Looking forward does not mean forgetting about the past: ERP evidence for the interplay of predictive coding and interference during language processing

Pia Schoknecht[1], Dietmar Roehm[1], Matthias Schlesewsky[2,3], Ina Bornkessel-Schlesewsky[2,3]

[1] Department of Linguistics, University of Salzburg, Salzburg, Austria

[2] School of Psychology, Social Work and Social Policy, University of South Australia, Adelaide, Australia

[3] Centre for Cognitive and Systems Neuroscience, University of South Australia, Adelaide, Australia

Correspondence concerning this article should be addressed to Pia Schoknecht, Department of Linguistics, University of Salzburg, Erzabt-Klotz-Strasse 1, 5020 Salzburg, Austria
E-Mail: pia.schoknecht@sbg.ac.at

Abstract

Interference and prediction have independently been identified as crucial influencing factors during language processing. However, their interaction remains severely underinvestigated. Furthermore, despite the growing body of behavioral studies investigating interference during sentence processing, the neurobiological basis of cue-based retrieval and retrieval interference remains insufficiently understood. Here, we addressed these issues with an ERP experiment that systematically examined the interaction of interference and prediction during language processing. We used the neurobiologically well-established predictive coding framework for the theoretical framing of our study. German sentence pairs were presented word-by-word, with an article in the second sentence constituting the critical word. We analyzed mean single trial EEG activity in a standard N400 time window and found an interaction between interference and prediction as measured by an offline cloze probability test. Under high predictability, sensory input was explained away (in the sense of predictive coding). Hence, no retrieval operations were necessary and no interference effects were observable. In contrast, under low predictability, model updating engendered memory retrieval, including the evaluation of distractor items, thus leading to interference effects. We conclude that interference should be included in predictive coding-based accounts of language because prediction errors can trigger retrieval operations and, therefore, induce interference.

Keywords: Interference, predictive coding, ERPs, N400, memory retrieval, German

1. Introduction

To successfully process a sentence or discourse, it is necessary to access previously encountered words, so that new words can be integrated with ones that have already been processed. The cognitive mechanisms underlying this process have been modeled in detail within the cue-based parsing framework (Lewis, Vasishth, & Van Dyke, 2006; Van Dyke & Lewis, 2003). It views memory retrieval as a cue-based process, which compares retrieval cues generated at the retrieval site with the features of the words in memory. Each new word or constituent is encoded into memory as a feature bundle (e.g. case, number, gender etc.). Importantly, not only the constituent itself is represented, but also predicted constituents that are required to form a grammatical sentence. Thus, when the noun *toy* is encountered in example one (from Lewis et al., 2006), two memory representations are constructed: one depicting its features and the other a prediction of the corresponding verb. When the verb *arrived* is processed, retrieval cues are generated to integrate it into the sentence context. The cues are matched in parallel against all (recent) memory representations. A sufficient match results in retrieval of the verb prediction and the integration of verb and noun.

(1) Melissa knew that the toy from her uncle in Bogotá arrived today.

But retrieval and, hence, comprehension difficulty can arise due to similarity-based interference from cue overlap, i.e. an overlap between the retrieval cues and features of more than one word / prediction recently stored in memory (Lewis et al., 2006). The correct to-be-retrieved word / prediction is called the target of the retrieval and other words / predictions in memory may function as interfering distractors during the retrieval process. This perspective is supported by numerous behavioral studies, in which interference was induced either through distracting words within a sentence or by extra-sentential memory items (see Jäger, Engelmann, and Vasishth (2017) for a recent review and meta-analysis). For example, Van Dyke and McElree (2006) used a dual-task paradigm to study interference from extra-sentential memory items (see 2a) during sentence comprehension. They presented it-

clefts (see 2b) that required a retrieval operation to integrate the matrix sentence verb (*sailed*/*fixed*) with its object (*boat*). The object is the target of the retrieval.

(2a) Memory load set: table – sink – truck

(2b) It was the boat that the guy who lived by the sea <u>sailed</u> / <u>fixed</u> in two sunny days.

In load conditions, three nouns that functioned as distractors (cf. 2a) were presented before a self-paced sentence reading task (for sentences such as 2b). In high interference conditions, the distractors were plausible objects of the matrix verb, and in low interference conditions, they were implausible (i.e. it is possible to fix, but not to sail a table / sink / truck; thus, interference is high for matrix verbs such as *fix*, but low for verbs such as *sail*). To ensure that there were no confounding differences in the comprehension of the verbs, no load conditions were included. In these conditions, the participants just read the sentences without an additional task. Reading times for the critical verb carrying the interference manipulation showed that, within the memory load conditions, high interference conditions led to longer reading times. In addition, comprehension accuracy was influenced by both memory load and interference, with responses being more accurate for low interference conditions and for no load conditions than for their high interference / load counterparts, respectively. These results were interpreted as showing that retrieval is the locus of interference during sentence comprehension rather than encoding, as the retrieval context was manipulated while encoding was held constant (Van Dyke & McElree, 2006).

Evidence for interference via distractors from within the sentence was provided, for example, by Gordon, Hendrick, and Johnson (2001), who found that similar noun phrase types within a sentence lead to higher processing difficulty. Van Dyke (2007) showed that distractors that match only a subset of retrieval cues could nevertheless influence retrieval and cause processing difficulties. Furthermore, it was found that distractors in-between target and retrieval site (retroactive interference) have more impact on processing than distractors that are processed before the target (proactive interference) (Martin & McElree, 2009).

However, the neurobiological basis of cue-based retrieval and interference in sentence processing remains insufficiently understood. Initial ERP evidence was presented by Martin, Nieuwland, and Carreiras (2012), who studied how distractors affect the processing of (un)grammatical NP ellipses in Spanish by manipulating the (mis)match between the gender of retrieval cues, targets and distractors. They reported a sustained, broadly distributed negativity for grammatical ellipses with a mismatching distractor compared to grammatical ellipses with a matching distractor and ungrammatical ellipses. It was concluded that the mismatching distractor interrupted the processing of grammatical ellipses via cue-based interference. In a follow-up study, Martin, Nieuwland, and Carreiras (2014) changed the syntactic structure of their material to render the antecedents of the ellipses (the target of the retrieval) less distinct. In this study, grammatical ellipses with matching distractors elicited an early anterior negativity compared to grammatical ellipses with mismatching distractors and ungrammatical ellipses. This was regarded as an effect of similarity-based interference. Thus, while both studies found negativities, they were elicited by different conditions (matching vs. mismatching distractors), had different distributions and latencies. The neurocognitive correlates of interference therefore clearly require further investigation. In addition, no attempt has been made to date to elucidate the possible neurobiological mechanisms underlying retrieval interference in language.

With the present study, we aimed to deepen the neurobiological understanding of interference during language comprehension by studying the interplay of interference and prediction within the neurobiologically well-established predictive coding framework. Predictive coding states that prediction is a general principle of brain function, with generative predictive models serving to explain the causes of sensory input (Friston, 2005). Predictive processing is organized within a hierarchically structured cortical architecture, within which anatomical feedback connections (from higher to lower levels) communicate predictions. Feedforward connections (from lower to higher levels), by contrast, communicate sensory input or – more precisely – only the parts of the input that are divergent from the prediction (Clark, 2013; Rao & Ballard, 1999). Thus, prediction errors essentially serve as a proxy for sensory information (e.g. Feldman & Friston, 2010), since the feed-forward activity associated with sensory input is "silenced" when predicted (i.e. excitatory

bottom-up activity is countered by inhibitory top-down activity in this case; cf. Friston, 2010). Overall, the system strives to reduce prediction errors, but when they do occur, they are used to update the current model so that more accurate predictions can be generated in the future.

According to a recent proposal by Bornkessel-Schlesewsky and Schlesewsky (2019), the N400 event-related potential provides a window on predictive coding during language comprehension. Specifically, this account posits that the N400 indexes *precision-weighted* prediction errors, with precision defined as the inverse of variance (Feldman & Friston, 2010) and – in the context of language – equating to the relevance of a particular information source for the form-to-meaning mapping. This offers an explanation for the presence versus absence of N400 effects for the same, prediction-error inducing phenomenon across different languages, depending on the language specific relevance (or "validity", cf. MacWhinney, Bates, & Kliegl, 1984) of cues such as word order, animacy or case marking (see Bornkessel-Schlesewsky et al. (2011) for empirical evidence, and Bornkessel-Schlesewsky and Schlesewsky (2019) for detailed discussion). In summary, while N400 amplitude shows a general tendency to decrease for predicted stimuli and increase for stimuli that involve prediction error (Dröge, Fleischer, Schlesewsky, & Bornkessel-Schlesewsky, 2016; Frank, Otten, Galli, & Vigliocco, 2015; Kuperberg, 2016), the magnitude of the N400 effect is assumed to be proportional to the degree to which the prediction error actually leads to a model update. In the present study, we will draw on this proposal to formulate our hypotheses with regard to the relation between prediction and interference during language processing.

The interplay of prediction and interference has — to the best of our knowledge — only been examined by one study to date. Campanelli, Van Dyke, and Marton (2018) used the Van Dyke and McElree (2006) paradigm (cf. example 2, above) with an added predictability manipulation (see example 3). By varying the subject noun (*person*/*choreographer*), the authors manipulated the cloze probability, and thereby the predictability, of the critical matrix sentence verb.

(3a) Memory load set: website – handbag – password

(3b) It was the dance that the person / choreographer who lived in the city performed / created early last month.

Campanelli et al. (2018) observed (marginal) reading time differences related to predictability and interference only in the sentence-final spillover region (e.g. 'early last month'). In the load conditions, residual log reading times were fastest in interference conditions with high predictability and slowest in interference conditions with low predictability of the critical verb. This result provides an initial indication that interference effects are canceled out when the word at the retrieval site is highly predictable. However, Campanelli et al. (2018)'s design was not fully crossed, as they investigated two levels of prediction (high, low) within the interference conditions, but only low prediction within no interference conditions. Thus, it is not possible to fully characterize the interplay between prediction and interference on the basis of these results.

We can, nevertheless, put forward a hypothesis regarding the relationship between prediction and interference within the context of a predictive coding framework. Recall from our discussion of predictive coding above that predictions essentially correspond to top-down activity within a hierarchically organized processing architecture, and that this activity cancels out bottom-up, stimulus-related activity when all features of a stimulus are fully predictable (silencing of prediction errors). By contrast, interference arises when bottom-up information calls for a retrieval and the features of the retrieval target are too similar to those of other items currently held in memory. It thus appears plausible that retrieval operations are only required in the absence of perfect predictability, i.e. when it is not possible to fully cancel out stimulus-related activity. In other words, when an upcoming word is fully predicted, it does not contain any new information that could serve to initiate a retrieval operation. We sought to examine this hypothesis with the present study.

1.1 The Present Study

The present study examined the interplay between prediction and interference in German, using a design that fully crossed the two factors. We focused on German articles as critical stimuli within a broader (two) sentence context, because they

overtly mark case, number and gender and are thus ideal for manipulating interference. In view of the design opportunities afforded by German for the examination of interference, it is somewhat surprising that only one study has investigated retrieval interference during German sentence processing to date (Nicenboim, Vasishth, Engelmann, & Suckow, 2018). Here, we aimed to provide further evidence for interference effects in German in addition to presenting what is – to the best of our knowledge – the first ERP study to investigate the interplay of predictive coding and interference. Specifically, we manipulated gender interference and prediction during the processing of an article (see example 4). In German, this word class carries (only) case, number and gender information.

(4) In der Garage steht ein Auto und vor dem Haus steht ein Wohnwagen. Peter verkauft <u>das</u> Auto an einen Kollegen.

'In the garage there stands a car[neuter] and in front of the house there stands a camper[masculine]. Peter sells the[neuter] car to a colleague.'

The target ('Auto' [neuter], engl.: car) and a distractor ('Wohnwagen' [masculine], engl.: camper) were introduced in a context sentence. At the critical article position within the second (target) sentence ('das' [neuter], engl.: the), the language processing system should seek to establish coreference with a noun in the preceding context (cf. Schumacher, 2009). We hypothesized that this process would be influenced by both the predictability of the article (and of the following noun) as well as by the degree of interference from the distractor noun. Materials were constructed so that they varied in terms of both target article and noun cloze probability and the degree of interference (see Methods for further details). By extending the distance between the retrieval site and the to-be-retrieved word across a sentence boundary, we hoped to gain insights into the scope of potentially interfering distractors.

Our hypotheses were as follows: when the article is fully predictable, prediction confirmation at this position further supports the existing prediction of the following noun. In cases of prediction disconfirmation at the article, by contrast, the internal model needs to be adjusted to prevent further prediction errors and the prediction for the upcoming noun needs to be revised. This revision necessarily involves

memory retrieval of potential candidates and we hypothesized that it should be influenced by interference. We expected similar effects at the position of the following noun. Again, if the specific noun was predicted, there will be prediction confirmation and no further processing is required (silencing of prediction errors). If it was not (fully) predicted, its integration calls for retrieval of the corresponding memory entry as well as an adjustment of the predictive model to prevent further prediction errors. Following Bornkessel-Schlesewsky and Schlesewsky (2019), we expected to find N400 effects for prediction errors versus prediction confirmation at both the article and following noun that and hypothesized that these would be modulated by interference.

2. Materials and Methods

2.1. Participants

Forty-four right-handed participants (8 male; mean age: 22.75 years, range: 18 – 35) from the University of Salzburg gave written informed consent and received either course credit or 20 Euros for their participation. All were native (Austrian) German speakers with normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. None of the participants took part in the questionnaire pretest. Four participants had to be excluded because of technical issues, three were excluded because their response accuracy was below 75 % and five participants were excluded because of excessive artifacts (more than 25 % of the critical sentences were affected). The data of thirty-two participants (5 male, mean age: 22.34 years, range: 18 – 35) was used for final analyses.

2.2. Materials

Eighty-five sets of four sentence pairs were constructed (see Table 1 for an example). Two noun phrases (NPs), which either matched or mismatched according to their gender features, were introduced in a context sentence and one of them was repeated in the target sentence. Target sentences were identical across conditions. The repeated NP was always the object of the target sentence and its article was our primary critical position for ERP analyses, though we also report additional analyses of the noun position. In German, articles and corresponding nouns must have

congruent gender, number and case features. In the high interference conditions (1/2), the article could grammatically be followed by each of the two nouns in the context, while, in the low interference conditions (3/4), there is only one compatible noun in the context. Additionally, the distance between the target noun and the retrieval site (the article) was manipulated via the word order in the context sentence.

2.3. Questionnaire Pretest

A sentence completion pretest was conducted to obtain cloze values for the critical article and the following noun in context (Kutas & Hillyard, 1984; Taylor, 1953). The material was split into four lists, with only one sentence pair from each lexical set occurring per list. In addition, there were two versions of each list: one truncated the target sentence before and one after the critical article, resulting in eight different questionnaires. The different lists / versions were constructed to ensure that the pretest participants stayed naïve towards the manipulation. No fillers were included in the pretest. A total of 144 students of the University of Salzburg participated in the pretest (16 to 20 per questionnaire version, mean age = 22.5 years, age range: 18 to 40 years, 39 men). Pretest participants were native speakers of (Austrian) German and were instructed to intuitively complete the sentence pairs. None of the pretest participants also took part in the main study.

The obtained cloze probability (CP) values ranged from 0 to 1 for both the article and noun. The results of the pretest and pairwise comparisons are shown in Table 2. The CP of the article in the high interference conditions was higher overall than in the low interference conditions. This is a logical consequence of the fact that, no matter with which of the two nouns of the high interference context the participants chose to continue the sentence, the article had the same form because the nouns shared the same gender feature. In contrast, the CP values for the article in the low interference conditions were lower because the article form varied based on the decision regarding which of the context nouns should follow. In both interference conditions, the participants showed a recency effect, i.e. they chose an article congruent to the last-mentioned noun more often than to the first-mentioned noun (cf. the higher CP values for a short distance to referent across conditions). In the high interference conditions, this recency effect is difficult to interpret as the nouns share the same

gender and might be due to an implicit bias related to verb semantics, i.e. in the way that one of the nouns is closer associated with the verb.

The CP values for the target noun are lower overall in the high interference conditions compared to low interference conditions. This reflects the fact that the article is ambiguous in the high interference conditions and that, in the low interference conditions, the article only matches one noun. Like the article results, the noun CP results show a recency effect.

In summary, both the article and the noun CPs obtained by our pretest reveal an offline recency effect. The participants expect the last-mentioned noun to be repeated in the target sentence. Also, for both the article and the noun, we see that the participants use the likelihood of an article and noun with a specific gender feature to complete the sentence pair.

2.4. Procedure

Eighty of the original eighty-five items were presented in the main experiment. The exclusion was done to achieve an even number of items and was not based on pretest results. Two pseudo-randomized lists were constructed so that each participant saw only two versions of one item (either version 1 and 4 or version 2 and 3). To avoid order effects, half of the participants were presented a backward version of the lists. Each of the lists consisted of 240 sentence pairs (160 critical and 80 filler sentence pairs). The first sentences of the filler sentence pairs were identical to the structure of the critical sentence pairs. The second sentence of the fillers was constructed to be coherent to the first sentence but did not include a referential link to it. All sentence pairs (both critical and filler pairs) were grammatical. A third of all sentence pairs in both lists (80 out of 240) was followed by a "yes"/"no" comprehension question. Half of the questions required a "yes" answer indicated by a right or left mouse click (balanced across participants).

Participants were instructed to silently read the sentence pairs, which were presented on the center of a computer screen in a word-by-word manner. Words were presented in white letters on a black background and participants were asked to minimize eye movements, blinks and general movements. A trial began with a 400 ms fixation cross followed by 200 ms blank screen. Words were presented for 400 ms

each with a 100 ms inter-stimulus interval between words; after the last word of the first sentence in the pair, there was a 1000 ms inter-stimulus interval. To encourage participants to read carefully, a third of the trials was followed by a yes/no comprehension question (i.e. 'Steht das Auto vor dem Haus?', engl.: 'Is the car in front of the house?'), which the participants had five seconds to answer. After that (or in trials without a task), there was a self-paced inter-trial interval, during which participants could blink as much as needed before pressing a button to start the next trial. All sessions began with a five-sentence-pair practice block to familiarize participants with the procedure. The following six experimental blocks were separated by short breaks. Each experimental session lasted approximately 2.5 hours; participants spent approximately 75 minutes under task.

2.5. Electroencephalogram recording and preprocessing

The electroencephalogram (EEG) was recorded from 58 active scalp electrodes positioned according to the standard 10-20 system and attached to an elastic cap (ActiCap from Easycap GmbH, Herrsching, Germany). EEG recordings used an ActiCHamp amplifier (Brain Products GmbH, Gilching, Germany) with a sampling rate of 500 Hz and AFz as the ground electrode. Six electro-oculogram (EOG) electrodes were positioned above and below both eyes and at the outer canthus of each eye. Impedances were kept below 10 kΩ. All electrodes were online referenced to the left mastoid and re-referenced to linked mastoids offline. All preprocessing steps were carried out using the Brainvision Analyzer 2. The raw data were filtered with a bandpass of 0.1 – 30 Hz. Artifact rejection was carried out semi-automatically, based on Brainvision Analyzer 2's Max-Min criterion. According to this criterion, intervals are rejected in which the six EOG channels show a difference of more than 100 μV within 200 ms windows, plus an additional 200 ms window before and after the artifact.

ERPs were computed in epochs of -200 to 1550 ms relative to critical article onset (noun onset was 500 ms after article onset). While we present grand average ERPs (see Figure 1) for comparability to previous research, the current design required single trial-based analyses, as we outline below.

2.6. Statistical Analysis

We performed linear mixed model analyses using the `R` (R Core Team, 2017) package `lme4` (Bates, Maechler, Bolker, & Walker, 2015) on the basis of the mean activity of EEG single trial data between 300 and 500 ms post onset of the article and noun. We chose the standard N400 time window based on theoretical considerations and our hypotheses, rather than via visual inspection of ERP waveforms. For both words, offline cloze probability values of the article and following noun were included as continuous predictors, together with the topographical factors sagittality (anterior: F7/8, F5/6, F3/4, F1/2, Fz, FC5/6, FC3/4, FC1/2, FCz; central: T7/8, C5/6, C3/4, C1/2, Cz, TP7/8, CP5/6, CP3/4, CP1/2, CPz; posterior: P7/8, P5/6, P3/4, P1/2, Pz, PO7/8, PO3/4, POz), and laterality (left: F7, F5, F3, FC5, FC3, T7, C5, C3, TP7, CP5, CP3, P7, P5, P3, PO7, PO3; medial: F1/2, Fz, FC1/2, C1/2, Cz, CP1/2, CPz, P1/2, Pz, POz; right: F8, F6, F4, FC6, FC4, T8, C6, C4, TP8, CP6, CP4, P8, P6, P4, PO8, PO4). In addition, all models included random intercepts for both subjects and items. While the use of maximal random effects structures has been encouraged for psycholinguistic research by some authors (Barr, Levy, Scheepers, & Tily, 2013), we chose to employ random intercepts only for the present study for two reasons. Firstly, models with maximal random effects structures did not converge for our data or were not computable within a manageable time frame. Secondly, as already noted by Barr and colleagues (2013), their suggestion explicitly targets *confirmatory* research rather exploratory investigations such as ours (see also Alday, Schlesewsky, & Bornkessel-Schlesewsky, 2017).

Beyond the commonalities of the article and noun models outlined above, the two models differed in the coding of the factors interference and distance. In the article model, Interference was modeled as a three-level factor (high, low with short distance, low with long distance between target and retrieval site) because, in the high interference conditions, the target noun cannot be determined at the article and therefore its recency is unclear until the noun position (i.e. both conditions are identical up to the position of the article). In the noun model, we modeled the two factors separately: Interference (high, low) and Distance between target and retrieval site (short, long). Using the `R` package `car` (Fox & Weisberg, 2011), categorical variables were encoded with sum contrasts, such that the estimate for a given fixed effects level represents the difference between this level and the grand mean. For

brevity, we present Type-II Wald χ^2 tests (also provided by the `car` package (Fox & Weisberg, 2011)) instead of the model summaries. We visualize effects of interest based on effects matrices created using the `R` package `effects` (Fox, 2003) and plotted using `ggplot2` (Wickham, 2016). Prior to the analyses, we excluded two items that were inconsistently constructed, leaving us with the data for 78 items.

3. Results

3.1 Comprehension Question Accuracy

Participants were encouraged to answer the comprehension questions correctly and within the time limit (five seconds). As noted above, three participants were excluded for accuracy rates below 75%. For the remaining participants, comprehension accuracy was good (mean: 85.4 %, range: 76.25 to 95 %), thus demonstrating that they processed the sentences attentively. Generalized linear mixed models with one or both of our condition factors (Interference and Distance between target and retrieval site) as predictors showed no improvements compared to a model without predictors (comparison basic model to Interference model: $\chi^2 = 3.1$, $p = 0.08$; comparison basic model to Distance model: $\chi^2 = 0.01$, $p = 0.9$; comparison basic model to interaction of Interference and Distance model: $\chi^2 = 3.9$, $p = 0.3$), hence comprehension accuracy was affected neither by Interference nor by Distance. We also checked whether there was a speed accuracy tradeoff. The corresponding generalized linear mixed model (with log reaction time, Interference, Distance and their interactions as predictors) revealed that correct answers were associated with a faster response time than incorrect answers (Estimate= -1.90217, SE = 0.23882, z = -7.965, p < 0.001). No other effects in this model reached significance.

3.2. Traditional ERP waveforms

For sake of completeness, we present traditional grand average ERP waveforms (N= 32) in Figure 1. Article onset is at 0 ms and noun onset at 500 ms. The traditional ERP methodology cannot account for the complexity of our design with continuous predictors (cf. Alday et al., 2017). Thus, while visual inspection of the grand average ERP waveforms is suggestive of some condition-based differences in the N400 time

windows at the article and noun positions, we will not interpret these further as they do not reflect our full design. Converging support for this assumption is presented in Table 3, which shows that, for both the article and noun positions, models including cloze probability values show a better fit to our data than models without them. This justifies our complex design and, in addition, our departure from the analysis of effects apparent in the ERP waveforms.

3.3. Electrophysiological Data: Article

Type-II Wald test results for the linear mixed model for the article are shown in Table 4. We focus on the four-way interaction between Sagittality, Interference, Article CP and Noun CP (χ^2 (4) = 17.54, p = 0.001), which is visualized in Figure 2. At posterior electrode sites, effects of interference emerged only when both Article CP and Noun CP were low and manifested themselves in the form of increased negativity for the high interference and low interference / long target distance conditions in comparison to the low interference / short target distance condition.

3.4. Electrophysiological Data: Noun

Type-II Wald tests for the linear mixed model for the noun are shown in Table 5. We focus on the Sagittality x Interference x Target distance x Noun CP interaction (χ^2 (2) = 29.54, $p < 0.001$), which is visualized in Figure 3. When Noun CP is low, at anterior electrode sites, the low interference / short target distance condition is more negative than the other conditions and the low interference / long target distance is more positive than the others. The high interference conditions do not differ from each other and show an amplitude that is intermediary between the two low interference conditions.

4. Discussion

We have presented an ERP study that investigated the interplay of prediction and interference within sentence pairs. In addition to the experimental manipulations, we included offline cloze probability (CP) values of the critical article and following noun in linear mixed model analyses of mean single trial EEG activity

within the N400 time window (300–500 ms post stimulus onset). At the position of the critical article, we found a posterior negativity for conditions with a high-interference distractor and for conditions with a close, low-interference distractor (i.e. long distance to target) in comparison to conditions with a distant, low-interference distractor (i.e. short distance to target). These effects only emerged when the CP of the article itself and of the following noun were low. At the noun position, interference and distance effects also manifested themselves when CP was low, but only the CP of the noun itself mattered. Here, the noun in the low interference condition with a short distance to the target elicited an anterior negativity compared to the other conditions. Amplitudes were less negative for the high interference conditions, which did not differ depending on target distance. Finally, conditions with a low-interference distractor and a long distance to target showed the least negative / most positive responses at the noun position.

In the following, we will first discuss the results for the article and noun positions, in turn. Subsequently, we will consider potential theoretical implications of our findings, before offering some conclusions.

4.1 Article

Effects at the article emerged only when the CP of both the article and noun were low. Thus, under conditions of high predictability, the distractors were essentially "ignored" by the processing system. We suggest that this aligns well with a predictive coding perspective, according to which high predictability leads to all (or almost all) of the sensory input for a current item being "explained away" (cf. Friston, 2010; and see the introduction for further discussion).

In contrast, when predictability is low, bottom-up information gains more weight. Since the input is not "explained away" under these circumstances, bottom-up sensory information (corresponding to prediction error), is propagated up through the levels of the hierarchically organized predictive processing architecture. Prediction errors induce model updating, which – we assume – requires an evaluation of all (other) possible candidate referents for the article and upcoming noun. Since this evaluation necessarily involves retrieval of the candidates, it is a natural locus for interference. We propose here that interference modulates the complexity of the

model update and, thereby, N400 amplitude. In sentences of the type employed in our experiment, the required model update involves retrieving the correct referent for the upcoming noun. This retrieval operation is facilitated under conditions of low interference: since the gender information provided by the article uniquely matches exactly one noun in the context, this scenario involves a low risk of future prediction errors in spite of the fact that the CP of noun and article was low prior to processing of the corresponding stimuli. In essence, retrieval is easy due to the lack of alternatives – possibly due to the restricted experimental context. In the low interference, short target distance condition, retrieval of the target noun is additionally supported by recency (see section 2.3 for a discussion of recency effects in our materials as revealed by the pretest). By contrast, when recency information contradicts gender information during referent retrieval, as is the case in the low-interference, long distance to target condition, model updating is more complex, due to the two conflicting information sources. Hence, we observe a more pronounced N400 effect under these conditions. Finally, under conditions of high interference, a low CP article induces a prediction error but, at the same time, provides no additional evidence to guide a model update (since it matches both referents in the context sentence). While a model update is clearly required here, it is associated with a high degree of uncertainty, again correlating with a more pronounced N400 amplitude.

Thus, in line with Bornkessel-Schlesewsky and Schlesewsky's (2019) proposal, our results at the position of the critical article suggest that N400 amplitude during incremental sentence comprehension does not reflect the presence/absence of a prediction error per se, but rather the effects of a prediction error on internal model updating. The present findings suggest that interference may play an important role in this regard, as the presence of interfering distractors not only impacts memory retrieval (i.e. the likelihood of retrieving the correct item from memory), but also the complexity of model updating. In addition, if an account along these lines is correct, it suggests that prediction error within a hierarchically organized predictive coding architecture might trigger memory retrieval. We return to this point in section 4.3 below.

4.2 Noun

At the noun position, effects are also only apparent under low (noun) CP. As already argued for the article position, we assume that highly predicted nouns do not engender model updating – and, hence, additional retrieval operations – as they are fully compatible with top-down predictions and thus lead to a silencing of prediction errors.

Under low noun CP, the most pronounced negativity was observable for the low interference / short target distance condition. In line with our discussion of the article effects, this could reflect the falsification (reflected in the low CP) of a high-precision prediction that was built even before the article was processed (arising from the absence of distractors and recency of the target). This leads to a pronounced prediction error response, i.e. increased N400 amplitude.

Both high interference conditions are less negative than the low interference / short target distance condition, irrespective of target distance. This could also be attributable to the precision-weighting of prediction errors (cf. Bornkessel-Schlesewsky & Schlesewsky, 2019): the article('s gender information) is a cue with low precision in these conditions, because its compatibility with both nouns in the context gives rise to two highly likely predictions. Accordingly, the precision of these predictions is reduced (see Todd et al., 2014, for discussion of precision modulation for predictions as opposed to sensory input) and N400 amplitude is reduced for a prediction error in comparison to the low interference / short target distance condition. This provides converging support for the idea (cf. Bornkessel-Schlesewsky & Schlesewsky, 2019) that the precision-based weighting of cues and predictions is not only language-specific, but also situation-, i.e. context-sensitive. Gender is normally a valid cue forming dependencies between articles and nouns in German, but the high interference context renders it invalid.

The low interference / long target distance condition was least negative at the noun position. This observation is more difficult to derive from our assumptions and we can thus only offer some speculative attempts as a post-hoc explanation. We assume that, in this condition, prediction precision is lower than in the low interference / short target distance condition because here, the gender cue is in conflict with the recency cue. However, this raises the question of why prediction precision should be even lower here than in the case of the high-interference conditions, as would be required to explain the extremely reduced (or even absent)

prediction error response in the N400. One possibility as to why this could be the case is that, in the low interference / long target distance condition, there are two different competing information sources (cues). In the high interference conditions, by contrast, the competing information stems from the same type of cue, namely gender. However, to test whether this explanation indeed holds, it would need to be rendered more precise via a quantitative model of the relative weighting (cue validity) of gender and recency in German and examined in a confirmatory study.

Finally, we would like to note that the topography of the noun effect is not typical for an N400 in the visual modality. It is, however, similar to that found by Martin et al. (2014), who reported an anterior negativity for ellipses with a matching distractor. But it is difficult to further connect these previous findings to our results, as our high interference conditions, which are most directly comparable to Martin and colleagues' critical condition, are associated with an intermediary amplitude of the interference-related anterior negativity effect for low CP nouns.

4.3 Theoretical implications

Our work has potential implications for all of the theoretical frameworks under discussion: predictive coding, cue-based parsing (retrieval) and the functional interpretation of N400 amplitude as reflecting model updating via precision-weighted prediction error signals. We will briefly discuss each of these in turn.

Within a hierarchically organized predictive coding architecture, prediction errors trigger model updating. Our results provide some initial indications as to how this updating process may be shaped by different information sources in the context of a complex domain such as higher-order language processing. Specifically, the updates required in our experimental conditions cannot depend entirely on incoming sensory information, but also require recourse to previous input, which needs to be accessed via retrieval operations. Retrieval can, in turn, be influenced by interference, which lowers the probability of correctly retrieving the target or, in a worst-case scenario, may even favor retrieval of an incorrect (distractor) item. If the incorrect item is retrieved, the priors of the new model are incorrect, thereby inevitably leading to prediction error in the future. If interference manifests as multiple possible targets for the retrieval, new predictions that are based on this retrieval situation are

less trusted because of the increased risk of prediction error. The resulting model thus has less reliability due to the lower-precision predictions. In summary, our findings suggest that: (a) prediction errors trigger memory retrieval when the incoming sensory information is not sufficient to guide model updating; and (b) retrieval interference during model updating diminishes the ability of the model update to minimize future prediction errors.

Under the cue-based parsing framework, memory representations not only comprise already processed words, but also predictions for upcoming words or constituents. The former are retrieved when new words need to be integrated with them (e.g. to establish co-reference or to saturate dependencies between heads such as verbs and their dependent arguments). By contrast, the latter are retrieved when the predicted words are actually encountered. The retrieval operation investigated with the presented study might be comparable to those during reflexive-/reciprocal-antecedent or ellipsis processing. Rather than retrieving an argument-verb dependency, for example, it calls for the retrieval of the correct (co-)referent for the current item. But there are no structural constraints (e.g. c-command in Chomskyan frameworks), which function as additional retrieval cues during the retrieval of our referent.

There appears to be a conceptual discrepancy between the memory representation of a predicted word that needs to be retrieved in cue-based parsing and predictions in the sense of predictive coding. In cue-based parsing, the memory representation of the representation is shifted out of attentional focus when other words are processed (see McElree, 2006). Thus, when the predicted word is encountered, the prediction must be retrieved in order to allow for an integration of the predicted, newly encountered word with the corresponding "old" word. Within the predictive coding architecture, by contrast, the prediction of an upcoming stimulus of a specific kind would be (repeatedly?) communicated via feedback connections until the stimulus is encountered, at which point the prediction would "explain away" the sensory input. However, given the propensity for non-adjacent dependencies in language, all stimuli that are encountered between the generation of the prediction and the predicted stimulus' appearance would lead to prediction error. This could either engender model updating and subsequent dismissal of the prediction, or – what is perhaps more likely – to a generalization of the prediction

(e.g. the prediction that a noun phrase should be followed, at some point, by a verb). This may lead to maintenance or even strengthening of the prediction (cf. anti-locality effects, (Husain, Vasishth, & Srinivasan, 2014; Konieczny, 2000)), which appears incompatible with the notion that a predicted memory representation disappears from the focus of attention and becomes active again later. One way to operationalize predictions for non-adjacent elements in a predictive coding framework may be to assume that these correspond to higher levels – and hence longer timescales – within the hierarchically organized predictive coding architecture (cf. Bornkessel-Schlesewsky & Schlesewsky, 2019; Bornkessel-Schlesewsky, Schlesewsky, Small, & Rauschecker, 2015).

The N400-as-model-updating account by Bornkessel-Schlesewsky and Schlesewsky (2019) posits that N400 amplitude reflects precision-weighted prediction errors and, hence, the consequences of a prediction error for model updating rather than prediction-error-related activity per se. Initially, the notion of precision-weighting was introduced to account for N400 modulations via the relevance (cue validity) of a particular feature within a given language. The results of the present study provide evidence to suggest that this notion could be extended to varying validity of a cue according to the specific context. Interference contexts reduce the validity of a cue as already described by the term of cue diagnosticity (Nairne, 2002). A cue is diagnostic if it uniquely picks out the target and excludes the distractors (see Martin et al., 2012 for detailed discussion). Thus, interference during a retrieval operation as part of a model update is closely related to the precision of future predictions.

Finally, the present findings provide further converging support for the notion that predictive processing during language comprehension may – at least under certain circumstances – lead to the anticipation of the *form* of upcoming input elements. This idea was prominently put forward by DeLong, Urbach, and Kutas (2005), who reported evidence for an anticipation of the English articles “a” versus “an” on the basis of a predicted, following noun. A recent large-scale replication study failed to replicate this effect at the position of the article (Nieuwland et al., 2018), thus leading to a controversial discussion of whether the language comprehension system indeed predicts upcoming form-based features. By showing effects of prediction and interference at an article (i.e. a function word that does not vary in its

compatibility with the context unless viewed in conjunction with the following noun), our study supports the notion that form-based predictions can indeed be generated. This adds further support to a growing body of literature in languages other than English, e.g. (noun-)gender-based N400 effects at the position of pre-nominal articles in spoken Spanish (Wicha, Bates, Moreno, & Kutas, 2003; but see Wicha, Moreno, & Kutas, 2004 for a P600 effect in a similar design) and pre-nominal adjectives in Polish (Szewczyk & Schriefers, 2013), and handshape-based N400 effects prior to the onset of a critical sign in German sign language (Hosemann, Herrmann, Steinbach, Bornkessel-Schlesewsky, & Schlesewsky, 2013). By further demonstrating that these pre-nominal prediction effects are modulated by a range of influences that have not been examined to date – most notably interference, but also the interaction of article and noun CP – our study provides an initial indication as to why previous studies may have shown apparently contradictory results.

5. Conclusion

The results of the present ERP study suggest that prediction outweighs interference. Under high predictability, interference effects do not arise, because the sensory input at the retrieval site is "explained away" in the sense of predictive coding. Hence, retrieval is not necessary, thus eliminating the possibility of interference. We conclude that interference may play a crucial role in a hierarchically organized, cortical predictive coding architecture for language as it influences the complexity of model updating and the precision of future predictions — particularly when predictability is low.

Acknowledgements: The authors would like to thank Jonas Diekmann for helpful discussions related to the research reported here. Thanks are also due to Anja Bergmair, Nadine Furtner and Veronica Baldin for their contributions to the experimental materials and to Barbara Sophie Hartl and again Anja Bergmair for their assistance during data acquistion. This research was supported by a research grant awarded to P.S. by the University of Salzburg. I.B.-S. acknowledges the support of an Australian Research Council Future Fellowship (FT160100437).

References

Alday, P. M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2017). Electrophysiology Reveals the Neural Dynamics of Naturalistic Auditory Language Processing: Event-Related Potentials Reflect Continuous Model Updates. *eNeuro, 4*(6). doi:10.1523/ENEURO.0311-16.2017

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J Mem Lang, 68*(3). doi:10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1 — 48. doi:10.18637/jss.v067.i01

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genc, S., Philipp, M., Roehm, D., & Schlesewsky, M. (2011). Think globally: cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain Lang, 117*(3), 133-152. doi:10.1016/j.bandl.2010.09.010

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Towards a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology, 10*. doi:10.3389/fpsyg.2019.00298

Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L., & Rauschecker, J. P. (2015). Neurobiological roots of language in primate audition: common computational properties. *Trends Cogn Sci, 19*(3), 142-150. doi:10.1016/j.tics.2014.12.008

Campanelli, L., Van Dyke, J. A., & Marton, K. (2018). *The Modulatory Effect of Expectations on Memory Retriebal During Sentence Comprehension.* Paper presented at the 40 th Annual Conference of the Cognitive Science Society, Austin, Texas.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci, 36*(3), 181-204. doi:10.1017/S0140525X12000477

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat Neurosci, 8*(8), 1117-1121. doi:10.1038/nn1504

Dröge, A., Fleischer, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2016). Neural mechanisms of sentence comprehension based on predictive processes and decision certainty: Electrophysiological evidence from non-canonical linearizations in a flexible word order language. *Brain Res, 1633*, 149-166. doi:10.1016/j.brainres.2015.12.045

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front Hum Neurosci, 4*, 215. doi:10.3389/fnhum.2010.00215

Fox, J. (2003). Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software, 8*(15), 1 — 27.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (2 ed.). Thousand Oaks CA: Sage.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain Lang, 140*, 1-11. doi:10.1016/j.bandl.2014.10.006

Friston, K. (2005). A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci, 360*(1456), 815-836. doi:10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat Rev Neurosci, 11*(2), 127-138. doi:10.1038/nrn2787

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *J Exp Psychol Learn Mem Cogn, 27*(6), 1411-1423.

Hosemann, J., Herrmann, A., Steinbach, M., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Lexical prediction via forward models: N400 evidence from German Sign Language. *Neuropsychologia, 51*(11), 2224-2237. doi:10.1016/j.neuropsychologia.2013.07.013

Husain, S., Vasishth, S., & Srinivasan, N. (2014). Strong expectations cancel locality effects: evidence from Hindi. *PLoS One, 9*(7), e100986. doi:10.1371/journal.pone.0100986

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language, 94*, 316-339.

Konieczny, L. (2000). Locality and parsing complexity. *J Psycholinguist Res, 29*(6), 627-645.

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Lang Cogn Neurosci, 31*(5), 602-616. doi:10.1080/23273798.2015.1130233

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends Cogn Sci, 10*(10), 447-454. doi:10.1016/j.tics.2006.08.007

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior, 23*(2), 127 — 150. doi:10.1016/S0022-5371(84)90093-8

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: evidence from verb-phrase ellipsis. *J Exp Psychol Learn Mem Cogn, 35*(5), 1231-1239. doi:10.1037/a0016271

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *Neuroimage, 59*(2), 1859-1869. doi:10.1016/j.neuroimage.2011.08.057

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2014). Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. *Brain Lang, 135*, 42-51. doi:10.1016/j.bandl.2014.05.001

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation*. San Diego: Academic Press.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory, 10*(5-6), 389-395. doi:10.1080/09658210244000216

Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and Confirmatory Analyses in Sentence Processing: A Case Study of Number Interference in German. *Cogn Sci*. doi:10.1111/cogs.12589

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., Meziere, D., Barr, D. J., Rousselet, G. A., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., Donaldson, D. I., Kohut, Z., Rueschemeyer, S. A., & Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *Elife, 7*. doi:10.7554/eLife.33468

R Core Team. (2017). R: A Language and Environment for Statistical Computing. Retrieved from https://www.R-project.org/

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci, 2*(1), 79-87. doi:10.1038/4580

Schumacher, P. B. (2009). Definiteness Marking Shows Late Effects during Discourse Processing: Evidence from ERPs. In S. L. Devi, A. Branco, & R. Mitkov (Eds.), *Anaphora Processing and Applications* (Vol. 5847 Lecture Notes in Artifical Intelligence, pp. 91-106). Berlin, Heidelberg: Springer.

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language, 68*(4), 297 —314. doi:https://doi.org/10.1016/j.jml.2012.12.002

Taylor, W. L. (1953). „Cloze procedure“: A new tool for measuring readability. *Journalism Quarterly, 30*, 415 — 433.

Todd, J., Heathcote, A., Mullens, D., Whitson, L. R., Provost, A., & Winkler, I. (2014). What controls gain in gain control? Mismatch negativity (MMN), priors and system biases. *Brain Topogr, 27*(4), 578-589. doi:10.1007/s10548-013-0344-4

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *J Exp Psychol Learn Mem Cogn, 33*(2), 407-430. doi:10.1037/0278-7393.33.2.407

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49*, 285 — 316.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *J Mem Lang, 55*(2), 157-166. doi:10.1016/j.jml.2006.03.007

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neurosci Lett, 346*(3), 165-168.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *J Cogn Neurosci, 16*(7), 1272-1288. doi:10.1162/0898929041920487

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Table 1: Example material.

Context sentences	
(1) high interference / long target distance	In der Garage steht ein Auto und vor dem Haus ein Motorrad. 'In the garage there stands a car (neutr.) and in front of the house a motorbike (neutr.).'
(2) high interference / short target distance	Vor dem Haus steht ein Motorrad und in der Garage ein Auto. 'In front of the house there stands a motorbike (neutr.) and in the garage a car (neutr.).'
(3) low interference / long target distance	In der Garage steht ein Auto und vor dem Haus ein Wohnwagen. 'In the garage there stands a car (neutr.) and in front of the house a camper (masc.).'
(4) low interference / short target distance	Vor dem Haus steht ein Wohnwagen und in der Garage ein Auto. 'In front of the house there stands a camper (masc.) and in the garage a car (neutr.).'
Target sentence	
for all conditions	Peter verkauft <u>das</u> Auto an einen Kollegen. 'Peter sells the (neutr.) car (neutr.) to a colleague.'

Critical word is underlined.

Table 2: Cloze probability values for the critical article and following noun.

Article	long distance	short distance	long - short
high interference	0.74	0.81	-0.07^*
			(−0.107 - −0.033)
low interference	0.39	0.49	-0.1^*
			(−0.172 - −0.029)
high - low	0.35^*	0.32^*	
	(0.29 - 0.41)	(0.27 - 0.37)	
Noun	long distance	short distance	long - short
high interference	0.45	0.57	-0.12^*
			(−0.195 - −0.046)
low interference	0.89	0.95	-0.06^*
			(−0.091 - −0.043)
high - low	-0.44^*	-0.38^*	
	(−0.49 - −0.38)	(−0.44 - −0.33)	

95 % confidence intervals in parentheses.

* $p < 0.05$

Table 3: Model comparisons for models of article and noun with and without cloze probabilities.

	df	AIC	BIC	logLik	deviance	χ^2	χ df	Pr(>χ^2)
Article								
sag.lat.interf	30	1438467	1438777	−719204	1438407			
sag.lat.interf.artCP	57	1438452	1439040	−719169	1438338	69.2	27	< 0.001
sag.lat.interf.artCP.nounCP	111	1438406	1439550	−719092	1438184	224	81	< 0.001
Noun								
sag.lat.interf.dist	39	1457968	1458371	−728945	1457890			
sag.lat.interf.dist.nounCP	75	1457724	1458498	−728787	1457574	316	36	< 0.001
sag.lat.interf.dist.nounCP.artCP	147	1457694	1459211	−728700	1457400	490	108	< 0.001

Model names reflect the predictors used: sag = sagittality, lat = laterality, interf = interference, dist = target distance, artCP = article cloze probability, nounCP = noun cloze probability.

Table 4: Type II Wald χ^2 test results of the article model.

	χ^2	Df	Pr(>χ^2)
sag	3863.41	2	< 0.001
lat	1443.89	2	< 0.001
interf	190.79	2	< 0.001
artCP	10.29	1	0.00134
nounCP	11.50	1	< 0.001
sag : lat	732.64	4	< 0.001
sag : interf	20.26	4	< 0.001
lat : interf	31.68	4	< 0.001
sag : nounCP	32.53	2	< 0.001
interf : nounCP	12.69	2	0.00176
artCP : nounCP	5.17	1	0.02295
sag : interf : artCP	22.09	4	< 0.001
lat : interf : artCP	23.23	4	< 0.001
lat : interf : nounCP	12.85	4	0.01206
sag : artCP : nounCP	17.76	2	< 0.001
lat : artCP : nounCP	13.19	2	0.00136
interf : artCP : nounCP	6.55	2	0.03779
sag : interf : artCP : nounCP	17.54	4	0.00151

Sag = sagittality, lat = laterality, interf = interference, artCP = article cloze probability, nounCP = noun cloze probability.

Table 5: Type II Wald χ^2 test results of the noun model.

	χ^2	Df	Pr(>χ^2)
sag	7579.38	2	< 0.001
lat	539.12	2	< 0.001
dist	162.03	1	< 0.001
nounCP	12.61	1	< 0.001
sag : lat	834.98	4	< 0.001
sag : dist	12.58	2	0.00186
lat : dist	24.88	2	< 0.001
interf : dist	7.00	1	0.00813
dist : artCP	14.68	1	< 0.001
sag : nounCP	14.02	2	< 0.001
interf : nounCP	13.68	1	< 0.001
sag : interf : nounCP	19.24	2	< 0.001
sag : dist : nounCP	6.92	2	0.03141
interf : dist : nounCP	81.50	1	< 0.001
sag : artCP : nounCP	14.77	2	< 0.001
interf : artCP : nounCP	17.13	1	< 0.001
lat : interf : dist : artCP	7.20	2	0.02729
sag : interf : dist : nounCP	29.54	2	< 0.001
lat : interf : dist : nounCP	10.79	2	0.00454

Sag = sagittality, lat = laterality, interf = interference, dist = target distance, artCP = article cloze probability, nounCP = noun cloze probability.

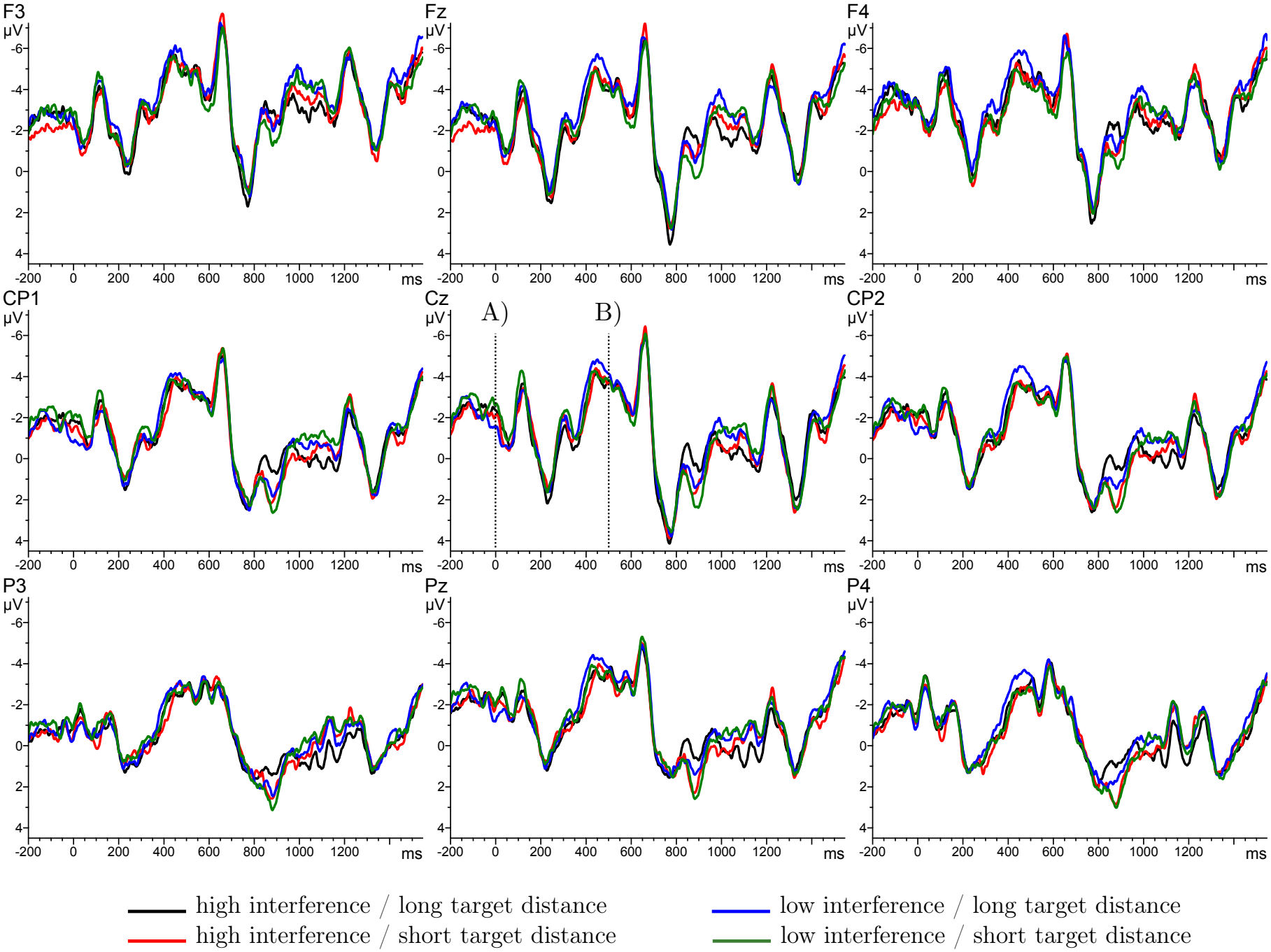

Figure 1: Grand averages at six electrode sites elicited by critical words (A) Article onset, B) Noun onset). Negativity is plotted upwards.

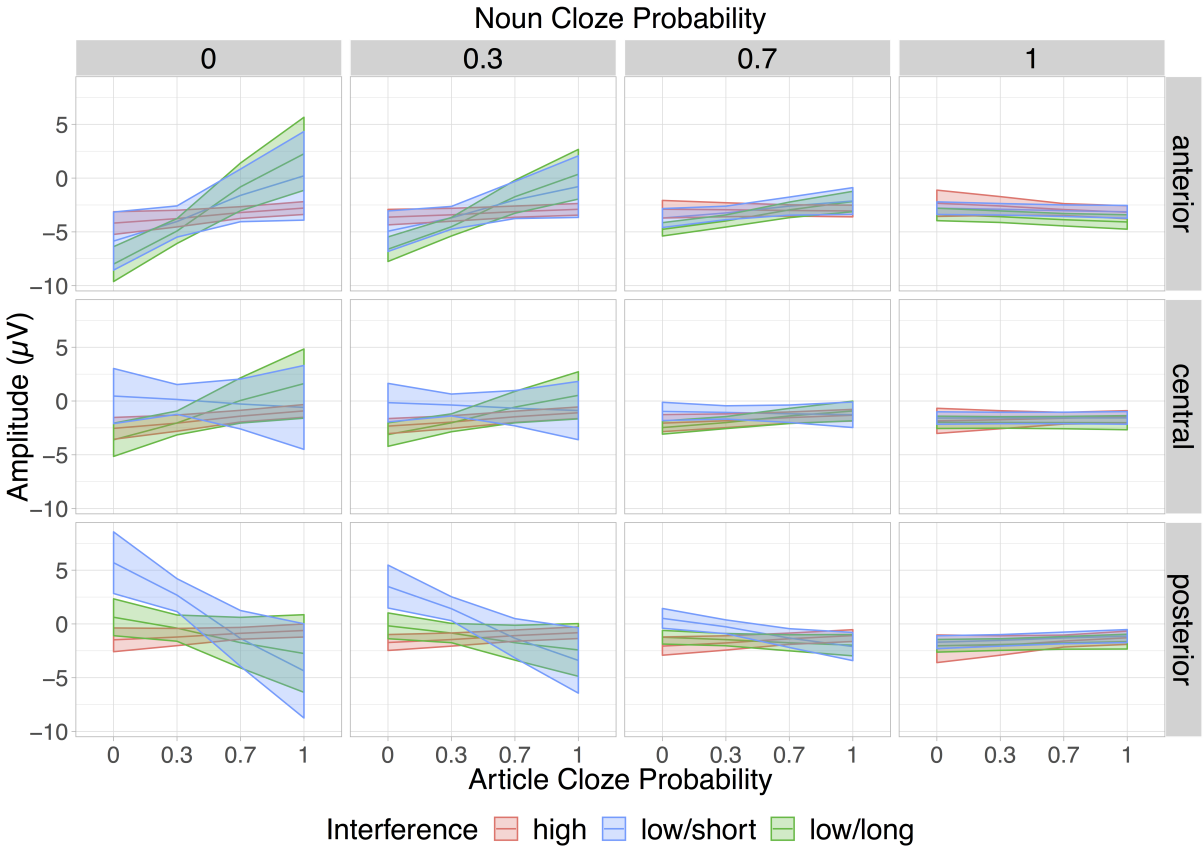

Figure 2: Effects of the Sagittality x Interference x Article CP x Noun CP interaction at the article position in the time window 300 - 500 ms post onset. Shaded areas indicate 95 % confidence intervals.

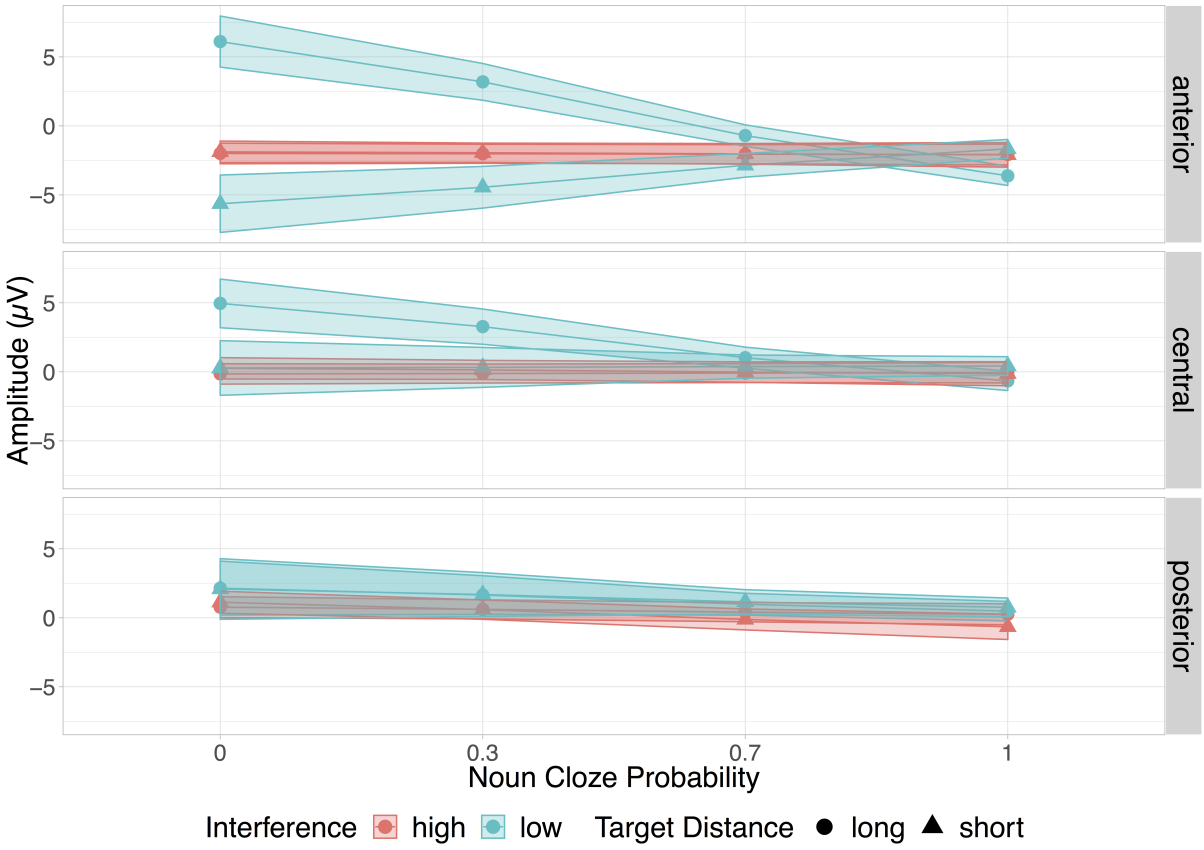

Figure 3: Effects of the Sagittality x Interference x Target Distance x Noun CP interaction at the noun position in the time window 300 - 500 ms post onset. Shaded areas indicate 95 % confidence intervals.