whether activation of object shape information during language processing, previously mainly observed in meta-linguistic tasks (such as lexical decision or sentence picture verification) or with an extended preview of visual stimuli (as in most previous visual world studies) could be observed when visual stimuli were only presented for a brief period (Experiment 1) or not at all (Experiment 2).

Experiment 1 used the visual world paradigm. The visual stimuli were shown only 500 ms before the onset of the critical spoken word. On critical trials, the final word of the spoken sentences could readily be predicted on the basis of the preceding sentence context. We found that participants were more likely to look at objects that were related in shape to the predictable target than at unrelated objects. Importantly, these fixations were initiated before the predicted word itself was processed. Thus, the shape of the referent object was anticipatorily activated. Experiment 2 was an EEG study. Participants listened to sentences ending in a highly predictable word, or in a semantically anomalous word with a referent that had a shape similar to the referent of the expected word, or in an unrelated word. Both anomalous conditions elicited N400 effects relative to the correct condition. Importantly, the N400 effect was attenuated by shape similarity between the referents of the expected and the encountered words. Our findings that shape similarity

influenced both anticipatory eye movements and N400 amplitude provide strong converging evidence for the notion that object shape representations can form part of the contents of listeners' predictions for upcoming meaning.

The experimental results also have implications for the generality of shape effects. The fact that we saw shape effects in both experiments provides evidence that activation of perceptual attributes in language comprehension is neither limited to meta-linguistic tasks nor to a particular type of online method (visual world eye-tracking or ERPs). The present results extend previous research by showing that a relevant visual context is not required for such effects to occur. Of course, our findings do not imply that shape representations, or other representations of physical properties of the objects mentioned, are always activated during language processing. Determining when listeners and readers are likely, or unlikely to activate such information is an important issue for further research.

In addition to clarifying the types of information that can be predicted, the observed shape effect may also be useful in distinguishing between concept-specific prediction versus graded prediction of a range of concepts. In the case of functional semantic features such as edibility, anticipatory eye movements (Altmann & Kamide, 1999) could arise from specific prediction (e.g., in their example, from prediction of cake, especially when a cake is present in the display), or it could arise from prediction of a range of candidates (e.g., edible things) that could plausibly be mentioned next. This is because it is exactly this functional feature (being edible) that allows these objects to play a given role in the context. Either or both of these processes – prediction of one specific concept or of a range of candidates – could also have driven the reduction of N400

amplitude to words semantically related to an expected word in the Federmeier and Kutas (1999) study. In contrast, object shape representations are less likely to be shared across a class of expected objects, because having a similar shape does not necessarily allow different objects (e.g., tomato and moon) to play the same role in a given sentence. Our experiments show that when a specific object (e.g. the moon) is predictable, then this can affect the processing of objects with a similar shape. This evidence for shape pre-activation could potentially be a useful index of the degree of activation of a specific candidate concept (e.g., the moon).

Why does pre-activation of object shape information occur? Further research is required to understand the functionality of the prediction of object shapes. One can speculate that the pre-activation of shape information might represent a link between language and the visual environment, where activating a shape representation would be a useful preparation for object recognition. In visual environments, where speakers sometimes refer to objects that are present, predicting what the shape of an object is could help in directing visual attention to the object before it is actually referred to. Just as prediction of words is thought to speed up language comprehension (e.g., Kutas & Federmeier, 2000), the prediction of shape attributes of objects might, in some contexts, speed up language-vision interactions.

The processing mechanisms underlying the pre-activation of shape information also need to be investigated in further research. Shape information could be activated because the pre-activation of the expected word or concept causes activation to spread to knowledge connected to the lexical and conceptual representation corresponding to the word, which would include shape attributes of the referent (cf. Mahon & Caramazza,

2008). In addition, some form of mediated priming might contribute to the visual anticipation effect we observed (cf. Bar, 2007). For instance, the concept moon might prime the concept round which in turn primes the concept tomato. As Mani and Huettig (in press) have argued, predictive language processing may draw on several mechanisms. An important task for further research is to identify these mechanisms and their impact in language processing.

4.4 Conclusions

In sum, two experiments showed, in quite different ways, that listeners anticipated the shape of objects that were about to be mentioned. The results converge with previous findings from other paradigms and show that meta-linguistic tasks or a visual environment displaying the referents is not necessary for obtaining effects of anticipated shape. Predictive language processing is not limited to functional semantic information but can involve the activation of visual representations.

Acknowledgements

We thank Alma Veenstra for lending her voice to our materials, Tilman Harpe for drawing pictures for Experiment 1, Marloes Wensink and Sophie Kirkels for help with association norms, and three anonymous reviewers for comments.

References

- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73(3), 247-264.
- Altmann, G. T. M., & Mirkovic, J. (2009). Incrementality and Prediction in Human Sentence Processing. *Cognitive Science*, 33(4), 583-609.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7), 280-289.
- Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457-474.
- Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.1).
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84–107.
- Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychological Bulletin & Review*, 12, 455-459.
- de Groot, A. M. B., & de Bil, J. M. (1987). *Nederlandse woordassociatienormen met reactietijden*. Lisse: Swets & Zeitlinger.

- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117-1121.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491-505.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469-495.
- Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133-146.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*(1), 1-76.
- Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes*, *8*(4), 379-411.
- Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, *19*(4), 286-300.
- Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, *15*(8), 985-1018.

- Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, 57(4), 460-482.
- Huettig, F., & McQueen, J. M. (2011). The nature of the visual environment induces implicit biases during language-mediated visual search. *Memory & Cognition*, 39, 1068-1084.
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137, 151-171.
- Junghöfer, M., Elbert, T., Tucker, D. M., & Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, 37(4), 523-532.
- Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, 2, 647-670.
- Kellenbach, M. L., Wijers, A. A., & Mulder, G. (2000). Visual semantic features are activated during the processing of concrete words: event-related potential evidence for perceptual semantic priming. *Cognitive Brain Research*, 10(1-2), 67-75.
- Kounios, J. & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 804-823.
- Kutas, M., DeLong, K. A., Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In: M. Bar (Ed.), *Predictions in the Brain: Using Our Past to Generate a Future*. Oxford University Press.

- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12), 463-470.
- Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, 62, 621-647.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203-205.
- Kutas, M., & van Petten, C. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M.A. Gernsbacher (ed.), *Handbook of Psycholinguistics* (pp. 83–143). San Diego, CA: Academic.
- Laszlo, S. and Plaut, D. C. (2011). Simulating event-related potential reading data in a neurally plausible parallel distributed processing model. *Proceedings of the 33th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Lautelager, M., Schaap, T., & Schieveld, D. (1986). *Schriftelijke woordassociatienormen voor 549 zelfstandige naamwoorden*. Lisse: Swets & Zeitlinger.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.
- Luck, S. J. (2005). Ten simple rules for designing ERP experiments. In T. C. Handy (Ed.), *Event-Related Potentials: A Methods Handbook* (pp. 17-32). Cambridge, MA: MIT Press.

- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology - Paris*, 102(1-3), 59-70.
- Mani, N., & Huettig, F. (in press). Towards a complete multiple-mechanism account of predictive language processing: Commentary on Pickering & Garrod. *Behavioral and Brain Sciences*.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177-190.
- McCallum, W. C., Farmer, S. F., & Pocock, P. V. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology*, 59(6), 477-488.
- Mitchell, D. C. (2004). On-line methods in language processing: Introduction and historical review. In M. Carreiras, & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP and beyond* (pp. 15–32). New York, NY: Psychology Press.
- Moss, H. E., McCormick, S. F., & Tyler, L. K. (1997). The time course of activation of semantic information during spoken word recognition. *Language and Cognitive Processes*, 12, 695-731.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, 2011, 9.

- Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*(4), 713-719.
- Otten, M., & van Berkum, J. J. A. (2008). Discourse-Based Word Anticipation During Language Processing: Prediction or Priming? *Discourse Processes*, *45*(6), 464-496.
- Pecher, D., Zeelenberg, R., & Raaijmakers, J. G. W. (1998). Does pizza prime coin? Perceptual priming in lexical decision and pronunciation. *Journal of Memory and Language*, *38*, 401-418.
- Perrin, F., & García-Larrea, L. (2003). Modulation of the N400 potential during auditory phonological/semantic interaction. *Cognitive Brain Research*, *17*, 36-47.
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105-110.
- Rossion, B., Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*, 217-236.
- Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America*, *57*, 1030-1033.
- Schreuder, R., Flores d'Arcais, G. B., & Glazenborg, G. (1984). Effects of perceptual and conceptual similarity in semantic priming. *Psychological Research*, *45*, 339-354.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632-1634.

- Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly*, *30*, 415-433.
- van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(3), 443-467.
- van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*(7), 967-985.
- van den Brink, D., Brown, C. M., & Hagoort, P. (2006). The cascaded nature of lexical selection and integration in auditory sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 364-372.
- van Loon-Vervoorn, W. A., & van Bekkum, I. J. (1991). *Woordassociatielexicon*. Amsterdam/Lisse: Swets & Zeitlinger.
- van Petten, C., & Luka, B. J. (2011). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*, 176-190.
- Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272-1288.
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168-171.

Table 1

Example sentences

Condition	Example
Correct	In 1969 zette Neil Armstrong als eerste mens voet op de <u>maan</u> . <i>In 1969 Neil Armstrong was the first man to set foot on the <u>moon</u>.</i>
Shape	In 1969 zette Neil Armstrong als eerste mens voet op de <u>tomaat</u> . <i>In 1969 Neil Armstrong was the first man to set foot on the <u>tomato</u>.</i>
Unrelated	In 1969 zette Neil Armstrong als eerste mens voet op de <u>rijst</u> . <i>In 1969 Neil Armstrong was the first man to set foot on the <u>rice</u>.</i>

Note. English translations are provided in Italics. Critical words are underlined.

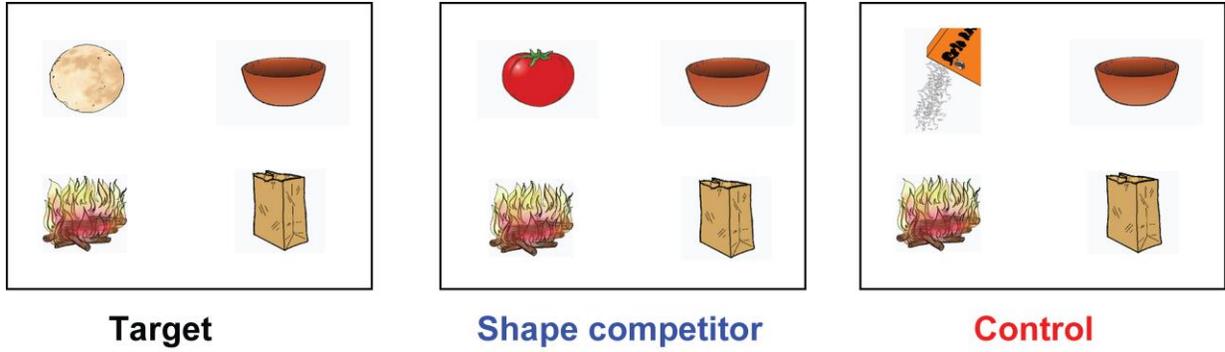


Figure 1. Example displays for each of the three conditions for the sentence "In 1969 Neil Armstrong was the first man to set foot on the moon". Shown along with the three distractors (a bowl, a fire, and a bag): the Target condition (with a moon), the Shape competitor condition (with a tomato), and the Control condition (with rice).

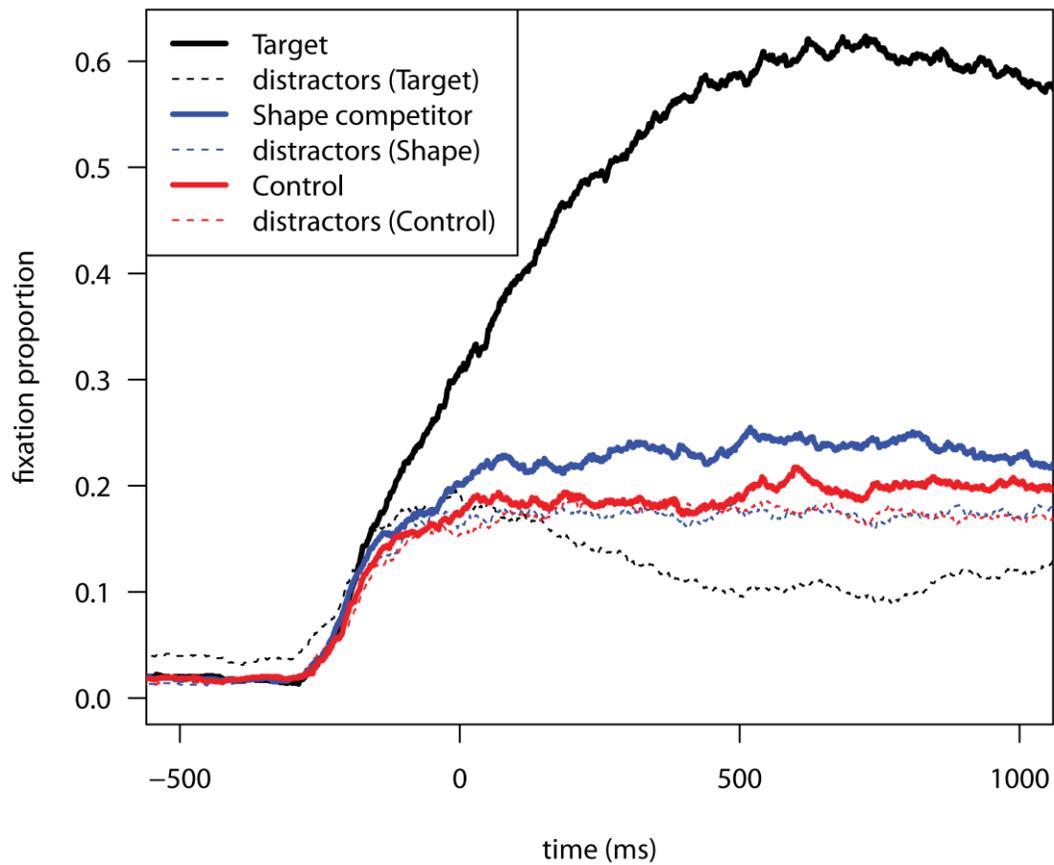


Figure 2. Results of Experiment 1. Time-course graph showing fixation proportions to Targets, Shape competitors, and Control objects (solid lines) along with fixation proportions averaged across the three corresponding unrelated distractors (dashed lines). Display onset was at -500 ms, time zero indicates critical word onset.

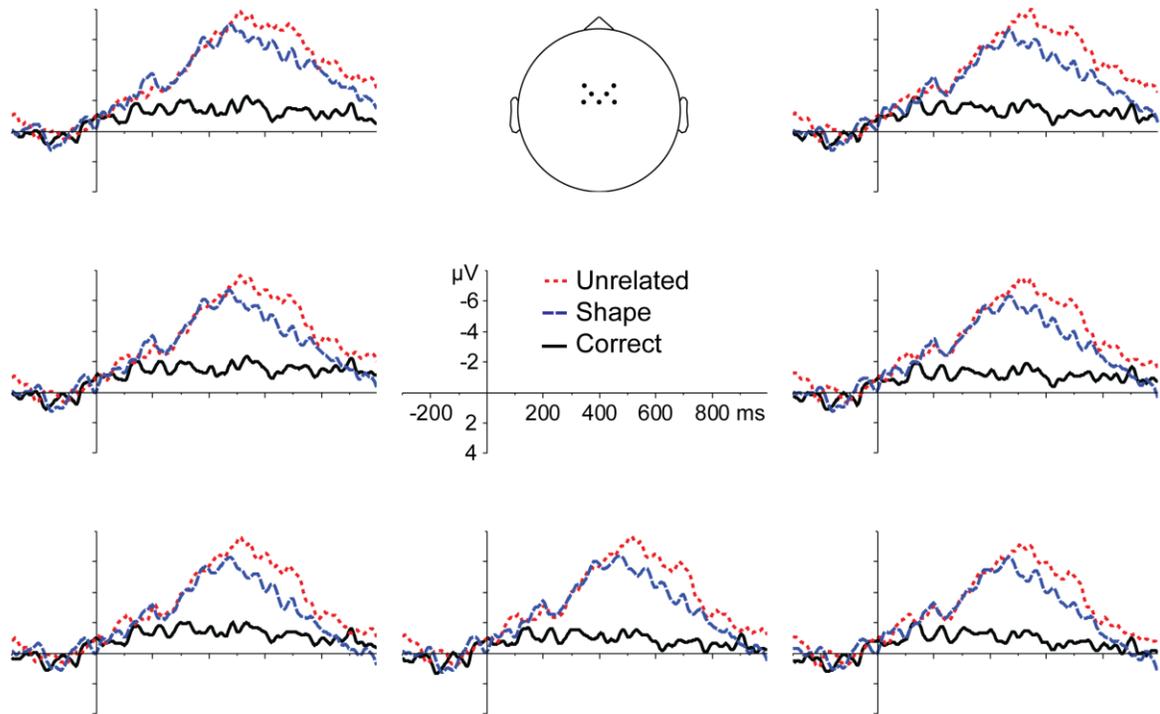


Figure 3. ERP waveforms at seven electrodes of which the locations are indicated on a head map (top), from left to right, top to bottom: FC1, FC2, FCC1h, FCC2h, C1, Cz, C2. Word onset is at zero ms. Negative is plotted up.

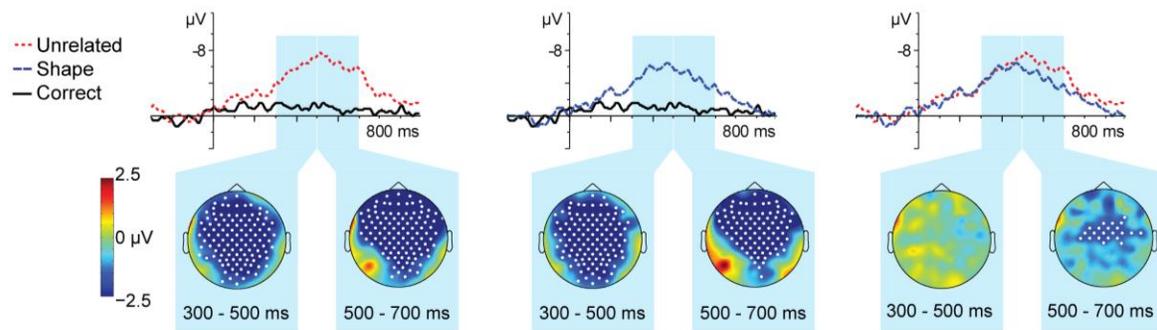


Figure 4. Scalp topographies of the mean difference between each of the conditions (indicated by the two relevant waves taken from the Cz electrode): Unrelated - Correct, Shape - Correct, and Unrelated - Shape. Two time windows are shown. White dots indicate electrodes included in a significant cluster.

Appendix A

Example of one of the twelve sets of 24 sentences in the materials of Experiment 1

Condition	Examples
Correct	<p>Nadat de chocoladeliefhebber haar boterham had besmeerd met boter bestrooide ze hem met flink veel <u>hagelslag</u>.</p> <p><i>After the chocolate lover had spread the bread with butter she covered it with a large amount of <u>chocolate sprinkles</u>.</i></p> <p>In 1969 zette Neil Armstrong als eerste mens voet op de <u>maan</u>.</p> <p><i>In 1969 Neil Armstrong was the first man to set foot on the <u>moon</u>.</i></p> <p>Het echtpaar had pech met het weer en trouwde in de stromende <u>regen</u>.</p> <p><i>The couple was unlucky weather-wise and got married in the pouring <u>rain</u>.</i></p> <p>De Chinees gebruikte altijd eetstokjes bij het eten van zijn witte <u>rijst</u>.</p> <p><i>The Chinese person always used chopsticks when eating his white <u>rice</u>.</i></p> <p>De keukenhulp verkreeg jus d'orange door het persen van een stuk fruit genaamd <u>sinaasappel</u>.</p> <p><i>The kitchen assistant obtained orange juice by squeezing a piece of fruit called <u>orange</u>.</i></p> <p>Het was winter en er vielen uit de lucht dikke witte vlokken <u>sneeuw</u>.</p> <p><i>It was winter and from the sky fell big white flakes of <u>snow</u>.</i></p> <p>Pastasaus krijgt zijn rode kleur door een groente genaamd <u>tomaat</u>.</p> <p><i>Pasta sauce gains its red color from a vegetable called <u>tomato</u>.</i></p> <p>Het was warm want gedurende de hele dag scheen de <u>zon</u>.</p> <p><i>It was hot because all day the <u>sun</u> shone.</i></p>

- Shape Nadat de chocoladeliefhebber haar boterham had besmeerd met boter bestrooide ze hem met flink veel regen.
- After the chocolate lover had spread the bread with butter she covered it with a large amount of rain.*
- In 1969 zette Neil Armstrong als eerste mens voet op de tomaat.
- In 1969 Neil Armstrong was the first man to set foot on the tomato.*
- Het echtpaar had pech met het weer en trouwde in de stromende hagelslag.
- The couple was unlucky weather-wise and got married in the pouring chocolate sprinkles.*
- De Chinees gebruikte altijd eetstokjes bij het eten van zijn witte sneeuw.
- The Chinese person always used chopsticks when eating his white snow.*
- De keukenhulp verkreeg jus d'orange door het persen van een stuk fruit genaamd zon.
- The kitchen assistant obtained orange juice by squeezing a piece of fruit called sun.*
- Het was winter en er vielen uit de lucht dikke witte vlokken rijst.
- It was winter and from the sky fell big white flakes of rice.*
- Pastasaus krijgt zijn rode kleur door een groente genaamd maan.
- Pasta sauce gains its red color from a vegetable called moon.*
- Het was warm want gedurende de hele dag scheen de sinaasappel.
- It was hot because all day the orange shone.*
- Unrelated Nadat de chocoladeliefhebber haar boterham had besmeerd met boter bestrooide ze hem met flink veel zon.

After the chocolate lover had spread the bread with butter she covered it with a large amount of sun.

In 1969 zette Neil Armstrong als eerste mens voet op de rijst.

In 1969 Neil Armstrong was the first man to set foot on the rice.

Het echtpaar had pech met het weer en trouwde in de stromende sinaasappel.

The couple was unlucky weather-wise and got married in the pouring orange.

De Chinees gebruikte altijd eetstokjes bij het eten van zijn witte maan.

The Chinese person always used chopsticks when eating his white moon.

De keukenhulp verkreeg jus d'orange door het persen van een stuk fruit genaamd regen.

The kitchen assistant obtained orange juice by squeezing a piece of fruit called rain.

Het was winter en er vielen uit de lucht dikke witte vlokken tomaat.

It was winter and from the sky fell big white flakes of tomato.

Pastasaus krijgt zijn rode kleur door een groente genaamd sneeuw.

Pasta sauce gains its red color from a vegetable called snow.

Het was warm want gedurende de hele dag scheen de hagelslag.

It was hot because all day the chocolate sprinkles shone.

Note. The same sentence contexts and critical words were recombined to create the different conditions. English translations of the Dutch sentences are provided in Italics. In one case in the English translation a critical word consists of two words (*chocolate sprinkles*), but the actual Dutch critical words were all single words. Critical words are underlined.

Appendix B

In supplementary analyses of Experiment 2 suggested by reviewers of the paper, we treated the shape-relatedness of the words' referents to the Correct words' referents in the Shape and in the Unrelated condition as a continuous variable and examined whether there was a correlation between this variable and the amplitude of the N400. If the N400 difference between the Shape and Unrelated condition is due to a difference in the relatedness of the shapes of the referents, such a correlation might be seen. The measure for the degree of shape relatedness between a competitor and a target was the average rating given by the participants in the norming study. All analyses were performed on the EEG signal from the shape cluster reported in the main text.

The first analysis was a within-subjects correlation. For each participant we calculated the correlation coefficient between the shape similarity ratings of the items (stemming from other participants) and the corresponding EEG signal. This was done for all items in the Shape and Unrelated conditions and involved no averaging across trials. The average correlation was $r = -0.007$ (r range -0.328 to 0.269) and did not differ significantly from zero, $t(23) = -0.237$, $p = 0.815$.

The second analysis was an item analysis. We computed the N400 amplitude for each item across participants and correlated it with the average similarity rating. This correlation was also close to 0 ($r = .055$, $p = .642$). Note that each item was heard by only eight participants in the main experiment and rated by different participants in the rating study.

Finally, we sorted the items, according to the average shape similarity ratings, into four 16-item bins of increasing relatedness and assessed whether the amplitude of the

N400 increased across the bins. However, we found that the linear trend was not significant, $F(1,23) = 0.002, p = .966$.

In our view, these results are not very informative. They do not support our assumption that the N400 difference between the Shape and Unrelated condition indeed arose because the items differed in their shape similarity to the targets. However, being null-results, they do not undermine our argument either. Power was low given that in the analyses above each data point stemmed from maximally 1, 8, or 16 trials, respectively. Future research could obtain ERPs with a higher signal-to-noise ratio using many more items (e.g., 30-60; Luck, 2005) and/or participants (e.g., over 100; Laszlo & Plaut, 2011).