

Published in final edited form as:

*Neuropsychologia*. 2014 April ; 56: 147–166. doi:10.1016/j.neuropsychologia.2014.01.007.

## Cross-linguistic variation in the neurophysiological response to semantic processing: Evidence from anomalies at the borderline of awareness

Sarah Tune<sup>1</sup>, Matthias Schlesewsky<sup>2</sup>, Steven L. Small<sup>3</sup>, Anthony J. Sanford<sup>4</sup>, Jason Bohan<sup>4</sup>, Jona Sassenhagen<sup>1</sup>, and Ina Bornkessel-Schlesewsky<sup>1</sup>

<sup>1</sup>Department of Germanic Linguistics, University of Marburg, Marburg, Germany

<sup>2</sup>Department of English and Linguistics, Johannes Gutenberg-University, Mainz, Germany

<sup>3</sup>Department of Neurology, University of California, Irvine, CA

<sup>4</sup>School of Psychology, University of Glasgow, Glasgow, UK

### Abstract

The N400 event-related brain potential (ERP) has played a major role in the examination of how the human brain processes meaning. For current theories of the N400, classes of semantic inconsistencies which do not elicit N400 effects have proven particularly influential. Semantic anomalies that are difficult to detect are a case in point (“borderline anomalies”, e.g. “After an air crash, where should the survivors be buried?”), engendering a late positive ERP response but no N400 effect in English (Sanford, Leuthold, Bohan, & Sanford, 2011). In three auditory ERP experiments, we demonstrate that this result is subject to cross-linguistic variation. In a German version of Sanford and colleagues' experiment (Experiment 1), detected borderline anomalies elicited both N400 and late positivity effects compared to control stimuli or to missed borderline anomalies. Classic easy-to-detect semantic (non-borderline) anomalies showed the same pattern as in English (N400 plus late positivity). The cross-linguistic difference in the response to borderline anomalies was replicated in two additional studies with a slightly modified task (Experiment 2a: German; Experiment 2b: English), with a reliable LANGUAGE × ANOMALY interaction for the borderline anomalies confirming that the N400 effect is subject to systematic cross-linguistic variation. We argue that this variation results from differences in the language-specific default weighting of top-down and bottom-up information, concluding that N400 amplitude reflects the interaction between the two information sources in the form-to-meaning mapping.

© 2014 Elsevier Ltd. All rights reserved

Address correspondence to: Ina Bornkessel-Schlesewsky Department of Germanic Linguistics University of Marburg Wilhelm-Roepke-Strasse 6A 35032 Marburg Germany phone: +49 - (0)6421 - 2824675 fax: +49 - (0)6421 - 2824558 iboke@staff.uni-marburg.de.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

<sup>4</sup>Frequency classes are computed in relation to the most frequent word found in the corpus for a particular language. For example, if a word is placed in frequency class 11 this means that the most frequent word has 2<sup>11</sup> times the number of occurrences of the selected word.

## Keywords

Language processing; cross-linguistic differences; borderline anomalies; shallow processing; N400; P600; late positivity; bidirectional coding account; top-down; bottom-up

---

## 1. INTRODUCTION

In everyday life, we use language to express our thoughts and to comprehend those around us. We make use of language in such a natural and seemingly effortless way that we are mostly unaware of the complex cognitive system that makes this possible. When processing speech or written language, we are faced with a difficult task, requiring us not only to combine words to form complex meanings, but also to assess whether the state of affairs described is consistent with what we already know about the world.

While the matching of linguistic meaning to world knowledge may appear *prima facie* to be straightforward, it is not always performed completely. Rather, under certain circumstances, we miss violations of our real world knowledge. A case in point is the so-called Moses illusion (Erickson & Matteson, 1981), a relatively robust failure to detect a distorted meaning in cases where a locally implausible phrase nevertheless exhibits a close fit to the global context. Erickson and Matteson asked people the now famous question “How many animals of each kind did Moses take on the Ark?” and reported that most people answered the question with “two” in spite of the fact that it was Noah, not Moses, who built and sailed the ark.

This type of “semantic illusion” has given rise to a great deal of research in theoretical and psychological linguistics, aiming to shed light on the linguistic basis of such illusions and the mechanisms involved in processing them (e.g. Ferreira, Ferraro, & Bailey, 2002; Sanford & Sturt, 2002; Sanford & Graesser, 2006). While the studies concerned with this particular phenomenon have employed a variety of materials and paradigms, there are several common results: First is that the Moses illusion effect generalises to other sentence materials (e.g. the “survivors illusion” in (1), cited from Sanford et al., 2011). Further, the illusion occurs at comparable rates independent of the number of times it is presented (detection rates at approximately 60%) or the task demands, i.e., incidental detection or an explicit judgement task (e.g. Reder & Kusbit, 1991; Barton & Sanford, 1993; Daneman, Reingold, & Davidson, 1995; Hannon & Daneman, 2001; Hannon & Daneman, 2004). However, detection rates are subject to more substantial variation when linguistic factors such as focus, sentence structure or semantic relatedness are manipulated (Shafto & McKay, 2000; Büttner, 2007). In accordance with the terminology in Sanford et al. (2011), we shall refer to sentences constructed in the spirit of the Moses Illusion (such as 1) as “borderline anomalies”, as an abbreviation of “anomalies at the borderline of awareness”.

(1) When an airplane crashes on a border with debris on both sides, where should the survivors be buried?

From the perspective of sentence understanding, a main interest in examining borderline anomalies such as (1) relates to questions about depth of processing. Specifically, it has been argued that referents with a good fit to the global discourse context (such as *survivors* in the context of an airplane crash) give rise to *shallow processing*, i.e. are not as deeply probed for their meaning in comparison to referents with a lower degree of contextual fit (Sanford & Garrod, 1998). In support of this proposal, Barton and Sanford (1993) found that the “survivor-anomaly” in (1) is detected much more readily in the context of a bicycle crash

than in the context of an airplane crash, since, statistically, the word *survivors* is much more likely to be used in the latter case.

More recent studies have examined how borderline anomalies are processed during on-line comprehension, focusing particularly on whether they disrupt processing even when they are not detected. Results from both eye tracking (Bohan & Sanford, 2008) and event related brain potentials (Sanford et al., 2011) suggest that this is not the case: neither eye movement nor event-related potential (ERP) records reveal differences between the non-detected borderline anomalies and their plausible counterparts. On the basis of their results, Sanford and colleagues conclude that borderline anomalies are indeed subject to shallow processing, arguing against an alternative account in which such anomalies disrupt processing, but not enough to reach conscious awareness. A sample item from Sanford et al. (2011) is given in (2). ERPs were measured at the underlined word, with the context words differentiating between the borderline anomaly and the plausible control given in italics and curly brackets.

(2) Child abuse cases are being reported much more frequently these days. In a recent trial, a 10-year {sentence / care order} was given to the victim, but this was subsequently appealed.

Of particular interest is that the detected anomalies in Sanford and colleagues' (2011) study engendered a late positivity but no N400 effect, when compared to control stimuli. These findings may contribute to a better understanding of N400 effects more generally, an important issue that is the subject of active debate, particularly related to the on-line processing of sentence meaning. Since first reported by Kutas and Hillyard (1980), the N400 has been viewed as a correlate of lexical-semantic processing. However, there are differing perspectives on the reasons for this correlation (for a recent review, see Lau, Phillips, & Poeppel, 2008). According to the "integration" view, N400 amplitude reflects the ease or difficulty with which a new word can be semantically integrated into an existing sentence context (e.g. Hagoort & van Berkum, 2007; Hagoort, 2008). By contrast, the "lexical pre-activation" view is that the N400 reflects the ease with which that word can be accessed in semantic memory (e.g. Kutas & Federmeier, 2000; Lau et al., 2008; Brouwer, Fitz, & Hoeks, 2012; Stroud & Phillips, 2012). Sanford et al.'s (2011) findings appear to support the lexical view: in the borderline anomalies, the critical word that would be considered "pre-activated" in light of its good lexical semantic fit to the global context induced an anomaly but no increased N400 effect. Similar conclusions follow from research on so-called "semantic reversal anomalies". In these sentences, exemplified by *For breakfast, the eggs would only eat toast and jam* (Kuperberg, Sitnikova, Caplan, & Holcomb, 2003) and *The hearty meals were devouring the kids* (Kim & Osterhout, 2005), the thematic roles and their arguments are misaligned (i.e. *eggs* and *hearty meals* are highly plausible Theme arguments of *eat* and *devour*, respectively, but implausible Agents). Like the borderline anomalies, semantic reversal anomalies have been shown to engender late positivity but not N400 effects in English (e.g. Kuperberg et al., 2003; Kim & Osterhout, 2005) and Dutch (e.g. Kolk, Chwilla, van Herten, & Oor, 2003; Hoeks, Stowe, & Doedens, 2004). This result, which sparked a great deal of discussion (for recent reviews, see Bornkessel-Schlesewsky & Schlewsky, 2008; van de Meerendonk, Kolk, Chwilla, & Vissers, 2009), appears to follow straightforwardly from the lexical preactivation account of the N400: as in the borderline anomalies, the critical word is lexically associated with the sentence context, but is anomalous within the sentence *per se*. The absence of an increased N400 effect for these sentences seems to suggest that lexical preactivation, rather than semantic integration or composition, is the critical factor determining N400 amplitude.

Interestingly, cross-linguistic variation in ERP responses to semantic reversal anomalies represents an additional complicating factor in characterizing the N400. In contrast to English and Dutch, German, Turkish and Chinese do show N400 effects for reversal

anomalies (Bornkessel-Schlesewsky et al., 2011; Schlesewsky and Bornkessel-Schlesewsky, 2009). In German, this N400 forms part of a biphasic response, incorporating an N400 followed by a late positivity.<sup>1</sup> Bornkessel-Schlesewsky and colleagues (2011) argue that the presence or absence of the N400 for reversal anomalies is determined by the extent to which sentence interpretation relies on word order (termed “sequence dependence” in Bornkessel-Schlesewsky et al., 2011). In English and Dutch, word order is by far the most important cue for sentence interpretation (MacWhinney, Bates, & Kliegl, 1984; Bates, Devescovi, & Wulfeck, 2001), while a variety of cues must be taken into account in German, Turkish and Chinese (including, for example, case marking and animacy).<sup>2</sup> These cross-linguistic results present a challenge for the lexical preactivation view of the N400, since all the sentences examined in each of these languages contained strongly associated nouns and verbs. From the cross-linguistic results, it appears that the N400 is sensitive to the differential weighting of information sources across languages. Moreover, this suggests that semantic inconsistencies are processed differently in languages that rely primarily on one information source during sentence comprehension (such as English) compared to languages which rely on more than one (such as German). Therefore, it may be the case that these “single source” languages (i.e. languages with one dominant cue) are more susceptible to a temporary “blindness” to semantic anomalies, as reflected by the absence of an N400 for detected anomalies.

In the present study, we aimed to examine whether this type of cross-linguistic variation does in fact generalise to borderline anomalies, which in English appear analogous to reversal anomalies. If borderline anomalies also engender a biphasic N400 - late positivity response in German, this would provide us with strong evidence against a purely lexical account of the N400.

## 2. EXPERIMENT 1

Experiment 1 was designed as a German version of Sanford et al.'s (2011) ERP study. Materials were kept as closely comparable to those used in the original experiment as possible (given that they had to be translated) and the experimental task and procedure was identical.

### 2.1 Materials and Methods

**2.1.1 Participants**—Twenty-nine monolingually raised native speakers of German participated in the experiment after giving informed consent (15 women, mean age 23.8, range 18-31). All were right-handed (as assessed by an adapted German version of the Edinburgh Handedness Inventory; (Oldfield, 1971), had normal or corrected-to-normal vision and no known neurological or auditory disorders. Six participants were excluded from the analyses: two due to excessive artifacts and/or incomplete recording of the EEG data and four because of exceptionally high detection rates of above 80%, leaving fewer than 15 artifact-free missed anomaly trials for the averaging procedures.

**2.1.2 Materials**—The materials used in the present study were a translated and adapted version of the English stimuli employed by Sanford et al. (2011). The pool of items contained both hard-to-detect (“borderline”) anomalies and more classic easy-to-detect

<sup>1</sup>The presence or absence of the late positivity for reversal anomalies is also subject to cross-linguistic variation, though along a different dimension to the N400. However, since it is the presence or absence of the N400 that is central to the present paper, we refer the interested reader to Bornkessel-Schlesewsky et al. (2011) for details on the variation of the positivity.

<sup>2</sup>This proposal was further supported by an experiment on Icelandic, in which Bornkessel-Schlesewsky et al. (2011) examined reversal anomalies with different verb classes, one of which called for strongly sequence-dependent interpretation, while the other did not. Strikingly, results revealed an English-type response (a monophasic late positivity with no N400) for the sequence-dependent verbs, but a German-type response for the other verb class (a biphasic N400 - late positivity pattern).

anomalies (i.e., words with a poor fit to the context) that served as filler trials. Some items needed to be excluded because the strength of the semantic illusion was weakened by translation to German or because they relied on knowledge that could not be presumed for German participants. Other items were modified in the sense that British characters, places and names were replaced by German equivalents to render the materials more relevant and applicable to the targeted test subjects.

All materials were pre-tested in two questionnaire studies. Questionnaire 1 (n=70) ensured that borderline anomalies were reliably missed some of the time (hence allowing for an analysis of both detected and undetected borderline anomalies in the ERP study) and that classic easy-to-detect (“poor fit”) anomalies were detected at least 95% of the time. For this purpose the stimuli were distributed across ten lists, each containing 13 borderline anomalies, 16 easy-to-detect anomalies and equal numbers of non-anomalous control items for both anomaly types (i.e. 58 stimuli in total per list). The lists were then pseudo-randomised and each final version was presented to seven participants, who were asked to indicate and explain any detected anomalies. Borderline anomalies detected at a rate of 75% or higher were modified or excluded. The results of the questionnaire study showed that 68% of the presented borderline anomalies were correctly judged as being implausible. To attain the same number of sentences used in the English ERP study by Sanford and colleagues, 20 new items were created to replace excluded trials.

The final set of materials was further subjected to an additional questionnaire study (Questionnaire 2), in which we tested the contextual fit of the critical word. As in Sanford et al.'s (2011) study, this was accomplished by asking participants to judge the relevance of the critical word to the situation on a 7-point Likert scale (1 = “does not fit”, 7 = “perfect fit”). Twenty participants rated the materials that were equally distributed across two lists. Borderline anomalies and poor fit anomalies yielded mean ratings of 5.02 and 2.20, respectively. Thus, borderline anomalies showed a significantly better contextual fit than their poor fit counterparts ( $t(210.201) = 18.54, p < 0.0001$ ). To account for unequal variances as indicated by Levene's test for homogeneity of variance, Welch's correction for the degrees of freedom was used. Importantly, the mean values for both anomaly types were highly comparable to those of Sanford and colleagues' materials (borderline anomalies: 5.16; poor fit anomalies: 2.17), thus demonstrating that contextual fit did not vary as a function of language.

In total, 215 stimulus pairs consisting of an anomalous condition and a corresponding plausible control condition were constructed, 135 pairs for the borderline anomalies and 80 pairs for the easy-to-detect “poor fit” anomalies. All stimuli were composed of two semantically connected sentences, with the first sentence providing context, and a second, critical sentence, containing a target word to which ERPs were timelocked. All critical sentences consisted of 17 words; however, sentence structures and linguistic methods of inducing the anomalies differed across anomaly types. In the following, the different layouts will be described on the basis of German and English examples.

**2.1.2.1 Borderline anomalies:** A sample borderline anomaly stimulus together with word-by-word translation from the present study is given in (3a).<sup>3</sup> The corresponding item from Sanford et al. (2011) is shown in (3b).

---

<sup>3</sup>Additional German and English examples as well as a detailed description of the critical words in terms of lexical category, case marking and grammatical function can be found in Appendix A.

(3a) Ein amerikanischer Jumbojet wurde von bewaffneten Extremisten gezwungen, in Kanada zu landen und Experten waren schnell vor Ort, um zu vermitteln.  
*An american jumbojet was by armed extremists forced in Canada to land and experts were quickly on site for to mediate.*

Die Lage beruhigte sich durch die {Verhandlungen/Kontaktaufnahme} der Behörden mit den verängstigten Geiseln, die im Flugzeug saßen.  
*The situation calmed itself by the {negotiations/communications} of-the authorities with the scared hostages who in-the plane sat.*

(3b) A North American jumbo jet was forced at gunpoint to land in Canada, experts were quickly on hand to help.

First of all the authorities' initial {*negotiations/communications*} with the scared and desperate hostages helped calm the situation.

For the stimuli containing borderline anomalies, the second sentence was a thematic continuation of the first sentence and contained two alternative local context words (highlighted in italics) and the critical target word (underlined). The target word was always the 13<sup>th</sup> word position and separated by five words from the contextual manipulation. For most items, the local context was altered by replacing one word only, while for a few items more words needed to be changed. It is the relation between local context word/phrase and target word that determines whether the latter is perceived to be anomalous. Therefore, upon encountering the target word, the listener/reader should be able to judge whether the sentence is plausible without needing any further input. However, borderline anomalies are often missed because the target word is highly associated semantically to the overall context, despite its implausibility for the meaning of the particular sentence in which it appears.

**2.1.2.2 Easy-to-detect anomalies:** Eighty easy-to-detect anomaly sentence sets were used. These sentence pairs were constructed in a similar manner to the borderline anomalies, but the internal structure of the critical sentences was less standardised because in these cases the anomaly is evoked by a single word with a very poor fit to both local and global context. The critical words (highlighted in italics) appeared in different positions across stimulus sentences. This ensured that participants had to pay close attention to the whole sentence and could not predict the critical region in the stimuli presented. An example is given below (4a), again with the corresponding item from Sanford et al. (2011):

(4a) Christian und Julia's gemeinsamer Abend in dem neuen Restaurant war ein richtiger

*Christian and Julia's shared evening in the new restaurant was a proper*  
 Reinformfall.  
*letdown.*

Zuerst servierte ihnen der {Maler/Kellner} die falschen Gerichte und dann wurde ihnen auch noch zu viel berechnet.  
*First served them the {painter/waiter} the wrong meals and then was them also still too much charged.*

(4b) Denise and Fred's date to the new restaurant was a complete disaster.

They were given the wrong meals by the {*painter/waiter*} and then they were overcharged for their meals.

The target words in all sentences were controlled for frequency (using the on-line Wortschatz corpus of the University of Leipzig) and length. The mean frequency class4 for the target words was 12.79 for borderline anomaly items, 13.47 for anomalous target words in easy-to-detect items and 12.37 for their plausible counterparts ( $F < 2$ ). There were also no significant differences in the average length of target words across anomaly types (mean length of target words: 8.1 letters for the borderline condition, 8.1 for poor fit to context anomalies and 7.5 for poor fit to context controls;  $F < 2$ ).



All auditory stimuli were recorded by a native female speaker of German who read the sentences with clear and natural intonation. To ensure equal volume levels, the stimuli were normalised digitally. After recording, trigger points for averaging ERPs were inserted at the spoken onset of each target word.

The stimuli were distributed across six lists, each composed of 215 sentence pairs, consisting of 135 borderline and 80 poor fit stimuli. While poor fit stimuli were divided evenly, of the 135 borderline items, 90 contained an anomaly and the remaining 45 were plausible controls. We employed this asymmetrical design, adopted from Sanford et al. (2011), to obtain a similar number of trials for each of the three experimental conditions (detected anomalies, missed anomalies, plausible controls) for the ERP analysis (based on the results of pre-test Questionnaire 1, in which the detection rate for borderline anomalies was ~60%). Within a final list, an item appeared as either anomalous or control condition, while across all lists, each condition of a stimulus pair was presented at least once. Thus, for borderline stimuli, this was achieved by rotating the materials over three lists with 135 stimuli each. Poor fit materials were divided into two lists consisting of 80 stimuli. Merging each borderline list with each poor fit list yielded the six final lists that were pseudo-randomised for presentation.

**2.1.3 Procedure**—For the experimental sessions, participants were seated in a dimly lit, sound attenuated booth, and listened to stimuli on loudspeakers. Participants were cued visually on a computer monitor. Each trial started with the presentation of a fixation asterisk in the centre of the screen, which was followed after 500 ms by the auditory presentation of the first sentence. After sentence offset, the asterisk remained on the screen for another 500 ms, after which participants were asked to press one of two active buttons on a hand-held game controller to initiate the presentation of the second sentence. Again, visual display of the fixation asterisk preceded auditory presentation by 500ms. After the second sentence ended, the asterisk was presented for another 1000 ms before being replaced by a question mark. The question mark served as a cue for the participants to indicate via button press whether they had detected an anomaly. The maximal response time was set to 3500 ms and the assignment of right and left buttons to the responses “plausible” and “implausible” was counterbalanced across participants. When a sentence was rated as plausible, the next trial started after a 2000 ms blank screen (the inter-trial interval). If participants judged a sentence as implausible, they were asked to verbally explain their decision to the experimenter, who wrote down the explanation and recorded via button press whether the anomaly was indeed detected. There was no time limit for the verbal explanations given for detected anomalies.

Participants were asked to fixate on the asterisk throughout the duration of its presentation (from 500 ms before sentence onset to 1000 ms after sentence offset) and to avoid movements and eye blinks during the presentation of the second sentence. Before the start of the actual experimental session, a training session was conducted to ensure that participants were familiar with the task. Each participant was presented with one of the six lists split into experimental sessions with seven blocks of 27 sentence pairs and a final block of 26 sentence pairs. Between blocks, participants took short breaks.

Since a successful detection of some of the borderline anomalies required a certain level of general knowledge, participants completed a post-experiment multiple-choice test to determine if they understood all borderline anomalies as being semantically implausible. Depending on the experimental list presented, participants answered 32–35 multiple choice questions that contained the critical word and asked for the correct local context word. Five answer options were given, including the presented, incorrect local context word (e.g. “Who built the ark?” A: Noah, B: Moses, C: Jona, D: Adam, E: I don't know). Questions to which

incorrect or no answers were given resulted in exclusion of the respective trial from subsequent analyses. A total of 108 trials (5.2%) and a mean of 4.7 (sd:2.2; range: 1–9) trials per participant were excluded.

**2.1.4 EEG recording and preprocessing**—The EEG was recorded from 25 Ag/AgCl scalp electrodes positioned according to the international 10/10 system by means of an elastic cap (EasyCap GmbH, Herrsching, Germany). The horizontal and vertical electrooculogram (EOG) was monitored by placing electrodes at the outer canthi of both eyes and above and below the right eye, respectively. All EEG and EOG channels were amplified with a BrainAmp amplifier (Brain Products, Gilching, Germany) and digitised at a rate of 500 Hz (ground: AFZ). In recording, the left mastoid served as the online reference electrode, but the EEG signals were rereferenced to linked mastoids offline. Scalp impedances were kept below 5 k $\Omega$ .

As a first step of processing, the EEG data were filtered with a 0.3–20 Hz band-pass filter to eliminate slow signal drifts. Automatic and manual rejections (with an EOG rejection threshold of 40  $\mu$ V) were carried out to discard trials containing EEG or EOG artifacts. Single-subject ERP averages were computed per experimental condition and electrode from –200 to 1200ms relative to the onset of the critical target word. Trials that contained false alarm responses to plausible control sentences, detected anomalies with incorrect explanations, missed easy-to-detect anomalies and items for which false answers were given in the multiple-choice post-test were excluded from the averaging procedure (resulting in an overall loss of approximately 8.7% of the trials). Finally, grand-averages were computed over all participants.<sup>5</sup> Despite the exclusion of six participants, the different experimental lists were still presented equally often (one list was only presented three times, all other lists four times).

**2.1.5 Data analysis**—For the statistical analysis of the ERP data, separate analyses were computed for borderline and easy-to-detect anomalies, since they contained different lexical material. In both cases, repeated-measures ANOVAs involving the factors ANOMALY (for borderline anomalies: detected anomalies vs. missed anomalies vs. plausible controls; for easy-to-detect anomalies: anomalous vs. non-anomalous) and region of interest (ROI) were calculated for mean amplitude values per time window per condition. There were four lateral ROIs consisting of 4 electrodes each: left-anterior (F3, F7, FC1, FC5), right-anterior (F4, F8, FC2, FC6), left-posterior (CP1, CP5, P3, P7) and right-posterior (CP2, CP6, P4, P8). For midline sites, each of the six electrodes (FZ, FCZ, CZ, CPZ, PZ, POZ) made up a ROI of their own. Analyses for lateral and midline ROIs were performed separately. Whenever statistical computation included a factor with more than one degree of freedom in the numerator and sphericity was violated, Huynh-Feldt-corrected significance values are reported (Huynh & Feldt, 1970).

## 2.2 Results

**2.2.1 Detection rates**—Analysis of the behavioural data showed that easy-to-detect anomalies were correctly judged as implausible at a rate of 95.5% (sd: 4.9%), whereas for borderline anomalies the detection rate was only 61.3% (sd: 8.8%). Clearly the participants had little difficulty in categorising easy-to-detect anomalies as implausible, but had more difficulty with the borderline anomalies.

**2.2.2 ERP data**—Figure 1 displays the grand-average ERPs time-locked to the target word for detected borderline anomalies compared to missed anomalies and plausible controls,

<sup>5</sup>Descriptive statistics for the number of trials averaged per condition and experiment are given in Appendix B.



while the comparison of anomalous and non-anomalous easy-to-detect items is shown in Figure 2.

As is apparent from both figures, both types of correctly detected semantic anomalies elicited a negativity between approximately 200–500 ms followed by a late positivity between approximately 600–1100 ms post-onset of the critical word.<sup>6</sup> However, no comparable effects were apparent for the comparison of missed borderline anomalies and plausible controls. Separate statistical analyses were carried out for both anomaly types as well as for lateral and midline regions of interest to confirm the impressions based on visual inspection.

**2.2.2.1 Borderline anomalies:** A repeated-measures ANOVA for lateral electrode sites in the time window of 200–500ms revealed a main effect of ANOMALY [ $F(2,44) = 7.21, p < 0.002$ ] as well as an interaction of ANOMALY x ROI [ $F(6,132) = 4.21, p < 0.003$ ]. Resolving the observed interaction by ROI showed significant effects of ANOMALY in all four lateral ROIs (min:  $F(2,44) = 3.41, p < 0.5$  for the left-anterior ROI; max:  $F(2,44) = 10.19, p < 0.001$  for the right-posterior ROI). We further analysed the main effect of ANOMALY in each of the ROIs by computing pairwise comparisons, correcting for multiple comparisons using a modified Bonferroni procedure (Keppel, 1991). The results showed no significant difference between missed anomalies and plausible controls [all  $F_s < 2.2$ ]. At the same time, detected anomalies differed from both missed anomalies and plausible controls in all lateral ROIs [all  $F_s > 5.2$ ].

The global ANOVA for midline electrodes showed comparable results: a main effect of ANOMALY [ $F(2,44) = 6.9, p < 0.01$ ] and an interaction of ANOMALY and ROI [ $F(10,220) = 3.8, p < 0.01$ ]. When resolving the interaction by ROI, significant effects of ANOMALY were found at all midline sites, with the strongest effect at posterior electrodes [min:  $F(2,44) = 4.6, p < 0.02$  at FCZ, max:  $F(2,44) = 10.4, p < 0.001$  at POZ]. Resolving the ANOMALY effect in each of the midline ROIs showed a pattern similar to that found for lateral ROIs: detected anomalies differed significantly from missed anomalies in all ROIs [all  $F_s > 7.4$ ], while the central and posterior electrodes CZ, PCZ, PZ and POZ also showed a significant distinction between detected anomalies and plausible sentences [all  $F_s > 7.2$ ; FZ and FCZ:  $F_s < 3.9$ ]. Importantly, no such difference was found for the contrast of missed anomalies and plausible sentences in any of the ROIs [all  $F_s < 2.9$ ].

In the 600–1100 ms time window, the statistical analysis revealed a main effect of ANOMALY for both lateral and midline electrode sites [lateral:  $F(2,44) = 29.3, p < 0.001$ ; midline:  $F(2,44) = 42.1, p < 0.001$ ], while only lateral sites showed an interaction ANOMALY x ROI [ $F(6, 132) = 4.6, p < 0.03$ ]. Again, all regions showed effects of ANOMALY, with the interaction due to more pronounced effects of ANOMALY at posterior electrode sites [max:  $F(2,44) = 38.8, p < 0.001$  at the left-posterior ROI; min:  $F(2,44) = 11.6, p < 0.001$  at the left-anterior ROI]. When the individual levels of ANOMALY were compared in a pairwise fashion for each of the lateral ROIs and across all midline electrodes, similar results were found: Detected anomalies differed significantly from both missed anomalies and plausible controls [lateral: all  $F_s > 24.3$  (detected vs. missed) and all  $F_s > 13.3$  (detected vs. plausible); midline:  $F(1,22) = 73.3, p < 0.001$  (detected vs. missed) and  $F(1,22) = 67.7, p < 0.001$  (detected vs. plausible)], while there was no difference between missed anomalies and plausible sentences [lateral: all  $F_s < 0.3$ ; midline:  $F(1,22) = 0.02, p = 0.89$ ].

<sup>6</sup>For detected borderline anomalies we also found an additional negative effect before the onset beginning of the critical word. See for Appendix C for a detailed analysis of this pre-onset negativity.

**2.2.2.2 Easy-to-detect anomalies:** In line with previous results for this type of anomaly, statistical analyses confirmed that implausible words elicited a considerably larger negativity than plausible words in the 200–500 ms time window [lateral:  $F(1,22) = 171.4, p < 0.001$ ; midline:  $F(1,22) = 168.6, p < 0.001$ ]. Interactions of ANOMALY  $\times$  ROI for both lateral and midline electrode sites [lateral:  $F(3,66) = 51.7, p < 0.001$ ; midline:  $F(5,110) = 75.8, p < 0.001$ ] reflected the centro-parietal distribution of the anomaly effect that is typical for an N400. For midline ROIs, the effect increased from anterior to posterior electrodes [min:  $F(1,22) = 47.6, p < 0.001$  at FZ; max:  $F(1,22) = 224.8, p < 0.001$  at POZ]. A similar pattern of results was observed in the analysis of lateral sites [min:  $F(1,22) = 49.5, p < 0.001$  for the left-anterior ROI; max:  $F(1,22) = 247.2, p < 0.001$  for the right-posterior ROI].

As is apparent from Figure 2, anomalous words also elicited a larger late positivity in a time window between 600–1100 ms. Statistical analyses confirmed a main effect of ANOMALY [lateral:  $F(1,22) = 39.8, p < 0.001$ ; midline:  $F(1,22) = 59.2, p < 0.001$ ] and an interaction of ANOMALY  $\times$  ROI [lateral:  $F(3,66) = 54.5, p < 0.001$ ; midline:  $F(5,110) = 32.7, p < 0.001$ ]. Resolving the interaction by ROI indicated that the positivity effect increased from anterior to posterior electrode sites for both lateral [min:  $F(1,22) = 4.9, p < 0.05$  for the left-anterior ROI; max:  $F(1,22) = 72.2, p < 0.001$  for the left-posterior ROI] and midline regions of interest [min:  $F(1,22) = 16.9, p < 0.001$  at FZ; max:  $F(1,22) = 77.9, p < 0.001$  at POZ].

In summary, detected borderline anomalies elicited an N400 effect followed by a late positivity in comparison to both missed anomalies and plausible controls. However, no differences were found between missed anomalies and plausible controls. Classic easy-to-detect anomalies triggered the emergence of an N400 followed by a late positivity.

**2.2.2.3 Comparison of N400 amplitude for borderline versus easy-to-detect anomalies:**

Figures 1 and 2 suggest that the N400 effect is considerably more pronounced for easy-to-detect anomalies (approximately  $-6 \mu\text{V}$ ) than for borderline anomalies (approximately  $-2 \mu\text{V}$ ). To examine whether there was indeed a difference in magnitude, we compared ERP amplitude differences (anomaly-control) in the N400 time window (200–500 ms) with an ANOVA including the factors ANOMALY-TYPE and ROI. This analysis revealed main effects of ANOMALY-TYPE [lateral:  $F(1,22) = 58.76; p < 0.001$ ; midline:  $F(1,22) = 59.38; p < 0.001$ ] and interactions of ANOMALY-TYPE and ROI [lateral:  $F(3,66) = 24.65; p < 0.001$ ; midline:  $F(5,110) = 28.95; p < 0.001$ ]. Resolving the interactions by ROI showed significant effects of ANOMALY-TYPE in all regions, with effects more pronounced at posterior sites [lateral min.:  $F(1,22) = 21.16, p < 0.001$  in the left-anterior region; max:  $F(1,22) = 85.21, p < 0.001$  in the right-posterior region; midline min:  $F(1,22) = 11.70, p < 0.01$  at FZ; max:  $F(1,22) = 93.07, p < 0.001$  at PZ]. Thus, easy-to-detect anomalies indeed showed an N400 effect with a larger magnitude than borderline anomalies and this difference in amplitude was most pronounced in those regions in which the N400 effect was maximal.

## 2.3 Discussion

In terms of detection rates, Experiment 1 showed very similar results to those observed by Sanford et al. (2011). As before, the mean detection rate for borderline anomalies was considerably lower than that for easy-to-detect anomalies. For these anomalies, we observed a biphasic N400 – late positivity pattern, as also observed for English. By contrast, the comparison of electrophysiological responses to detected and non-detected borderline anomalies and their plausible controls revealed a deviation from previous findings for English: in the present study, detected borderline anomalies elicited an N400 effect followed by a late positivity in contrast to missed anomalies and plausible controls, which did not differ from each other. Recall that in the case of closely matched English borderline

anomalies, the neural response to detected anomalies resembled that to non-detected and plausible stimuli in the N400 time range, with a differential effect arising only in the late positivity (Sanford et al., 2011). The results of Experiment 1 thus point to cross-linguistic differences in the processing of detected borderline anomalies (i.e., those that show a close fit to global context).

### 3. EXPERIMENT 2

The comparison between Experiment 1 and the previous findings by Sanford and colleagues (2011) indicates that the neural processing of borderline anomalies differs across languages: while German showed a biphasic N400 - late positivity response to detected anomalies, only a late positivity was observable in English. More recent results, however, suggest that it may be possible to induce N400-like effects for borderline anomalies in English, too, by manipulating task environment. Bohan, Leuthold, Hijikata, and Sanford (2012) report an ERP study using similar materials to those used in their original 2011 experiment, but employing visual presentation and an additional task. After judging whether a given passage was plausible or not (and, in the case of an “implausible” answer, reporting the nature of the anomalous content), participants rated how certain they were of their answer on a 6-point scale. Bohan and colleagues speculate that the difference between their results and the previous findings by Sanford et al. (2011) might be attributable to changes in task demands.

We shall return to the question of how task demands and the cross-linguistic differences proposed here might be integrated within one account of the N400 in the General Discussion. Before addressing this question, however, we sought additional support for task-independent, cross-linguistic variation in the electrophysiological response to borderline anomalies. To this end, we conducted two additional ERP studies (one in German, Experiment 2a; one in English, Experiment 2b) with completely parallel design and analysis procedures. In these experiments, we aimed to reduce the impact of the judgement task as much as possible in order to allow us to examine the “basic” pattern that emerges in each language when task influences are minimised. To this end, we modified the design of Experiment 1 in three ways: (a) context and target sentences were presented with a fixed inter-stimulus interval, thus eliminating participants' control over target sentence presentation; (b) the judgement task only comprised a button press (“plausible” versus “implausible”) but did not require participants to explain the nature of the anomaly following an “implausible” judgement; and (c) the number of trials in the experiment was decreased from 215 to 180 to reduce participants' exposure to the critical manipulation (i.e., the processing and classification of semantically anomalous and non-anomalous sentences).

#### 3.1 Experiment 2a

##### 3.1.1 Materials and Methods

**3.1.1.1 Participants:** Twenty-six monolingually raised native speakers of German participated in the experiment after having given informed consent (13 women, mean age 23.3, range 19–29). None of the participants had taken part in Experiment 1 and parameters for participant inclusion were the same as for Experiment 1. Four participants were excluded due to excessive EEG artefacts.

**3.1.1.2 Materials:** The materials were identical to those employed in Experiment 1 with the exceptions already noted above: the number of critical sentences was reduced from 215 to 180 by removing sentences from the easy-to-detect anomalies and the plausible borderline condition. In Experiment 2a, each participant thus heard 90 borderline anomaly sentences and 30 controls, as well as 30 sentences in each of the easy-to-detect anomaly and control conditions, respectively. The borderline anomalies were selected on the basis of detection

rates in Experiment 1, i.e. the items that were excluded were those that had shown the highest by-item detection rates in Experiment 1.

**3.1.1.3 Procedure:** The experimental procedure was identical to that in Experiment 1 with the two exceptions mentioned above: (a) context and target sentences were presented with a fixed ISI of 1000 ms; (b) the plausibility judgement consisted only of a button press, i.e. participants were not required to explain why they considered sentences implausible. The maximal reaction time was set at 2000 ms. In addition, in order to avoid anticipatory motor response preparation following the processing of the critical word, Experiment 2 did not employ a fixed assignment of push-buttons to the “plausible” and “implausible” categorisations per participant. Rather, the assignment of the left and right buttons to “plausible” and “implausible” responses varied on a trial-by-trial basis and was signalled by two smiley faces (one laughing and one frowning). Across each session, the assignment of the “plausible” and “implausible” categories to the left and right buttons was counterbalanced.

**3.1.1.4 EEG data recording and preprocessing:** EEG data recording and preprocessing was identical to Experiment 1 with the exception that eye movement artefacts were corrected using an independent component-analysis (ICA) based correction method. The ICA correction was employed in order to ensure that data analysis was comparable to Experiment 2b, in which it was necessary in order to avoid the loss of too many trials due to eye movement artefacts. To this end, we calculated an Extended Infomax ICA for each participant and subsequently selected template ICs for blinks and saccades, respectively. The two ICs best correlating with vertical and horizontal EOG templates were identified using an automatic procedure (Viola et al., 2009) and subsequently subtracted from the raw EEG data.

**3.1.1.5 Data analysis:** Data analysis was undertaken in an identical manner to Experiment 1.

### 3.1.2 Results

**3.1.2.1 Detection rates:** Analysis of the behavioural data showed that easy-to-detect anomalies were correctly judged as implausible at a rate of 95.6% (sd: 4.2%), whereas for borderline anomalies the detection rate was 71.2% (sd: 9.4%). To compare whether accuracy in Experiment 2a differed from that of Experiment 1, we computed a behavioural analysis for Experiment 1 that included only the materials that were used both German experiments. This yielded a detection rate for borderline anomalies of 63.5% (sd:9.3%) that differed significantly from that of Experiment 2a [ $t(42.878) = -2.79, p = 0.008$ ] and an accuracy of 95.3% (sd:4.6) for easy-to-detect anomalies that did not differ from that of Experiment 2a [ $t(42.871) = -0.21, p = 0.83$ ].

**3.1.2.2 ERP data: Borderline anomalies:** Figure 3 shows grand average ERPs timelocked to the critical word in the borderline anomaly conditions and the corresponding plausible controls. As is apparent from the Figure, the findings from Experiment 2a replicate those of Experiment 1: detected borderline anomalies engendered a biphasic N400 - late positivity response in comparison to missed anomalies as well as plausible controls.<sup>7</sup>

A repeated-measures ANOVA for lateral electrode sites in the time window of 200–500ms revealed a main effect of ANOMALY [ $F(2,42) = 11.48, p < 0.001$ ]. We further analysed the

<sup>7</sup>As in Experiment 1, for detected borderline anomalies an early negativity prior to the onset of the critical word was observed relative to non-detected and plausible control sentences. See Appendix C for analysis and discussion of this effect.

main effect of ANOMALY by computing pairwise comparisons, correcting for multiple comparisons using a modified Bonferroni procedure (Keppel, 1991). The results showed no significant difference between missed anomalies and plausible controls [ $F < 1$ ]. At the same time, detected anomalies differed from both missed anomalies [ $F(1,21) = 14.86, p < 0.001$ ] and plausible controls [ $F(1,21) = 24.06, p < 0.0001$ ].

The global ANOVA for midline electrodes showed comparable results: a main effect of ANOMALY [ $F(2,42) = 11.00, p < 0.001$ ] and an interaction of ANOMALY and ROI [ $F(10,210) = 2.03, p < 0.05$ ]. When resolving the interaction by ROI, significant effects of ANOMALY were found at all midline sites except FZ, with the effect strongest at CPZ [ $F(2,42) = 13.89, p < 0.0001$ ]. Resolving the ANOMALY effect in each of the midline ROIs showing a significant main effect of ANOMALY revealed a pattern similar to that found for lateral ROIs: detected anomalies differed significantly from missed anomalies in all ROIs [all  $F_s > 5.7$ ] as did detected anomalies and plausible sentences [all  $F_s > 4.8$ ]. As in Experiment 1, no difference was found for the contrast of missed anomalies and plausible sentences in any of the ROIs [all  $F_s < 1$ ].

In the 650–1100 ms time window,<sup>8</sup> the statistical analysis revealed a main effect of ANOMALY for both lateral and midline electrode sites [lateral:  $F(2,42) = 8.14, p < 0.01$ ; midline:  $F(2,42) = 9.27, p < 0.001$ ], while only lateral sites showed an interaction ANOMALY x ROI [ $F(6, 126) = 2.95, p < 0.05$ ]. Again, all regions showed effects of ANOMALY, with the interaction due to more pronounced effects of ANOMALY at posterior electrode sites [max:  $F(2,42) = 10.76, p < 0.001$  at the left-posterior ROI; min:  $F(2,42) = 3.24, p < 0.05$  at the left-anterior ROI]. When the individual levels of ANOMALY were compared in a pairwise fashion for each of the lateral ROIs and across all midline electrodes, similar results were found: Detected anomalies differed significantly from plausible controls in all regions [lateral: all  $F_s > 5.3$ ; midline:  $F(1,21) = 19.6, p < 0.001$ ] and from missed anomalies in posterior lateral ROIs [all  $F_s > 5.3$ ] as well as for midline sites [ $F(1,21) = 7.92, p < 0.05$ ]. A difference between missed anomalies and plausible sentences was observed in the left-posterior ROI [ $F(1,21) = 6.30, p < 0.05$ ].

**3.1.2.3 ERP data: Easy-to-detect anomalies:** ERP results for the easy-to-detect anomalies are shown in Figure 4. Here, the findings of Experiment 2a again replicate those of Experiment 1, with the semantically anomalous condition eliciting a biphasic N400 - late positivity pattern in comparison to the plausible control condition.

The statistical analyses in the 200–500 ms time window showed a main effect of ANOMALY [lateral:  $F(1,21) = 42.19, p < 0.0001$ ; midline:  $F(1,21) = 43.91, p < 0.0001$ ]. Interactions of ANOMALY x ROI for both lateral and midline electrode sites [lateral:  $F(3,63) = 19.1, p < 0.0001$ ; midline:  $F(5,105) = 21.8, p < 0.0001$ ] reflected the centro-parietal distribution of the effect. For midline ROIs, the effect increased from anterior to posterior electrodes [min:  $F(1,21) = 6.55, p < 0.05$  at FZ; max:  $F(1,21) = 74.10, p < 0.0001$  at POZ]. A similar pattern of results was observed in the analysis of lateral sites [min:  $F(1,21) = 4.40, p < 0.005$  for the right-anterior ROI; max:  $F(1,21) = 85.2, p < 0.0001$  for the right-posterior ROI].

For the late time window (650–1100 ms), statistical analyses showed a marginal main effect of ANOMALY for midline sites [ $F(1,21) = 3.80, p < 0.07$ ] and a significant interaction of ANOMALY x ROI [lateral:  $F(3,63) = 23.44, p < 0.0001$ ; midline:  $F(5,105) = 9.11, p <$

<sup>8</sup>The time window chosen for the late positivity in Experiment 2a (650–1100 ms) differed slightly from that in Experiment 1 (600–1100 ms) on account of visual inspection of effect onset in the grand average ERPs. This amounts to a reduction of the window size by 10% of the sample points.



0.0001]. Resolving the interaction by ROI indicated that the positivity effect only reached significance in posterior lateral ROIs [ $F_s > 9.3$ ;  $p_s < 0.01$ ] and for midline sites PZ and POZ [ $F_s > 10.4$ ;  $p_s < 0.01$ ].

## 3.2 Experiment 2b

### 3.2.1 Materials and Methods

**3.2.1.1 Participants:** Twenty-four monolingually raised native speakers of American English (students at the University of California, Irvine) participated in the experiment after giving informed consent (15 women, mean age 21.5, range 18–29). Parameters for participant inclusion were the same as for Experiments 1 and 2a. Six participants were excluded due to excessive EEG artefacts.

**3.2.1.2 Materials:** The materials were adapted from Sanford et al.'s (2011) stimuli for American participants (i.e. passages that required specifically British world knowledge were altered to fit into an American context and British expressions were replaced by appropriate counterparts in American English).<sup>9</sup> Materials were recorded by a trained speaker of American English using the same recording parameters as for Experiment 1.

As in Experiment 2a, each participant in Experiment 2 heard 180 passages in total: 90 borderline anomaly sentences and 30 controls, as well as 30 sentences in each of the easy-to-detect anomaly and control conditions, respectively.

**3.2.1.3 Procedure:** The experimental procedure was identical to that in Experiment 2a.

**3.2.1.4 EEG data recording and preprocessing:** The EEG data were recorded using an EGI net amps 300 amplifier and a 256-channel HydroCel Geodesic Sensor Net (Electrical Geodesics, Inc., Eugene, OR) with a 500 Hz sampling rate. The data were recorded using a vertex reference, but re-referenced to linked mastoids offline. Impedances were kept below 50 k $\Omega$ .

In order to ensure maximal comparability of the data analysis to Experiments 1 and 2a, the entire data preprocessing and analysis procedure was restricted to the 32 channels that were recorded in our previous studies. Data preprocessing was accomplished in an identical manner to Experiment 2a. For this data set, some participants showed significant EMG contamination at occipital electrodes; accordingly, additional IC components representing muscle artefacts (Jung et al., 2000) were also removed for some participants. These components were identified by their location and significant high-frequency content in their power spectra. On average, 7.4 (sd: 2.4) ICs were removed per participant. The mean weight at electrode CPZ for all artefact ICs was low (0.02), indicating that these ICs did not substantially represent or influence activity measured at centroparietal sites.

**3.2.1.5 Data analysis:** Data analysis was undertaken in an identical manner to Experiments 1 and 2a.

### 3.2.2 Results

**3.2.2.1 Detection rates:** Analysis of the behavioural data showed that easy-to-detect anomalies were correctly judged as implausible at a rate of 91.3% (sd: 4.9%), whereas for borderline anomalies the detection rate was 55.6% (sd: 11.4%). Comparing these results to the detection rates of Experiment 2a revealed significant differences for both anomaly types

---

<sup>9</sup>See Appendix A for a more detailed description of the differences between the materials used here and in Sanford et al. (2011).



[Borderline anomalies:  $t(32.81) = 4.66, p < 0.001$ , easy-to-detect-anomalies:  $t(33.67) = 2.97, p = 0.005$ ].

**3.2.2.2 ERP data: Borderline anomalies:** Figure 5 shows ERPs for borderline anomaly sentences and their plausible controls. The data pattern replicates that observed by Sanford et al. (2011): detected borderline anomalies elicited a late positivity in comparison to missed anomalies and plausible controls, but no N400 effect.

For the 200–500 ms time window, neither the lateral nor the midline electrodes showed a significant effect of ANOMALY [all  $ps > 0.11$ ] or an interaction of ANOMALY  $\times$  ROI [all  $Fs < 1$ ].

In the 650–1100 ms time window, the data showed a main effect of ANOMALY for the midline electrodes [ $F(2,34) = 5.09, p < 0.05$ ]. Pairwise comparisons between the three levels of anomaly type showed a significant difference between detected and plausible borderline anomalies [ $F(1,17) = 6.62, p < 0.05$ ] as well as between detected and missed borderline anomalies [ $F(1,17) = 6.87, p < 0.05$ ]. There was no difference between missed anomalies and plausible controls [ $F < 1$ ].

**3.2.2.3 ERP data: Easy-to-detect anomalies:** The ERP results for easy-to-detect anomalies and their plausible counterparts are shown in Figure 6. As in Experiments 1 and 2a as well as Sanford et al. (2011), these types of anomalies elicited a biphasic N400 - late positivity response in comparison to plausible controls.

In the 300–600 ms time window, the data showed a main effect of ANOMALY for the midline electrodes [ $F(1,17) = 7.99, p < 0.05$ ] and an interaction ANOMALY  $\times$  ROI [lateral:  $F(3,51) = 11.36, p < 0.0001$ ; midline:  $F(5,85) = 11.97, p < 0.0001$ ]. Analyses per ROI revealed significant effects of ANOMALY in posterior lateral ROIs [ $Fs > 10.4$ ] and for midline sites CZ, CPZ, PZ and POZ [ $Fs > 7.7$ ].

The analysis of the 650–1100 ms time window revealed an interaction of ANOMALY  $\times$  ROI [lateral:  $F(3,51) = 16.19, p < 0.0001$ ; midline:  $F(5,85) = 10.05, p < 0.0001$ ]. Analyses per ROI showed that the positivity effect for anomalous versus plausible sentences reached significance only at midline sites PZ [ $F(1,17) = 4.94, p < 0.05$ ] and marginal significance at POZ [ $F(1,17) = 3.61, p = 0.07$ ].

### 3.3 Cross-experiment analysis of Experiments 2a and 2b

In order to directly compare the German and English findings in Experiments 2a and 2b, we conducted an additional cross-experiment analysis including LANGUAGE as a between-participants factor. Note that, while main effects of LANGUAGE apparent in this analysis could in principle be due to the different ERP systems used in Experiments 2 and 3 (though this appears unlikely in view of the basic methodological foundations of event-related brain potentials), the predicted interactions between LANGUAGE and ANOMALY cannot be explained via a change in amplifier.

**3.3.1 Borderline anomalies—**In the N400 time window, the cross-experiment analysis of the borderline anomaly sentences showed an interaction of ANOMALY  $\times$  LANGUAGE [lateral:  $F(2,76) = 5.37, p < 0.01$ ; midline:  $F(2,76) = 3.40, p < 0.05$ ].

The analysis of the late positivity time window, by contrast, did not show an ANOMALY  $\times$  LANGUAGE interaction [ $ps > 0.3$ ], but only main effects of ANOMALY and LANGUAGE.

**3.3.2 Easy-to-detect anomalies**—In spite of the qualitatively similar data patterns observed for the easy-to-detect anomalies in German and English, the analysis of the N400 time window showed an interaction of ANOMALY  $\times$  LANGUAGE [lateral:  $F(1,38) = 14.03, p < 0.001$ ; midline:  $F(1,38) = 9.81, p < 0.01$ ]. This result indicates that the N400 effect for easy-to-detect anomalies was smaller in amplitude in the English experiment (Experiment 2b) as opposed to the German experiment (Experiment 2a).

In the late positivity time window, no interactions with LANGUAGE reached significance. Rather, we only observed a main effect of ANOMALY.

### 3.4 Discussion

The results of Experiment 2 replicate the findings of Experiment 1 and Sanford et al. (2011) for German and English, respectively. They thus demonstrate that the cross-linguistic difference suggested by the comparison of Experiment 1 and Sanford and colleagues' findings is indeed robust. This conclusion was supported by an additional cross-experiment analysis including the between-participants factor LANGUAGE, which revealed an interaction between ANOMALY and LANGUAGE in the N400 time window but not the late positivity time window for the borderline anomalies.

Interestingly, the cross-experiment analysis also showed an interaction with LANGUAGE in the N400 time window for the easy-to-detect anomalies, thus indicating that the magnitude of the N400 effect for this anomaly type varies across language. Specifically, it appears to be less pronounced for English (Experiment 2b) than for German (Experiment 2a); visual inspection suggests a similar difference in magnitude between Experiment 1 and the data in Sanford et al. (2011). We shall return to this issue in the General Discussion, where we suggest that differences in N400 effect magnitude for the easy-to-detect anomalies can potentially be explained by the same mechanism that accounts for the cross-linguistic variation in the borderline anomalies.

Finally, the findings of Experiments 2a and 2b – when viewed in comparison to the results of Experiment 1 and those by Sanford et al. (2011) – suggest that the detection rates for borderline anomalies are subject to a certain degree of inter-individual variability, rather than being influenced systematically by the choice of task or the language under investigation. While the direct comparison of the behavioural findings between the two German studies (Experiment 1 and Experiment 2a) appear to suggest, at a first glance, that the methodology employed in Experiments 2a and 2b engenders higher detection rates, this assumption is not compatible with the comparison between the corresponding English experiments (Experiment 2b and Sanford and colleagues' 2011 study), which showed a reversed effect (a 55% detection rate in Experiment 2b versus a 63% detection rate in Sanford et al's experiment). Overall, these various comparisons do not show systematic differences in detection rate depending on task or language, but rather indicate that the detection rate in a give experiment depends, at least in part, on the particular sample of participants under examination.

## 4. GENERAL DISCUSSION

The results of recent ERP studies indicate that the N400 effect elicited by semantic manipulations is subject to systematic cross-linguistic variation in some cases, as reflected in diverging electrophysiological responses to semantic reversal anomalies in English and German. Here, we investigated the processing of German and English borderline and easy-to-detect anomalies to test whether language-specific patterns would also be found in this case. We thereby aimed to provide new evidence regarding the functional mechanism(s) underlying the N400.

Three ERP experiments confirmed our predictions regarding cross-linguistic differences in the electrophysiological response to borderline anomalies. For German, Experiments 1 and 2a demonstrated a biphasic N400 - late positivity response to detected borderline anomalies in comparison to both non-detected anomalies and plausible controls. For English, by contrast, Experiment 2b replicated previous findings by Sanford et al. (2011) in showing only a late positivity effect for detected borderline anomalies versus both non-detected anomalies and controls, but no N400 effect. For classic easy-to-detect anomalies with a poor fit to the global context, all of our experiments showed a similar result (as also observed by Sanford et al., 2011), namely a biphasic N400 - late positivity pattern for anomalous versus plausible sentences.<sup>10</sup>

In view of these findings, the following discussion focuses mainly on the implications of the differential results found for English and German for current accounts of the lexical-semantic N400. We will propose an analysis that accounts for both the English and German patterns and also provides a potential explanation for task-dependent variation within a language. We will also touch briefly on the late positivity effects found in both German and English; however, the discussion is mainly centred around the N400, since this is the effect that differentiates the neural responses to borderline anomalies in the two languages.

#### 4.1 Cross-linguistic differences in the N400

##### 4.1.1 Challenges for preactivation-based (lexical) accounts of the N400—

Several accounts have been put forward with respect to the underlying mechanisms of the N400. The two most prominent theories link N400 modulations to (i) the costs of integrating new information into an ongoing meaning representation (integration view) (e.g. Hagoort & van Berkum, 2007; Hagoort, 2008); or (ii) to the accessibility of a word's lexical representation as determined by its “preactivation” (lexical preactivation view) (e.g. Kutas & Federmeier, 2000; Lau et al., 2008). Recently, lexical accounts have been advocated by a number of researchers on the basis of findings such as the fact that semantic reversal anomalies typically do not engender N400 effects in English and that N400 amplitude therefore does not appear to reflect message-level plausibility (e.g. Brouwer et al., 2012; Stroud & Phillips, 2012). Since the comparatively low detection rates for borderline anomalies appear to be connected to the close semantic fit of the critical word to the global context (Barton & Sanford, 1993), the absence of an N400 effect for English borderline anomalies also seemed best explained in terms of lexical preactivation. The findings of the current study, however, challenge this interpretation. As the stimuli used in the German experiments were kept as similar as possible to the English materials and the degree of contextual fit in the respective conditions was almost identical between the two languages (see section 2.1.2), the lexical preactivation perspective does not account for the presence of an N400 effect for detected borderline anomalies in German.

Moreover, the qualitative differences in the processing of German and English are further corroborated by the finding of comparable electrophysiological dissociations in two different domains: semantic reversal anomalies (Bornkessel-Schlesewsky et al., 2011) and the borderline anomalies reported in the present study. These results thus call for an account in which an N400 modulation reflects more than a purely top-down influence of contextually generated lexical preactivation.

<sup>10</sup>Note that, while semantic anomalies were traditionally associated with a monophasic N400 response (Kutas & Hillyard, 1980), a range of studies have now observed biphasic N400 - late positivity patterns in response to “classic” (easy-to-detect) semantic violations (e.g. Faustmann, Murdoch, Finnigan, & Copland, 2007; Gunter, Jackson, & Mulder, 1992; Roehm, Bornkessel-Schlesewsky, Rösler, & Schlewsky, 2007; Sanford et al., 2011) and to semantic reversal anomalies (Bornkessel-Schlesewsky et al., 2011; Bourguignon et al., 2012). It has not yet been shown conclusively under which conditions semantic incongruities engender a late positivity in addition to an N400, though van de Meerendonk, Kolk, Vissers, & Chwilla (2010) recently suggested that this may be related to the strength of the incongruity.

**4.1.2 The interplay of top-down and bottom-up information sources**—We propose that the cross-linguistic differences in question are most adequately explained within accounts that emphasise the interplay of top-down and bottom-up information sources in lexical-semantic N400 modulations (e.g. Federmeier, 2007; Lotze, Tune, Schlesewsky, & Bornkessel-Schlesewsky, 2011). Lotze and colleagues, for example, observed a change in N400 amplitude due to a purely form-based, bottom-up manipulation, which modulated neither lexical preactivation nor ease of integration (i.e. capitalisation of a semantically incongruous sentence-final word). Additional evidence in favour of their “bidirectional coding account” stems from studies that have demonstrated an influence of discourse and information structure (Burkhardt, 2006; Schumacher, 2009) or prosody (Schumacher & Baumann, 2010) on the N400. From a cross-linguistic perspective, the bidirectional coding account allows for a modulation of the proposed interactive mechanism by assuming that different languages vary with regard to their (default) relative weighting of top-down and bottom-up information sources. It also provides a potential explanation for task effects, assuming that task can modulate the top-down/bottom-up balance (e.g. in the sense of a rational adaptation to current task constraints; Howes, Lewis, & Vera, 2009). In the following, we will discuss the general assumptions of the account in more detail as well as how it applies to the processing of English and German borderline anomalies, respectively.

During sentence interpretation, language processing requires the use of various cues in the input. Top-down influences include semantic cues such as global contexts and lexical associations at a more local level. Additionally, there are grammatical cues, with position and word order having special status. Because language unfolds over time, word order is a cue that is equally accessible in all human languages, whereas availability of other grammatical cues is dependent on characteristics of the language in question (see Bornkessel-Schlesewsky et al., 2011, for discussion).

The semantic cues provided by context serve to activate potential referents and concepts. Concomitantly, grammatical cues focus the predictions for upcoming words (e.g., via category restrictions). If grammatical cues induce the anticipation of a noun, this can lead to decreased activation of verbs and consequently to stronger predictions. While this basic principle is assumed to hold for all languages, individual languages differ with respect to the balance of top-down and bottom-up influences. Specifically, the degree to which interpretation is driven by word order seems to play a crucial role: though German and English are closely related, they show substantial differences with regard to the extent of their dependency on word order in interpretation.

In English, rigid word order gives rise to a high degree of position-based predictability and to a dominant top-down influence. The importance of word order as a cue to sentence interpretation in English has been well studied, but most prominently by those within the framework of the competition model. MacWhinney et al. (1984) describe that word order clearly overrides agreement as a cue to sentence interpretation in English: “When given a sentence like ‘The pencil are kicking the cows’, English and Italian listeners make their decisions in entirely opposite directions” (MacWhinney et al., 1984, p. 144), i.e., English listeners choose a subject-verb-object (SVO) interpretation while Italian listeners interpret the structure as object-verb-subject (OVS). In terms of the bidirectional coding account, the pronounced positional predictability leads to strong expectations, and reduction in the probability of encountering certain words as opposed to others. The dominance of top-down information also means that the influence of potentially conflicting bottom-up information is significantly weaker: no problem is recognized unless there is a category error or failed expectation. For borderline anomalies, neither of these occurs as the context leads to strong lexical associations and preactivation (including that of the critical, locally implausible

word), with category expectations satisfied at the same time. This is reflected in the absence of an N400 effect for English borderline anomalies.

Importantly, the N400 observed for easy-to-detect anomalies can be explained by the same basic mechanism. In contrast to the borderline anomalies, the critical word in these semantic anomalies has a very poor fit to the global context and is therefore not preactivated during discourse processing. The lack of lexical preactivation leads to a conflict, and thus engenders an N400 effect for anomalous words.

German, by contrast, is a language that allows for more flexible word order, thus rendering positional predictability considerably weaker. As a result, bottom-up cues such as morphological case marking or agreement are more important for sentence interpretation (cf. MacWhinney et al., 1984, who use the terms “local” vs. “topological” cues for a similar distinction to that framed in terms of bottom-up vs. top-down cues here). Without strong top-down expectations, these bottom-up features must be matched against the sentence context to determine the relation of the current word to previous discourse. In borderline anomalies, the implausibility is introduced by a mismatch between the critical word and a preceding local context word. As a result of the stronger weighting of bottom-up information in German, this mismatch has a stronger impact as reflected by the presence of an N400 effect for detected borderline anomalies. This account further explains why the amplitude of the N400 effect is more pronounced for the easy-to-detect anomalies than for the borderline anomalies, since the former involve a stronger conflict between bottom-up and top-down information in the absence of lexical preactivation.

In summary, the interaction of top-down and bottom-up information is subject to cross-linguistic variation, with the importance of word order in interpretation constituting the crucial difference between German and English. Despite comparable preactivation for the critical word in borderline anomalies in German and English, only German requires strong focus on bottom-up information. The mismatch between the critical word and the local sentence context yields the N400 effect in German because of the grammatically motivated weighting of bottom-up information that is significant in German and negligible in English.

This account can also explain the cross-linguistic differences in the magnitude of the N400 effect for easy-to-detect anomalies. Recall that the statistical analysis comparing Experiments 2a and 2b revealed a less pronounced N400 effect for English in contrast to German, similar to the visual comparison of Experiment 1 and Sanford et al.'s (2011) findings. If the threshold for inducing a bottom-up mismatch between a word and its preceding context is higher in English than in German -- even if this threshold is exceeded by the easy-to-detect anomalies in both languages -- it is exceeded to a higher degree in German. Thus, in spite of comparable fit to the global context (or lack thereof) and virtually identical detection accuracy in both languages, the N400 effect for the easy-to-detect anomalies appears to be reliably larger in German in comparison to English. This could suggest that a smaller number of easy-to-detect anomaly trials engenders an N400 effect in English compared to German (i.e. the bottom-up-threshold is exceeded only in a certain number of cases). Note that this variation in the N400 is independent of detection accuracy, an issue which we will discuss in more detail in the following section.

#### **4.2 Neural correlates of anomaly detection and the late positivity**

It is important to note that the dominance of top-down influences in English does not imply that English speakers should be more susceptible to semantic illusions. The comparable anomaly detection rates for English and German (see section 3.4 for discussion) show that this is not the case. In other words, the presence or absence of an N400 effect is not directly correlated with the detectability of a distorted meaning, but reflects a language-specific



interaction of the cues that drive interpretation. Depending on the importance of specific cues for interpretation, a conflict may or may not be registered during this phase of processing. Detection may, however, occur later. Accordingly, we argue that anomaly detection is reflected in the late positivity that follows the N400, but not in the N400 itself. This view is supported by the presence of a significant correlation between individual detection rates for borderline anomalies and late positivity effect in Experiment 1 measured at electrode POZ [ $r(21) = .43, p < 0.05$ ], and the lack of such correlation between the mean anomaly detection rate per participant and the N400 effect [ $r(21) = -.11, p = 0.67$ ]. These results are also in line with findings by Kolk, van Herten and colleagues, which suggest that conflict detection correlates with positivity effects (e.g. Kolk et al., 2003; van Herten, Kolk, & Chwilla, 2005; van Herten, Chwilla, & Kolk, 2006). Since the presence of a late positivity for German borderline and easy-to-detect anomalies mirrors the results found in English, we refer to Sanford et al. (2011) for a more detailed discussion of this effect.

### 4.3 Outlook: Modulating the top-down/bottom-up balance

A question that arises from our interpretation of the present findings is whether and how the balance between top-down and bottom-up factors during language comprehension can be modified. As suggested by Bohan et al.'s (2012) findings, it may be possible to induce N400 effects for borderline anomalies in English with a suitable task manipulation. This indicates that the top-down/bottom-up balance is not fixed at a set level within a language, but can vary depending on the experimental environment.

Previous behavioural results suggest that manipulating sentence focus can increase detection rates of Moses-type illusions (e.g., by means of it-clefts such as “It was Moses who took two animals of each kind on the Ark. True or False?”, Brédart & Modolo, 1988), as can increasing the difficulty of a font in reading (Song & Schwarz, 2008). In addition, Wang, Hagoort and Yang (2009) observed an increased N400 effect for contextually inappropriate vs. appropriate continuations in Chinese when the critical word was in focus. Taken together, these results suggest that the manipulation of (certain) structural and physical properties may lead to an increased salience of bottom-up information; a comparable manipulation could thus help induce an N400 effect for borderline anomalies in English.

Another possibility lies in manipulating the linguistic content itself. For example, if the critical word were to induce a morphosyntactic mismatch (e.g., an agreement violation), this could increase the degree of bottom-up processing. Similar outcomes can also be achieved by information at the syntax-semantics interface, e.g. verbs with non-standard mappings from form to meaning. Bourguignon et al. (2012) observed N400 rather than late positivity effects for semantic reversal anomalies in English when these were induced by Experiencer-verbs instead of standard action (Agent-Patient) verbs. This suggests that a verb which requires such a non-standard form-to-meaning mapping could lead to an increased consideration of bottom-up information.

Finally, as suggested by Bohan et al.'s (2012) findings, task demands might also modulate the bottom-up/top-down balance. By focusing participants' attention on judgement accuracy, the importance of bottom-up information is increased. (For further evidence regarding task-based modulations of electrophysiological activity within the N400 time window, see Haupt, Schlesewsky, Roehm, Friederici, & Bornkessel-Schlesewsky, 2008).

In summary, we propose that the cross-linguistic variation reported here (and that previously observed for semantic reversal anomalies) reflects differences in the default weighting of top-down versus bottom-up information in a given language. In this framework, these weights are not fixed, but can vary depending on the contextual environment. Importantly,



however, in the majority of ecological situations, German and English call for a differential weighting of top-down and bottom-up information.

## 5. CONCLUSION

The present findings demonstrate that the N400 response to semantic anomalies is subject to cross-linguistic variation. We interpret this result as arising from the interplay between top-down and bottom-up factors during language processing and the importance of these different information sources in a given language. We suggest that in languages with a relatively strict word order (e.g. English) and concomitant top-down predictability, language comprehension is constrained much less by bottom-up factors than in a language in which item-based information is more directly relevant for sentence understanding (e.g. German, in which morphological case marking and animacy play important roles in interpretation). We assume that the N400 reflects the degree of match between top-down and bottom-up information sources and that this is why German shows an N400 effect for borderline anomalies while English does not.

## Acknowledgments

Parts of the research reported here were supported by a German Academic Exchange Service scholarship awarded to ST and by a grant from the German Research Foundation to IBS (BO 2471/3–2). We would like to thank Laura Maffongelli, Fritz Milde, Aidan Brennan and Fiona Weiß for assistance with the data acquisition. Some of this work was performed in the U.S. and partly funded by the NIH NIDCD under grant DC-R01-3378 to SLS.

## APPENDIX A. Additional examples and description of sentence materials used in Experiments 1, 2a and 2b

Tables A1 and A2 show additional examples for the borderline anomalies and easy-to-detect anomalies, respectively. A full set of experimental materials can be obtained from the corresponding author upon request. For Experiment 2b, the materials from Sanford et al. (2011) were used, of which some had to be modified for American participants. Importantly, changes were kept as minimal as possible and never affected the context words differentiating between plausible and anomalous condition of a given item or the critical target word to which ERPs were time-locked. For borderline anomalies, 23 items were adapted; for easy-to-detect anomalies, four items were changed. In some cases, changes were made to context or target sentences only, while other items required adaptation of both target and context sentence. Some examples of the modifications are provided in Table A3.

The comparison of the stimuli used in Experiments 2a and 2b shows that, for the easy-to-detect anomaly condition, 56 of 60 items were literal translations of the same sentence pairs. For the borderline anomaly condition, Experiment 2a used 101 literal translations of the 120 items of Experiment 2b, the remaining 19 items belonged to the pool of new stimuli added to Experiment 1. An overview of the target words divided by lexical category is given in Table A4. In addition, for noun phrases, presence/absence of case marking and grammatical functions are listed in Table A5.

**Table A1**

Additional examples for borderline anomalies in Experiments 1, 2a and 2b.

Borderline anomalies	
German (Exp. 1 and 2a)	English (Exp. 2b)
(1) Dorothea und Sascha hatten einige Freunde	(1) Dorothy and Sam were having a dinner party,

<b>Borderline anomalies</b>	
<b>German (Exp. 1 and 2a)</b>	<b>English (Exp. 2b)</b>
zum Abendessen eingeladen, allerdings waren sie noch nicht ganz fertig, als die ersten Gäste eintrafen. Dorothea zerdrückte schnell ein paar reife <i>{Artischocken/Avocados}</i> für die Zubereitung ihres Lieblingsdips <u>Guacamole</u> , der zuerst serviert wurde.	but their guests were arriving and they weren't quite ready. Dorothy quickly mashed up some fresh <i>{artichokes/avocados}</i> to make his favourite dip, <u>guacamole</u> , which she served first.
(2) Emma hatte Sarah schon gewarnt, dass sie beim Betreten des Zimmers jede Menge Schmutz erwarten würde. Trotzdem war sie verärgert über die <i>{weiße/schwarze}</i> Dreckschicht, die die Lieferung der <u>Kohlen</u> am Vormittag verursacht hatte.	(2) Emily warned Sarah to expect a large mess when she walked in to the living room. However, when she saw a fine <i>{white/black}</i> dust everywhere due to the <u>coal</u> delivery, she was angry.
(3) Draußen herrschten eiskalte Temperaturen und so entschied sich Jakob für seine wärmste Winterkleidung. Als erstes zog er seine <i>{neuen Winterschuhe/neue Winterjacke}</i> an und dann noch dicke <u>Socken</u> , um nicht zu frieren.	(3) It was an icy, cold day outside and Jack decided to put on his warmest clothes. He put on his new winter <i>{boots/jacket}</i> and then his thick woolly <u>socks</u> so he'd stay warm.
(4) Es war das größte und modernste Schiff seiner Zeit und niemand hatte erwartet, was passieren würde. Auf ihrer lang ersehnten Jungfernfahrt im <i>{Indischen/Atlantischen}</i> Ozean sank die bis heute berühmte <u>Titanic</u> innerhalb von wenigen Stunden.	(4) It was the biggest ship of its day and no one expected what was about to happen. On her maiden voyage in the <i>{Indian/Atlantic Ocean}</i> , an accident sunk the <u>Titanic</u> in a few hours.
(5) Peter hörte im Radio den neuesten Song von Lady Gaga und mochte ihn sehr. Er konnte einfach nicht mit dem <i>{Summen/Singen}</i> des im Grunde genommen albernen <u>Songtextes</u> ihrer neuen Single aufhören.	(5) Pete heard the new song by Lady Gaga on the radio and liked it a lot. He really could not stop himself <i>{humming/singing}</i> those quite silly and annoying <u>lyrics</u> for the whole day.

**Table A2**

Additional examples for easy-to-detect anomalies in Experiments 1, 2a and 2b.

<b>Easy-to-detect anomalies</b>	
<b>German (Exp. 1 and 2a)</b>	<b>English (Exp. 2b)</b>
(1) Harald wollte seiner Frau zum Hochzeitstag ein schönes Geschenk kaufen und entschied sich, ihr neue Schuhe zu schenken. Im Schuhgeschäft kaufte er ein Paar <u>Pedalen/Stiefel</u> und bat die nette Verkäuferin, sie als Geschenk zu verpacken.	(1) Harold wanted to buy his wife a lovely present and decided to buy her some shoes. In the shoe shop he bought her some <u>pedals/boots</u> and asked the assistant to gift wrap them.
(2) Sarah rief nach ihrem Ehemann, nachdem sie die Treppe im Haus heruntergestürzt war. Tom verband ihren Knöchel mit einem <u>Schrauben-schlüssel/Verband</u> aus dem Erste-Hilfe-Koffer und brachte sie anschließend ins nächste Krankenhaus.	(2) Sarah called for her husband Don to help her after she fell down the stairs. Don bandaged Sarah's ankle with the <u>spanner/bandages</u> from his first-aid box and then took her to hospital.
(3) Da er erst kürzlich einen Unfall auf dem Wasser gehabt hatte, war Johann bei seinen Segeltrip ein wenig nervös. Er lenkte sein Boot vorsichtig in das <u>Blumenbeet</u> /in den <u>Hafen</u> , und ankerte ohne Probleme am Ende des langen Steges.	(3) John was feeling nervous about sailing because he'd had an accident on the water recently. He sailed his boat carefully into the <u>flowerbed/harbour</u> and successfully moored alongside the pier without hitting anything.
(4) Johannes war es überaus wichtig, komfortabel und stilvoll zu reisen. Er zahlte zweitausend Euro für einen Flug erster Klasse nach Australien in einem nagelneuen,	(4) Travelling in comfort and style was so important to John. He paid two thousand dollars for a premium class flight to Australia on a newly refurbished

Easy-to-detect anomalies	
German (Exp. 1 and 2a)	English (Exp. 2b)
großen Schlauchboot/Airbus. (5 Die beide Wanderer waren durchgefroren und hungrig, und zu allem Übel war ihr Ziel noch weit entfernt. Sie waren offensichtlich im Kreis gelaufen, da sie wohl einen Fehler beim Lesen der <u>Diät/Karte</u> gemacht hatten.	dinghy/jet. The two hill walkers were cold, hungry and lost. They had been walking in circles for nearly the whole day because they had misread the <u>diet/map</u> .

**Table A3**

Examples of the modifications undertaken in adapting Sanford et al.'s (2011) materials for American participants. Modified words are printed in bold.

Changes	British Version (Sanford et al., 2011)	American Version (Exp. 2b)
context sentence	A pay dispute between <b>lorry</b> drivers and their employer reached a crisis in negotiation, even the professional mediators seemed very dejected.	A pay dispute between <b>truck</b> drivers and their employer reached a crisis in negotiation, even the professional mediators seemed very dejected.
target sentence	Television news reports of <b>British</b> soldiers {celebrating/weeping} in response to their enemy's victory have received many complaints.	Television news reports of <b>US</b> soldiers {celebrating/weeping} in response to their enemy's victory have received many complaints.
context and target sentence	<b>Scotland</b> has chronic levels of heart disease and obesity and <b>Scotland's</b> politicians want to change this. The <b>Scottish Executive</b> is hoping to {prevent/encourage} people from adopting a healthy lifestyle to halt this trend.	<b>The USA</b> have chronic levels of heart disease and obesity and <b>America's</b> politicians want to change this. The <b>Surgeon General</b> is hoping to {prevent/encourage} people from adopting a healthy lifestyle to halt this trend.

**Table A4**

Comparison of the word categories of the critical words in Experiments 2a and 2b.

Condition	Exp.	Noun	Proper noun	Adjective	Verb	Total
Borderline anomalies	2a	95	10	7	18	120
	2b	83	18	6	10	120
Easy-to-detect anomalies	2a	58	-	1	1	60
	2b	56	2	1	1	60

**Table A5**

For critical words that were nouns / noun phrases, overview of the number of case-marked nouns and of the grammatical functions of the critical items.

Condition	Exp.	Case-marking		Grammatical function				
		Yes	No	Subject	Direct object	Indirect object	Prepositional object	Other
Borderline anomalies	2a	98	7	15	18	1	25	46

Condition	Exp.	Case-marking		Grammatical function				
		Yes	No	Subject	Direct object	Indirect object	Prepositional object	Other
	2b	61	40	3	31	-	25	42
Easy-to-detect anomalies	2a	51	7	3	26	10	11	8
	2b	26	32	1	29	-	19	9

For English, target word were counted as case-marked when they were pronouns or part of prepositional phrases. The category *other* includes prepositional phrases, appositions, adverbial phrases and attributes.

## APPENDIX B. Number of trials included in the ERP analysis per experiment and condition

Table B1 provides an overview of the number of trials included in the final ERP analysis per experiment and condition.

**Table B1**

Trials analysed per experimental condition and experiment.

Condition		Exp. 1		Exp. 2a		Exp. 2b	
		Range	Average	Range	Average	Range	Average
Borderline Anomalies	detected	27–67	52.3 (10.6)	44–80	63.0 (8.9)	24–66	50.5 (10.4)
	non-detected	16–38	25.4 (5.7)	15–44	24.6 (8.0)	24–65	37.7 (10.2)
	plausible control	18–42	35.3 (6.1)	17–28	23.0 (3.0)	20–28	24.2 (2.4)
Easy-to-detect Anomalies	anomalous	24–42	37.2 (4.6)	22–30	28.2 (2.1)	20–30	25.6 (3.2)
	plausible control	28–40	37.0 (3.2)	24–29	27.0 (1.8)	24–29	27.3 (1.5)

## APPENDIX C. Analysis of the pre-onset negativity found in Experiment 1 and 2a

Figures C.1 and C.2 show grand average ERPs time locked to the the critical word in the borderline anomaly conditions and the corresponding plausible controls for Experiment 1 and 2a, respectively. As is apparent from the Figures, in addition to the N400 and late positivity, a further negative effect can be observed for detected anomalies relative to non-detected anomalies and plausible controls in time windows beginning before the onset of the critical word. While the effect is broadly distributed in Experiment 1, it appears to be restricted to parietal electrodes in Experiment 2a.

For Experiment 1, a time window of –350 to +50ms was chosen for statistical analysis of lateral and midline regions; the time window of –200 to +50ms was analysed for Experiment 2a. The results of the statistical analyses summarised below confirm the impression gained through visual inspection of the grand average ERPs: in Experiment 1, main effects of ANOMALY show that the negativity is broadly distributed across the lateral

and midline regions. In Experiment 2a, on the other hand, interactions of ANOMALY x ROI were observed, with main effects of ANOMALY only significant at PZ and POZ. The topographical maps provided in Figures C.1 and C.2 show that the pre-onset negativities are not only differentially distributed across the German experiments, but that their topography is also distinct from that of the respective N400 effects. Importantly, the deduction that the presence of the N400 effects is not dependant the pre-onset negativity is, on the one hand, supported by a significant negative correlation between the two negative effects at PZ [ $r(20) = -.68, p < 0.001$ ] and POZ [ $r(20) = -.70, p < 0.001$ ] in Experiment 2a and, on the other hand, by the lack of such a correlation in Experiment 1 [PZ:  $r(21) = .04, p = 0.85$ ; POZ:  $r(21) = .08, p = 0.71$ ], see also Figure C.3.

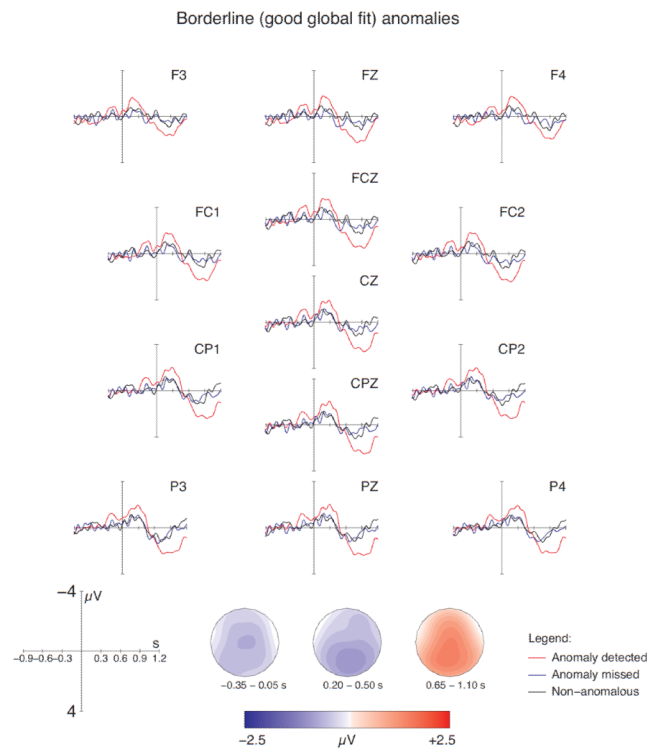
Furthermore, we tested whether the pre-onset negativity might have been caused by items in which the word preceding the critical word could have led to an early detection of the anomaly. For example, in the anomalous target sentence “First of all the authorities' initial negotiations with the scared and desperate hostages...” the attributes preceding the critical word might have already rendered a plausible continuation less likely. We carefully searched the materials for items with similar characteristics and computed an additional ERP analysis excluding these items. Visual inspection of the resultant ERPs, however, showed that the pre-onset negativity as well as the N400 effect were unaffected by this procedure. Taken together, the results speak against an interpretation of the observed pre-onset negativities in Experiment 1 and 2a as a systematic effect that might have been caused by the stimulus materials used. Moreover, the presence of the N400 found in the German studies cannot be explained by the occurrence on this earlier negativity. While there is not yet a clear explanation for the causes underlying the pre-onset negativity, it is possible that this effect is linked to inter-subject variability in terms of physiological parameters such as arousal, attention or overall mood that might have influenced cognitive performance (e.g. Lakatos et al., 2008; Kuipers & Thierry, 2011; Mathewson et al., 2011).

**Table C1**

Summary of the statistical analysis for the pre-onset negativity in Experiments 1 and 2a.

Exp. 1 (-350 – +50ms)	Exp. 2a (-200 – +50ms)
<b>ANOMALY</b>	<b>ANOMALY x ROI</b>
LAT: $F(2,44) = 3.95, p = 0.02$ detected vs. non-detected: $F(1,22) = 5.10, p = 0.034$ , marginally significant detected vs. control: $F(1,22) = 7.61, p = 0.011$	LAT: $F(6,126) = 5.73, p = 0.002$ ANOMALY - no effect in any ROI -
MID: $F(2,44) = 5.74, p = 0.006$ detected vs. non-detected: $F(1,22) = 7.80, p = 0.010$ detected vs. control: $F(1,22) = 10.80, p = 0.003$	MID: $F(10,210) = 4.97, p = 0.009$ ANOMALY PZ: $F(2,42) = 3.48, p = 0.040$ detected vs. control: $F(1,21) = 8.13, p = 0.009$ POZ: $F(2,42) = 4.15, p = 0.023$ detected vs. non-detected: $F(1,21) = 8.73, p = 0.007$ detected vs. control: $F(1,21) = 8.30, p = 0.009$

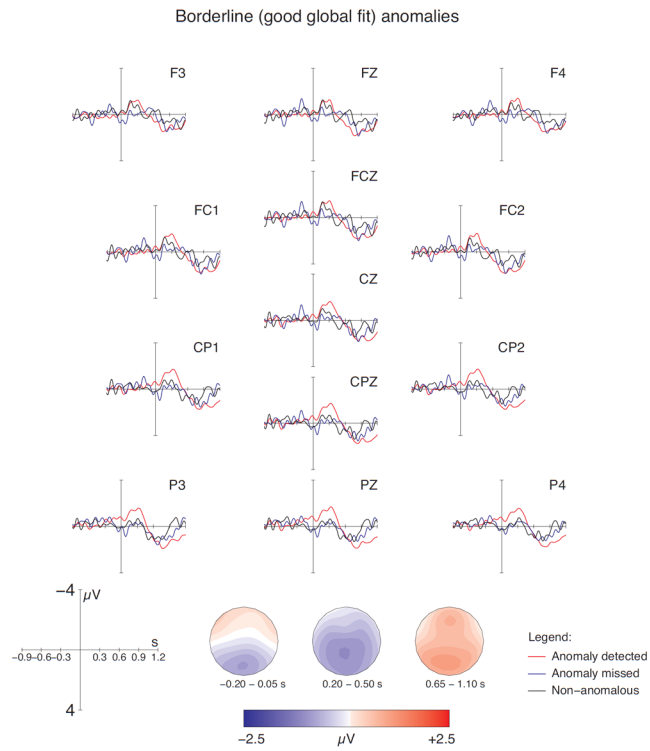
Only effects that reached significance are reported. Analogous to the analyses reported in the main text, a modified Bonferroni procedure was used to account for multiple testing, resulting in a corrected threshold of  $p = 0.033$  for pairwise comparisons.



**Figure C.1.**

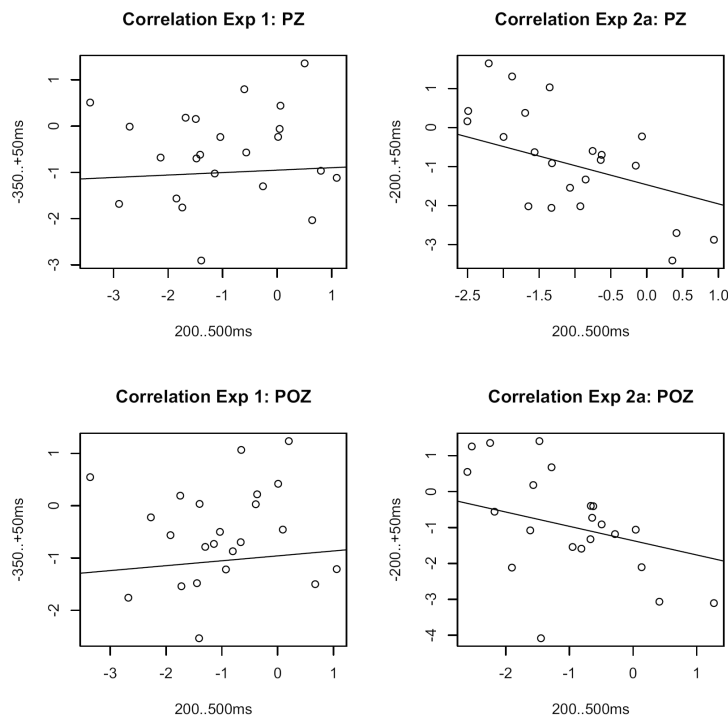
Grand average ERPs at the position of the critical word (onset at the vertical bar) in the borderline (good global fit) anomaly conditions at 13 selected electrodes in Experiment 1. The figure contrasts ERP responses to detected anomalies (red traces), missed anomalies (blue traces) and plausible controls (black traces). Negativity is plotted upwards. The topographical maps show the scalp distribution for the voltage difference between detected anomalies and plausible sentences in the pre-onset negativity, N400 and late positivity time windows, respectively.





**Figure C.2.**

Grand average ERPs at the position of the critical word (onset at the vertical bar) in the easy-to-detect (poor global fit) anomaly conditions at 13 selected electrodes in Experiment 2a. The figure contrasts ERP responses to anomalous (red traces) and non-anomalous sentences (blue traces). Negativity is plotted upwards. The topographical maps show the scalp distribution for the voltage difference between detected anomalies and plausible sentences in the pre-onset negativity, N400 and late positivity time windows, respectively.



**Figure C.3.** Scatterplots correlating the amplitude differences of detected borderline anomalies and plausible sentences in Experiment 1 and 2a for the pre-onset negativity and the N400 time windows at PZ and POZ.

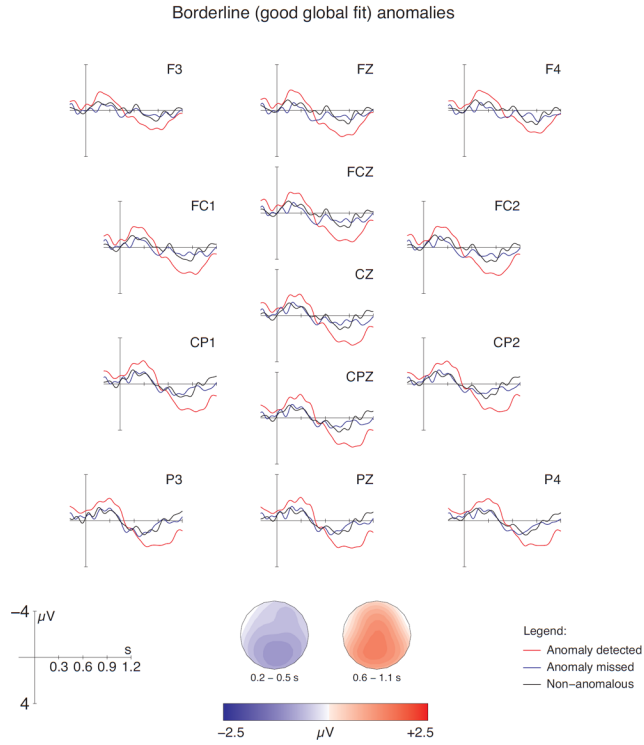
## REFERENCES

- Barton SB, Sanford AJ. A case-study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory and Cognition*. 1993; 21:477–487. [PubMed: 8350739]
- Bates E, Devescovi A, Wulfeck B. Psycholinguistics: A cross-language perspective. *Annual Review of Psychology*. 2001; 52:369–396.
- Bohan J, Sanford AJ. Anomaly detection at the borderline of consciousness: An eyetracking study. *Quarterly Journal of Experimental Psychology*. 2008; 61:232–239.
- Bohan, J.; Sanford, AJ.; Glen, K.; Clark, F.; Martin, E. Focus and emphasis devices modulate depth of processing as reflected in semantic anomaly detection. Poster presented at the 14th Architectures and Mechanisms for Language Conference; Cambridge, UK. 2008.
- Bohan J, Leuthold H, Hijikata Y, Sanford AJ. The processing of good-fit semantic anomalies: An ERP investigation. *Neuropsychologia*. 2012; 50(14):3174–3184. [PubMed: 22975191]
- Bornkessel-Schlesewsky I, Kretzschmar F, Tune S, Wang L, Genç S, Philipp M, et al. Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language*. 2011; 117:133–152. [PubMed: 20970843]
- Bornkessel-Schlesewsky I, Schlewsky M. An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*. 2008; 59:55–73. [PubMed: 18617270]
- Brédart S, Modolo K. Moses strikes again: Focalisation effect on a semantic illusion. *Acta Psychologica*. 1988; 67:135–144.
- Brouwer H, Fitz H, Hoeks JCJ. Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*. 2012; 1446:127–143. [PubMed: 22361114]
- Bourguignon N, Drury JE, Valois D, Steinhauer K. Decomposing animacy reversals between agents and experiencers: An ERP study. *Brain and Language*. 2012; 122:179–189. [PubMed: 22694997]

- Burkhardt P. Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*. 2006; 98:159–168. [PubMed: 16725188]
- Büttner AC. Questions versus statements: Challenging an assumption about semantic illusions. *Quarterly Journal of Experimental Psychology*. 2007; 60(6):779–789.
- Daneman M, Reingold EM, Davidson M. Time course of phonological activation during reading: Evidence from eye fixations. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1995; 21:884–898.
- Erickson TA, Matteson ME. From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*. 1981; 20:540–552.
- Faustmann A, Murdoch BE, Finnigan SP, Copland DA. Effects of advancing age on the processing of semantic anomalies in adults: Evidence from event-related brain potentials. *Experimental Aging Research*. 2007; 33:439–460. [PubMed: 17886018]
- Ferreira F, Ferraro V, Bailey KGD. Good-enough representations in language processing. *Current Directions in Psychological Science*. 2002; 11:11–15.
- Gunter TC, Jackson JL, Mulder G. An electrophysiological study of semantic processes in young and middle-aged academics. *Psychophysiology*. 1992; 29:38–54. [PubMed: 1609026]
- Hagoort P. The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society B*. 2008; 363:1055–1069.
- Hagoort P, van Berkum JJA. Beyond the sentence given. *Philosophical Transactions of the Royal Society B*. 2007; 362:801–811.
- Hannon B, Daneman M. Susceptibility to semantic illusions: An individual-differences perspective. *Memory and Cognition*. 2001; 29:449–461. [PubMed: 11407422]
- Hannon B, Daneman M. Shallow semantic processing of text: An individual-differences account. *Discourse Processes*. 2004; 37:187–204.
- Haupt FS, Schlesewsky M, Roehm D, Friederici AD, Bornkessel-Schlesewsky I. The status of subject-object reanalyses in the language comprehension architecture. *Journal of Memory and Language*. 2008; 59:54–96.
- Hoeks JCJ, Stowe LA, Doedens G. Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*. 2004; 19:59–73. [PubMed: 14972359]
- Howes A, Lewis RL, Vera A. Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review*. 2009; 116:717–751. [PubMed: 19839682]
- Huynh H, Feldt LS. Conditions under which the mean-square ratios in repeated measurement designs have exact F-distributions. *Journal of the American Statistical Association*. 1970; 65:1582–1589.
- Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, Sejnowski TJ. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*. 2000; 37(2):163–178. [PubMed: 10731767]
- Keppel, G. *Design and analysis*. 3rd ed. Prentice Hall; Englewood Cliffs, NJ: 1991.
- Kim A, Osterhout L. The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*. 2005; 52:205–225.
- Kolk HHJ, Chwilla DJ, van Herten M, Oor PJ. Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*. 2003; 85:1–36. [PubMed: 12681346]
- Kuipers JR, Thierry G. N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience*. 2011; 5:61. [PubMed: 21747766]
- Kuperberg GR, Sitnikova T, Caplan D, Holcomb P. Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*. 2003; 17:117–129. [PubMed: 12763198]
- Kutas M, Federmeier KD. Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*. 2000; 4:463–469. [PubMed: 11115760]
- Kutas M, Hillyard SA. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*. 1980; 207:203–205. [PubMed: 7350657]

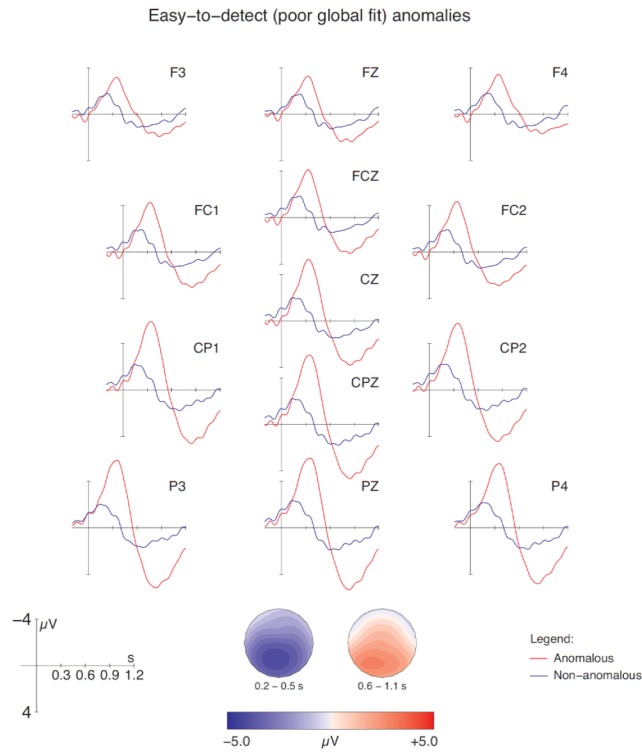
- Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*. 2008; 320:110–113. [PubMed: 18388295]
- Lau E, Phillips C, Poeppel D. A cortical network for semantics: (de)constructing the N400. *Nature Reviews. Neuroscience*. 2008; 9:920–933.
- Lotze N, Tune S, Schlesewsky M, Bornkessel-Schlesewsky I. Meaningful physical changes mediate lexical-semantic integration: Top-down and form-based bottom-up information sources interact in the N400. *Neuropsychologia*. 2011; 49:3573–3582. [PubMed: 21939678]
- MacWhinney B, Bates E, Kliegl R. Cue validity and sentence interpretation in English, German and Italian. *Journal of Verbal Learning and Verbal Behavior*. 1984; 23:127–150.
- Mathewson KE, Lleras A, Beck DM, Fabiani M, Ro T, Gratton G. Pulsed out of awareness: EEG alpha oscillations represent a pulsed-inhibition of ongoing cortical processing. *Frontiers in Psychology*. 2011; 2:99. [PubMed: 21779257]
- Oldfield RC. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*. 1971; 9:97–113. [PubMed: 5146491]
- Reder LM, Kusbit GW. Locus of the Moses Illusion: Imperfect encoding, retrieval or match? *Journal of Memory and Language*. 1991; 31:385–406.
- Roehm D, Bornkessel-Schlesewsky I, Rösler F, Schlesewsky M. To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*. 2007; 19:1259–1274. [PubMed: 17651001]
- Sanford AJ, Garrod S. The role of scenario mapping in text comprehension. *Discourse Processes*. 1998; 26:159–190.
- Sanford AJ, Graesser AC. Shallow processing and underspecification. *Discourse Processes*. 2006; 2:99–108.
- Sanford AJ, Leuthold H, Bohan J, Sanford AJS. Anomalies at the Borderline of Awareness: An ERP Study. *Journal of Cognitive Neuroscience*. 2011; 93:514–523. [PubMed: 19925201]
- Sanford AJ, Sturt P. Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences*. 2002; 6:382–386. [PubMed: 12200180]
- Schlesewsky M.; Bornkessel-Schlesewsky, I. When semantic P600s turn into N400s: On cross-linguistic differences in online verb-argument linking. In: Horne, M.; Lindgren, M.; Roll, M.; Alter, K.; Torkildsen, J. v. K., editors. *Brain Talk: Discourse with and in the brain. Papers from the first Birgit Rausing Language Program Conference in Linguistics. Birgit Rausing Language Program; Lund: 2009. p. 75-97.*
- Schumacher, PB. Definiteness marking shows late effects during discourse processing: Evidence from ERPs. In: Devi, SL.; Branco, A.; Mitkov, R., editors. *Anaphora processing and applications. Springer; Heidelberg: 2009. p. 91-106.*
- Schumacher PB, Baumann S. Pitch accent type affects the N400 during referential processing. *Neuroreport*. 2010; 21:618–622. [PubMed: 20489672]
- Shafiq M, McKay DG. The Moses, mega-Moses, and Armstrong illusions: Integrating language comprehension and semantic memory. *Psychological Science*. 2000; 11(5):372–378. [PubMed: 11228907]
- Song H, Schwarz N. Fluency and the detection of misleading questions: Low processing fluency attenuates the Moses illusion. *Social Cognition*. 2008; 26:791–799.
- Stroud C, Phillips C. Examining the evidence for an independent semantic analyzer: An ERP study in Spanish. *Brain and Language*. 2012; 120:108–126. [PubMed: 21377198]
- van de Meerendonk N, Kolk HHJ, Chwilla DJ, Vissers CTWM. Monitoring in language perception. *Language and Linguistics Compass*. 2009; 3:1211–1224.
- van de Meerendonk N, Kolk HHJ, Vissers CTWM, Chwilla DJ. Monitoring in Language Perception: Mild and Strong Conflicts Elicit Different ERP Patterns. *Journal of Cognitive Neuroscience*. 2010; 22:67–82. [PubMed: 19199401]
- van Herten M, Chwilla DJ, Kolk HHJ. When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*. 2006; 18:1181–1197. [PubMed: 16839291]
- van Herten M, Kolk HHJ, Chwilla DJ. An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*. 2005; 22:241–255. [PubMed: 15653297]

- Van Oostendorp H, DeMul S. Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica*. 1990; 74:35–46.
- Viola FC, Thorne J, Edmonds B, Schneider T, Eichele T, Debener S. Semi-automatic identification of independent components representing EEG artifact. *Clinical Neurophysiology*. 2009; 120:868–877. [PubMed: 19345611]
- Wang L, Hagoort P, Yang Y. Semantic illusion depends on information structure: ERP evidence. *Brain Research*. 2009; 1282:50–56. [PubMed: 19501067]

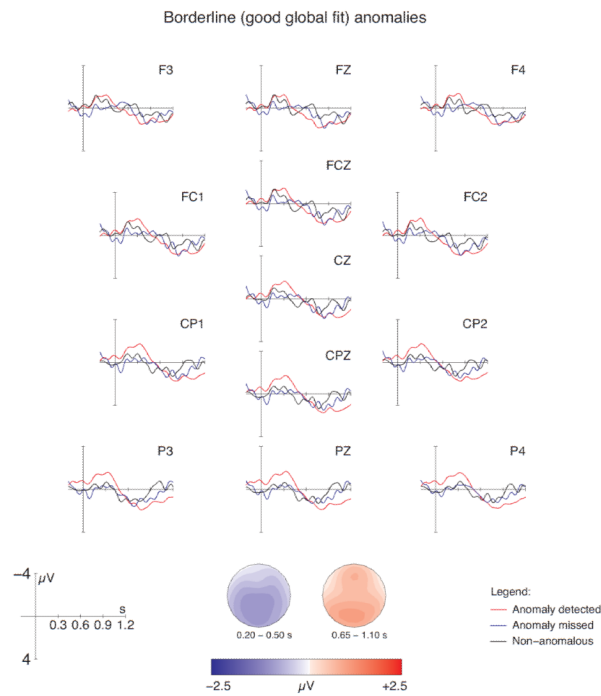


**Figure 1.** Grand average ERPs at the position of the critical word (onset at the vertical bar) in the borderline (good global fit) anomaly conditions at 13 selected electrodes in Experiment 1. The figure contrasts ERP responses to detected anomalies (red traces), missed anomalies (blue traces) and plausible controls (black traces). Negativity is plotted upwards. The topographical maps show the scalp distribution for the voltage difference between detected anomalies and plausible sentences in the N400 and late positivity time windows, respectively.

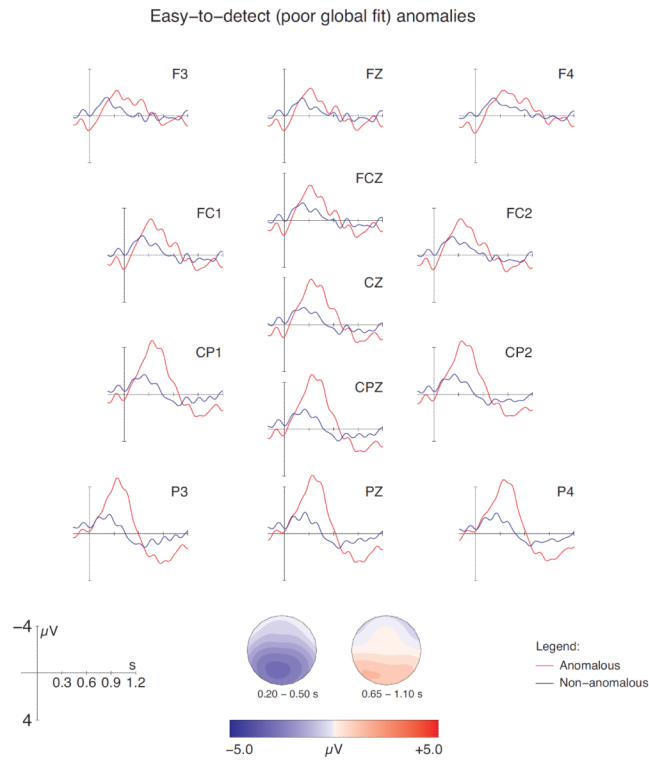




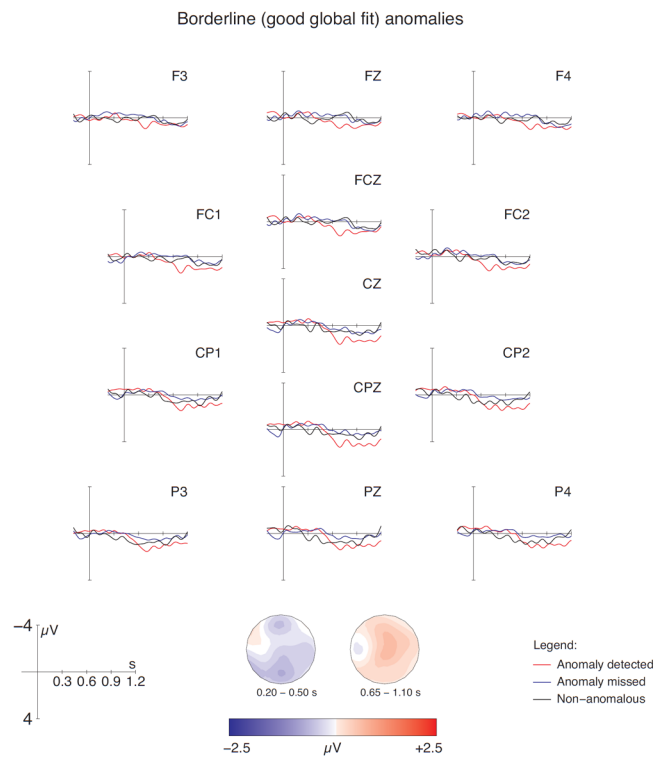
**Figure 2.** Grand average ERPs at the position of the critical word (onset at the vertical bar) in the easy-to-detect (poor global fit) anomaly conditions at 13 selected electrodes in Experiment 1. The figure contrasts ERP responses to anomalous (red traces) and non-anomalous sentences (blue traces). Negativity is plotted upwards. The topographical maps show the scalp distribution for the voltage difference between anomalous and plausible sentences in the N400 and late positivity time windows, respectively.



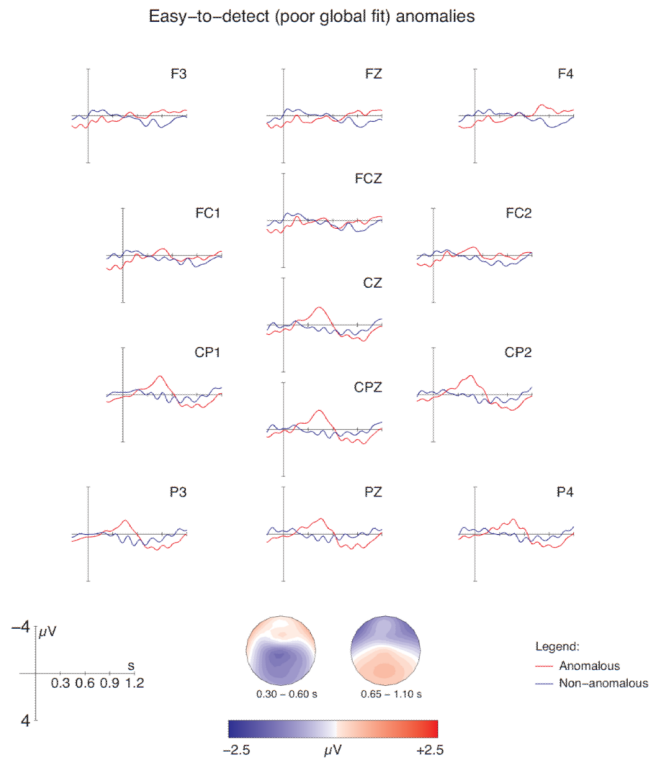
**Figure 3.** Grand average ERPs at the position of the critical word (onset at the vertical bar) in the borderline (good global fit) anomaly conditions at 13 selected electrodes in Experiment 2a. The figure contrasts ERP responses to detected anomalies (red traces), missed anomalies (blue traces) and plausible controls (black traces). Negativity is plotted upwards. The topographical maps show the scalp distribution for the voltage difference between detected anomalies and plausible sentences in the N400 and late positivity time windows, respectively.



**Figure 4.** Grand average ERPs at the position of the critical word (onset at the vertical bar) in the easy-to-detect (poor global fit) anomaly conditions at 13 selected electrodes in Experiment 2a. The figure contrasts ERP responses to anomalous (red traces) and non-anomalous sentences (blue traces). Negativity is plotted upwards. The topographical maps show the scalp distribution for the voltage difference between anomalous and plausible sentences in the N400 and late positivity time windows, respectively.



**Figure 5.** Grand average ERPs at the position of the critical word (onset at the vertical bar) in the borderline (good global fit) anomaly conditions at 13 selected electrodes in Experiment 2b. The figure contrasts ERP responses to detected anomalies (red traces), missed anomalies (blue traces) and plausible controls (black traces). Negativity is plotted upwards. The topographical maps show the scalp distribution for the voltage difference between detected anomalies and plausible sentences in the N400 and late positivity time windows, respectively.



**Figure 6.**

Grand average ERPs at the position of the critical word (onset at the vertical bar) in the easy-to-detect (poor global fit) anomaly conditions at 13 selected electrodes in Experiment 2b. The figure contrasts ERP responses to anomalous (red traces) and non-anomalous sentences (blue traces). Negativity is plotted upwards. The topographical maps show the scalp distribution for the voltage difference between anomalous and plausible sentences in the N400 and late positivity time windows, respectively.